

Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective *Arabidopsis*

Peter D. Hoffman, Jeffrey M. Leonard, Gerrick E. Lindberg, Stephanie R. Bollmann, and John B. Hays¹

Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, Oregon 97331-7301, USA

During the many cell divisions that precede formation of plant gametes, their apical-meristem and floral antecedents are continually exposed to endogenous and environmental mutagenic threats. Although some deleterious recessive mutations may be eliminated during growth of haploid gametophytes and functionally haploid early embryos (“haplosufficiency quality-checking”), the multiplicity of plant genome-maintenance systems suggests aggressive quality control during prior diploid growth. To test in *Arabidopsis* a hypothesis that prior mismatch repair (MMR) is paramount in defense of plant genetic fidelity, we propagated in parallel 36 MMR-defective (*Atmsh2-1*) and 36 wild-type lines. The *Atmsh2-1* lines rapidly accumulated a wide variety of mutations: fifth-generation (G5) plants showed abnormalities in morphology and development, fertility, germination efficiency, seed/silique development, and seed set. Only two *Atmsh2-1*, but all 36 wild-type lines, appeared normal at G5. Analyses of insertion/deletion mutation at six repeat-sequence (microsatellite) loci showed each *Atmsh2-1* line to have evolved its own “fingerprint,” the results of as many as 10 microsatellite mutations in a single line. Thus, MMR during diploid growth is essential for plant genomic integrity.

[**Keywords:** Plant mismatch repair; AtMSH2; mutational loading; *Arabidopsis*; haplosufficiency quality-checking]

Supplemental material is available at <http://www.genesdev.org>.

Received May 3, 2004; revised version accepted September 10, 2004.

Most organisms deploy multiple systems to maintain their genomes, particularly genome copies bequeathed to progeny. However, resources allocated to accurately replicate DNA and to remove and replace damaged or incorrect bases vary, resulting in different rates of mutation accumulation. Plants face unique obstacles to long-term genetic integrity. They lack reserved germ lines: gametes arise from meristem cells that have already divided many times. During meristem growth and subsequent floral development, DNA integrity is jeopardized by multiple opportunities for replication errors and for DNA damage by environmental mutagens from which plants cannot escape, such as solar UV-B light and genotoxic chemicals, and from endogenous DNA-damaging oxyradicals arising from photosynthesis by chloroplasts and oxygen metabolism in mitochondria. However, even long-lived plants do not show extraordinary mutational loading (Klekowski and Godfrey 1989). Thus, plants appear to employ efficient genomic-fidelity mechanisms. Plants can take advantage of haploidy in gametophytes

and functional haploidy for genes that are paternally silenced during growth early after fertilization to eliminate some cells with deleterious recessive mutations or aberrant chromosomes—“haplosufficiency quality-checking” (Walbot and Evans 2003). However, plants can also prevent mutagenesis during prior diploid growth of meristems and floral tissues. They encode systems similar to those that in most eukaryotes repair DNA-base damage: mismatch repair (MMR), base excision repair, nucleotide excision repair, and photoreactivation (*Arabidopsis* Genome Initiative 2000; Hays 2002). Indeed, for each of these repair systems, plants encode one or more extra proteins not found in most or all other eukaryotes. The relative contributions of haplosufficiency quality-checking and DNA-maintenance systems to plant genetic integrity are unknown at present. We suggest that the primary defense against mutation accumulation is rigorous DNA maintenance, with haplosufficiency quality-checking providing a valuable backup. To test the hypothesis that genome maintenance is indispensable for plant genomic integrity, we used a mutant *Arabidopsis thaliana* line deficient in MMR (Leonard et al. 2003). In all eukaryotes examined, and most prokaryotes, evolutionarily conserved MMR protein systems correct DNA replication errors—apparently includ-

¹Corresponding author.

E-MAIL haysj@science.oregonstate.edu; FAX (541) 737-1777.

Article and publication are at <http://www.genesdev.org/cgi/doi/10.1101/gad.1217204>.

ing those caused by endogenous oxidative damage (Earley and Crouse 1998)—and further maintain genomic stability by responding to certain DNA lesions, antagonizing homoeologous (diverged-sequence) recombination and promoting meiosis (for reviews, see Modrich 1997; Earley and Crouse 1998; Buermeyer et al. 1999; Kolodner and Marsischky 1999; Harfe and Jinks-Robertson 2000; Jiricny and Nystrom-Lahti 2000). Three eukaryotic protein heterodimers, composed of MSH subunits homologous to prokaryotic MutS proteins, recognize DNA mismatches with different but overlapping specificities: MSH2 · MSH3, MSH2 · MSH6, and (only in plants) MSH2 · MSH7 (Culligan and Hays 2000). A second heterodimer, MLH1 · PMS2, couples mismatch recognition to excision of the error-containing nascent DNA. Deficiencies in the constant eukaryotic subunit MSH2 or in prokaryotic MutS proteins thus dramatically increase mutation rates in bacteria, yeast, and mammals (see, for example, Schaaper 1993; Earley and Crouse 1998; Buermeyer et al. 1999; Harfe and Jinks-Robertson 2000). We compared mutation accumulation during seed-to-seed propagation in wild-type and *AtMSH2::TDNA* plants (Earley and Crouse 1998; Leonard et al. 2003), in which T-DNA insertion has disrupted the *AtMSH2* gene. In the absence of other reports of defective *AtMSH2* genes, we designate this allele *AtMSH2-1*. (Inactivation of *AtMLH1* or *AtPMS2* would most likely increase somatic mutation as well, but such deficiencies can severely impact fertility [Buermeyer et al. 1999]).

Although haplosufficiency quality-checking should be similar in wild-type and MMR-defective plants, the 36 *AtMSH2*-defective (*Atmsh2-1*) lines accumulated a variety of abnormalities in only five generations of seed-to-seed propagation. This suggests that MMR during meristem growth and floral development is essential for genetic fidelity. Our goal was not to generate mutant plants per se or to investigate particular phenotypes, but to demonstrate the effects of sustained mutational loading during multigenerational propagation: appearance of apparent single- and perhaps multiple-locus phenotypes and an overall trend toward decreased fitness. These specific lines, and information related to the general pattern of mutation accumulation, may be of value to other investigators.

Results

Accumulation of morphological/developmental mutations

To compare mutational loading—including possible emergence of multilocus phenotypes—in MMR-deficient and MMR-proficient plants, we propagated, in parallel pots, 36 lines of the previously described *Atmsh2-1* mutant, along with 36 wild-type lines. We screened each generation, as 20-d seedlings, for obvious morphological phenotypes, then thinned each pot (without regard to appearance) to four plants, one near each corner, and surveyed these during development of flowers and siliques. We collected and saved seeds from one plant chosen ran-

domly for propagation. If this plant was sterile, the line was scored as extinct, but seeds from all three siblings were saved. Visible mutations began to appear after two generations. At generation five (G5) we quantitatively scored germination of the seeds from each of the (single) G4 plants used to propagate the lines. We further scored survival of the G5 germinants to the thinning (20-d) stage, where we analyzed morphology and development in detail. We again thinned to four G5 plants without regard to appearance, but in addition retained obviously mutant plants in order to save the seeds. Plants were subsequently examined for flower and silique/seed aberrations. We also measured G5 repeat-sequence instability (see below), as described previously (Culligan and Hays 2000; Leonard et al. 2003). As before, a single plant in each line was randomly chosen to seed generation six.

All 36 sixth-generation *Atmsh2-1* lines are compared with representative wild-type seedlings in Figure 1, and phenotypes of G5 plants are compared in Tables 1, 2, and 3. Whereas no abnormalities were seen in any wild-type line, developmental or germination phenotypes were apparent in nearly one-third of the G5 *Atmsh2-1* lines. Since neither developmental abnormalities nor altered fertility were observed in G1 *Atmsh2-1* plants, here or

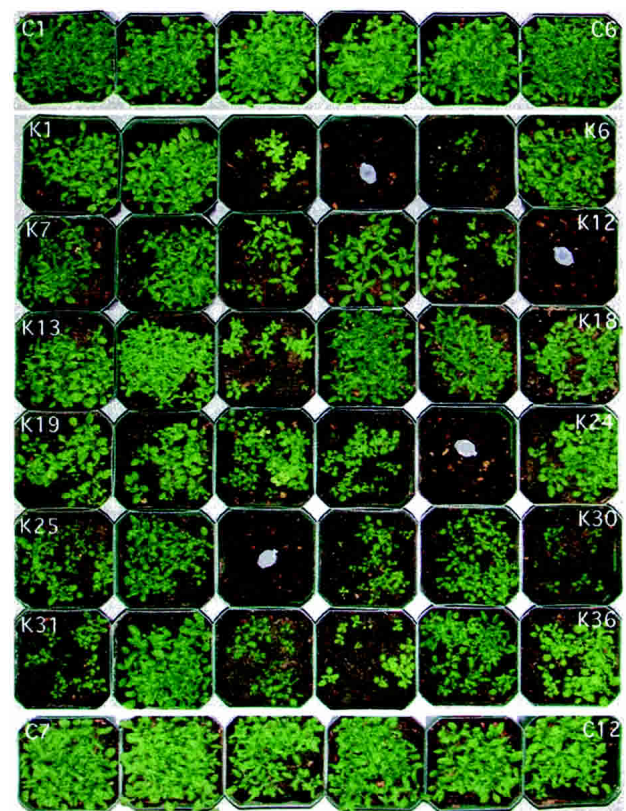


Figure 1. Generation-six (G6) 20-d seedlings. (Top row) Wild-type lines C1–C6 (left to right). (Rows 2–7) *Atmsh2-1* lines K1–K36 (left to right) in rows. (Bottom row) Wild-type lines 7–12. Empty pots mark lines (K4, K12, K23, K27) previously scored as extinct.

previously (Culligan and Hays 2000; Leonard et al. 2003), the phenotypes seen here should mostly reflect de novo mutations, rather than segregation of pre-existing recessives. The empty pots mark four lines (K12, K27, K4, K23) scored as “extinct” at generations G2, G3, or G5, respectively—that is, the single plant chosen randomly for propagation proved to be sterile. All sterile plants have failed to produce or release pollen. Seeds of the three siblings of sterile plants have been preserved; if these prove viable, the four lines can be continued. In other lines (K15, K23, K28), some G5 or G4 plants were sterile (developed no siliques), but by chance these were

not the ones selected for propagation. Table 1 summarizes morphological and other abnormalities. Selected morphology mutants are shown in Figure 2. An atlas of all lines—histories, descriptions, and pictures—is presented as Supplementary File 2. Mutations in lines K21 and K15, respectively, segregate as apparent recessives (3:1): *pale-leaf (plf)*—near albino true leaves (Fig. 2B); *over-eager (ove)*—early multiple flowering bolts, sterility, dwarfism (Fig. 2C). All plants of lines K14 show *crinkled leaves (cr1)* (Fig. 2D). All plants of lines K30 (Fig. 2E) and lines K22, K28, and K29 (data not shown) show stress-like dark-leaf purple-stem/vein phenotypes. The stress-

Table 1. Deficiencies (*X*) of G5 *Atmsh2⁻* plants: sterility (extinction) and/or abnormal germination, development, seed set, or siliques/seeds

Line	Deficiencies					Visible G5 abnormalities
	Extinction ^a	Germination	Development ^d	Seed set	Siliques/seeds ^c	
K1		X		X	X	
K2		X		X		
K3				X		[NS] <i>light green</i> (°)
K4	G5 ^a			X		
K5		X	X	X	X	
K6			X		X	
K7				X		(°)
K8		X		X	X	
K9		X			X	
K10		X	X	X		
K11		X	X		X	
K12	G2 ^a					
K13				X		
K14				X	X	[NS] <i>crinkled leaf</i>
K15		X		X		[S] <i>early flowers, dwarf, sterile^b</i>
K16				X		
K17		X	X			
K18						
K19		X			X	
K20		X		X	X	
K21				X	X	[S] <i>pale leaves</i>
K22			X	X		[NS] <i>stress-like</i>
K23	G5 ^a	X		X	X	
K24			X	X	X	
K25		X			X	
K26			X			
K27	G3 ^a					
K28		X	X		X	[NS] <i>stress-like</i> , [S] <i>sterile^b</i> ; (°)
K29			X		X	[NS] <i>stress-like</i>
K30		X	X		X	[NS] <i>stress-like</i>
K31			X	X	X	
K32		X		X	X	
K33				X		
K34			X	X	X	[NS] <i>light green</i>
K35		X		X		(°)
K36						

Visible abnormalities at generation G5 are briefly described and abbreviated (*italics*), and Mendelian segregation [S], i.e. 24%–29% mutant fraction, or nonsegregation [NS] indicated. Where indicated (°), abnormalities were seen in previous generations but not carried forward.

^aPlant randomly selected for propagation at end of indicated generation was sterile.

^bTo date all sterile plants have failed to produce/release pollen.

^cSee Supplemental Material (Supplementary Fig. S2) for G3 or G4 morphology, growth, germination or sterility phenotypes not seen in G5.

^dSurvival to thinning.

^eSee Table 3, column 5.

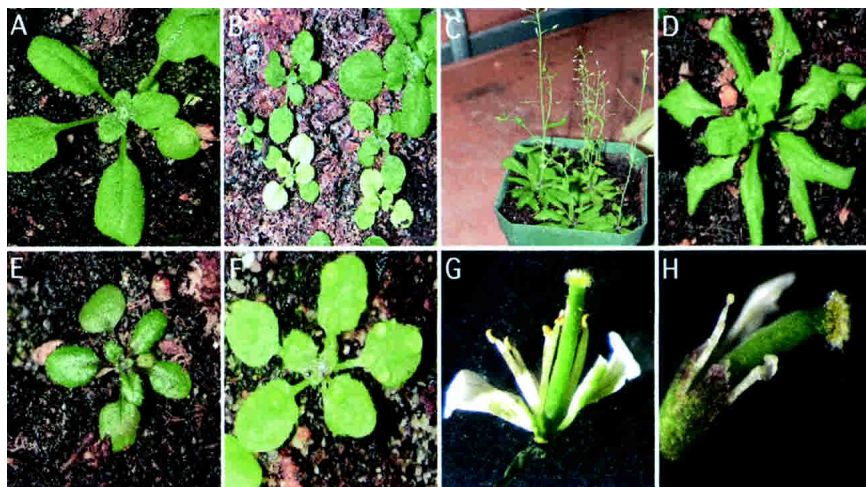


Figure 2. Representative morphology and growth abnormalities. (A) Col-0 wild type. (B) (K21) Segregating (3:1) for *pale leaf* (*plf*). (C) (K15) Segregating (3:1) for early flowering, dwarf, sterile—*overeager* (*ove*). (D) (K14) Nonsegregating for *crinkled leaf* (*cr1*). (E) (K30) Nonsegregating for constitutive stress response, purple veins, dark color. (F) (K34) Nonsegregating for *light-green* (*ltg*). (G) Dissected Col-0 wild-type flower. (H) (K4) Segregating for sterile flowers—*defective anthers* (*dea*).

like phenotypes were also apparent in G6 seedlings, that is, were heritable. Leaves of all plants in lines K34 and K3 are uniformly *light green* (*ltg*) (Fig. 2F). Line K4 segregates for *defective anthers* (*dea*) (Supplemental Material). In line K3, small near-albino plants segregated (3:1) out of the all-light-green background at G3, but (by chance) were not propagated further (Supplementary Fig. S2). Mutations manifested in all G5 plants in a given line were most likely already homozygous at G4.

Of nine morphological/developmental mutants identified at G5, only three showed 3:1 segregation (Table 1; Supplemental Material). All MMR-defective (but not wild-type) plants in a given generation of any line could show a particular phenotype if some unexpected environmental stress, present for the first time at that generation, elicited a response from (1) all (MMR-defective) plants or from (2) all plants in a particular pot, or if a particular line were already homozygous for a new mutation at that generation.

In seven G5 lines (K2, K14, K22, K28, K29, K30, and K34), all plants in the pot showed the same phenotype (K28 segregated for sterility, but *all* its G5 plants showed a stress-like phenotype.) These lines showed different phenotypes: stress-like response by lines K22, K28, K29, and K30 (and no other lines); *pale-green* only by lines K3 and K34 (but noted first at different generations); *crinkled-leaf* only by K14. Thus, explanation 1 does not apply. All of these lines showed the same respective phenotypes at generation six (Supplemental Material), ruling out explanation 2. For explanation 2 to hold, the phenotype of a G4 progenitor already homozygous for a recessive mutation, or at least heterozygous for a rare dominant-negative mutation, would need to have escaped detection. In the case of a heterozygous dominant-negative G4 progenitor, we might have failed to distinguish 75% from 100% phenotypic G5 expression before thinning, then by chance thinned out all phenotypically abnormal plants. While we tried to avoid such omissions, we cannot unequivocally rule them out. A major strength of this approach is that seeds of all generations remain available to investigate the history and etiology of any

mutant lines, specifically including those above that do not show segregation.

We cannot distinguish here between phenotypes that reflect single segregating mutations and more complex genotypes, for example, homozygosity for one or more earlier but cryptic mutations revealed by a final segregating mutation. In lines K3 and K28, a second segregating mutation showed its phenotype in the clearly apparent phenotypic background of a first (homozygous) mutation, but we do not know whether the first is essential for the second phenotype or the two are independent.

We measured efficiencies of (G5) seed germination, and post-germination growth to the stage where plants were thinned (about 20 d), for all *Atmsh2-1* and wild-type lines (Table 2). The mean for wild-type plants was 48 ± 2.5 germinated G5 seeds per pot. The standard deviation (SD) of 2.5 may partly reflect variations in numbers of seeds actually planted, because ungerminated *Arabidopsis* seeds are not readily detected in soil. *Atmsh2-1* germination yields varied widely—from 9% to 102% of the wild-type mean. We arbitrarily designate as abnormal *Atmsh2-1* germination yields that are two or more (wild-type) SDs below the wild-type mean—the bottom 2.5% in a random distribution. Only two of 36 wild-type lines, but 17 of 34 *Atmsh2-1* lines, showed such reductions. Most wild-type germinants (mean $98 \pm 5\%$) progressed to the thinning stage (20 d), and only two fell below 98% by more than two (wild-type) SD. The *Atmsh2-1* mean survival to thinning was ($85 \pm 13\%$) ($p = 0.00002$ vs wild type); 13 of 33 fell two or more (wild-type) SD below the wild-type mean.

Seeds and/or siliques were abnormal in at least 19 *Atmsh2-1* lines (Table 3). Phenotypes included segregating embryonic lethality (Fig. 3B), segregating and apparent late embryonic lethality (whitish seeds) (Fig. 3C), segregating apparent early lethality (Fig. 3D), segregating dark centers (Fig. 3E), abnormal seed colors and shapes (Fig. 3F), abnormal silique sizes and shapes (Fig. 3G), and early loss of seeds (Fig. 3H). Of 19 lines with identified G5 seed/silique abnormalities (Table 3), six clearly showed 3:1 segregation, nine appeared to show the same

Table 2. Germination of wild-type (*wt*) and *Atmsh2*⁻ mutant lines (K1–36) was scored as visible plantlets in each pot

Line	Germination		Survival to thinning	
	Rel. to ^a wt mean	SDs below ^b wt mean	Rel. to ^c wt mean	SDs below ^d wt mean
K1	0.78	-2	0.92	0
K2	0.74	-3	1.06	1
K3	1.02	0	0.91	-1
K4	0.91	-1	0.95	0
K5	0.54	-4	0.76	-4
K6	0.87	-1	0.85	-2
K7	0.91	-1	0.98	0
K8	0.57	-1	0.92	-1
K9	0.46	-1	0.90	-1
K10	0.65	-4	0.87	-2
K11	0.72	-3	0.73	-4
K12	Extinct line			
K13	0.85	-1	1.02	0
K14	0.87	-1	0.98	0
K15	0.61	-4	0.89	-1
K16	0.85	-1	1.00	0
K17	0.78	-2	0.80	-3
K18	1.06	1	0.90	-1
K19	0.83	-2	0.89	-1
K20	0.80	-2	0.95	0
K21	0.96	0	N/D	-
K22	1.00	0	0.78	-3
K23	0.09	-11	1.00	0
K24	0.98	0	0.82	-2
K25	0.67	-4	0.90	-1
K26	0.85	-1	0.79	-3
K27	Extinct line			
K28	0.76	-2	0.43	-10
K29	0.91	-1	0.78	-3
K30	0.57	-4	0.65	-6
K31	0.96	0	0.70	-5
K32	0.80	-4	0.89	-1
K33	0.85	-1	1.05	1
K34	0.85	-1	0.79	-3
K35	0.57	-4	0.96	0
K36	0.91	-1	0.88	-1

Survival-to-thinning ratio is the fraction of plantlets still apparently viable when lines were thinned to four plants for further propagation.

^aRelative to mean number of germinants per wild-type line (48 ± 2.5).

^bDeviation of number of mutant germinants from wild-type mean, in units of standard deviations of wild-type plants (2.5) from wild-type mean (48).

^cRatio of number of apparently viable mutant seedlings at thinning stage (~21 d) to number of germinated seeds, divided by analogous ratio for wild-type plants (0.98 ± 0.05).

^dDeviation of survival-to-thinning ratios of mutant 21-d seedlings from mean wild-type ratio, in units of standard deviations for wild-type plants (0.05) from wild-type mean (0.99).

respective phenotypes in all seeds/siliques, and four (lines K1, K8, K23, and K29) showed more complex phenotypes: intermediate segregation ratios or different ratios for different characteristics. The phenotypes of the G4 progenitors, which might explain some of the G5

phenotypes, were not determined. However, these could be investigated in the future using stored G4 or earlier seeds. The average number of seeds per silique (seed set) for three *Atmsh2-1* G5 plants in each line was reduced below the wild-type mean (51 ± 6) seeds by two or more wild-type standard deviations in half the 34 nonextinct G5 lines. Some lines showed high plant-to-plant variation in seed number—as much as 67% of the (three-plant) mean (Table 3, column 3).

Remarkably, only two *Atmsh2-1* lines (K16, K36) appeared normal at generation G5 by the above criteria (Table 1). Lines K13 and K16 appeared to grow slightly more robustly than wild type. We measured the summed wet weights for 20 plants in each of six pots of these and wild-type plants. The six-pot means, for K13 and K16, 0.834 ± 0.09 and 0.814 ± 0.10 g, respectively, did exceed the wild-type mean of 0.787 ± 0.11 , but the differences were not statistically significant (double-sided *p* values, 0.42 and 0.66, respectively). Thus, MMR deficiency results in spontaneous generation of a wide variety of mutations that affect different developmental pathways and show different patterns of segregation. Seeds of all generations of all lines are freely available to other investigators, to recapitulate development of suspected multi-locus phenotypes, to investigate segregation patterns, or to characterize mutants of interest in detail.

Mutation rates in *Atmsh2-1* plants

Obvious mutant phenotypes first appeared among *Atmsh2-1* plants at generation two, and by G5 ~28 morphological/developmental or seed/silique mutations had been identified (Tables 1, 3; Supplementary File 2). Does the delay in appearance of most mutations simply reflect a relatively low rate of mutation loading, so that even single-hit mutations accumulate gradually, or a relatively rapid mutation rate but mostly multiple-hit phenotypes? First, we note that another seven lines (K2, K7, K10, K13, K16, K26, K33) show quantitative abnormalities, mostly quite marked, in one or more traits—germination, average development rate, seed set (Tables 1, 2, 3; Supplementary File 2). Thus, visible phenotypes reflect only one aspect of mutational loading. To obtain a quantitative estimate of mutational loading, we determined the DNA sequence of two ~1-kbp regions in the (arbitrarily selected) *AtRAD30* gene, and smaller targets from two additional transcribed genes containing A₁₀ mononucleotide runs in each of the surviving 34 G5 *Atmsh2-1* lines, a total of 94,348 bp. The screening/scanning technique (see Materials and Methods, DNA Sequence Analysis) was such that only homozygous mutations (bp changes) could be reliably detected. We found no apparent homozygous mutations, either base substitutions or insertions/deletions. Two targets included A₁₀ runs, expected to mutate relatively rapidly in MMR-defective cells (Tran et al. 1997; Duval and Hamelin, 2002).

What rates of mutation accumulation might be expected, considering the replication-error rates expected for MMR-deficient cells— 10^{-7} to 10^{-6} errors per base pair per duplication, based on DNA polymerase proper-

Table 3. Seed set and seed/silique morphology of *Atmsh2*⁻ lines

Line	Seeds per silique ^a			Seed/silique abnormalities ^c
	Mean	Std. dev.	Relative seed set ^b	
K1	33	±8	0.65	[NS] AS; [S] LL, WS
K2	25	±14	0.50	
K3	34	±8	0.67	
K4	38	±8	0.75	
K5	31	±21	0.61	[NS] AS, LL, multi-color seeds
K6	49	±3	0.97	[S] EL
K7	27	±5	0.53	
K8	29	±13	0.57	[1/2] small seeds, [NS] EL, LL, WS
K9	42	±3	0.83	[S] late-dark seeds
K10	33	±6	0.67	
K11	43	±5	0.85	[S] EL
K12	(extinct)			
K13	29	±16	0.57	
K14	26	±15	0.51	[NS] EL, partial-fill siliques
K15	18	±17	0.36	
K16	30	±6	0.59	
K17	50	±3	0.98	
K18	56	±1	1.10	
K19	40	±2	0.79	[NS] few seeds
K20	29	±7	0.57	[S] LL, WS
K21	36	±8	0.71	[NS] short, fat siliques
K22	38	±5	0.74	
K23	21	±10	0.42	[>1/2] aberrant seeds, LL
K24	23	±10	0.30	[NS] EL, LL, WS
K25	40	±12	0.78	[S] LL
K26	43	±8	0.85	
K27	(extinct)			
K28	44	±12	0.87	[S] EL, LL WS
K29	53	±3	1.04	[>1/2] aberrant seeds, LL, WS
K30	43	±10	0.84	[NS] very few seeds
K31	23	±3	0.45	[NS] AS, LL
K32	19	±4	0.37	[NS] very few seeds
K33	38	±23	0.74	
K34	27	±15	0.52	[NS] very few seeds
K35	29	±9	0.56	
K36	44	±10	0.86	
Wt11	47	±4	0.93	
Wt17	51	±4	1.01	
Wt29	53	±1	1.04	
Wt33	52	±5	1.02	

Silques of 6-wk plants were examined for obvious abnormalities and opened for examination and counting of seeds. Seed set includes abnormal seeds (see Table 2 for germination efficiencies).

^aAverage number of seeds in five siliques was scored for three plants in each line. Means and standard deviations for the three plant averages are shown.

^bMean seeds/silique for each *Atmsh2*⁻ line, divided by mean value for wild-type lines (50.7).

^c[S] segregating; [NS] nonsegregating; (AS) aberrant seeds; (LL) late lethal; (EL) early lethal; (WS) whitish seeds.

ties and analyses of other MMR-deficient organisms? To roughly estimate mutation accumulation rates during plant propagation, we have derived expressions for the probabilities of being homozygous or heterozygous (either homolog) for a mutation at a given base pair (Supplemental Material). After one generation, $Pr\{homozygous\} \equiv \mu = m (s - 2 + f)/16$, and $Pr\{heterozygous\} \equiv \tau = m (s - 2 + f)/8 + m (d_m + h_m + d_f + h_f)/2$, where m is the assumed constant mutation rate per base pair per chromosome duplication, s and f are the average number of apical-meristem divisions and floral-meristem divisions, respectively, in the ancestry of the last non-sexually differentiated floral precursor cell that gives rise to the male and female lines, d_m and d_f are the average number of premeiotic diploid divisions in the male and female lines, respectively, and h_m and h_f are the average number of post-meiotic haploid gametophyte divisions leading to anthers and ovules, respectively. For a sample calculation, we took values likely to be at the middle to high end of the range: $s = 28$, $f = 6$, $(d_m + h_m + d_f + h_f) = 16$. This yielded $\mu = 2m$ and $\tau = 6m$. To calculate the cumulative probability of mutations in zygotes at the end of the g th generation, we developed the formulas $g\mu + \tau [g - 2 + (1/2)^{g-1}]$ for homozygosity and $2\tau [1 - (1/2)^{g-1}]$ for heterozygosity (Supplementary File 1). For the values of μ and τ calculated above, these formulas yield $Pr\{homozygous\} \approx 28m$, and $Pr\{heterozygous\} \approx 12m$. The homozygosity probability would increase to $36m$ for G6. Thus, for m of 10^{-7} to 10^{-6} , the expectation of homozygosity for any mutation after five generations of growth (here G0 through G4), would be 3×10^{-6} to 3×10^{-5} , or one base-substitution mutation in every 30,000–300,000 bp. Unless our assumed numbers of cell divisions in the average gamete lineage are much too low, absence in 94,000 bp of sequence of any base-substitution mutations or insertion-deletion mutations at modest-repeat-length mononucleotide runs is thus consistent with the error rates of 10^{-7} to 10^{-6} seen for replicative DNA polymerases in vitro and mutation rates in other MMR-defective organisms. In an *Arabidopsis* genome of 120 Mbp, there should thus be an average of 400–4000 base-substitution mutations after five generations. The next section describes mutation frequencies observed at long dinucleotide repeat sequences (microsatellites).

Microsatellite instability

We previously observed high rates of dinucleotide-repeat-sequence insertion-deletion mutation (microsatellite instability [MSI]) in *Atmsh2-1* plants (Culligan and Hays 2000; Leonard et al. 2003), similar to rates in MMR-defective mice. Here, we first analyzed MSI at six dinucleotide-repeat loci (12 targets) in one G5 plant selected at random from each of 34 *Atmsh2-1* and 18 wild-type lines. Since decreases or increases in repeat length (shifts) after three or more G5 DNA duplications would not be detectable in bulk G5 DNA (Leonard et al. 2003), these analyses necessarily measure the phenotypes of particular G4 cells that gave rise to G5 seeds. The

Figure 3. Representative silique/seed abnormalities. See Table 3 for a complete listing. For photography through a dissecting scope (10 \times –100 \times), wild-type Col-0 and *Atmsh2-1* lines (K1, K8, K11, K9, K5, K23, K34) were illuminated from above (A–G) or back-lit (H). Siliques in A–E were partially opened by dissection. (A) Wild-type Col-0 silique: uniform seed color and size. (B) (K1) Segregating (average 3:1) for seed lethality. (C) (K8) White seeds (WS) with late lethality (LL). (D) (K11) Segregating (average 3:1) for early lethality (EL). (E) (K9) Segregating (average 3:1) for dark internal pigmentation. (F) (K5) Aberrant seeds (AS)—abnormal colors and shapes. (G) (K23) Aberrant siliques—shorter, fatter, bumpy appearance. Normal silique dimensions are $\sim 2 \times 20$ mm. (H) (K34) Partially filled mature siliques (brown color normal for this back-lighting).



summed cumulative single-plant data for all 34 lines and all six loci (12 alleles) reveal at least one shift since G0 in every *Atmsh2-1* line, and up to 10 in some (Fig. 4, blue bars), but no shifts in any wild-type plants (data not shown). Deletions were roughly twice as frequent as additions (data not shown). In contrast, the first generation (G0) of growth of *Atmsh2-1* and other MMR-defective mutants generated mostly zero, one, or two shifts in the 12 target alleles (Fig. 4, red, orange, yellow, green, blue bars).

We further analyzed the six loci (12 target alleles) in 16 plants from each of two G5 wild-type lines, and each of six randomly selected (<http://www.random.org/inform.html>) G5 *Atmsh2-1* lines (K1, K4, K6, K20, K25, and K29), to distinguish homozygosity from heterozygosity at the respective loci in the G4 plants (Fig. 5). G4 plants initially homozygous for a particular shifted allele would be expected to bequeath mostly this allele to the (16) G5 progeny that we analyzed, plus new shifts that occurred during G4 growth and by chance ended up in the seeds that gave rise to the G5 progeny examined. Thus, G4 baseline homozygosity for either wild-type (zero repeat-unit shifts) or shifted alleles is signified in Figure 5 by long bars indicating 1.0 or slightly lower fractions of the total G5 progeny target alleles (2×16) analyzed for each locus. All lines analyzed except K29 show by this criterion baseline homozygosity for new repeat lengths at up to three loci: NGA6 (–1) in K1; NGA8 (–2) and NGA139 (–2) in K4; NGA8 (+1), NGA151 (–1), and NGA139 (+3) in K6; NGA151 (+1) and NGA6 (–1) in K20; NGA139 (–2), NGA8 (–1), NGA151 (–1) in K25. For all lines tested, one or more of the 16 G5 progeny appeared to have shifted their repeat lengths at these loci up or down from these new baseline lengths during G4 growth. Situations where G4 plants were initially heterozygous for two alleles (two repeat lengths) at a microsatellite locus are more complicated. Without further shifts during G4 growth, Mendelian segregation would yield one each G5

progeny plants homozygous for a baseline allele, and two heterozygous progeny. Thus, 16 of the 32 targets scored for each locus among the 16 plants would show one repeat length and 16 would show the other. Shifts of either baseline allele to other repeat lengths in G4 cells that gave rise to G5 plants would decrease the fraction of this allele below 0.5, but shifts from one baseline repeat length to the other would increase the fraction of the latter to above 0.5. Line K29 shows similar frequencies for two alleles at loci NGA6 (0, –2) and NGA8 (0, +1), as does line K25 [NGA6 (0, –1) and NGA139 (–2, –3)]; other roughly similar allele frequencies suggesting baseline heterozygosity are apparent in Figure 5. No wild-type lines showed new baseline repeat lengths. The apparent frequencies of shifts away from the deduced homozygous or heterozygous baseline repeat lengths during G4 growth (as evidenced by G5 progeny) were used to estimate the rates of G4 \rightarrow G5 shifts (data not shown). The average G4 \rightarrow G5 frequency of unique shifts, for 12 targets (six loci) and 96 plants (six lines), was 25-fold higher than the average for all wild-type experiments (here [G4 \rightarrow G5] and previously [Leonard et al. 2003]), and thus similar to G0 \rightarrow G1 values for *Atmsh2-1* (20-fold), *AtMSH2*-RNAi (18-fold), *AtMLH1::TDNA* (26-fold), and *AtPMS2::TDNA* (10-fold) (Leonard et al. 2003; our unpublished data). Again, no shifts were observed in any wild-type plant in these experiments. Thus, no additional new genome-maintenance defects affecting MSI appear to have arisen during generations G1 through G4 in these six lines. Seeds for entire lines of descent (G1–G5) are available to study microsatellite-expansion/contraction and allelic drift in these isolated rapidly mutating populations.

Discussion

The remarkable profusion of mutations after only six generations of MMR-defective growth (including one

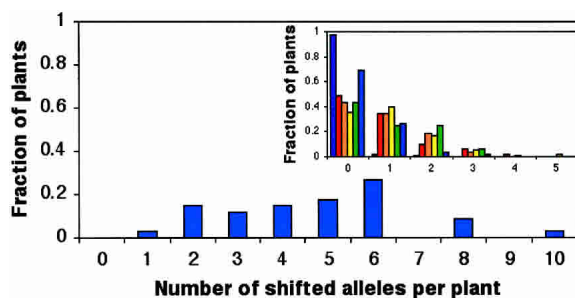


Figure 4. Representative microsatellite genotypes for all *Atmsh2-1* (*AtMSH2::TDNA*) lines. One G5 plant from each line was analyzed for MSI at six loci (12 alleles) and the fraction of these plants showing the indicated numbers of cumulative shifts was determined (dark blue). (Inset) Similar analyses for G0 → G1 shifts of wild type (purple), *AtMSH2-1::TDNA* (red), *AtMSH2-RNAi* (orange), *AtMLH1::TDNA* (green), and *AtPMS2::TDNA* (blue), and for *Atmsh2⁻* G4 → G5 (yellow).

generation [G0] to bulk up seeds from the original mutant line) suggests that MMR during diploid meristem growth and floral development is indispensable for genetic stability in plants and thus for species integrity. Other DNA repair activities are likely to be important when plants are exposed to environmental mutagens. The respective contributions to the high mutation rate of *Atmsh2-1* plants of inability to correct DNA replication errors and failure to avert consequences of miscoding by oxidized DNA bases (Earley and Crouse 1998; Ni et al. 1999; Leonard et al. 2003)—byproducts of the vigorous oxygen metabolism of plants—remain to be determined. Although haplosufficiency quality-checking in gametes and early embryos (Walbot and Evans 2003) seems not to be the single most decisive factor in genetic stability, it undoubtedly provides an additional layer of protection. Indeed, some of the *Atmsh2-1* seed deficiencies/aberrations (Table 3) may reflect unmasking of deleterious recessive maternal alleles, for example lines K8, K32, and K34 (Supplemental Material).

The experiment described here—MMR-defective descent with random choice of progenitor plants—might provide an accelerated model for long-term mutational loading of populations under sustained genotoxic stress, from environmental pollution or global climate change, for example. Our observations suggest that mutants with altered fitness and/or fertility may appear relatively early, especially among self-pollinating species, from segregation of recessive alleles into homozygous-mutant progeny and perhaps occasionally from dominant-negative mutations. In some ecosystems, sustained mutagenic stress might thus alter relative populations of plant species with different genome-maintenance capabilities, affecting as well populations of animals that feed on them.

MMR-defective “mutator” plants may offer advantages for both plant breeding and basic studies of plant genetics. First, a wide variety of mutations is generated. Multiple studies of both forward mutation spectra and specific reversion rates in MMR-defective bacteria,

yeast, and mammalian cells reveal very large increases in rates of loss or gain of one or two units (frameshifts) at mono- or dinucleotide repeats of six or seven or more units, and substantial increases in both A:T → G:C and G:C → A:T transitions, modestly preponderant over all four transversions (for reviews, see Schaaper 1993; Buermeier et al. 1999; Harfe and Jinks-Robertson 2000). Analysis of *coding* sequences only in the *Arabidopsis* genome reveals adenine runs monotonically decreasing from 3743 A₇ runs to 13 A₁₄ runs, and 91 G₇ to two G₁₄ runs. In contrast, the common technique of treating seeds with ethylmethane sulfonate yields almost exclusively G:C → A:T transitions, even when treatments are so drastic as to result in high embryonic lethality after one generation of selfing (McCallum et al. 2000; Colbert et al. 2001). Genes encoding mononucleotide repeat sequences are particularly unstable in MMR-deficient organisms (Tran et al. 1997; Duval and Hamelin 2002). Also, since mutations accumulate gradually during propagation of MMR-defective plants, early-appearing mutations might not be accompanied by as many potentially confounding additional mutations as when plants are heavily treated with chemical or physical mutagens. This may be important when alteration of a multiple-locus trait is desired, because mutagen treatment sufficient to introduce the necessary multiple mutations might increase lethality to unacceptable levels. Furthermore, mutants with improved fitness may arise before loads of deleterious mutations become too high. Line 13 shows a trend to increased wet weight (6% more than wild type) that, although not statistically significant here, would be economically advantageous in a crop plant if reproducible. Generation-to-generation accumulation of mutations engendering multilocus phenotypes (in line K3, for example) is accessible to study; it can be interrupted by crossing to an MMR-proficient line or recapitulated with stored seeds. Similarly, the frequencies

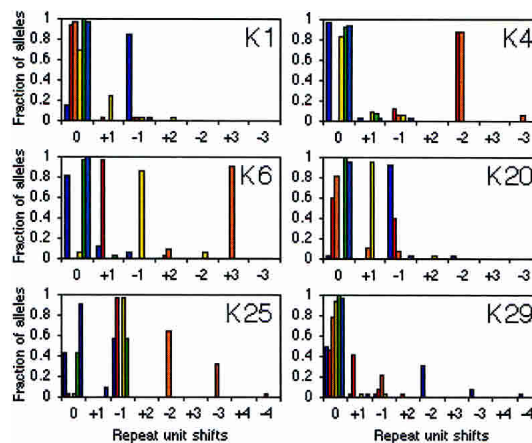


Figure 5. Microsatellite analysis of G5 *Atmsh2-1* plants. Microsatellite “fingerprints” for indicated *Atmsh2-1* lines: fractions of the 32 alleles (16 plants) that are shifted by the indicated numbers of dinucleotide repeat units relative to G0 *Atmsh2-1* plants (purple, NGA6; red, NGA8; orange, NGA139; yellow, NGA151; green, NGA172; blue, NGA1107).

of microsatellite polymorphisms available as mapping markers could be increased by MMR-deficient growth. Investigators using these and other MMR-defective plants (for example, *AtMLH1::TDNA* [J.M. Leonard, A.C. Torres, and J.B. Hays, unpubl.] and *AtPMS2::TDNA* [A.C. Torres, J.M. Leonard, and J.B. Hays, unpubl.]) will need to avoid multiple generations of mutation-prone propagation. We suggest that MMR-deficient mutants be outcrossed with wild-type plants before study of new mutations of interest.

In a previous study with *Caenorhabditis elegans*, cultures of MSH6-deficient animals were reported to readily segregate out visible mutants, and analyses of mutations in endogenous and transgenic reporter genes revealed the expected base substitutions and mononucleotide-run frameshifts, as well as MSI (Duval and Hamelin 2002; Tijsterman et al. 2002). In another study, multigeneration propagation of parallel MSH2-deficient *C. elegans* subcultures resulted in relatively rapid extinctions, accumulation of microsatellite shifts, and elevated reversion of a particular (dominant) point mutation (Degtyareva et al. 2002). In *Arabidopsis*, only a few of the thousands of seeds produced by a single plant need be viable, so emergence of a wide variety of developmental and fertility phenotypes over many generations may be observed. Not only can mutants of interest be recovered from stored seeds, but the mutation-accumulation process itself can be recapitulated and studied at will.

Materials and methods

Propagation of plant lines

We grew all plants in 64-cm² pots containing potting soil, 18 pots per flat, under cool-white lamps (PAR 70 $\mu\text{mol m}^{-2}\text{sec}^{-1}$). *Atmsh2-1* lines (K1–K36) were begun with seeds bulked up (generation G0) from a Columbia-0 line from the Salk Institute (Salk_002708), which was homozygous for a T-DNA insertion in (*AtMSH2::TDNA*) and showed MSI (Leonard et al. 2003). We designate this insertion mutation *Atmsh2-1*. Col-0 wild-type lines were similarly initiated from a single plant. For each generation of each line, we manually counted out 50 seeds, thoroughly mixed them with ~0.5 mL dry sand, and uniformly spread them on the surface of a single pot. After vernalization at 4°C for 48 h, plants were grown at 21°C under a 16/8-h light/dark cycle. At day 10 after germination, pots were examined under a magnifying glass to score morphological abnormalities, and plants of interest were photographed. Germination was scored as any visible cotyledon development, by different observers on successive days. At day 15 of generation five (G5), each individual *Atmsh2-1* pot was photographed alongside a wild-type pot. At day 21, plants were thinned out randomly—neither favoring nor disfavoring obvious mutants (but see below)—such that one plant remained near each corner of the pot (typically *not* the outermost plants), and the number thinned was recorded for comparison to the number of germinants. (At G5, the mutant plants of interest were additionally retained in the interior of the pot, for eventual harvest of seeds.) The four corner plants—designated NW, NE, SE, and SW—were grown to day 28, when one corner selected by random lot and the corresponding plant in each pot was used for further propagation. If in

any pot that plant proved sterile, the line was scored as extinct, although seeds from all siblings were harvested and stored.

Silique and seed morphologies and seed set

At day 35, we harvested from all *Atmsh2-1* lines siliques 8–12 (counting from the bottom) from the primary inflorescences of the three final plants not used for propagation, and similarly harvested siliques from four wild-type lines selected at random. Siliques were examined for morphological abnormalities and photographed where warranted, before and/or after manual dissection. The numbers of seeds per silique (seed set) were scored, and subnormal seeds were noted and photographed.

Purification of plant DNA

DNA was purified from plant tissues essentially as described (Draper and Hays 2000). The procedure involves grinding ~100 mg tissue to a fine powder in liquid N₂, proteinase K digestion in 400 μL buffered high-salt sodium dodecyl sulfate, extraction with phenol-chloroform with the aid of Phaselock-Heavy (VWR), and isopropanol precipitation and washing, yielding about 1 μg DNA in 200 μL buffer.

Analysis of MSI

Samples (~5 ng) of plant DNA (see Purification of Plant DNA above) were analyzed as previously described (Leonard et al. 2003). Briefly, six dinucleotide microsatellite loci (NGA6, NGA8, NGA139, NGA151, NGA172, and NGA1107) (18) were PCR-amplified using one fluorescently labeled and one unlabeled primer for each and product yields estimated by agarose-gel electrophoresis and ethidium staining. Multiplex mixtures containing roughly equal amounts of product from each of the six reactions were analyzed by capillary gel electrophoresis and quantitative fluorescence detection using an ABI Prism 3100 Genetic Analyzer and associated software (Applied Biosystems). Electrophoretic profiles for each locus for each sample were used to detect increases or decreases (shifts) in the number of dinucleotide repeats.

In one experiment, MSI in one plant from each of the 34 nonextinct G5 lines was analyzed (Fig. 4). In another experiment, six G5 *Atmsh2-1* lines were randomly chosen for the standard MSI analysis (K1, K4, K6, K20, K25, and K29). We analyzed 16 progeny plants from each line and determined for each of the six loci the fraction of the 32 allele targets showing a particular repeat number (expressed as “shifts” of one or more dinucleotide repeat lengths observed in G0 wild-type plants). Each line had apparently established new “baseline” homozygous or heterozygous genotypes at various loci (Fig. 5) in the G4 plant that seeded the G5 progeny. Using this assumption, we estimated the respective frequencies of further G4 → G5 shifts away from the baseline genotypes by the PCR-electrophoresis analysis described above. These estimates agreed well with previous measurements of G0 → G1 shift frequencies (see text).

DNA sequence analysis

DNA was extracted from the 34 *Atmsh2-1* lines remaining at G5 as described above under Purification of Plant DNA. DNA from the following genes was amplified by PCR (primer pairs, product size indicated): *AtDINB* (primers F2/R2, 1180-bp product and F4/R4, 934-bp product); *At1g16060.2*-putative ovule development gene, containing an A₁₀ run (primers OvF/OvR, 584-bp product); *At1g19350.3*, putative brassine steroid-signaling positive regulator, containing an A₁₀ run (primers BrF/BrR, 394-

bp product). Primer sequences were as follows: F2, CAAA GATTGAGCATATGCGTGATCG; R2, CGAATGTCATGAC AGTTGATCGGTC; F4, GGAAATGGTCTACAGATCGGTG GAATT; R4, AAACAGGAAGTTCTGCCTTCAACAGCTT; OvF, GGTCAGTCTATACATACACACA; OvR, AGGTGAAG TAGCAAAAATGTTTGG; BrF, CACTGTAAGTTACGTACA CCACC; BrR, CCGCTCTCTTCTCCGATTGTTT.

PCR amplification using 2.5 units *Taq* polymerase and standard 20- μ L reaction mixture containing 0.2- μ m primers, 200- μ m dNTPs, 1.5 mM MgCl₂, 50 mM KCl, and 10 mM Tris-HCl at pH 8.4, employed 5-min initial denaturation at 94°C, followed by 30 cycles of 94°C denaturation (30 sec), 55°C annealing (30 sec), and 72°C extension (90 sec), followed by 5 min at 72°C. Product yields were estimated by agarose-gel electrophoresis of aliquots, and product mixtures were purified using a Qiagen PCR Purification Kit. For Big-Dye thermal-cycle sequencing using an ABI 3730 apparatus, the OSU Central Services Laboratory used 1 μ L of product with one PCR primer and the reaction conditions recommended by the manufacturer. A total of 94,348 bp were sequenced (F2/R2, 40,120; F4/R4, 31,756; OvF/OvR, 18,632; BrF/BrR, 3840).

Acknowledgments

We thank John Bowman, Andrew Buermeyer, Kevin Culligan, David Earhardt, Adela Torres, and Virginia Walbot for careful reading of the manuscript and suggestions. This work was supported by NSF grants MCB-0078262 and MCB-0345061 to J.B.H.

Note added in proof

A group at Morphotek Inc., using a different strategy to inhibit MMR in *Arabidopsis*, also observed a strong mutator phenotype (Q. Chao, C.D. Sullivan, J.M. Getz, K.B. Gleason, P.M. Sass, N.C. Nicolaides, and L. Grasso, in prep.).

References

- Arabidopsis* Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- Buermeyer, A.B., Deschenes, S.M., Baker, S.M., and Liskay, R.M. 1999. Mammalian DNA mismatch repair. *Annu. Rev. Genet.* **33**: 533–564.
- Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., MacCallum, C.M., Comai, L., and Henikoff, S. 2001. High-throughput screening for induced point mutations. *Plant Physiol.* **126**: 480–484.
- Culligan, K.M. and Hays, J.B. 2000. *Arabidopsis thaliana* MutS-homolog proteins—atMSH2, atMSH3, atMSH6, and a novel atMSH7 protein—form three distinct heterodimers with different specificities for mismatched DNA. *Plant Cell* **12**: 991–1002.
- Degtyareva, N.P., Greenwell, P., Hofmann, E.R., Hengartner, M.O., Zhang, L., Culotti, J.G., and Petes, T.D. 2002. *Caenorhabditis elegans* DNA mismatch repair gene *msh-2* is required for microsatellite stability and maintenance of genome integrity. *Proc. Natl. Acad. Sci.* **99**: 2158–2163.
- Draper, C.K. and Hays, J.B. 2000. Replication of chloroplast, mitochondrial and nuclear DNA during growth of unirradiated and UVB-irradiated *Arabidopsis* leaves. *Plant J.* **23**: 255–265.
- Duval, A. and Hamelin, R. 2002. Mutations at coding sequences in mismatch repair deficient human cancers: Toward a new concept of target genes for instability. *Cancer Res.* **62**: 2447–2454.
- Earley, M.C. and Crouse, G.F. 1998. The role of mismatch repair in the prevention of base pair mutations in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **95**: 15487–15491.
- Harfe, B.D. and Jinks-Robertson, S. 2000. DNA mismatch repair and genetic instability. *Annu. Rev. Genet.* **34**: 359–399.
- Hays, J.B. 2002. *Arabidopsis thaliana*, a versatile model system for study of eukaryotic genome-maintenance functions. *DNA Repair* **64**: 1–22.
- Jiricny, J. and Nystrom-Lahti, M. 2000. Mismatch repair defects in cancer. *Curr. Opin. Genet. Dev.* **10**: 157–161.
- Klekowski, J.E.J. and Godfrey, P. 1989. Aging and mutation in plants. *Nature* **340**: 389–391.
- Kolodner, R.D. and Marsischky, G.T. 1999. Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.* **9**: 89–96.
- Leonard, J.M., Bollmann, S.R., and Hays, J.B. 2003. Reduction of stability of *Arabidopsis* genomic and transgenic DNA-repeat sequences (microsatellites) by inactivation of AtMSH2 mismatch-repair function. *Plant Physiol.* **133**: 328–338.
- McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. 2000. Targeted screening for induced mutations. *Nat. Biotechnol.* **18**: 455–457.
- Modrich, P. 1997. Strand-specific mismatch repair in mammalian cells. *J. Biol. Chem.* **272**: 24727–24730.
- Ni, T.T., Marsischky, G.T., and Kolodner, R.D. 1999. MSH2 and MSH6 are required for removal of adenine misincorporated opposite 8-oxoguanine in *S. cerevisiae*. *Mol. Cell* **4**: 439–444.
- Schaaper, R.M. 1993. Base selection, proof reading, and mismatch repair during DNA replication in *Escherichia coli*. *J. Biol. Chem.* **268**: 23762–23765.
- Tijsterman, M., Pothuf, J., and Plasterk, R.H.A. 2002. Frequent germline mutations and somatic repeat instability in DNA mismatch-repair-deficient *Caenorhabditis elegans*. *Genetics* **161**: 651–660.
- Tran, H.T., Keen, J.D., Krickler, M., Resnick, M.A., and Gordonin, D.A. 1997. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol. Cell. Biol.* **17**: 2859–2865.
- Walbot, V. and Evans, M.M.S. 2003. Unique features of the plant life cycle and their consequences. *Nat. Rev. Genet.* **4**: 369–379.