AN ABSTRACT OF THE DISSERTATION OF

Buck W. L. Wilcox for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on March 13, 2012.

Title: Effects of DNA Mismatch Repair Inhibition in *Arabidopsis thaliana*.

Abstract approved:

John B. Hays

Genomic instability underlies diseases of unregulated cell growth that result in cancers and developmental abnormalities in humans. Similar genome destabilizing mechanisms are used to create genetic variety in crops for use in breeding and trait development. Errors that occur during DNA replication may cause mutations if they are not corrected before further cell divisions. DNA mismatch repair (MMR) corrects misinsertions and insertion/deletion DNA loop-outs that arise during DNA replication in plants, animals, prokaryotes, and some archaea, all of which incur mutations at rates 100 to 1,000-fold greater when subjected to inherited or somatic-mismatch repair deficiencies. An understanding of the effects of mismatch repair on somatic and germ-line cells in *Arabidopsis thaliana* is

critical to the development of this plant as a model system for the study of genomic instability. Insertions and deletions of multiples of two base pairs in dinucleotide repeat sequences (microsatellites) occur more frequently in the absence of mismatch repair, and the mismatch-repair status of an individual, tissue, or cell may be inferred on the basis of microsatellite mutation frequency. Single-template PCR analysis measured microsatellite mutation frequencies in leaves and shoot-apical-meristem stem cells, and allowed me to address for the first time an important question: Do plants relax mismatch repair in vegetative tissues relative to meristematic germ-line and floral tissue? Analyses of four microsatellite loci in mismatch repair-deficient and wild type plants surprisingly suggest that there is little difference in mismatch repair activity between leaves and seeds. Mismatch-repair-deficient leaves displayed only two-fold higher microsatellite mutation frequency compared to wild type, and wild-type leaves also displayed a two-fold higher microsatellite mutation frequency compared to shoot-apical-meristems. The high frequency of microsatellite mutation in these wild-type tissues is unexpected, and it suggests that plants relax mismatch repair in differentiated tissues while maintaining genetic fidelity in a small set of stem cells in the shoot apical meristem (SAM). Genome sequencing of *msh2*$^{-/-}$ mutation accumulation *A. thaliana* lines provides an estimated germ-line mutation rate of 3.9×10^{-7} in the absence of mismatch repair. Comparison of the rates of base substitution mutation per chromosome in mismatch repair-deficient plants with rates reported for wild-type plants suggests mismatch repair is more efficient on

chromosome 5 than on chromosomes 1-4. Bias towards G:C → A:T mutations among transitions is maintained but increased nearly 100-fold in the absence of mismatch repair.

©Copyright by Buck W. L. Wilcox

March 13, 2012

All Rights Reserved

EFFECTS OF DNA MISMATCH REPAIR INHIBITION IN *Arabidopsis thaliana*

by
Buck W. L. Wilcox

A DISSERTATION

Submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented March 13, 2012
Commencement June 2012

Doctor of Philosophy dissertation of Buck W. L. Wilcox presented on March 13, 2012.

APPROVED:

Major Professor, representing Molecular and Cellular Biology

Director of the Molecular and Cellular Biology Program

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Buck W. L. Wilcox, Author

ACKNOWLEDGEMENTS

I would like to acknowledge and extend my sincere gratitude to my advisor Dr. John Hays for sharing his passion for research and for providing the opportunity to study plant genetics, mutation, and mismatch repair. I would like to thank my committee members, Drs. Michael Freitag, Andrew Buermeyer, Tom Wolpert, and Daniel Sudakin for their guidance and feedback throughout the development of this dissertation. I am grateful for Drs. Robert Mason and Lynda Ciuffetti for graciously providing teaching assistantships that allowed me to fully experience the rewards of teaching and to develop a teaching style.

I owe much gratitude to Christopher Sullivan, Dr. Todd Mockler, Dr. Brian Knaus, Sanjuro Jogdeo, Erin Bredeweg, and Kyle Pomraning for their invaluable advice and patience in helping me to use the software they graciously provided, and for help in using the top-level computational tools maintained in the Center for Genome Research and Bioinformatics at Oregon State University. I want to thank Dr. Walt Ream, Larry Hodges, and Dr. Vidyasagar Sathuvalli for the useful, lively discussions and for graciously sharing access to their equipment. I extend my appreciation of the past and present members of the Hays lab, particularly Dr. Marc Curtis and Peter Hoffman, for the useful discussions and for sharing their knowledge and expertise. Thank you, Anantnoor Brar, for all of your assistance in the lab and for the help and encouragement through the preparation of this document. The inspiration, support, and camaraderie of my friends and

fellow students in the Molecular and Cellular Biology Program and its affiliated departments is heartwarming, encouraging, and much appreciated.

Finally, I would like to express my heartfelt gratitude for my family and thank them for their encouragement, never-ending love, and support. Thank you, Mother, for your unwavering optimism; thank you, Father, for teaching me the value of utilitarianism; thank you, Jesse, for your determination; and thank you, Margaret, for always being there and for the hours you spent assisting me in the laboratory.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Table | Page

Effects of DNA Mismatch Repair Inhibition in *Arabidopsis thaliana*

1 Introduction

DNA mismatch repair is highly conserved; the system detects premutagenic DNA lesions and removes 99% of uncorrected DNA polymerase misinsertions. Mismatch repair lowers the base substitution rate per round of replication from 10^{6}-10^{7} to 10^{9}-10^{10} [1]. Mismatch repair protects humans and mice from increased mutation rates and cancer risks [2], and genetic manipulation of the system in agricultural crops might lead to the generation of desirable mutations [3,4,5]. Advances in the field of mismatch repair research benefit medicine, agriculture, and genetic sciences. *In situ* measurement of the effects of mismatch repair is key to developing *Arabidopsis thaliana* as a DNA-repair model, but existing mutation reporters cannot detect rare somatic mutations, or they require insertion of a different reporter gene construct into the plants genome for each mutation pathway to be measured [6,7,8,9]. Two reporter constructs are necessary to determine the rate of transitions – one reporter to measure G:C → A:T point mutations and another to measure A:T → G:C – and four are required to measure the four transversion mutations. Somatic microsatellite mutation frequencies have been determined by measuring microsatellite mutation in plants containing transgenic mutation reporters, and germ-line microsatellite mutation frequencies have been determined by measuring endogenous microsatellite mutation among siblings in a progeny analysis. Expected germ-line mutation frequencies in wild-

type and mismatch repair-defective plants, paired with curiously small gains in somatic mutation frequency in mismatch repair-deficient leaves, suggest that plants relax mismatch repair in their leaves while maintaining genomic fidelity across generations [1]. Plants do not contain a reserved germ line as in animals, where the fate of cells destined to undergo meiosis is determined early in development. Instead, plants determine their germ line (beginning with the megaspore mother cells that undergo meiosis to create haploid gametophytes) at the end of sporophyte development [10,11,12]. The true germ line in plants is the haploid gametophyte, but there is a loose lineage of cells in the sporophyte shoot apical meristem that have equal, fractional probabilities of contributing to gamete production. This small group of cells may exhibit increased mismatch repair activity, compared to most of the sporophyte, to prevent mutations that occur during somatic growth from accumulating across generations (mutational loading). An alternative model suggests that mutation loading is prevented when deleterious mutations in genes necessary for gametophyte development and growth that have accumulated during somatic growth of the sporophyte are lost during haploid growth in the absence of a complementing functional allele [13]. Herein I apply to plants a tissue-specific approach used in animals to detect rare mutations to determine if shoot apical meristems (SAMs) harbor cells that exhibit lower mutation frequencies than the more differentiated leaf tissue. I also report the first estimate of the germ-line base substitution rate in mismatch repair-deficient eukaryotes based on whole-genome sequence data. Base substitutions occur 100-

fold more frequently, and they are more uniformly distributed within the genome of mismatch repair-deficient Arabidopsis mutation accumulation (MA) than in wild-type MA lines.

1.1 Mutation

The diversity of life arises from genomic alterations in individuals that generate the genetic variation that drives species evolution. DNA damage response (DDR) and damage-avoidance mechanisms counter mutational processes to ensure tissue homeostasis and accurate passage of genetic material between generations [14]. Mutations result from both endogenous and exogenous insults to the genome. Accurate transmission and propagation of traits through various reproductive modes is central to the idea of species as it relates to organisms. This fidelity is countered by mutation, which is implicated in precipitating aging, cancer, genetic diseases, and it underlies genetic variation. Understanding the balance between DNA repair and mutation is important for a complete understanding of human disease, genetics, and the breeding of organisms for desirable traits.

A significant internal source of mutation is spontaneous deamination of DNA bases creating abasic sites that must be repaired by base excision repair (BER). BER is also responsible for preventing mutations arising from a wide variety of oxidized and alkylated bases. Additional endogenous sources of mutation are DNA polymerase errors and reactive nitrogen and oxygen species generated by intracellular signaling, metabolic reactions, or incident highly energetic particles.

Radiation causes a large proportion of DNA damage when high-energy particles bombard DNA or when they pass though tissue, creating reactive chemical species that in turn damage DNA. UV radiation directly damages DNA, creating cyclobutane pyrimidine dimers and 6-4 pyrimidine-pyrimidone photoproducts. These lesions are reversed via direct reversal by photolyases, and they are removed by nucleotide excision repair (NER). If they impede DNA replication they may be bypassed by translesion DNA synthesis (TLS) which is often mutagenic [15,16]. DNA containing double-strand DNA breaks (DSBs) is repaired by either homologous recombination (HR) or non-homologous end-joining (NHEJ), dependent upon cell cycle. Rejoining DNA ends separated by DSBs by homologous recombination requires a sister chromatid to template repair, and thus chromosome breaks incurred in the G1 phase between cycles of DNA synthesis are repaired by NHEJ.

1.2 Mismatch repair

Mismatch repair provides the fundamental roles of error checking, correction, and activating DNA damage signaling pathways, focused on repairing single base mismatches and short insertion and deletion loop-outs that arise during DNA replication and homeologous recombination (recombination that occurs between DNA that is not perfectly complementary). Homologs of the core set of mismatch repair genes are found in most species, and their functions are well conserved. *Escherichia coli* mismatch repair uses a unique form of nascent-strand recognition – GATC hemimethylation – not found in most species. In

eukaryotes, diversification of MutS and MutL led to two families of proteins and allowed subfunctionalization and the creation of specialized proteins with varying and complementary lesion substrate specificities and cell cycle-specific regulation.

In *E. coli*, mismatch repair is dependent on the functions of the MutS and MutL homodimers and a nicking endonuclease MutH. MutS recognizes non-canonical DNA nucleotide base paring (mismatches) as well as short single-stranded bulges or loops in the helix that result from 1-3 base pair insertions or deletions (indels) that arise primarily during DNA replication and have escaped DNA polymerase error checking and correction. MutL then binds the MutS·lesion complex and signals MutH incision of the nascent strand. The nascent strand and MutH incision site (5′- or 3′- to the lesion) is the unmethylated strand in the hemimethylated palindrome GATC. Adenine methylation in GATC sequences does not occur immediately following DNA replication in *Escherichia coli*. Instead, for a brief period of time following DNA replication, only the parental strand contains GmeATC sequences. Excision of the nascent strand beginning at the nick site through the lesion to an upstream nick creates a ssDNA gap up to ~1000 bases in length that is then filled to create dsDNA by extension of the 3' end through the previously damaged area by the replicative (high fidelity) DNA polymerase.

In plants, as in most other eukaryotes, the MutS homodimers are replaced by heterodimers of the MSH (MutS homolog) family of proteins. They interact with heterodimers composed of MLH (MutL homolog) and PMS (Post Meiotic

Segregation) proteins that replace MutL. Plants contain MSH1, MSH2, MSH3, MSH4, MSH5, MSH6, and MSH7 (Figure 1.1). MutSα (MSH2·MSH6) and MutSγ (MSH2·MSH7) differ in their base-base mismatch recognition specificities, and MutSα binds single nucleotide loop-outs. MutSβ (MSH2·MSH3) is specialized to recognized all loop-outs up to about 10 base pairs (bp) in length. Latent endonuclease activity of MutLα (MLH1·PMS2 in plants), which results in strand incision from a preexisting nick, is activated upon interaction with MutSα, PCNA (likely providing strand discrimination), and the RFC clamp loader [15,17]. The strand-recognition signal in eukaryotes and many prokaryotes has not been identified, but it is likely associated with 5′- and 3′-ends of daughter strand DNA created during replication, and the signal is conveyed to MSH proteins by PCNA clamp proteins. MSH4 and MSH5 are specific to meiosis, and MSH1 functions in mitochondrial and plastid DNA metabolism.

Aberrant proteins overexpressed from certain mutant alleles of mismatch repair genes prevent mismatch correction. Expression of the 134 N-terminal amino acids of the 862 amino acid human PMS2 protein, in the presence of active, wild-type PMS2, causes an increased frequency of mutations associated with mismatch repair deficiency [18]. The dominant negative mutator phenotype was also induced in plants with overexpression of the human PMS-134 fragment (Chao 2005). Dominant-negative mutants of Msh2 have also been described [19]. It is unclear how these mutant proteins inhibit mismatch repair, but understanding of this phenomenon may be beneficial to development of crop breeding techniques.

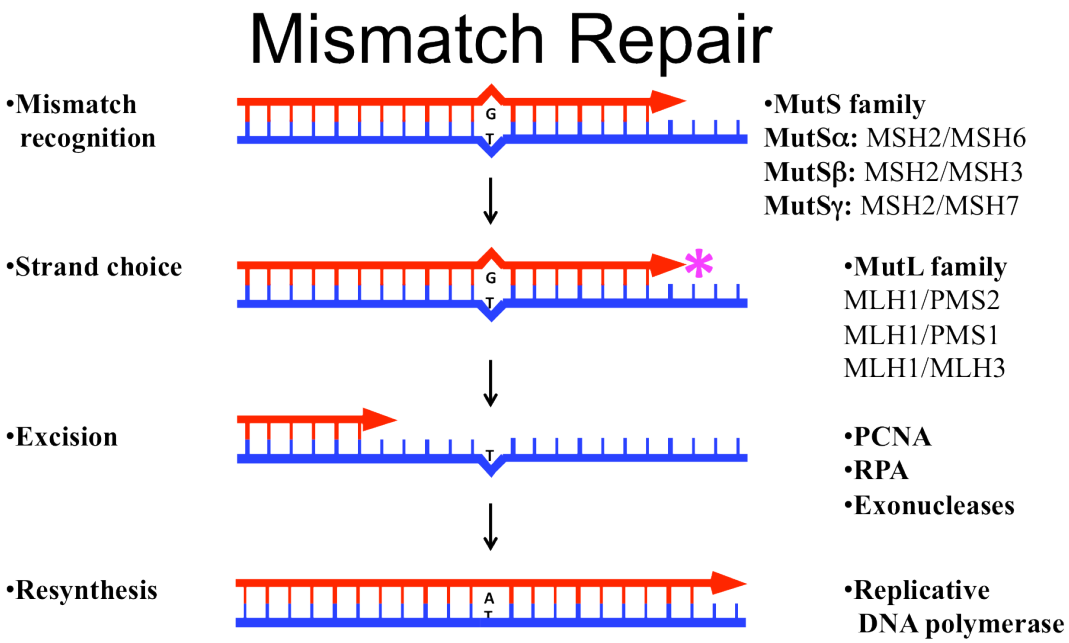

Figure 1.1. Eukaryotic DNA mismatch repair corrects mismatched bases that arise during DNA replication. Mismatches are recognized by heterodimers MutS α, β, and γ, which display differential affinities for various mismatched substrates. Excision of the nascent strand from a preexisting nick is dependent on MutL heterodimer interacting with MutS heterodimer, the PCNA clamp loader, the single-strand binding protein RPA, and exonucleases. The error is corrected when DNA polymerase fills in the remaining gap.

Dominant-negative proteins may bind to lesions but inhibit mismatch repair by lacking competency to commit to repair progression, or they may sequester factors such as Msh3 and Msh6 necessary for mismatch repair activity. Mutant yeast Msh2 and *E. coli* MutS protein amino-acid substitutions in either the helix-turn-helix domain, which mediates changes in interactions between the two subunits of heterodimers or homodimers, or in the ATPase domain lower mismatch repair activity when overexpressed [20,21].

1.3 *Arabidopsis thaliana*

Arabidopsis thaliana (L.) Heynh is a model angiosperm well suited for genetics and genomics research because it has a short life cycle, many progeny, is physically small with minimal growth requirements, and it has a relatively small diploid (1C ≈ 125 Mbp) nuclear genome. Genetic transformation is efficiently mediated by *Agrobacterium tumefaciens* [22,23,24,25]. The sequence of the Arabidopsis genome was published in 2000. There are extensive mutant collections including a T-DNA knockout library that covers most of the roughly 27,000 protein coding genes, data repositories, and computational tools available through TAIR (The Arabidopsis Information Resource, http://arabidopsis.org) [26].

1.3.1 Shoot apical meristem development

The shoot apical meristem (SAM) in plants is the self-replenishing source of undifferentiated stem cells [27]. As cells differentiate and become part of organ primordia and tissues, the stem cells divide to maintain a constant stem cell

population. In addition to stem cells in the shoot apical meristem that are progenitors of cells in stems, leaves, and flowers, plants maintain a population of stem cells in the tip of the growing root that are the progenitors of the cells in the roots. Animal stem cells, too, divide infrequently but indefinitely, and they replenish an assortment of differentiated cell lineages that have limited capacity for division. Whereas stem cells are capable of becoming a variety of cell types, the differentiated daughter cells are restricted in the type of cell they may become. Stem cells in plants are similar in that they divide infrequently, and they are the progenitors of all other cells in the plant. However, unlike animals, plants are unique in that they can continuously grow new organs season after season, and some species grow for over one hundred years. In *A. thaliana*, the mature shoot apical meristem contains about 35 stem cells defined by *CLAVATA3* expression in the central zone (CZ) (Figure 1.2), where CLAVATA3 has been shown to be inhibitory to cell division rate [27,28]. In mature seeds, it is estimated that there are about two stem-cell initials in the embryo that will eventually give rise to all gametes [12,29]. The number of so-called permanent stem cells is an estimate of what is called the genetically equivalent cell number (GECN) as estimated by mutant sector analysis. Chemical mutagens, often combined with ionizing radiation, were applied to mature seeds,and yellow sectoring resulting from mutations affecting chlorophyll biosynthesis was scored in the mature organs of the mutagenized plants and in their progeny (Figure 1.3). If a seed has one cell in the mature embryo that will divide to give rise to all gametes, and that one cell

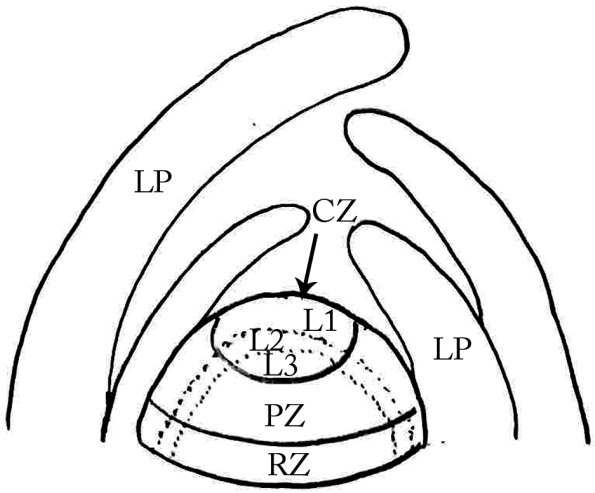

Figure 1.2. Arabidopsis shoot apical meristem. The zones of the shoot apical meristem are divided into three layers – L1 is the epidermal layer. The central zone (CZ) is the most apical group of cells, and it contains permanent and transiently amplifying stem cells that express CLAVATA3. The peripheral zone (PZ) contains rapidly dividing cells that differentiate to form leaves. The rib zone (RZ) is the basal layer of the shoot apical meristem, and divisions of cells in this region contribute to growth of the stalk. Leaf primorida (LP) are built in the meristem, and expand primarily by cell enlargement.

incurs a recessive heterozygous mutation, then the wild-type phenotype and the recessive mutant phenotype will segregate 3:1 in the progeny. In Arabidopsis mutation-sector analysis yielded about a 7:1 segregation of the recessive yellow sector phenotype indicating that two cells are the progenitors of all gametes – it is improbable that both cells would incur similar heterozygous mutations affecting chlorophyll, thus the 7:1 segregation pattern emerges from an undamaged cell whose progeny are all wild-type and a second cell with a heterozygous mutation that segregates 3:1. In determinate plants, like Arabidopsis, the stem cells in the shoot apical meristem will divide indefinitely, but once the shoot apical meristem adopts floral meristem identity, the stem cells in the now floral meristem are restricted to a finite number of divisions which result in the formation of microspore mother cells and megaspore mother cells, and eventually the male and female gametophytes, respectively. Alternatively, in indeterminate plants, the number of new organs and the number of cell divisions within meristems that produce these organs is unlimited.

If mutations occur in stem cells, they may end up in daughter cells or progeny. The accumulation of mutations (mutation loading) from parent to offspring in self-fertilizing plants such as Arabidopsis must be minimized, so to maintain genomic integrity and species identity. One potential source of protection may be haplosufficiency quality checking where the requirements of growth and development in the haploid gametophytes select against recessive deleterious mutations in genes needed during the brief period (a few divisions) of

gametophyte growth. In this manner, the population gets rid of deleterious mutations that have a gametophytic effect [13]. Alternatively, active mismatch repair could be responsible for limiting the accumulation of mutations in plant germ lines. Unlike in animals where a germ line is identified early in embryogenesis and cell divisions are reserved for the purpose of producing gametes, in plants one population of cells is responsible for producing the cells of differentiated tissues and for initiating the germ line gametophyte. Chlorophyll mutant sector analysis suggests that cell fates in the plant germ line are not fixed [29], and that cells of the germ line may re-emerge from differentiated cells or they may be lost during growth. Thus it is stated often that plants do not have a reserved germ line.

In mismatch repair-deficient plants containing a GUS reporter gene construct out of frame due to a $(G)_7$ mononucleotide repeat, only a five-fold increase in somatic mutation frequency was seen when compared to wild type [1]. In comparison, similar experiments in yeast resulted in at least 400-fold increases in mutation frequency [30]. In plants, germline fidelity, measured as numbers of microsatellite mutations in endogenous microsatellites among siblings, was more affected by deficiencies in mismatch repair than was somatic genome fidelity measured as microsatellite frameshift mutations in GUS reporters. The effect of the absence of mismatch repair on the rate of inherited mutations was in agreement with similar experiments performed in mice [31,32]. Mismatch repair appears to protect plant germ lines from mutation accumulation, yet leaves display

paradoxically high levels of mutation. The small increase in mutation in leaves when mismatch repair was inhibited could be attributed to high background levels of mutation in wild type leaves due to relaxation of mismatch repair, or a frameshift mutation process that is not corrected by mismatch repair. If plants exhibit relaxed mismatch repair in their somatic tissue (leaves) yet transmit low mutation frequencies to their progeny, then the sporophyte probably contains a reserved sub-population of cells with high-fidelity mismatch repair and genome maintenance. The small population of cells might be thought of as acting as an "equivalent germ line" in the shoot apical meristem of the sporophyte.

The mutation frequencies in the leaves and seeds of Arabidopsis were compared for both *msh2*$^{-/-}$ and wild-type Arabidopsis with the prediction that wild-type plants would display a differential mutation frequency in leaves and seeds, and the mismatch repair-deficient plants would show little difference in mutation frequency between leaves and seeds. Mutation frequency was also assessed in shoot apical meristems of *cal1*$^{-/-}$ *ap1*$^{-/-}$ Columbia-0 Arabidopsis. The *cal1 ap1* double mutant displays cauliflower-like florets on the inflorescence where flowers are usually found. The florets are clusters of shoot apical meristems that form from reversion of floral meristems to shoot apical meristems. The target population of putative stem cells and likely candidates for representing the sporophyte germ line consists of about 35 cells per meristem. These cells are defined by their expression of the meristem marker CLAVATA3 (CLV3) in the central zone (CZ) of the shoot apical meristem. CLV3 is an extracellular protein

ligand that activates a receptor on the surface of cells in the adjoining rib meristem (RM) region, and receptor activation inhibits expansion the rib meristem. The *ap1 cal1* plants yield enough shoot apical meristem tissue to facilitate recovery of small, specific populations of stem cells that express a fluorescent protein marker driven by the CLV3 promoter.

1.3.2 Plant Germ Line

Cell lineages that span generations define plant germ lines. Throughout most of the growth of the Arabidopsis sporophyte, the germ-line cell lineage lies within the shoot apical meristem. More specifically, it lies in a small number of stem cell in the central zone of the meristem [29] (Figure 1.3). It is estimated that about two meristem cells in the Arabidopsis embryo contribute to seed production and constitute the germ line in the embryo [29]. The pool of stem cells that may be thought of as a germ line is not stable and cells with various lineages are rapidly incorporated or eliminated. That is, mutant sectors seen early in vegetative growth are rarely passed on to progeny. The instability of the germ line is perhaps due to the small size of the meristem, or it may just appear that way as a consequence reservation by the meristem of different populations of stem cells to proliferate at different stages of growth. Current evidence shows that germ-line status does not ensure that a particular cell's direct progeny will be passed on to the following generation and chance events may dictate whether a cell's descendants are passed to progeny. The plant germ line is unlike the animal reserved germ line in which a specific population of cells is fated during embryogenesis to produce only

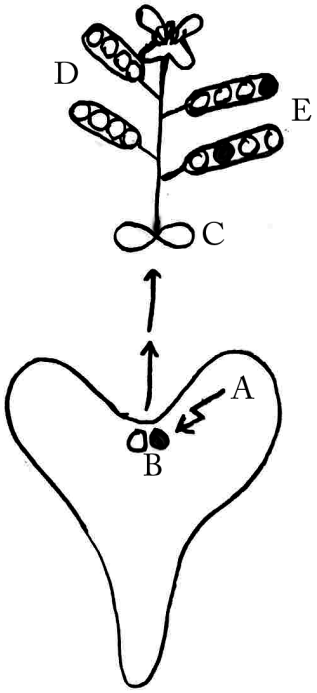

Figure 1.3. Sporophyte germ line. The mature embryo contains about two cells that are the direct progenitors of all gametes and progeny. Mutagenesis (EMS and X-rays) (A) induces a heterozygous mutation in one of the two stem cells (B) that affects chlorophyll synthesis which causes yellow sectors in tissue that is made by cell divisions of the mutagenized cell. The 7:1 segregation of the mutant phenotype among progeny indicates the presence of two stem cells at the time of mutagenesis – the undamaged cells produce all wild-type progeny (D, unfilled circles) and the damaged cell (darkened) produces progeny segregating 3:1 (E).

germ tissue [33]. Plant meristem stem cells (initials) divide infrequently, producing meristem tissue. After several cell divisions within the meristem, organ primordia form and develop, and the cells differentiate and form the mature tissues of leaves, stems, and flowers. In determinate plants, like the weedy cress *Arabidopsis thaliana*, early growth of the shoot meristem produces shoot organs before the shoot meristem transforms into an inflorescence meristem and then a floral meristem. The floral meristem produces reproductive tissues with its final sporophyte cell divisions differentiating to becoming megaspore mother cells and microspore mother cells (megasporocytes and microsporocytes). Megasporocytes and microsporocytes divide by meiosis to form the megaspore and microspore gametophytes. Sperm are formed by two mitotic divisions of the microspore, and the egg is formed by three mitotic divisions of the megaspore [34,35]. This system is in contrast to that found in animals where the germ line is identified during early embryogenesis and thereafter does not contribute cell divisions to the growth of somatic tissue.

1.4 Microsatellites

Microsatellites, also known as simple sequence repeats (SSRs) or short tandem repeats (STRs), are tracts of repetitive DNA composed of sequences one to five base pairs in length repeated in tandem [36]. Microsatellites are found throughout the genomes of eukaryotes and prokaryotes, and in plants they are enriched in regulatory regions of the genome [37]. In humans, variation of microsatellites in genes is associated with and may result in cancers and neurological diseases. In

Arabidopsis there are about 104,000 microsatellites in the genome. Mononucleotide polyA (polyT on the complementary strand) and dinucleotide $(AT)_n$ and $(AG)_n$ repeats, where n is the number of times the motif is repeated, are most prevalent, while trinucleotide repeats are most common in rice [38,39]. Arabidopsis 5' untranslated regions (UTRs) upstream of genes contain a higher density of mono-, di-, tri-, and tetranucleotide microsatellites than expected by chance, but only trinucleotide repeats are found in higher-than-expected frequencies in exons.

Microsatellites are susceptible to insertion and deletion mutations that occur during DNA replication (Figure 1.4). Primer mispairing and DNA polymerase slippage are thought to underlie microsatellite mutability, whereby a single-strand DNA loop on either the template strand or the nascent strand results in the deletion or insertion, respectively, of one or several repeat units [40]. Insertions or deletions in the short microsatellites that sometimes reside in coding regions may cause frameshift mutations, which are often deleterious.

Premutagenic single-strand DNA loops that occur in microsatellites during DNA replication are recognized and corrected by mismatch repair, and deficiencies in mismatch repair result in increases of microsatellite mutation frequency up to two orders of magnitude in bacteria and yeast [41]. The loop-outs are the target of mismatch repair surveillance. Single-base insertions and deletions, as in mononucleotide repeat sequences, are recognized by MutSα while single-strand DNA loops greater than one base are identified by MutSβ.

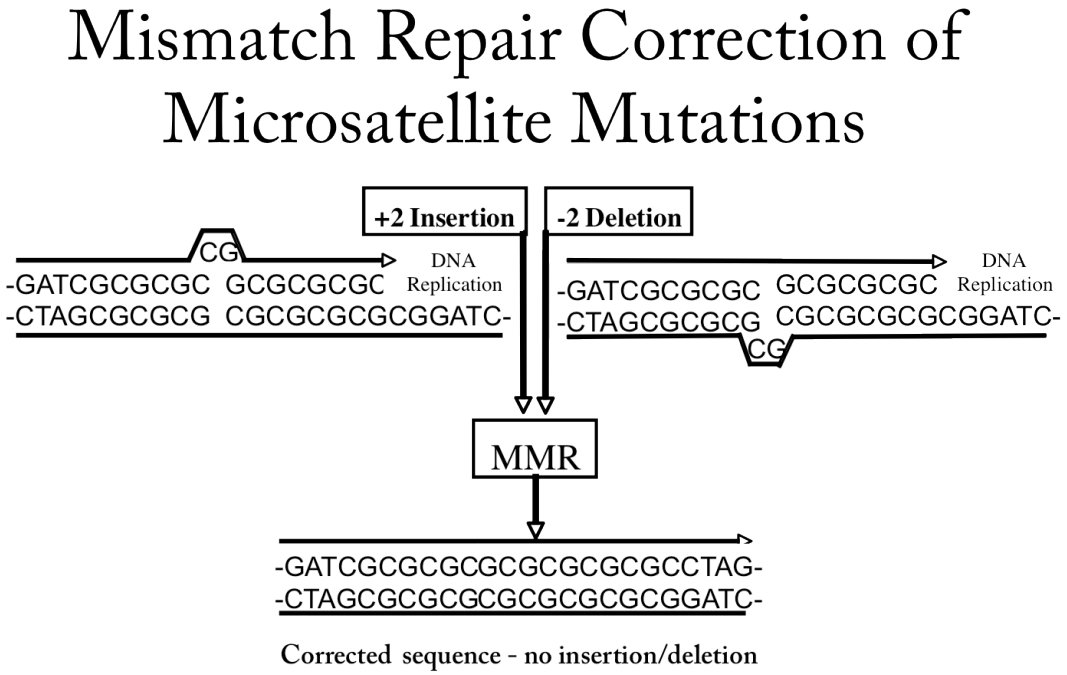

Figure 1.4. Microsatellite mutation. Mutations in dinucleotide repeats are expected to occur in multiples of two base pairs – one repeat unit. DNA polymerase slipping and primer mispairing likely result in extra-helical loops of single stranded DNA. A loop on the nascent strand will result in an insertion mutation another round of replication, and a loop on the template strand will result in a deletion. Mismatch repair detects the extra-helical DNA loops after replication, and the system corrects them by a mechanism similar to what occurs for correction of base-base mismatches (Figure1.1).

After identification by MutSα or MutSβ, the nascent strand is incised, excised, and resynthesized. Microsatellite instability is defined by an elevated level of microsatellite mutations, and mismatch repair has such a dramatic effect on microsatellite stability that increased microsatellite mutation is considered a "hallmark" of mismatch repair deficiency. Microsatellite instability is used extensively as an indicator of mismatch repair deficiency in a variety of species including yeast, Arabidopsis, and humans [42].

Microsatellite instability analysis uses PCR and capillary electrophoresis to selectively amplify and determine the length of several microsatellite loci to determine the frequency of length variants in a tissue. Microsatellite analysis is used to determine mismatch repair status of cancer tissues, and it has been applied in plants to determine the mutation rates and the effects of mismatch repair on somatic and germ-line mutation [1,4,6]. Common assays include frameshift reporters in which an insertion or deletion in a synthetic microsatellite supplied on a transgenic vector will correct the reading frame of the reporter gene, often β-glucuronidase (GUS), which when active will effect a color change in its substrate, 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) [1]. Direct detection of microsatellite mutations by PCR may provide a more physiologically relevant view of microsatellite mutation, but its analysis is complicated by what is termed PCR stutter. PCR stutter is an artifact of PCR amplification of microsatellites, likely caused by the same mechanisms that produce microsatellite mutations in cells during DNA replication. It is thought that DNA polymerase slipping during

DNA synthesis of repeat sequences followed by a restart that is out of phase by a multiple of the size of repeat units (two bp in dinucleotide repeat) causes single-stranded DNA loopouts. The artifactual PCR products interfere with identification of legitimate mutant microsatellite alleles, especially if the mutant allele accounts for less than one-quarter of all alleles in the tissue sample [1].

The frequency of microsatellite mutation is highly dependent upon the number of repeats units, the composition of the repeat unit, and the position of the microsatellites in the genome [38,39]. Microsatellites with higher numbers of repeat units mutate more frequently, and microsatellites with longer repeat units are less mutable [30,43]. Because of their propensity to mutate, and the relative ease of detecting insertions and deletions in DNA as opposed to detecting single nucleotide changes, variation of repeat number in microsatellite loci is a commonly used genetic marker in forensic science and breeding. One approach that allows the detection of rare microsatellite alleles by PCR uses very dilute DNA samples to template the PCR. If only one or two DNA molecules containing target microsatellite loci are present and amplified in one PCR, then any rare allele in that product is easily identified. Many such single-template reactions must be performed to ensure a complete analysis and accurate sampling of tissue-specific allele frequencies. PCR amplification of individual molecules of DNA, called digital PCR, has been used to identify rare microsatellite alleles and determine tissue-specific microsatellite mutation rates in humans and mice [2,44,45].

The effects of mismatch repair-deficiency on microsatellite mutation frequency in MSH2-null plants were determined by GUS transgene reversion (by insertion or deletion in a $(G)_7$ mononucleotide repeat) assays and PCR analysis [1,4]. Inherent to the PCR analysis communicated in these reports was low sensitivity of detecting rare alleles. To assess the amplitude of infrequent mutation in microsatellite loci these studies relied on a progeny analysis approach in which multiple progeny were analyzed for mutation. The progeny analysis detects rare alleles that occur in a progenitor's germ line if they are passed on to progeny. Inheritance of a heterozygous mutation is easily detectable, as the mutant allele will represent half of the alleles in the plant. The mutation frequency obtained by this method is analogous to germ-line mutation frequencies measured in animals, but the approach does not allow for determination of somatic or tissue-specific mutation frequencies. The need to determine tissue-specific microsatellite mutation frequency was suggested by the results of comparing GUS reporter microsatellite mutation frequencies in wild-type and mismatch repair-deficient Arabidopsis leaves to similar frequencies determined in yeast and mice deficient in mismatch repair [1].

Microsatellite instability analysis has proved its utility through years of use in clinical analysis and experimental use in a variety of model systems including yeast, mice and Arabidopsis, in forward, endogenous and transgenic mutation reporters.

1.5 Digital PCR

Experiments in mice examining microsatellite mutation and mismatch repair have used a single-template PCR strategy to address the low sensitivity in detecting multiple, rare mutations. In single-template PCR, the template DNA is diluted so that individual PCR reactions amplify an average of one (Poisson-distributed) target DNA molecule (Figure 1.5). Under such conditions, each target locus is amplified in isolation from all other targets and is easily distinguished from other alleles of the same locus despite background stutter; thus, more mutant alleles can be identified directly from one individual. If enough PCR reactions are performed, rare mutations can be detected, increasing the sensitivity of the method. Important considerations when using PCR to amplify DNA from one template are the effects of contamination and non-specific reactions. Both of these effects are exaggerated by the high number of reaction cycles needed to produce enough DNA to be visualized in an agarose gel. The solution is a two-step hemi-nested PCR strategy, where the first round of 18 reaction cycles amplifies DNA that is empirically diluted to single template levels. Products of the first round are diluted into a second round 18-cycle PCR where the reverse primer is labeled and nested within the region flanked by the reverse primer used in the first round.

I have employed microsatellite loci used previously in the Hays lab in a single-template analysis. Nested primer pairs for each microsatellite have been designed so that a single set of reaction conditions can be used for all loci of interest, and

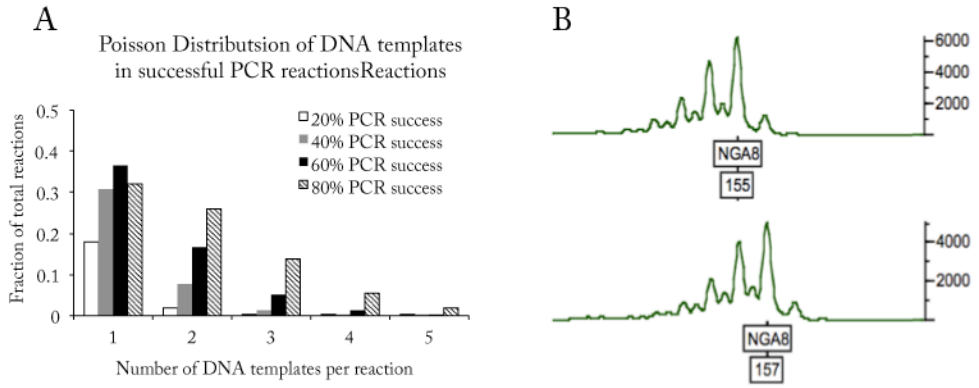

Figure 1.5. Single-template microsatellite analysis. (A) Digital PCR refers to the on/off nature of PCR reactions where each reaction contains zero or one DNA template. The Poisson distribution describes the distribution of DNA templates among many parallel, independent PCR reactions, and the number of templates in each successful reaction can be determined statistically. For example, when 40% of PCR reactions produce product, 30% of the total reactions contained one template and about 8% contained two templates. (B) Capillary electrophoresis of PCR amplified microsatellites displays characteristic stutter exemplified here by the peaks on both sides of the tallest peak. The tallest peak is legitimate and not stutter. The top trace shows the wild-type NGA8 allele of 155 base pairs, and the bottom trace shows a two-base pair insertion mutant allele of 157 base pairs.

a protocol is now established to empirically and reproducibly dilute genomic DNA to the single copy level. Currently, we have the ability to analyze seven microsatellite loci at the single-genome level in Arabidopsis.

1.6 Research Objectives

Mismatch repair prevents the accumulation of mutations during plant growth and reproduction. Plant breeders seeking novel variety desire mutations, and trait introgression is often dependent on mismatch-repair-controlled recombination between closely related species. Trait stability is also important in crop production, but the common horticultural practice of vegetative propagation is prone to mutation and somaclonal variation that may cause loss of desired characteristics. Modulating mismatch repair activity in somatic tissue and germ lines may simultaneously address the needs of plant breeders seeking novelty and plant propagators seeking stability. To better understand the role of mismatch repair in preventing mutation loading, I aimed to determine the genomic base substitution rate and spectra in mismatch repair-defective Arabidopsis, and to determine the effect of dominant-negative *msh2* alleles on microsatellite mutation frequency in Arabidopsis. To further understand the relative contributions of mismatch repair and haploid gametophyte growth in preventing mutation loading in selfing populations, I also aimed to determine tissue specific microsatellite mutation frequencies in somatic tissues and putative germ-line tissue. Somatic microsatellite mutation frequencies measured in plant vegetative tissues are high, and they do not increase dramatically in the absence of mismatch repair, the

pathway whose function it is to prevent microsatellite mutations. In marked contrast, when looking in the progeny of mismatch repair-deficient plants highly increased frequencies of microsatellite mutation is observed as a result of mutation in the progenitor, and the progeny of wild-type mismatch repair-proficient plant lines show almost no mutation. Do plants relax mismatch repair in their vegetative tissues while maintaining it in the germ line, or do they have a mechanism that limits the number of times stem cells divide between embryonic identification and gametophyte formation? If mismatch repair is relaxed in somatic tissues I expect that in mismatch-repair-defective plants the microsatellite mutation frequency in leaves and equivalent-germline meristem tissue will be equal, and in mismatch-repair-proficient plants the mutation frequency will be greater in the leaves than in the shoot apical meristem. Here I use high-throughput genome sequencing, digital PCR, and fluorescence assisted cell sorting to address this question and to estimate the germ-line base-substitution rate in mismatch repair-defective plants. I also report that heterologous overexpression of dominant negative *msh2* alleles may provide a method to produce chronically increasing mutation frequency in plants.

2 Materials and Methods

2.1 Mutation accumulation

2.1.1 Plant materials and propagation

As described in reference [4], 36 independent lines of *msh2*$^{-/-}$ (SALK_002708) mutants (K1-K36) were initiated from bulked-up seed of a single line that showed microsatellite instability (MSI). Wild-type (WT) control lines (C1-C36) were also planted using seed from one Colombia-0 individual. Each line was established and propagated by hand-counting 50 seeds, mixing them with ~0.5 mL dry sand, and dusting the seed and sand mixture over a single pot of pre-moistened potting mix (Figure 2.1). Seeds were vernalized at 4 °C for two days, after which time they were placed under cool-white fluorescent lamps for 16-/8-hour light/dark cycles at 21 °C. Germination was scored two weeks after placement in the light by counting the number of individuals with visible cotyledons. Three weeks after germination, the plants were again counted to determine the fraction of plants capable of developing true leaves. They were thinned out to four plants per pot without bias, but with the requirement that there be one plant in each quadrant (NW, NE, SE, or SW) of each pot. One quadrant was selected at random (by coin toss), and in each case the plant in that quadrant was designated to be the progenitor of the following generation. Lines were scored as "extinct" when the chosen quadrant contained a sterile plant. Otherwise, about seven weeks after planting, siliques 8-12 up from the base of the primary inflorescence on the chosen progenitor for each line are collected, and the

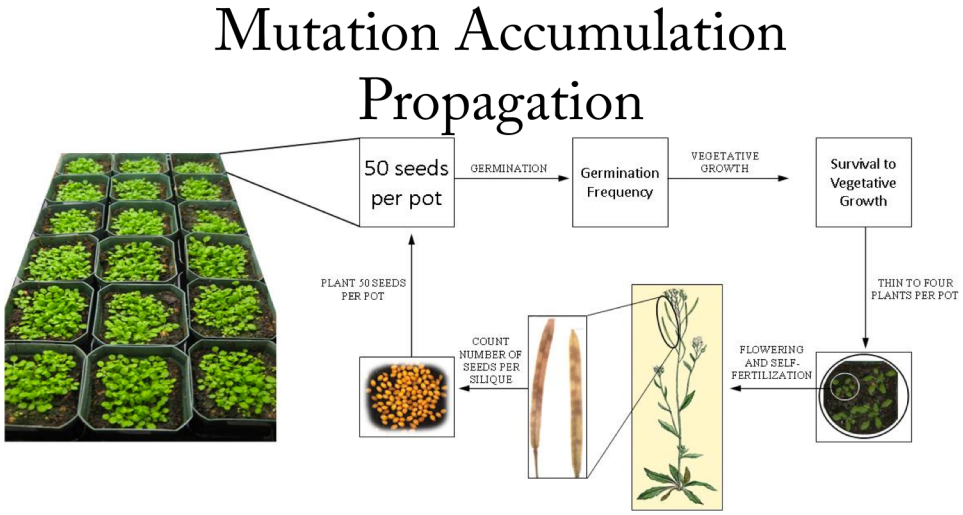

Figure 2.1. Mutaiton accumulation propagation. Mutation accumulation lines create bottlenecks every generation, which limits selective pressures. Arabidopsis mutation accumulation lines were initiated using 50 seeds per line. The seeds were mixed with about 1 mL of sand and evenly distributed over moist soil. After three days at 4 °C, the seeds are placed under fluorescent lamps set to 16 hr:8 hr on:off regimen. After two weeks germination was scored by counting seedlings. Two weeks later the plants are thinned so that for each line only four plants remain – one in approximately each quadrant of the pot. The number of plants removed at thinning is recorded to track the ability of the plants to grow vegetatively. Once the plants bolt (set an inflorescence) and senesce, the seeds are collected from a randomly chosen plant in each line, and the fertility of the plants is determined by counting the number of seeds in five siliques (beginning with the eighth silique up from the base of the plant. Each successive generation is initiated using 50 seeds from one randomly chosen progenitor per line, and a line was removed from propagation (considered to be "extinct") when the randomly chosen progenitor of the successive generation was sterile. Not all plants in that line need be sterile for it to be scored extinct.

number of seed in each was recorded. The remaining seeds were collected and used to repeat the process beginning again with 50 seeds for each line.

2.1.2 PCR genotyping of transgenic plants

DNA was prepared from single cotyledons using the method described in reference [46] and resuspended in 100 μL DNA buffer (10 mM Tris, pH 7.5, 1 mM EDTA). PCR was templated with 1 μL DNA, and the AtMSH2-specific primer OMSH30 was paired with the T-DNA specific primer OLBA1 or with the AtMSH2 specific primer OMSH5 (Table 2-1). PCR was carried out using Taq DNA polymerase under cycling conditions of 98 °C for 3 minutes followed by 32 cycles of 98 °C for 15 seconds, 55 °C for 30 seconds, 72 °C for 1 minute, and a final cycle of 72 °C for 5 minutes. PCR products were analyzed by 1% agarose gel electrophoresis.

2.2 Determination of base-substitution frequency in *msh2*$^{-/-}$ Arabidopsis

2.2.1 Plant materials

I sequenced pools of genomic DNA extracted from multiple individuals of two independent *msh2*$^{-/-}$ mutation-accumulation lineages obtained from 17 generations of selfing and single-seed-descent. The *msh2*$^{+/-}$ mutant line obtained from the Salk Institute is the product of T-DNA insertional mutagenesis of the Arabidopsis Col-0 ecotype. Seeds saved from the stock used to initiate the mutation-accumulation lines (Generation 0) and from two Generation-17 mutation-accumulation lines (K6 and K10) were surface-sterilized by successive

Table 2.1. Genotyping primers. OMSH30 and OLBA1 produce a PCR product in the presence of wild-type Arabidopsis MSH2, and OMSH5 paired with OLBA1 guide selective amplification of the junction between the T-DNA vector and MSH2 in the T-DNA insertional knockout. The presence of DN1 and DN2 *msh2* alleles is confirmed by the presence of a PCR product resulting from the use BVUP paired with MSH2RC850.

Primer	Genotyping Primers	Annealing temperature
MSH2 wild-type		
OMSH30	CCAGTTGCCCTACTCCATACTG	55°C
OLBA1	TGGTTCACGTAGTGGGCCATC	
msh2$^{-/-}$		
OMSH5	TTTCAGTGTCAATGTGAGCG	55°C
OLBA1	TGGTTCACGTAGTGGGCCATC	
DN1 and DN2		
MSH2RC850	CATATCTGCGGATTGTGAAGTTCCC	55°C
BVUP	CGTGTAGGACATGGCAACC	

washes in 70% ethanol and 3% sodium hypochlorite, followed by five rinses in sterile water. Sterile seeds were sown on 0.5× MS medium (0.8% agar) in Petri plates. Parafilm-sealed Petri plates were kept at 4 °C for 3 days prior to placement under cool-white fluorescent lamps at 22°C set to maintain a 16:8 light:dark photoperiod. Plants with about 4 leaves were harvested and frozen at -80 °C. DNA was extracted from individual whole plants, quantified by PicoGreen assay, and mixed equally by mass.

2.2.2 Genomic DNA sequencing library preparation

Five-base-pair "barcodes" used as unique identifiers were included in the adaptor oligonucleotides that were ligated to the ends of genomic DNA fragments (Table 2.2). In addition to the barcodes, which allow multiple DNA samples to be sequenced in one reaction, the adaptors contain sequences that hybridize to oligonucleotides that prime DNA synthesis in the final steps of library preparation and in the DNA sequencing reactions. Adaptors were prepared by mixing equal masses of 100 mM HPLC-purified adapter oligonucleotides, boiling the mixture for 2 minutes, and cooling the adaptors to room temperature, allowing the adaptors to anneal. Adaptors were stored in a -20 °C freezer.

A modified Illumina Sequencing library preparation protocol was adapted to use DNA isolated from whole plants frozen in liquid nitrogen and powdered with a plastic pestle in 1.5 mL microcentrifuge tubes [47]. DNA was isolated using Qiagen Plant Genomic DNA Extraction Kits (Qiagen, Valencia, CA) as per

Table 2.2. Oligonucleotide adaptors and primers used in Illumina sequencing library preparation. Mutation-accumulation lines K10 and K6 were sequenced together using uniquely identifiable "barcode" sequences (bold-face) placed in the adaptors. The adaptor provides hybridization sites used to prime PCR and DNA synthesis during sequencing. The barcode is sequenced with genomic DNA fragments and used to match sequences to the original DNA sample. Once the sequences are sorted computationally based on the barcodes, the barcodes are removed from the sequence data. The G0 progenitor was not barcoded, and only one adaptor set was used in the preparation of the sequencing library.

	Barcoded Oligonucleotides
SE_Adaptor1_ACGTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACGTT**
SE_Adaptor1-P_ACGTT	Phos/ACGTAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
SE_Adaptor2_CGTAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CGTAT**
SE_Adaptor2-P_CGTAT	Phos/TACGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
SE_Adaptor3_GTACT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GTACT**
SE_Adaptor3-P_GTACT	Phos/GTACAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
SE_Adaptor4_TACGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TACGT**
SE_Adaptor4-P_TACGT	Phos/CGTAAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
	Non-barcode Olgionucleotides
PE_Adaptor1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
SE_Adaptor1-P	Phos/GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
	Library Amplification Primers
SE_PCR1.1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
SE_PCR2.1	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT

manufacturer recommendations with two 50-μL elutions of the DNA from the silica matrix columnin using Qiagen buffer EB. Recovered genomic DNA was diluted in 600 μL Tris-Cl (pH 7.5) followed by 60 minutes ultrasonic irradiation ("sonication") at 4 °C in a Bioruptor circulating water bath (Diagenode, Denville, NJ) set to maximum power with a 30 second on/off cycle to shear the DNA to fragments 200-1000 bp in length. Sonification conditions were determined empirically depending on the DNA concentration and test tube size. Sheared DNA was concentrated to 40 μL volume using Microcon 100 membrane filters (Millipore, Billerica, MA), quantified using PicoGreen(Life Technologies, Grand Island, NY) and a SpectraMax Gemini fluorescent plate reader (Molecular Devices, Sunnyvale, CA); 2 μg was diluted to 30 μL in water. Enzymes, reaction buffers, and nucleotides used in the subsequent end-repair reactions, A-overhang addition reactions, adaptor-ligation reactions, and PCR were purchased from New England Biolabs (Ipswich, MA). The purification and clean-up steps were performed using silica matrix microcentrifuge column-based kits purchased from Qiagen.

End-repair reactions convert ends of sheared DNA into 5'-phosphorylated blunt ends. For 30 μL (2 μg) of sheared DNA, end-repair reaction mixtures contained 1× T4 DNA ligase buffer (with 1 mM ATP), 0.4 mM dNTPs, 7.5 units T4 DNA polymerase, 2.5 units exonuclease+ Klenow polymerase (DNA pol1), and 25 units T4 polynucleotide kinase in a 50 μL reaction volume, and mixtures were incubated at room temperature for 30 minutes. Products of the end-repair

reaction were purified using Qiagen PCR purification kits, and were eluted from the silica matrix column with 32 μL Qiagen buffer EB.

The A-overhang addition reaction adds a 3' A-overhang onto the genomic DNA to facilitate subsequent ligation of adaptors with 5' T-overhangs. The A-overhang reaction mixture contained 30 μL DNA eluted after the end repair reaction, 1× Klenow buffer, 0.1 mM dATP, and 15 units exonuclease- Klenow fragment, was incubated at 37 °C for 30 minutes. The products of the A-overhang addition reaction were purified with Qiagen MINELUTE kits with the final elution in 10.5 μL Qiagen buffer EB.

The subsequent adaptor ligation reaction adds double-stranded DNA adaptors (hybridized oligonucleotides), which contain priming sites used during DNA sequencing and, in some cases, unique identifiers (barcodes). The adaptor ligation reaction mixtures were incubated for 30 minutes at room temperature, and they included all DNA recovered from the purification after A-overhang addition, plus 1× T4 DNA ligase buffer, 5 μM adaptors, 20 units T4 DNA ligase, and enough water to make up 25-μL reactions. Products of adaptor ligation were separated by agarose electrophoresis in a 2% Nu-Seive agarose, ethidium bromide stained TAE gel. After electrophoresis, gel slices were removed to isolate DNA estimated to be between 250-350 bp in length based on a 100 plus bp ladder (Fermentas, Glen Burnie, MD). The gel slices were weighed, the DNA within them was purified away from the agarose gel using Qiagen gel extraction kits as

per manufacturer instructions (with 2 washes of Qiagen buffer PE), and DNA was eluted from the silica matrix column with 30 μL Qiagen buffer EB.

Adaptor-library amplification was accomplished in 50-μL PCR reactions, but variation in the efficiency of adaptor ligation affected the number of cycles necessary to amplify the library. Empirical determination of the number of PCR cycles needed to amplify the library was based on the minimum number of cycles required to visualize 5 μL of PCR product in a 1.2% TAE agarose ethidium bromide stained gel after electrophoresis. For each library, at least three independent PCR amplification were performed, and then equal amounts of product from each was combined to help counter any random bias in a given PCR reaction. Combined PCR products were purified using Qiagen PCR purification kits, and eluted DNA was quantified using PicoGreen assays. If barcoded adaptors were used, equal masses of four different libraries (one library per barcode) were pooled (multiplexed) and submitted to the Oregon State University Center for Genome Research and Biocomputing core laboratory for Illumina cluster generation and sequencing.

2.2.3 Bioinformatics

For two mutant lines and one generation zero progenitor, I obtained a sequence coverage depth of 10-20x using the Illumina GAIIX and HiSeq2000 machines (Illumina, San Diego, CA) to produce sequence reads of 40 bp and 58 bp. The two generation-17 samples were barcoded (4 × 5 bases) and sequenced together in one single-end flow cell on both the GAIIX (40 bp) and the

HiSeq2000 (58 bp) sequencing machines. After sorting sequences into different groups based on their barcodes, the barcode sequences were removed, and the resulting 35-bp, 53-bp, and 58-bp sequences were aligned to the most recent version (10) of the *Arabidopsis thaliana* reference genome maintained by The Arabidopsis Information Resource (TAIR 10), using the Short Oligonucleotide Alignment Program 2 (SOAP2) software package [48]. Single-base substitutions were identified by comparing each sequenced genome to the reference. The lists of base substitutions for each sequenced genome were then compared to each other to remove single nucleotide polymorphisms (SNPs) in common, with the assumption that shared SNPs were present in the progenitor of the mutation-accumulation experiment. Putative base substitutions were identified after applying three *filtering* criteria: (1) 4× minimum site coverage and (2) 100x maximum coverage, and (3) near-unanimous consensus base of A,T,C, or G, and quality score ≥ 15 [49]. The mutation rate that is reported here is the per generation rate, and it is defined as the number of base substitutions per interrogated site (number of sites that met bioinformatics quality-filtering criteria) per generation.

2.3 Tissue-specific microsatellite mutation frequency

2.3.1 Plant materials

WT and *msh2*$^{-/-}$ mutant *Arabidopsis thaliana* seeds (from the mutation-accumulation experiment) were sown on moist potting mix (Sungrow SB40) and

vernalized at 4°C for 3 days, at which time they were placed under cool-white fluorescent lamps with a 16/8 light/dark cycle at 21 °C.

Venu Reddy University of California, Riverside) supplied *Arabidopsis thaliana* seed for CLV3 meristem-marker plant lines in the *ap1 cal1* background. These plants express GFP in the cells of the shoot apical meristem central zone (CZ) driven by the *CLAVATA3* (*CLV3*) promoter. The *ap1 cal1* double mutant background was used because its phenotype allows the collection of many shoot apical meristems in a short period of time from the cauliflower-like inflorescence meristems that have reverted to a shoot apical meristem identity [28].

2.3.2 DNA extraction and pooling

Arabidopsis thaliana young leaves, seeds, and meristem protoplasts (stored at -80 °C until use) were initially frozen with liquid nitrogen and powdered in 1.5 mL microcentrifuge tubes using plastic pestles. DNA was then extracted using Qiagen Plant DNAeasy kits. DNA was quantified by PicoGreen assays. Equal masses of DNA from 5 individuals were mixed to create pools, and pooled DNA was then diluted to 1 ng/μL in 1 ng/μL herring sperm DNA (purified by sequential phenol/chloroform extractions). The herring sperm DNA is incorporated to prevent loss of plant DNA by adhesion to plastic tube walls. Control PCR reactions without template plant DNA were performed to ensure there was no amplification of the herring DNA.

2.3.3 Shoot apical meristem protoplast isolation

I collected the cauliflower-floret-like shoot apical meristem clusters from about 50 plants, and placed 70-100 mg of tissue in a 70-μm cell strainer (Benton, Dickinson and Company, Franklin Lakes, NJ) in one well of a six-well culture plate. I confirmed correct expression of the CLV3:GFP fluorescent marker by epifluorescent microscopic observation (400x, FITC). Six mL Protoplasting Solution A (20 mM MES, 20 mM KCl, 0.4 M D-mannitol, 0.1% BSA, 10 mM $CaCl_2$, pH 5.7) with 1.25% (w/v) cellulose (Yakult) and 0.3% (w/v) Macerozyme (Yakult) was added to the tissue in strainers to digest cell walls and release protoplasts into solution. The tissue was digested for 45 minutes at room temperature with shaking at 200 RPM. Every ten minutes, I rinsed the tissue with the enzyme solution to wash freed protoplasts through the mesh strainer [50]. Protoplast tonicity was monitored by bright-field microscopic (100x) observation of a 10-μL aliquot of the protoplast suspension. When the cell suspension became cloudy it was transferred to a 50-mL Falcon conical tube, and then centrifuged at 500x g for 10 minutes at 4 °C. The supernatant was poured off and the pellet was resuspended in 600 μL ice-cold Protoplasting Solution A without enzymes.

Flow sorting was performed on a MoFlo sorting flow cytometer (Cytomation, Fort Collins, CO) operated by Samuel Bradford in National Institutes of Health Science Center flow cytometry core facility at Oregon State University. Setting a "gate," referred to as gating, that defines target cells to be selectively captured is necessary for accurate cell sorting. GFP-expressing

protoplasts fluoresce green (530 nm) more intensely than non-expressing cells, and form a clear population when compared with wild-type cells in a red vs. green dot-plot of fluorescence intensities of each detection event. Non-GFP cells were used to define the limits of autofluorescence intensities (Figure 2.2). The population of cells was centered on the diagonal in a green versus red fluorescence dot plot by adjusting photomultiplier tube detector voltages and gains. GFP-positive protoplasts which lie off of the diagonal in the direction of increased green signal intensity were sorted into 400 μL of lysis buffer from a Qiagen Plant Genomic DNA Kit, frozen under liquid nitrogen, and thawed at 60 °C. DNA isolation continued as per manufacturer recommendations with two final elutions of the DNA from the silica matrix columns using 50 μL Qiagen buffer EB.

2.3.4 Digital PCR

To determine empirically the dilution needed for digital PCR, pooled DNA samples in 1 ng/μL herring sperm DNA (Tris pH 7.5) were diluted $1:10^3$, $1:10^4$, $1:10^5$, and $1:10^6$. At least 16 hemi-nested PCR reactions were performed in parallel for each dilution for each DNA pool. The second-round PCR products were analyzed by gel electrophoresis, and the ratio of successful to unsuccessful PCR reactions was used to calculate the average templates per well in the successful reactions. I found that the dilutions were reproducible, and that all loci observed were roughly equal in amplification efficiency.

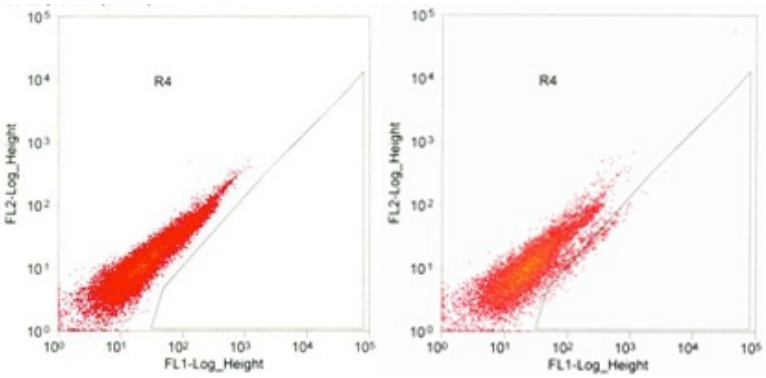

Figure 2.2. Fluorescence-assisted cell sorting of protoplasts. Gating to select protoplasts expressing GFP is in the dot plot of red versus green fluorescence. GFP-negative protoplasts are used to determine where to draw the gate based on autofluorescence (left panel), and protoplasts expressing GFP fluoresce more intensely green channel. Autofluorescence in non-expressing cells was used to define the gate such that only GFP (FLI) positive cells are identified and collected.

Once empirical determinations of single-template concentrations were made, DNA pools were again diluted in 1 ng/μL herring sperm DNA to the appropriate concentration. At each microsatellite locus, more than 100 PCR reactions were performed for each DNA pool. After 18 cycles, 1 μL of each PCR was used to template a second round of PCR, which acted as a labeling and signal amplification reaction. The second-round PCR was similar to the first round except that a nested, fluorescently labeled (HEX or FAM) primer was paired with the forward primer (Table 2.3) (Figure 2.2). The first round of PCR incorporates unlabeled primers to selectively amplify each locus, and in the second round of PCR the reverse primer is replaced by a fluorescently-labeled, nested primer that primes DNA synthesis beginning nearer to the target than the first-round reverse primer. Reactions that received as template the product of an unsuccessful first-round PCR were also unsuccessful. Importantly, only products successfully amplified from single template molecules during the first round of PCR were labeled and further amplified in the second round of PCR. PCR amplification of single DNA templates used Phusion DNA polymerase (New England Biolabs) in 10 μL reactions typically with 3% DMSO. Cycling conditions were 98 °C for 3 minutes, followed by 18 cycles of 98 °C for 10 seconds, annealing temperature (see tables 2-3 and 2-4 for primer sequences and annealing temperatures) for 15 seconds, and 68 °C for six seconds, followed by an additional cycle of 68 °C for 5 minutes.

Table 2.3. Oligonucleotide primers used in the first round of digital PCR. Primers shown are paired with the unlabeled primers shown in Table 2.4 for the first round of "nested" PCR.

Locus	Pimers	Annealing temperature
NGA6	GAAAACACAGAGAGAAACCCCTA	61°C
NGA8	GCCGCTGACACTTGTCACTA	57°C
NGA139	CCAAATTTGAAATGAATCGAAGT	56°C
NGA151	TTAAGCTTCAATGAAAGTGATGG	56°C

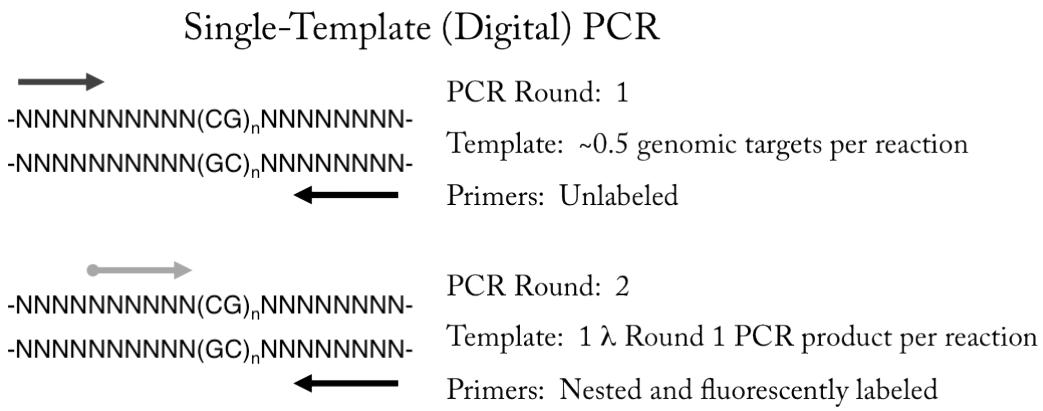

Figure 2.3. Hemi-nested single-template PCR. To address the low sensitivity of detecting multiple rare mutations, hemi-nested single-template PCR is used to amplify an single DNA molecules in each successful reaction. Two rounds of 18-cycle PCR are used to amplify DNA after empirical dilution to single-template levels. Products of the first round are diluted into a second-round 18-cycle PCR where the reverse primer was fluorescently-labeled and nested within the region originally flanked by the reverse primer used in the first round.

2.3.5 Capillary electrophoresis and data analysis

After 18 cycles, one-half volume of each reaction was analyzed by agarose gel electrophoresis, and the remaining product was frozen (-20°C) for further analysis. Fluorescently labeled PCR products from successful reactions were multiplexed for electrophoresis on the basis of product size and fluorescence excitation and emission spectra. Multiplexed PCR products were diluted empirically based on fluorescence intensity of the DNA in the ethidium bromide-stained agarose gel and prior experience (generally 1:50 to 1:200) to so that HEX and FAM fluorescence intensities were within the detection range of the equipment used for PCR product size detection based on.

Diluted, multiplexed PCR products were analyzed by capillary electrophoresis fragment analysis using an ABI 3730 capillary DNA sequencer using ROX400 standards (Applied Biosystems and Life Technologies)

Data were analyzed using ABI GeneMapper software. I inspected by eye the automated sizing calls produced by GeneMapper using manufacturer recommended default settings for dinucleotide repeat analysis. Data were omitted when the signal was too low (less than 60 fluorescence intensity units) or when the background was too high.

Two microsatellite mutation frequencies are reported herein. The unique mutation frequency is determined by dividing the number of mutant alleles detected by the total number of alleles observed. The total mutation frequency is

obtained by dividing the total number of times a mutant allele is detected by the number of alleles observed.

2.4 Progeny analysis of microsatellite mutation frequency in Arabidopsis carrying dominant-negative *msh2* alleles

The mutated residues in the MSH2-DN1 and MSH2-DN2 proteins are in the ATPase domain (Gly833Asp) and helix-turn-helix domain (Gly671Asp), respectively, and they are homologous in Arabidopsis of yeast dominant-negative proteins. Jeff Leonard created the Arabidopsis mutants. Aly Mohamed fused the proteins to C-terminal MYC epitope tags and placed the tagged protein constructs downstream of the Gelvin "Super Promoter" which incorporates *$(Aocs)_3$ Amas Pmas* elements from the octopine synthase and mannopine synthase genes and the tobacco-etch-virus translational leader in the binary vector pE1803 [51,52]. The constructs were transformed into wild-type (Col-0) and *rdr6-15* Arabidopsis [53]. The *rdr6-15* mutation confers resistance to silencing of overexpressed transgenes [54]. The presence of the transgenes was verified by PCR. I analyzed progeny of eight individual transformants (two each of DN1 and DN2 in both wild-type and *rdr6-15* Col-0 backgrounds) and three WT Col-0 plants for microsatellite instability at four endogenous dinucleotide microsatellite loci. For each transgenic (DN1 and DN2) and wild-type (Col-0) line, seeds were sown on 0.5 × MS agar, and plants were collected after 2-3 weeks of growth in incubators illuminated by fluorescent lights 16 hours per day. DNA was extracted by the Edwards method (described above under "PCR Genotyping" section 2.1.2), aliquoted, and stored at

-20°C until PCR analysis [53]. Two to three individuals from each line were PCR-genotyped to verify the presence or absence of transgenes. Genotyping was performed as described above for *msh2*$^{-/-}$ except only one primer pair was used (Table 2.1). The presence of the DN1 and DN2 transgenes was verified by PCR using primers specific to the T-DNA vector harboring mutagenized *msh2*. Primers BVUP and MSH2RC850 were included in a PCR reaction of 1 cycle at 98 °C, 32 cycles of 98 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute, and 1 cycle at 72 °C for 5 minutes. PCR reactions were analyzed by 1% agarose TAE gel electrophoresis.

Duplicate, independent PCR reactions amplifying the four microsatellite loci were performed for each of the DNA samples, using fluorescently-labeled microsatellite-specific primers (Table 2-4). For each individual plant DNA sample all four amplification products were mixed and diluted 1:60 for capillary electrophoresis fragment analysis on an ABI 3730 capillary DNA sequencer using ROX400 standards (ABI). Data were analyzed using ABI GeneMapper software. Automated sizing calls were inspected and corrections made where appropriate to omit noisy data. PCR to amplify microsatellite loci was performed using Phusion DNA polymerase in similar reactions as those used for digital PCR, but here only one 32-cycle reaction was used. It employed the fluorescent-labeled primer (see Table 2.4 for primer sequences and annealing temperatures.)

Table 2.4. Oligonucleotide primers used to amplify microsatellites in Arabidopsis. Primer pairs shown are used in progeny analysis microsatellite PCR. In the case of two-round digital PCR, the fluorescently-labeled (HEX and FAM) primers are replaced by the primers listed in Table 2.3. The second round of digital PCR uses the primer pairs show here.

Locus	Primers	Annealing temperature	Product size (bp)
NGA6 $(AG)_{31}$	GACTAAAGTGGGTCCCTTGG (HEX)CACACCCAAAACTCGTAAAGC	61°C	265
NGA8 $(AG)_{25}$	TGGCTTTCGTTTATAAACATCC (HEX)GAGGGCAAATCTTTATTTCGG	57°C	154
NGA139 $(TC)_{33}$	GGTTTCGTTTCACTATCCAGG (HEX)AGAGCTACCAGATCCGATGG	56°C	177
NGA151 $(TC)_{29}$	CCAGAGCTTGTTTTGGGAAG (FAM)TTTGATGAAACGGAATATAGAAAGC	56°C	251

3 Results

To determine the mutation rate and spectra in the absence of mismatch repair in plants, I sequenced the nuclear genomes of three lines from a collection of mismatch-repair-deficient Arabidopsis mutation-accumulation (MA) lines. I also conducted a progeny analysis to determine the effect of dominant-negative *msh2* alleles on microsatellite mutation frequency in plants. The mismatch repair-deficient mutation-accumulation lines were initiated from one *msh2*$^{-/-}$ individual and propagated by single-seed descent for 18 generations [4]. Plants carrying the dominant-negative alleles were maintained for 3 generations.

3.1 Modulation of Mismatch Repair in Plants

3.1.1 Mutation-accumulation in mismatch repair-defective Arabidopsis

As reported in reference [4], the progenitor of the mutation-accumulation lines was acquired from the SALK T-DNA knockout collection as a heterozygote. Seeds from a homozygous *msh2*$^{-/-}$ individual, obtained after the heterozygote was allowed to self-fertilize, were grown. Seeds from a single individual were used to create 36 parallel mutation-accumulation lines. A set of 36 control lines initiated from one WT Col-0 individual was also started. Results after five generations of single-seed descent were published [4]. I continued propagating the *msh2*$^{-/-}$ and wild-type mutation-accumulation lines beginning at Generation Ten (G10), and I tracked the germination, development, and fertility of the plants. By G18, all but five *msh2*$^{-/-}$ lines had become "extinct" (see section 2.1.1) during the propagation

process (section 2.1.1), whereas all of the original wild-type lines remained (Figure 3.1). I compared germination and the ability of the plants to develop to the four-leaf stage at G5 [4] and at G18 (Table 3.1). By Generation 18, seven of the putative mismatch-repair-deficient lines were now actually wild-type, due either to outcrossing or seed contamination. The remaining "extinct" lines were infertile *msh2*$^{-/-}$ lines. One mismatch repair-deficient line developed severe germination defects at G18; only one of 50 seeds planted grew.

3.1.2 Base substitution rate in MSH2-deficient Arabidopsis

I sequenced the genomes from two plants after 17 generations of single-seed descent from one *msh2*$^{-/-}$ progenitor, whose genome I also sequenced. One of the G17 lines sequenced (K10 in table 3.1) contained wild-type *MSH2* by Generation 9, and the other G17 line, referred to as K6, was still mismatch repair-deficient. The mismatch repair-deficient progenitor is hereafter referred to as the G0 plant.

About 100 million sites in the 120 million base-pair (Mb) genome met stringency filters requiring that a polymorphism be supported by a minimum of four reads, be covered by a maximum sequencing depth of 100 reads, be at sites where only near-unanimous consensus calls of A, T, G or C were made, and be supported with a minimum quality score of 15, which provides about 97% accuracy [55]. Base-substitution mutations in each line were identified by a "consensus" approach [56] so that mutations arising during single-seed descent were differentiated from preexisting mutations by comparing all mutations in lines K10 and K6, and the G0 plant.

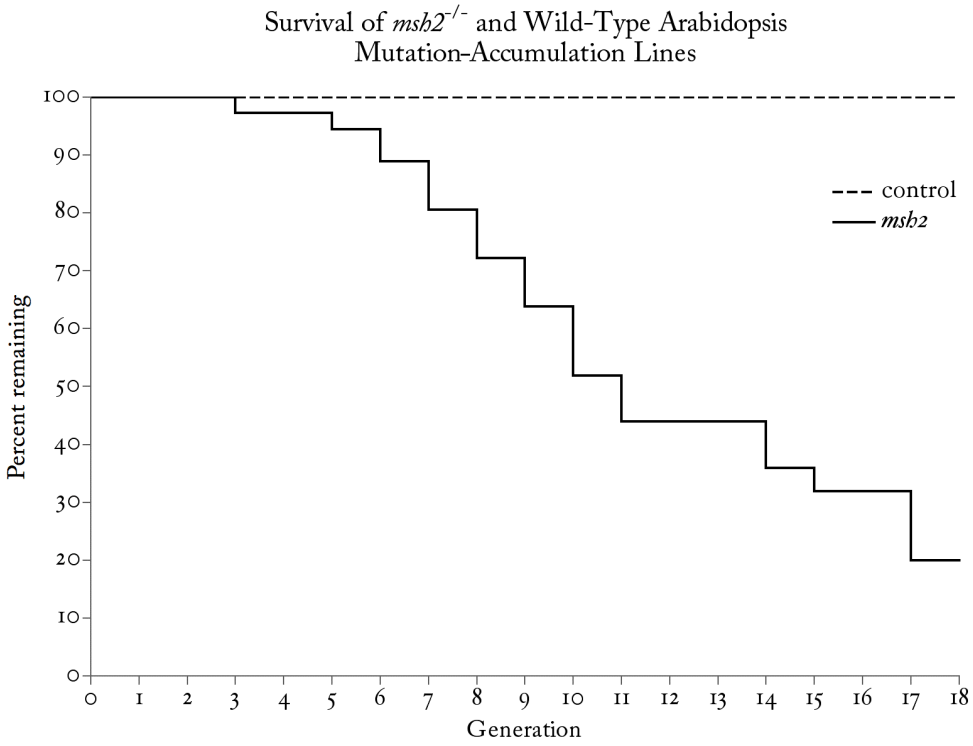

Figure 3.1. Survival of *msh2*$^{-/-}$ and wild-type Arabidopsis propagated seed-to-seed. The percentage of the original 36 wild-type and *msh2*$^{-/-}$ lines (corrected for the seven *msh2* lines that lost the mutant *msh2* allele) at each generation of the mutation-accumulation experiment that have not been removed due to infertility in the randomly chosen progenitor of the following generation.

Table 3.1. Mutation accumulation lines. Comparison of Generation 5 [4] and Generation 18 germination and survival to thinning. X indicates extinction of the line, and * indicates loss of mutant *msh2* allele and replacement with wild-type MSH2.

	Germination				Survival to thinning			
	Rel. to WT mean		SDs below WT mean		Rel. to WT mean		SDs below WT mean	
Line	Gen. 5	Gen. 18	Gen. 5	Gen. 18	Gen. 5	Gen. 18	Gen. 5	Gen. 18
K1	0.78	X	4	X	0.92	X	1	X
K2	0.74	0.95	5	*	1.06	1.02	-2	*
K3	1.02	X	0	X	0.91	X	1	X
K4	0.91	X	2	X	0.95	X	1	X
K5	0.54	X	9	X	0.76	X	4	X
K6	0.87	0.89	2	4	0.85	1.04	3	-2
K7	0.91	X	2	X	0.98	X	0	X
K8	0.57	X	8	X	0.92	X	1	X
K9	0.46	1.03	10	-1	0.9	0.94	2	1
K10	0.65	1.03	7	*	0.87	0.96	2	*
K11	0.72	1.03	5	*	0.73	1.02	5	*
K12	X	X	X	X	X	X	X	X
K13	0.85	X	3	X	1.02	X	-1	X
K14	0.87	X	2	X	0.98	X	0	X
K15	0.61	X	7	X	0.89	X	2	X
K16	0.85	X	3	X	1	X	0	X
K17	0.78	0.80	4	7	0.8	0.85	4	3
K18	1.06	1.01	-1	*	0.9	0.87	2	*
K19	0.83	X	3	X	0.89	X	2	X
K20	0.8	X	4	X	0.95	X	1	X
K21	0.96	X	1	X	N/D	X	N/D	X
K22	1	X	0	X	0.78	X	4	X
K23	0.09	X	17	X	1	X	0	X
K24	0.98	X	0	X	0.82	X	3	X
K25	0.67	X	6	X	0.9	X	2	X
K26	0.85	0.99	3	*	0.79	0.93	4	*
K27	X	X	`	X	X	X	X	X
K28	0.76	X	5	X	0.43	X	11	X
K29	0.91	X	2	X	0.78	X	4	X
K30	0.57	X	8	X	0.65	X	7	X
K31	0.96	X	1	X	0.7	X	6	X
K32	0.8	0.02	4	35	0.89	1.04	2	-2
K33	0.85	1.01	3	0	1.05	0.89	-1	2
K34	0.85	0.99	3	*	0.79	0.93	4	*
K35	0.57	X	8	X	0.96	X	0	X
K36	0.91	1.01	2	*	0.88	0.87	2	*
	WT mean=48	WT mean=48.5	WT SD=2.5	WT SD=1.34	WT mean=.98	WT mean=.96	WT SD=.05	WT SD=.04

Base substitutions found in two or three of these lines were eliminated from data used to determine mutation-accumulation rate and spectra. This was done to ensure that mutations preexisting in the progenitor, but not in the reference genome, were excluded. Thus rate determinations were based only on mutations that arose since the establishment of the mutation-accumulation lines. It was not expected that all preexisting mutations found in G0 would be found in K6 or K10 because of differences in sequencing coverage. After removing any base substitutions found in more than one of the three sequenced lines, I found 718 base substitutions in the K6 line, 185 base substitutions in the K10 line, and 48 base substitutions in the G0 progenitor plant (Table 3.2). From the K6 data, I estimate the mutation-accumulation rate to be 3.9×10^{-7} base substitutions per site per generation. A rate determination is not impossible for the G0 plant because there is no information on how many generations that line has been propagated. No determination for K10 at G17 is possible because it was mismatch-repair-proficient after G10. However, based on the mutation rate in wild-type plants, only about four mutations are expected to arise in the seven generations of propagation with functional mismatch repair [56].

In all three lines about one-third of the base substitutions were in intergenic regions, and in K6 more than two-thirds of the mutations in coding regions were nonsynonymous (caused amino-acid substitutions (Table 3.3)) This suggests that there was little selection against nonsynonymous mutations during the 17 generations of propagation. The three mutations in non-message RNA

Table 3.2. High-throughput genome sequencing statistics. Three lines from a *msh2*$^{-/-}$ mutation-accumulation experiment were sequenced (generations of propagation are in parentheses). The G0 line is the progenitor of the experiment, the K10 line was propagated for 17 generations by single-seed descent, but inadvertently replaced by wild-type MSH2 by generation 9, and the K6 line remained mismatch repair-defective.

	MA line (generation)		
	K6 (17)	K10 (17)	Progenitor (0)
Reads aligned ($\times 10^7$)	4.4	4.6	8.8
Percent coverage	96	95	96
Depth	10	9.3	25
Sites interrogated ($\times 10^8$)	1.08	1.0	1.14

Table 3.3. Base-substitution distribution among genome regions. Number of base substitutions identified in each line sequenced, and their distribution among functional classes of genomic regions.

	MA line		
	K6	K10	Progenitor
Intergenic	220	55	18
Coding	167	45	10
Synonymous	38	10	7
Nonsynonymous	129	35	3
Introns	119	18	5
UTR	23	5	3
RNA	3	0	0
Transposable Elements	179	59	11
Pseudogenes	7	3	1
Total base substitutions	718	185	48

genes were in non-coding RNAs. In K6, 13 of the UTR mutations were in 5' UTRs and the remaining 10 were in 3' UTRs.

In the G0 plant, chromosome five is devoid of base substitutions (Figure 3.2), and in K10 there is a region on each chromosome (0-10 Mb on Chr. 1,15-20 Mb on Chr. 2, 1-5 Mb and 19-24 Mb on Chr. 3, 12-17 Mb on Chr. 4, and 13-18 Mb on Chr. 5) that is devoid of mutations. This may indicate that this line outcrossed with a wild-type plant (Figure 3.3). Base substitutions are distributed across all five linkage groups in K6, and they are equally dense along the lengths of the chromosome arms (Figure 3.4) making chromosomal mutation rates roughly equal (Figure 3.5).

The ratio of the rates of transition to transversion mutations (Ts:Tv) for K6 is 4:1 with G:C → A:T ~5-fold higher than A:T → G:C. (Figure 3.6).

3.2 Effect of dominant-negative *msh2* alleles on microsatellite mutation frequency in Arabidopsis

MSH2-deficient plants exhibit altered phenotypes, and accumulate base substitutions at rates consistent with DNA polymerase error rates [57]. Mismatch repair may also be inhibited in plants by the presence of dominant-negative *msh2* alleles.

I determined the microsatellite mutation frequency at dinucleotide repeats in Arabidopsis that had been transformed to carry either helix-turn-helix (HTH) or the ATPase dominant negative *msh2* allele. Microsatellite mutation frequency at four loci in four DN1 and four DN2 lines, two each in a wild-type (Col-0) and

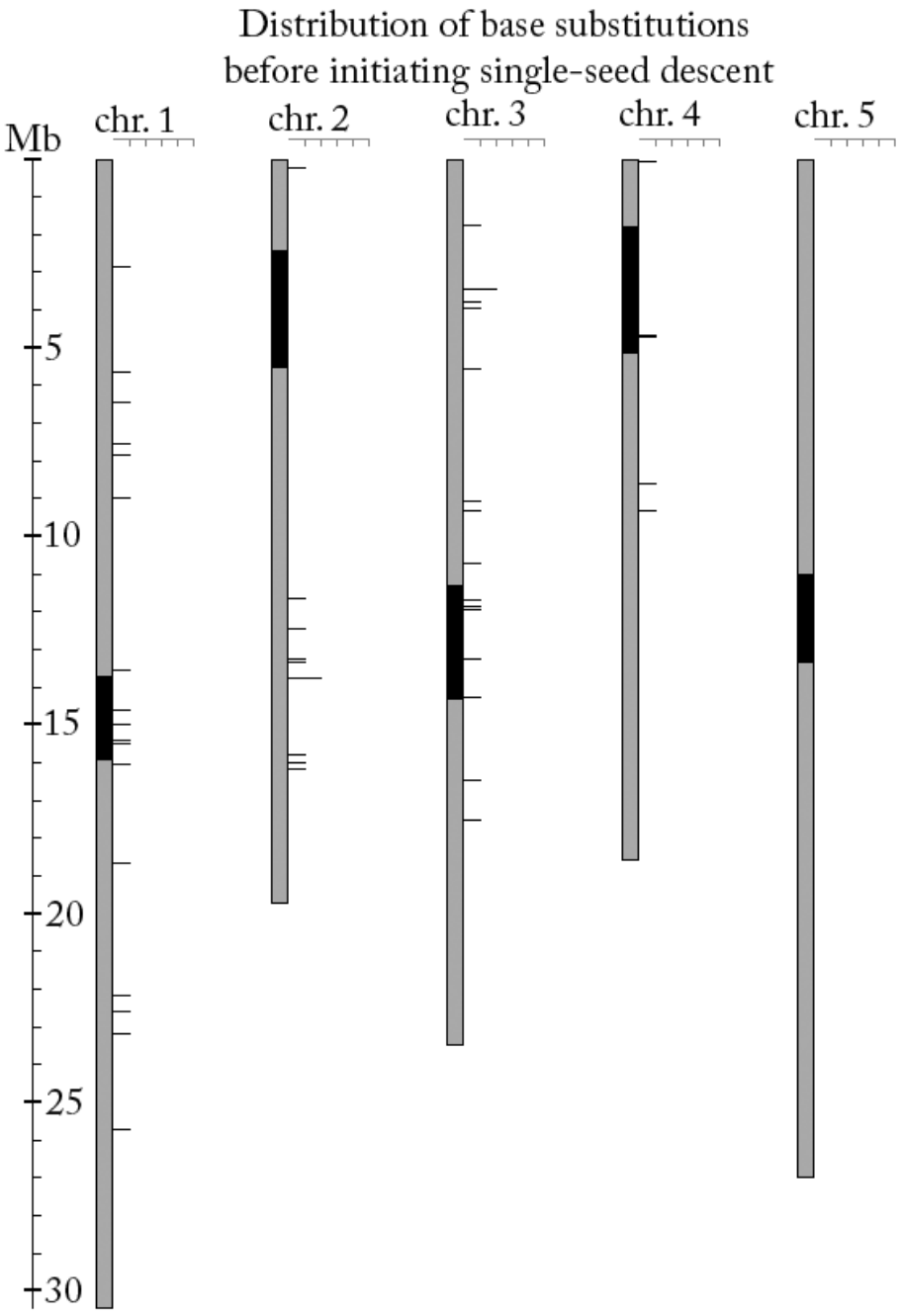

Figure 3.2. Distribution of base substitutions across chromosomes in the mismatch repair-defective G0 progenitor of the mutation-accumulation propagation experiment. Number of mutations in 200 kb sliding windows calculated every 50 kb. Vertical scale is in millions of base pairs (Mb), and black segments represent centromere regions.

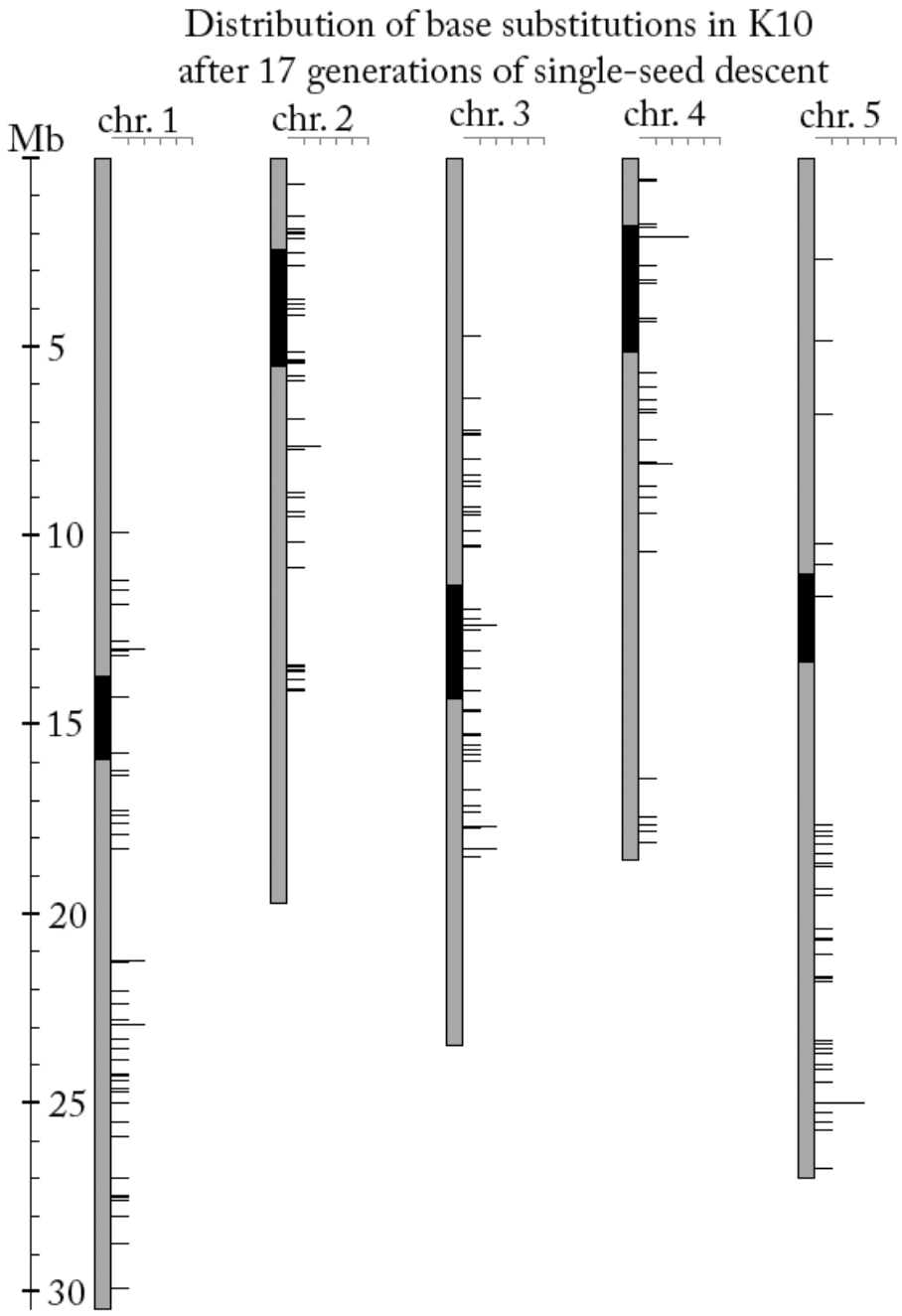

Figure 3.3. Distribution of base substitutions across chromosomes in the K10 mutation-accumulation line. Mutations accumulated over ten generations of propagation as mismatch-repair-deficient and 7 generations of mismatch-repair-proficient propagation. This line perhaps contains a recombinant mixture of DNA from K10 and DNA from the putative mismatch-repair-proficient line with which it crossed. Number of mutations in 200 kb sliding windows calculated every 50 kb. Vertical scale is in millions of base pairs (Mb), and black segments represent centromere regions.

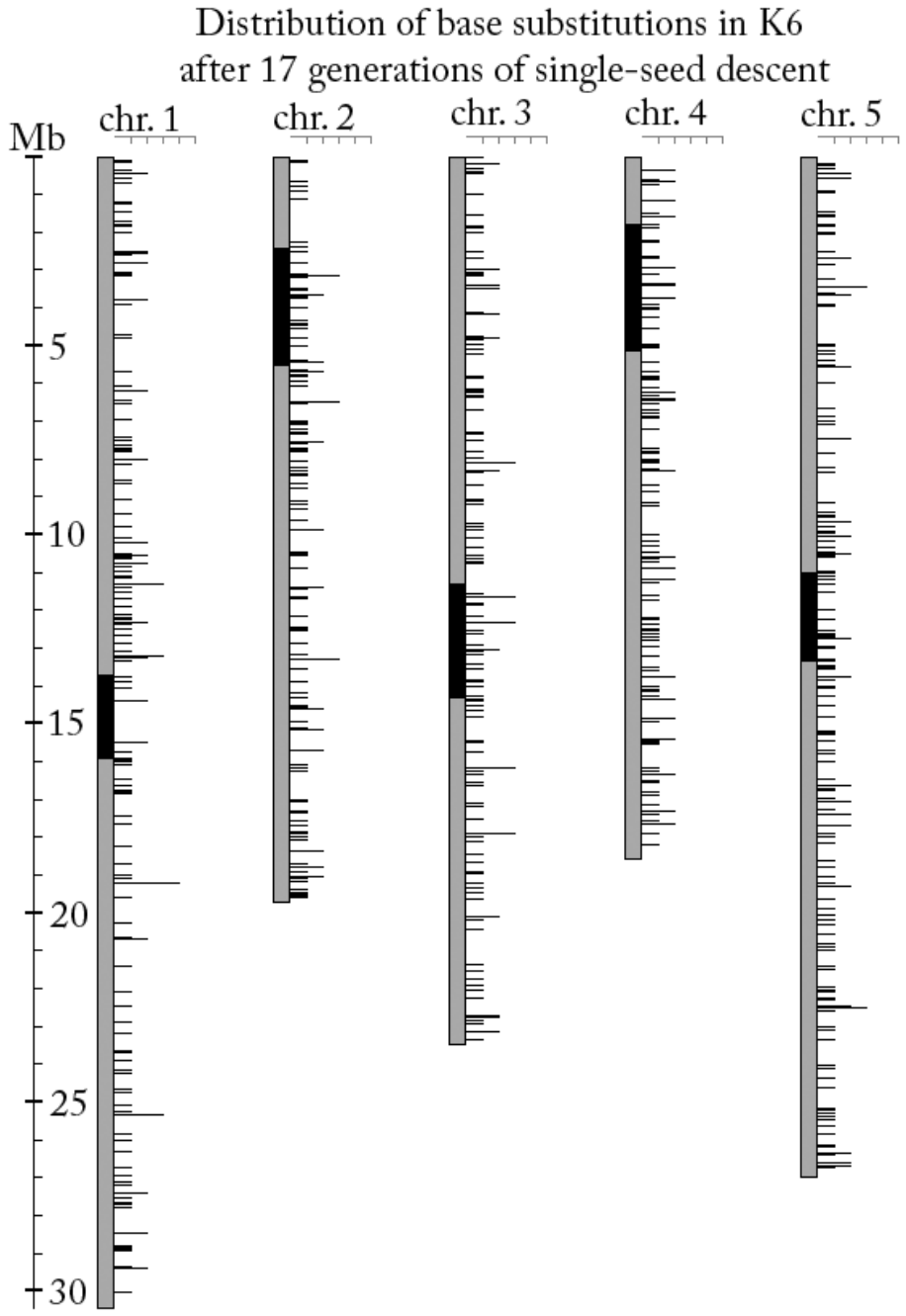

Figure 3.4. Distribution of base substitutions across chromosomes in K6 after 17 generations of propagation by single-seed descent. Number of mutations in 200 kb sliding windows calculated every 50 kb. Vertical scale is in millions of base pairs (Mb), and black segments represent centromere regions.

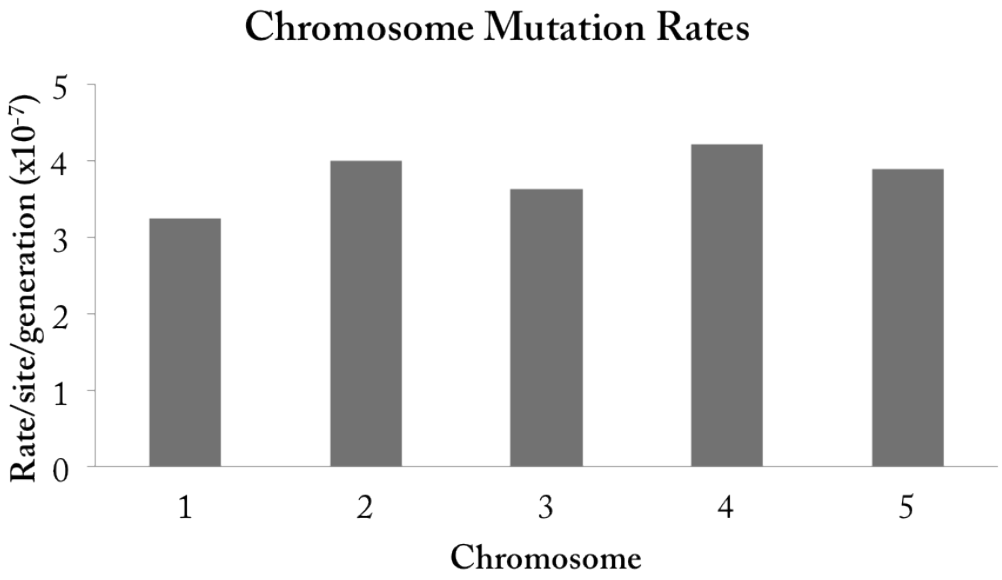

Figure 3.5. Chromosome base substitution rates. Mutation rates per chromosome per generation in 17 generations of mismatch repair-defective propagation.

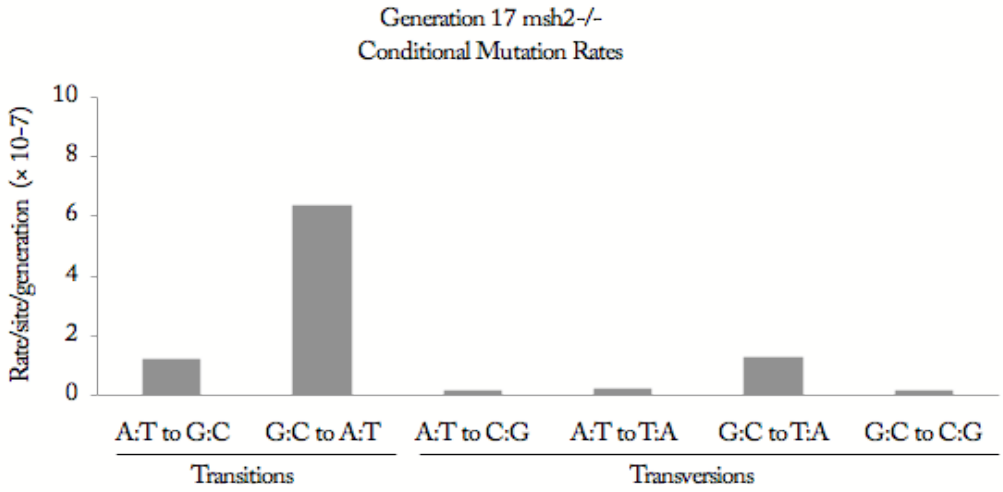

Figure 3.6. Conditional mutation rates per base pair per generation. Conditional mutation rates for each of the six transitions and transversions are calculated using the number of sites in the genome where the particular mutation may occur.

rdr6-15 mutant background, propagated for 3 generations with the transgene were compared to three wild-type lines (Table 3.4). In the DN1 Col-0 and DN2 Col-0 lines there was a two-fold increase in the unique mutation frequency over WT (Figure 3.7). The increase in total mutation frequency over WT was marked in three lines, DN1-1 Col-0, DN2-1 Col-0 and DN1-1 *rdr6-15*, although the latter displayed unique mutation frequencies similar to wild-type. The data, however, do not support elevated mutation frequencies every generation. Comparison to mutation frequencies in *msh2*$^{-/-}$ lines [1] suggest that mutation frequencies may have initially increased in the presence of DN1 and DN2, but then returned to wild-type mutation rates.

3.2.1 Digital PCR

Determining tissue specific microsatellite mutation frequencies required developing a single-molecule, or digital, PCR assay. Modifications to our standard microsatellite progeny analysis protocol included utilizing a high-fidelity DNA polymerase that is fused to an extended DNA binding domain, a hemi-nested primer strategy with two rounds of PCR, and the inclusion of herring-sperm DNA as a carrier. Incorporation of Phusion polymerase, first into the progeny analysis method, and then into the single-template method reduced stutter formation (see section 1.4) and increased the uniformity of the stutter pattern among parallel reactions. The presence of 3% DMSO in the reaction with Phusion is also helpful in limiting stutter. Use of dITP in place of or supplementing dGTP had no positive effect on stutter formation. The lengths of

Table 3.4. Progeny analysis of microsatellite mutation frequency in Arabidopsis carrying dominant-negative *msh2*. The two helix-turn-helix mutant lines (DN2), two ATPase mutant lines (DN1) (both wild-type and the *rdr6-15* mutant), and three wild-type Col-0 lines were analyzed for microsatellite instability by progeny analysis. In every progeny plant, two alleles (diploid) of the microsatellite loci NGA6, NGA8, NGA139, and NGA151 were analyzed. The number of the unique length mutations observed out of the total alleles observed gives the unique mutation frequency, and the number of all mutant alleles observed divided by the total alleles assayed gives the total mutation frequency. Unique mutation frequency is likely more accurate because it is not subject to effects due to number of cell divisions since a mutation arose.

	Col-0				*rdr6-15*				Col-0		
	DN1-1	DN1-2	DN2-1	DN2-1	DN1-1	DN1-2	DN2-1	DN2-1	WT-1	WT-2	WT-3
Number of progeny	39	33	40	37	36	38	44	44	72	72	55
Total alleles	306	252	310	290	274	304	340	344	568	544	378
Unique mutants	2	0	2	1	1	1	1	1	0	2	0
Total mutants	23	0	45	3	63	1	1	1	0	3	0
Unique frequency	0.007	0	0.006	0.003	0.004	0.003	0.003	0.003	0	0.004	0
Total frequency	0.075	0	0.145	0.010	0.230	0.003	0.003	0.003	0	0.006	0

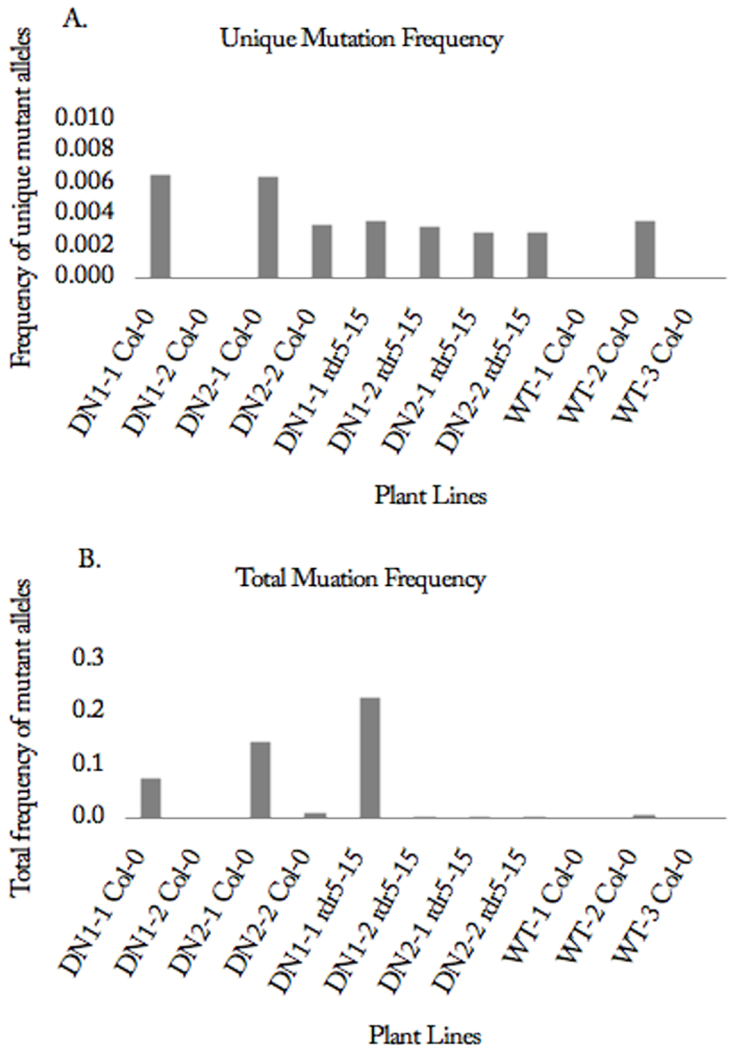

Figure 3.7. Effect of dominant-negative *msh2*. Microsatellite mutation frequency determined by progeny analysis for Arabidopsis carrying dominant negative *msh2* constructs. (A) shows unique mutation frequency and (B) shows the total mutation frequency.

the dinucleotide loci used (>20 repeats) are ultimately the source of extensive stutter. Although I never eliminated stutter from the microsatellite mutation assays, the use of Phusion DNA polymerase and DMSO made stutter patterns more uniform, which allowed reliable scoring of alleles. The conditions that I determined to increase the reliability of PCR microsatellite analysis when amplifying from single target molecules is also applicable to routine microsatellite PCR analysis used in progeny analysis, where the reactions are templated by many target DNA molecules.

Four *Arabidopsis thaliana* microsatellite loci were interrogated at effective single-template DNA concentrations with comparable efficiency in a two-round, hemi-nested PCR strategy (Figure 3.7). The success of PCR reactions for the four dinucleotide-repeat tracts displays similar dependence on DNA template concentration. DNA stocks stored at 1 ng/uL in Herring Sperm DNA at the same concentration and subjected to serial ten-fold dilutions (down to $1:1 \times 10^6$) reproducibly exhibited single-molecule concentrations (by Poisson Distribution of negative PCR reactions). At lower concentrations, the number of positive PCR reactions deviated from what would be expected based on higher concentration DNA PCR, but these differences are likely due to stochastic effects, such as sampling error. The long repeat tracts (~25 dinucleotide repeats) result in considerable polymerase slippage and stuttering.

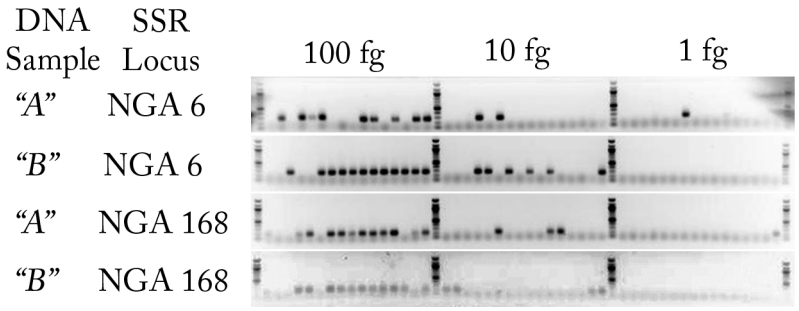

Figure 3.7. Empirical determination of single-template DNA concentration. The Poisson distribution of successful PCR reactions at limiting amounts of DNA was used to estimate the number of target loci in each PCR reaction. The Arabidopsis haploid genome is approximately 100 fg.

3.3 Tissue-Specific Microsatellite Mutation Frequency in Plants

To establish whether mismatch repair is relaxed in somatic plant tissues, and if plants contain the equivalent of a reserved germ-line, I determined the microsatellite mutation frequencies in leaves and in shoot apical meristem in Arabidopsis. Central-zone meristem cells (containing the stem cell population) were collected by fluorescence assisted cell sorting (FACS) protoplasts isolated from *ap1-1 cal1-1* plants carrying a fluorescent reporter to mark meristem cells [28]. The FACS instrument was operated in purity mode to prevent collection of non-GFP cells. The reporter construct, pCLV3::mGFP5-ER, directs expression of green fluorescent protein (GFP) in the central zone cells of the shoot apical meristem, and the *ap1-1 cal1-1* double mutant background causes the conversion of floral meristems into shoot meristems, which results in each plant producing several hundred shoot apical meristems.

Single molecule (digital) PCR analysis of mismatch repair-proficient and -deficient leaves and seeds and mismatch-repair-proficient CLV3-expressing shoot apical meristem cells produced about 100 PCR products in each tissue (Table 3.5). Analysis of these products revealed the following: 1) An approximately 2.5-fold increase in the unique microsatellite mutation frequency at dinucleotide repeats in wild-type leaves compared to wild-type (CLV3-positive) SAM protoplasts. 2) A unique mutation frequency in mismatch-repair-deficient leaves that was almost two-fold higher than wild-type leaves and about 2/3 that in either wild-type or mismatch repair-deficient seeds. 3) Equally high mutation

frequencies in wild-type and mismatch-repair-deficient seeds (Figure 3.9). The total mutation frequencies follow a different pattern: the wild-type SAM cells and wild-type leaf frequencies are equal, and the total microsatellite mutation frequency in wild-type seeds is half that in mismatch-repair-deficient seeds (Figure 3.10).

Here both unique and total mutation frequencies are reported, but unique frequencies are more likely a reliable indicator of mutation frequency because early mutations that predominate after many cell divisions are counted as only one mutation event.

Table 3.5. Tissue-specific digital PCR. DNA was extracted from seeds, leaves, and 35,000 stem cell protoplasts isolated by fluorescently assisted cell sorting and diluted so that on average only single template DNA molecules were amplified in PCR. PCR products were analyzed by capillary electrophoresis, and unique length fragments for NGA6, NGA8, NGA139, and NGA151 were identified. Total mutation frequencies were calculated by dividing the total number of mutant alleles observed by the total number of alleles analyzed. Unique mutation frequencies were calculated by dividing the number of unique lengths for each locus by the total number of alleles analyzed.

	Seed		Leaf		SAM
	msh2	WT	*msh2*	WT	*cal1 ap1*
Total alleles assayed	127	92	99	107	131
Unique mutants	11	8	6	4	2
Total mutants	38	14	17	4	4
Unique frequency	0.087	0.087	0.061	0.037	0.015
Total frequency	0.299	0.152	0.172	0.037	0.031

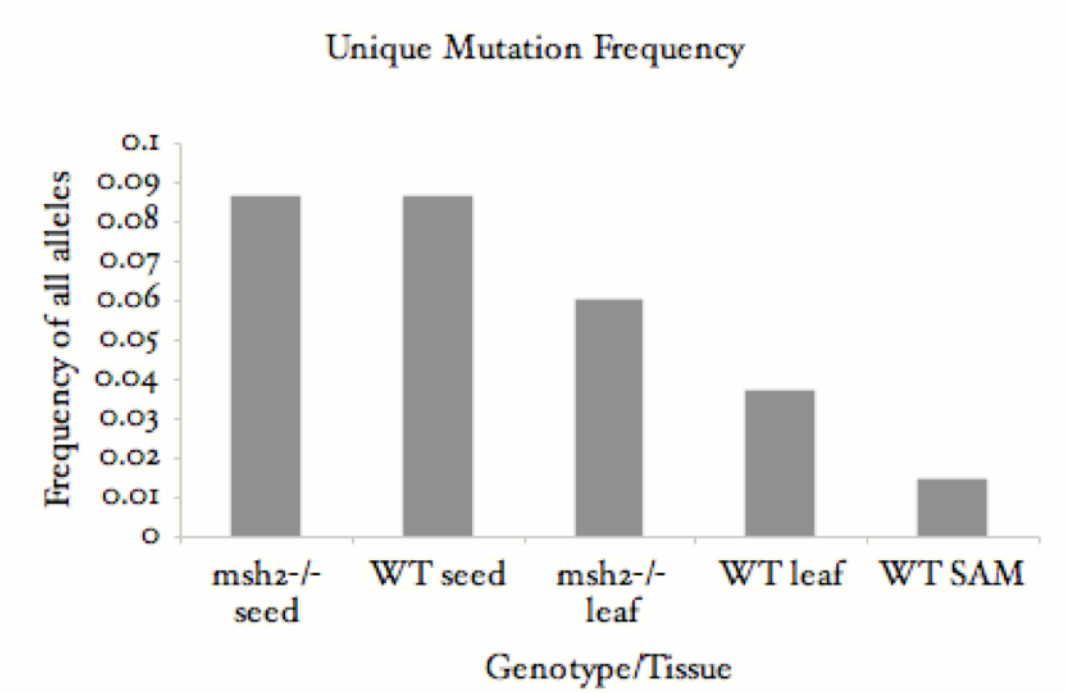

Figure 3.8. Tissue-specific unique microsatellite mutation frequency. Digital PCR analysis of microsatellite mutation in wild-type and *msh2*$^{-/-}$ seed and leaves and in wild-type central zone meristem stem cells.

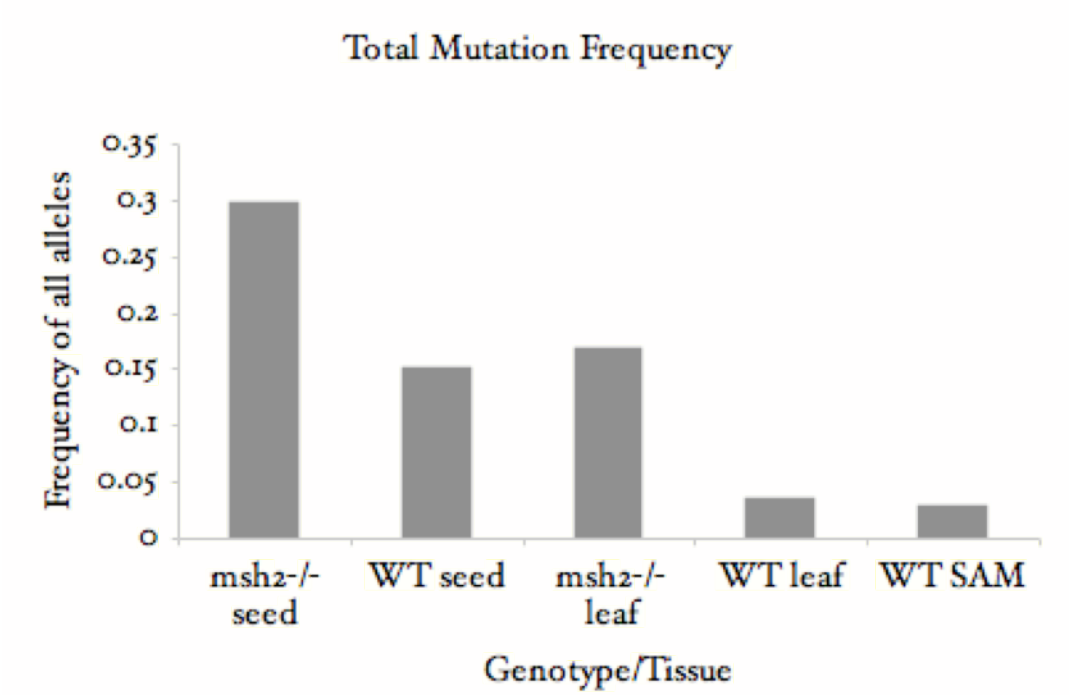

Figure 3.9. Tissue-specific total microsatellite mutation frequency. Digital PCR analysis of microsatellite mutation in wild-type and *msh2*$^{-/-}$ seed and leaves and in wild-type central zone meristem stem cells.

4 Discussion

Plants are sessile and exposed to the environment, typically without shelter. Their genetic material is exposed to assaults on its integrity by both endogenous and exogenous sources, for example, reactive oxygen species (ROS) and UV irradiation, respectively. Despite the lack of a canonical reserved germ line, plants are able to limit the transmission of spontaneous mutations to progeny, and maintain genomic integrity over many generations. Growth and development requirements of the haploid gametophyte may limit passage of some deleterious recessive mutations from one generation to the next (haplosufficiency quality-checking), but data from *msh2*$^{-/-}$ mutation-accumulation lines suggest that mismatch repair is necessary to prevent the accumulation of mutant phenoypes over multiple generations of single-seed descent [4].

If mismatch repair does more to prevent mutation loading than haplosufficiency quality-checking we would expect to see elevated mutation rates in mismatch repair-defective plants. Here I present the first direct estimate of spontaneous mutation rates in mismatch repair-defective plants, and confirm the hypothesis that the absence of mismatch repair in plant meristems increases the rate of spontaneous mutation to a rate similar to estimated DNA polymerase error rates.

4.1 Mutation-accumulation in mismatch repair-deficient Arabidopsis

I continued the mutation-accumulation line propagation for an additional 13 generations while assessing several criteria of fitness. We tracked the ability of the

plants to germinate, develop vegetative structures, and set seed. Lines that are eliminated from the propagation are those that fail to set seed and have therefore likely acquired mutations that have deleterious effects on fertility. Seven mismatch repair-defective lines were "replaced" by plants with wild-type MSH2. This could easily be due to seed contamination, or it could be due to outcrossing with nearby wild-type plants. Outcrossing occurs at a low frequency in Arabidopsis (does not exceed 2% in the wild), but it is expected that this frequency may be significantly higher during mismatch repair-deficient mutation-accumulation propagation because abnormal flowers (as reported in [4]) may facilitate non-self fertilization if they open before pollen maturation [22]. The number and pattern of base substitutions suggests that outcrossing caused the loss of the defective *msh2* allele in the K10 line (see section 4.2.3).

4.2 Base substitution in mismatch repair-deficient Arabidopsis

High-throughput sequencing has allowed direct estimates of spontaneous mutation to be made in human cancers and a variety of species including worms, yeast, and flies [56,58,59,60,61]. Interestingly, mutation rates are correlated with genome size, and there is a clear predominance of G:C → A:T transition mutations in most species [62]. Genome-wide mutation rates in the absence of DNA repair have not been previously reported, but mismatch repair-deficient *Caenorhabditis elegans* showed a ~100-fold increase in mutation rate, based on sequence from 24 randomly selected PCR amplified genomic regions of about 300

to 1300 bp [63,64]. The estimated mutation rate in *msh2*-deficient *C. elegans* was 6.4×10^{-7}.

Here I report an estimate of the base substitution mutation rate in *msh2*$^{-/-}$ *Arabidopsis thaliana* mutator plant lines based on whole genome sequence data. The numbers and positions of base substitutions were determined for three *msh2*$^{-/-}$ Arabidopsis mutation-accumulation lines: one propagated for 17 generations (K6), one propagated for nine generations before outcrossing followed by eight generations of mismatch repair-proficient propagation (K10), and the progenitor (G0) that was already homozygous *msh2*$^{-/-}$ for two generations when the mutation-accumulation experiment began.

Data for K6 and K10 have not been averaged. K10 appears to have outcrossed at generation nine of the mutation-accumulation propagation and become wild-type at the MSH2 locus.

4.2.1 Base substitution rate

In reference [56] it is estimated that the base substitution rate in wild-type Arabidopsis was 5×10^{-9} to 7×10^{-9} in five wild-type mutation-accumulation lines propagated for 30 generations. The corresponding rate for *msh2*$^{-/-}$ mutants elucidated here is 70-fold higher, similar to the increase seen in *C. elegans* [64]. After 17 generations of single-seed descent of *msh2*$^{-/-}$, I sequenced two independent lines derived from a single ancestor. Analysis reveals an average base substitution rate of 3.9×10^{-7} base substitutions per site per generation, which is

two orders of magnitude higher than the rate reported for wild type *A. thaliana* [56], and similar to replicative DNA polymerase error rates. DNA polymerase nucleotide selectivity provides for an error rate of about 10^{-5}, and polymerase proofreading further reduces the error rate by about two orders of magnitude [65]. Mismatch repair lowers the mutation rate by another two orders of magnitude.

4.2.2 Base substitution spectrum

In the mismatch repair-defective lines, ratios of G:C → A:T vs. A:T → G:C are similarly high in mismatch-repair-defective and wild-type [56]. Why might A:T → G:C mutations be not as prevalent as G:C → A:T in the wild-type or *msh2*$^{-/-}$ lines? There is no basis to assume that DNA polymerase exhibits asymmetric proofreading or nucleotide selection such that template-A:nucleotide G would be recognized differently than template-G:nucleotide A. In the presence of mismatch repair, mutations due to UV, oxidative damage, and cytosine deamination are a larger proportion of the total mutations than in mismatch repair-defective plants and may skew the transition:transversion (Ts:Tv) ratio. In the absence of mismatch repair, insertions and deletions are expected to be a significant fraction of total mutations, although their abundance is not assessed here.

Arabidopsis *msh2*$^{-/-}$ and wild-type both display relatively low rates of G:C → T:A transversions in contrast to the rates in wild-type *C. elegans* and yeast [59,60]. Once likely cytosine deaminations are removed from wild-type base substitutions (see reference [56]) the G:C → A:T versus A:T → G:C ratios are roughly the

same (5-fold) in wild-type and *msh2*$^{-/-}$. Thus, if the ratios reflect high UV mutation at PyrimidineC sites and high oxidative mutation at G, mismatch repair would appear to correct this as well. Polymerase misinsertions likely result in both transition mutations equally often, and thus the rate of A:T → G:C may be considered an upper limit for the contribution of polymerase errors to the rate of G:C → A:T (Figure 3.6). Additionally, oxidative damage to guanine likely results in more G →T than G → A, and thus the rate of G:C → T:A may be considered an upper limit for the contribution of oxidative damage to the rate of G:C → A:T. When these individual rates are subtracted from the G:C → A:T rate the remaining mutation rate that may now be attributed to UV is at least 5× the rate considered attributable to oxidative damage (the rate of G:C → T:A). This indirect evidence from Figure 3.6 suggests that UV-induced mutation predominates among the G:C → A:T transitions.

4.2.3 Distributions of base substitutions

A phenomenon that appears to be due to super-efficient mismatch repair is the reduced base substitution in chromosome five in wild-type Arabidopsis [56]. Remarkably, in K6 the base substitution rates are quite similar in all five chromosomes. Notable in this analysis is that the G0 plant shows little clustering of mutations around centromeres as suggested previously for wild-type plants, and chromosome five is devoid of mutations [56]. It is suggested in reference [56] that the deficit of mutations in chromosome five is due to differential distribution of mutations near centromeres in the five chromosomes. Here I show definitively

that the differential mutation rate in chromosome five is abolished in the absence of mismatch repair. Thus, sites resistant to repair would need to be differentially distributed near to centromeres. Figure 3.6 shows a quite uniform base substitution distribution in contrast to wild-type [56] and G0.

In the K10 line outcrossing with wild-type plants is assumed in contrast to contamination with wild-type seed, because the K10 line has retained at Generation 17 many of its accumulated base substitutions, there are regions on each chromosome devoid of mutations (0-10 Mb on Chr. 1, 15-20 Mb on Chr. 2, 1-5 Mb and 19-24 Mb on Chr. 3, 12-17 Mb on Chr. 4, and 13-18 Mb on Chr. 5). These are inferred to be due to crossing over with wild-type chromosomes (compare Figures 3.5 and 3.6 and see section 3.1.2).

The relative abundances of base substitutions in the different regions of the genome also appear to be affected by mismatch-repair status. In the absence of mismatch repair, the rate of mutation per site per generation increased 100-fold in coding regions, 150-fold in introns, and only 30-fold in intergenic regions, based on my data for *msh2* plants and data for wild-type plants from reference [56]. To compare the relative numbers of mutations in genic, intergenic, and transposable element regions in wild-type and *msh2* lines I set the number of genic mutations to one and calculated relative amounts of mutation in the other two genomic regions. In wild-type plants there are 2.0 times as many mutations in intergenic regions as there are in genic regions, and there are 0.54 as many mutations in transposable elements as in genic regions [56]. In *msh2* lines there are 0.71 as many intergenic

mutations as there are genic, and only 0.58 as many in transposable element regions. The three-fold difference between the two relative interginic:genic ratios is independent of the sites sampled in wild-type and *msh2*$^{-/-}$. To oversimplify, there are three times as many mismatch-repair-resistant sites in intergenic as in genic regions. Note, however that the respective mismatch-repair efficiencies are about 99% and 97%. Coding regions may be marked epigenetically to communicate with DNA repair processes to ensure low mutation rates. It may be that intergenic regions actively escape mismatch repair surveillance due to the timing of DNA replication in those regions (late replication or early condensation of nucleosomes after S-phase shutting down mismatch repair) or due to DNA replication in specific DNA replication factories which may have different proteins and repair factors [66,67,68,69].

4.3 Modulation of microsatellite mutation frequency by dominant-negative *msh2* alleles in Arabidopsis

Dominant negative mutants of MSH2 and PMS2 inhibit mismatch repair in animals, yeast, and Arabidopsis. In yeast two independent amino acid substitutions in MSH2 dominantly inhibit mismatch repair. Overexpression of dominant mutant mismatch repair proteins in plants could be used to induce the accumulation of mutations useful in crop breeding programs. Once desired traits are obtained, the dominant mutant allele, maintained as a heterozygote, may be crossed out resulting in a plant that is free from heterologous gene constructs. Homologous mutations in the helix-turn-helix ATPase domains of Arabidopsis

MSH2, non-synonymously substituted, cloned into a Gelvin Super Promoter [53] and transformed into *A. thaliana* by Jeff Leonard and Aly Mohamed. I measured by PCR progeny analysis the effects of mutant alleles of *Atmsh2* on microsatellite mutation frequency. One *Atmsh2* allele, designated Dominant-Negative 1 (DN1), glycine 833 has been replaced by aspartic acid in the ATPase domain. In the second allele, DN2, glycine 671 has also been replaced by aspartic acid in the Helix-Turn-Helix (HTH) domain. Each allele was placed downstream of the Gelvin Super Promoter [53] in the pE1803 binary vector. Aly Mohamed transformed each construct into wild-type Col-0 Arabidopsis, as well as the *rdr6-15* mutant line of Col-0, by *Agrobacterium*-mediated transformation (floral-dip method). The *rdr6-15* mutation confers resistance to silencing of overexpressed recombinant DNA constructs, but the data here show that this may not be important. Mr. Mohamed sowed potential transformed seeds on hygromycin MS agar to select for positive transformants – which he then transplanted to soil, grew to maturity, and collected seeds from. Putative transformants were verified to contain the transgene by PCR.

Putative over-expression of DN1 and DN2 alleles of AtMSH2 is associated with increased microsatellite mutation frequency in Arabidopsis. DN1 and DN2 in the Col-0 background clearly show more microsatellite mutations in one of their respective independently transformed lines. DN1 in *rdr6-15* (a3-13) shows modest or significant total mutation frequencies, but modest unique frequencies,

which may represent a single early mutation that was inherited by multiple progeny

Unique microsatellite mutation frequencies are preferred in reporting results of progeny analysis because it is impossible to determine when mutations arise – more recent germ-line mutations are likely to be in fewer progeny than older mutations, and the total frequency of recent mutations will be lower in a population of progeny. The unique mutation frequency is considered to be dependent on the rate of mutation while the total mutation frequency depends on when mutations arise and at what rate.

The average frequency of inherited unique germ-line microsatellite mutations in wild-type Arabidopsis used for controls in the progeny analysis was 0.001 – a frequency within the range expected for wild-type plants. This low frequency is in contrast to apparently high somatic-mutation frequencies reported in reference [1]. I attempted to reconcile this difference by determining the mutation frequency in putative germ-line cells within the shoot apical meristem.

4.4 Digital PCR

In reference [4] the frequency of formation of shifted alleles over five generations of propagation of self-fertilizing plants that were either mismatch repair-deficient or wild-type is described. Microsatellites were PCR amplified using fluorescently labeled primers, and then sized by capillary electrophoresis. A

limitation of this technique is that it is difficult to discern multiple low-abundance mutant alleles.

Polymerase slippage during PCR amplification of repeat sequences produces artifactual shifted products called stutter bands. If a plant tissue contains a mutant allele that comprises less than about 25% of all alleles for a locus, it is impossible to consistently distinguish differences between mutations and PCR artifacts using standard PCR procedures [4].

To address this limitation the authors of reference [4] assayed the microsatellite mutation frequency of an individual plant indirectly, using DNA from 16 or 32 progeny of that plant. Progeny analysis limits artifacts from somatic mutations: even if a mutant allele is present in only one or a few cells in a mature shoot apical meristem, if those particular cells give rise to some of the gametes and thus to progeny their alleles will likely be present in all cells of several progeny.

The frequency of rare microsatellite mutations has been assayed in mice using single-template PCR or small-pool techniques. Genomic DNA in the PCR reactions is diluted so PCR reactions contain about one amplifiable template on average; such that each allele is amplified in isolation from all others. If enough individual alleles are interrogated, this approach greatly increases the sensitivity of the PCR to infrequent mutations.

The single-template technique allows us to address for the first time an important question: Do plants relax mismatch repair in vegetative tissues relative

to meristematic "germ-line" and floral tissue? Preliminary experiments examining two loci in Msh2-null and wild type plants unexpectedly indicated that there was no difference in microsatellite instability between leaves and seeds. Mismatch repair-deficient plants showed only two-fold higher mutation frequency compared to wild type. The high frequency of mutation in vegetative tissues of wild-type plants was unexpected and will be examined further. Repeated and enlarged data sets are necessary to verify this trend. These may guide plant propagation and mutation breeding techniques.

4.5 Microsatellite mutation in Arabidopsis leaves, seeds, and shoot apical meristems

Analysis of microsatellite-repeat-number mutations in vegetative tissues by single-template approach has revealed higher than expected mutation frequencies. At most there was a two-fold difference in unique mutation frequency at dinucleotide repeats between MSH2 proficient and deficient leaves. These data are consistent with the idea that mismatch repair is relaxed in wild-type leaves (or alternatively that there is a process producing a high level of microsatellite mutation that is not corrected by mismatch repair).

A direct approach is to analyze only DNA from putative germ-line cells of the shoot apical meristem. Venu Reddy and his colleagues have developed Arabidopsis shoot apical meristem fluorescent marker lines which allow various portions of the shoot apical meristem to be sorted and collected by fluorescence assisted cell sorting (FACS) after protoplasting whole tissue [28]. To produce the

amount of shoot tip tissue needed FACS and for the PCR assays, it was necessary to use *cal1 ap1* plants. They are prolific producers of shoot apical meristems, since floral meristems are converted to shoot apical meristems. It is assumed, but not certain, that the stem cells present in the converted floral meristems of *cal1 ap1* mutants are not de-differentiated, but have been maintained in a pluripotent state in the stem cell lineage leading to the floral meristem and destined to create megaspores and microspores. The stem cells may also have undergone increased numbers of cell divisions in rebuilding SAMs after transitioning to flowering. Of about 35 CLV3-expressing "stem" cells in a meristem, only a few are "permanent stem cells" that are carried at the apex. The remaining cells are not necessarily always stem cells and usually differentiate as they are incorporated into organ primordia. Cells at the periphery of the pool of CLV3-expressing cells may have undergone several cell divisions more than might be expected for putative pure "permanent stem cells." The germ line must be considered probabilistically. It may consist of cells that move in and out of the stem cell population, although infrequently. Thus the CLV3-expressing cells in total may define the germ line.

4.6 Plant germ line

The germ-line mutation frequency determined by progeny analysis (Table 3.4, Col-0 columns) is about 4-fold lower than the digital-PCR-determined frequency of microsatellite mutations in shoot apical meristem central-zone cells, which contain the stem cell population. The elevated mutation frequency in collected

central zone meristem cells may reflect contamination by differentiated cells, or it may be a consequence of the altered meristem development in the *ap1 cal1* double mutant used to facilitate the capture of stem cells [28]. Selection against aberrant recombination during meiosis, DNA mismatch repair during vegetative growth and development, and counter selection during haploid gametogenesis likely limit the passage of gross chromosomal rearrangements and more localized mutations from one generation to the next, but these activities do little to prevent mutations in differentiated somatic tissues.

Seeds were expected to exhibit less mutation than leaves due to a simplistic hypothesis that the number of cell divisions required to build a seed was less than the number needed to build leaves. Seeds, however, are composed of a sporophyte seed coat, triploid endosperm, which nourishes the developing embryo, and the sporophyte embryo contained within a gametophyte embryo sac.

In plants, meiosis does not directly produce gametes, but haploid products of meiosis grow by two to three mitotic divisions; some progeny become gametes and others become haploid vegetative cells. Requirements for growth in the haploid stage are thought to provide selection against deleterious mutations, and thus help prevent accumulation of mutations in progeny. In the seed, the embryo accounts for 500 cells of a total 6000-7000 of cells (exclusive of the seed coat), and the embryonic shoot and root meristems account for at most a tenth of embryo cells. Two of these protomeristem cells are thought to represent the germ line, because they are the progenitors of all gametes [12] .

The embryo does not contain a shoot meristem until approximately mid- to late-torpedo stage [70,71]. There is little advantage to protecting the endosperm and non-embryonic seed from mutations since these are "dead-end" tissues that do not contribute directly to the next generation. At that point, the meristem contains about two stem cells whose descendants will eventually include the gametes. Mutations that occur in the parent and are inherited wholly by offspring will be represented in all tissues of the offspring. This implies more or less random selection from among a much larger (>>2) pool of cells that are the ancestors of progeny. Recent mutations, however, will not end up evenly distributed throughout the plant. A mutation carried in one of the two stem cells in the embryo that will give rise to the "permanent" stem cells will, at most, populate one-half of the plants' tissues, germ cells, and progeny [29]. Later mutations in the enlarged dividing stem cell population end up in fewer offspring. Additionally, the stem cell population is not fixed; that is, stem cells may leave the stem cell pool, and meristematic differentiated cells, may join the stem cell pool, carrying any new mutations with them. Plants likely maintain higher levels of mismatch repair in both the permanent stem cells (when they (rarely) divide) and the 35 or so dividing stem cells. My data show that the mutation frequency in the shoot apical meristem of wild-type plants may be half the frequency as in wild-type leave. The increase in microsatellite mutation frequency in leaves due to putative mismatch-repair deficiency is much less than the corresponding elevations in mismatch-repair-defective yeast and mice [30,72]. It is, however, 15 times the low

wild-type microsatellite mutation rate measured in Arabidopsis by progeny analysis. Two factors that may contribute to the differences in mutation frequency among shoot apical meristems, leaves, and seeds are elevated mismatch repair activity in stem cells and reduced cell division in a lineage of "permanent stem cells" that produce microspore mother cells and megaspore mother cells.

Differentiated cells in Arabidopsis undergo DNA replication without cell division in a process called endoreduplication, in which a cell may end up with 30 copies or more of its genome [73]. Mismatch repair may not be crucial in these DNA duplications and its absence may lead to increased numbers of small sectors of mutation in vegetative tissue. Mutations in stem cells in the germ line will be propagated with each subsequent cell division of the mutant stem cell and with the divisions of its descendants as they build tissues. Thus mutations arising in these parental cells will be widespread in the plant. Conversely, a mutation may arise in transiently amplifying cells or in differentiating cells in the peripheral zone of the meristem. Mutations arising here, outside of the permanent stem cells, will be less widespread in tissues, resulting in lower total mutation frequencies than mutations that arise in stem cells. Therefore only unique mutations are meaningful. (However, tissues may accumulate more mutations as they grow).

The shoot-apical-meristem microsatellite-mutation frequency (assayed by digital PCR) was 0.015. The germ-line microsatellite mutation frequency (assayed by progeny analysis) was 0.001. This may not be a strictly valid comparison, but it indicates that the microsatellite mutation frequency obtained for SAM stem cells

by single template PCR is too high to explain the low germ-line mutation frequency obtained by progeny analysis. Mismatch repair may relax after cells leave the pool of permanent stem cells and become dividing stem cells that give rise to leaves. Alternatively, several factors may contribute to limit passage of mutations to progeny.

Mismatch repair is only one of a variety of integrated systems that protect the genome. In plants, the lack of an early-reserved germ line as in animals may be a protective measure that allows cell fates to be determined late in development, to avoid the eventual transmission to progeny cells that have sustained damage during growth. Mutation frequencies in offspring may reflect mismatch repair, late haplosufficiency quality checking that eliminates aneuploid cells that survived meiosis, and perhaps an early quality check of some kind before cells enter the germ line. It has been shown in root apical meristems that programmed cell death eliminates grossly damaged cells from the stem cell population, and perhaps a similar mechanism protects the germ line from substantial damage such as double strand breaks in chromosomes [74,75,76]. Progeny analysis and genome sequencing data indicate that mismatch repair reduces the frequency of inherited microsatellite mutations about 14-fold and reduces the rate of inherited base substitutions in Arabidopsis ~100-fold. Conversely, the shoot apical meristem exhibits elevated levels of microsatellite mutation frequency – about equal to one-half the germ-line (progeny analysis) mutation frequency in mismatch repair-defective plants [1]. This is possibly due to heavy contamination, but even if there

were no contamination, the structure and maintenance of the meristem could be responsible. Between ten and thirty-five stem cells divide infrequently to produce meristem (transiently amplifying) cells that divide to build the tissues of organ primordia [28,29]. Thus, the majority of the cells that divide to build the meristem and produce organs are not stem cells, although they originated from stem cells.

The permanent and dividing stem cells sit atop the growing stalk in the upmost layer of the apical meristem supported by the peripheral zone and rib zone, which are both responsible for tissue formation and bolting [77]. In this way, a few protected permanent stem cells may be carried by multiple divisions of their few direct descendants until the protected permanent stem cells are selected to become megaspore mother cells or microspore mother cells. Genome integrity is protected in stem cells that rarely replicate their DNA because there are fewer opportunities for mutation. The core set of dividing stem cells is not fixed: some stem cells may leave and differentiated meristem cells may dedifferentiate and rejoin, but these event are rare.

I determined shoot apical meristem mutation frequencies based on flow-sorted GFP-positive cells putatively from the central zone (CZ) of the shoot apical meristem. Even if the collected sample were 100% CZ, mutations may arise during divisions of the 30 or so stem cells that maintain this zone and provide daughter cells that eventually give rise to organs. Their mutation frequency would be higher than in the permanent stem cells, and would increase with each cell

division. The structure of the shoot apical meristem, with its multiple layers and apically located stem-cell initials, may support the function of genome preservation by limiting the number of cell divisions required of "permanent" stem cells before gametophyte germ line determination.

It appears that mutation avoidance processes relax soon after a cell leaves the "stem-cell niche" (the ~35 permanent and dividing stem cells). Reestablishment of stem cell niches and apical meristems, as occurs during vegetative propagation and callus formation [78,79,80], may eventually halt progression of mutation, but will not reverse pre-existing mutations during vegetative propagation and callus formation. A solution to this problem might be to create transgenic plants that constitutively express all necessary components of mismatch repair in somatic tissues, but it is unclear the effect this may have on plant growth. Additionally, other DNA repair processes such as nucleotide excision repair and base excision repair may be reduced in somatic tissues, and may also have roles in preventing somaclonal variation.

Mismatch-repair defects cause increases in the rates of inherited mutations, and might be inhibited to effect agronomically useful mutations and traits. Mutation breeding is the use of mutagens to induce mutations in plants that may be useful in crop production. Mutagens are usually applied in one or several doses. Mutation breeding relies on alkylating agents such as ethyl methanesulfonate (EMS), gamma-ray and X-ray irradiation, solar radiation, and colchicine to induce mutations, genome rearrangements, and changes in ploidy.

Inhibiting mismatch repair does not induce mutations or damage DNA; instead it is a chronic condition in which spontaneous errors by the replicative DNA polymerase go uncorrected. The spectrum of mutations in *msh2* plants is largely G:C → A:T, similar to but not as extreme as the spectrum induced by the popular mutagen EMS, but A:T → G:C and G:C → T:G are substantial (as are indels at short mononucleotide repeats). The mutation accumulation rate in the absence of mismatch repair is then dependent on factors such as polymerase error rate. Engineering the replicative polymerase or affecting nucleotide pools would make it possible to alter the spontaneous base substitution spectrum, or even to further increase the rate of mutation. Perhaps mutagens or base analogs with desired specificities and resultant mutation spectra could be used in combination with mismatch-repair inhibition.

5 Conclusion

Mutation breeding techniques are used to obtain desirable traits by selecting from populations that genetic variation that has been induced by mutagen treatment or propagation techniques that induce somaclonal variation. Current methods utilize EMS, UV, azide, colchicines, gamma-irradiation, as well as cosmic rays it induce mutations in crops. The spontaneous mutations that result from uncorrected DNA polymerase errors may also be useful in breeding programs. The effects of polymerase errors, C → U deamination, and UV-photoproducts cause single base substitutions and short nucleotide insertions and deletions, but the base-base mismatches and insertion and deletion loops that occur as a byproduct of incorrect DNA polymerization are corrected by mismatch repair.

My data suggest that mismatch repair corrects UV-induced mutations, although correction of U:G in nonreplicationg DNA is unlikely. We have confirmed an ~100-fold increase in mutation rate in the absence in mismatch repair, and this estimate agrees with the error rates for replicative DNA polymerases. Inhibition of mismatch repair by dominant-negative proteins may be a useful means to generate chronic mutation that will accumulate in progeny. Dominant-negative repression of mismatch repair is promising because it offers an avenue to genetically transform polyploidy crops that may have multiple copies of target genes. Only one copy of the transgene is necessary, and targeting multiple gene copies in the genome is unnecessary. Once traits have been identified for marker assisted selection and further breeding the inactivating construct may be crossed

out of the plants to create a mutagenized population that ultimately does not contain any transgene DNA.

Meristem germ-line mutation frequencies measured here by single-template PCR do not reconcile the meager five-fold increase in mutation frequency seen when comparing GUS homopolymer frameshift reversion in wild-type and mismatch repair-defective Arabidopsis with low rates of germ line transmission of mutations in wild-type plants. My inability to detect lower mutation frequencies in meristem protoplasts may be attributed to plant structure and development. Perhaps the protoplast isolation was contaminated with differentiated cells, and perhaps the mutation frequency in the meristem cells descended from dividing stem cells is elevated due to relaxed mismatch repair in this population. A definitive answer as to whether plants relax mismatch repair in somatic tissue must wait until a *msh2* mutant is crossed into the *ap1 cal1* background to allow the collection of mismatch repair-defective shoot apical meristem tissue. With that tissue one can ascertain whether genomes in shoot apical meristems and leaves mutate equally in the absence of mismatch repair. In the dormant seed there are approximately two germ-line cells, and these two germ cells are descendants of two germ-line cells (sperm and egg formed by two and three mitotic divisions following meiosis) in the previous gametophyte. This reserved packet of protected growth has been shown to express mismatch-repair and other DNA-repair genes at elevated levels [28]. Conversely, somatic tissue in plants endoreduplicate to produce up to 30 copies of the genome per cell, and one may consider that the germ-line is

"protected" from the standpoint that it doesn't undergo any more replications than necessary. There are at least five method plants may utilize to protect the genome in their germ lines: 1) Mismatch repair activity in permanent and dividing stem cells corrects base substitutions and short insertions and deletions associated with DNA replication; 2) permanent stem cells may be severely limited in the number of cell divisions they undergo before producing gametophytes thus limiting exposure of the genome to replication; 3) Programmed cell death may remove cells with badly damage chromosomes from the stem cell population; 4) meiosis eliminates cells with gross chromosomal rearrangements that fail to segregate correctly; and 5) haplosufficiency quality checking may eliminate any cells with deleterious dominant mutations and cells without the necessary full chromosome complement required for gametophyte growth. The elevated level of spontaneous mutation can be achieved in germ-line cells expressing dominant-negative alleles of *MSH2*. Sustained, uncorrected spontaneous mutations can be used to create variety for breeding.

As we increase what we know about mismatch repair, we gain greater understanding of what it may take to manipulate the system for further understanding and utility. Ironically the very mutation processes we seek to limit in ourselves are what we desire of our breeding stock for trait generation and selection. Modulating mismatch repair for medicine and agriculture will prove fruitful in time and with continued research.

Bibliography

1. Leonard JM, Bollmann SR, Hays JB (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.
2. Yao X, Buermeyer AB, Narayanan L, Tran D, Baker SM, et al. (1999) Different mutator phenotypes in Mlh1- versus Pms2-deficient mice. Proc Natl Acad Sci U S A 96: 6850-6855.
3. Tam SM, Hays JB, Chetelat RT (2011) Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato. Theor Appl Genet 123: 1445-1458.
4. Hoffman PD, Leonard JM, Lindberg GE, Bollmann SR, Hays JB (2004) Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective Arabidopsis. Genes Dev 18: 2676-2685.
5. Dekker M, de Vries S, Aarts M, Dekker R, Brouwers C, et al. (2011) Transient suppression of MLH1 allows effective single-nucleotide substitution by single-stranded DNA oligonucleotides. Mutat Res 715: 52-60.
6. Bollmann SR, Tominey CM, Hoffman PD, Hoffman TM, Hays JB (2011) Reversion-reporter transgenes to analyze all six base-substitution pathways in Arabidopsis. Plant Physiol 155: 1286-1300.
7. Van der Auwera G, Baute J, Bauwens M, Peck I, Piette D, et al. (2008) Development and application of novel constructs to score C:G-to-T:A transitions and homologous recombination in Arabidopsis. Plant Physiol 146: 22-31.
8. Yoshihara R, Nakane C, Takimoto K (2006) A new system for detecting mutations in arabidopsis thaliana and the mutational spectra resulting from ethylmethanesulfonate treatment. J Radiat Res (Tokyo) 47: 223-228.
9. Kovalchuk I, Kovalchuk O, Hohn B (2000) Genome-wide variation of the somatic mutation frequency in transgenic plants. EMBO J 19: 4431-4438.
10. Berger F, Twell D (2011) Germline specification and function in plants. Annu Rev Plant Biol 62: 461-484.
11. Twell D (2011) Male gametogenesis and germline specification in flowering plants. Sexual plant reproduction 24: 149-160.
12. Rédei GP (1992) Classical Mutagenesis. In: Koncz C, Chua N-H, Schell J, editors. Methods in Arabidopsis Research. Singapore: World Scientific Publising Co. Pte. Ltd. pp. 482.
13. Walbot V, Evans MM (2003) Unique features of the plant life cycle and their consequences. Nat Rev Genet 4: 369-379.
14. Hoeijmakers JH (2009) DNA damage, aging, and cancer. N Engl J Med 361: 1475-1485.

15. Bray CM, West CE (2005) DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity. New Phytol 168: 511-528.
16. Britt AB (1996) DNA Damage And Repair In Plants. Annu Rev Plant Physiol Plant Mol Biol 47: 75-100.
17. Pluciennik A, Dzantiev L, Iyer RR, Constantin N, Kadyrov FA, et al. (2010) PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. Proc Natl Acad Sci U S A 107: 16066-16071.
18. Nicolaides NC, Littman SJ, Modrich P, Kinzler KW, Vogelstein B (1998) A naturally occurring hPMS2 mutation can confer a dominant negative mutator phenotype. Mol Cell Biol 18: 1635-1641.
19. Studamire B, Price G, Sugawara N, Haber JE, Alani E (1999) Separation-of-function mutations in *Saccharomyces cerevisiae* MSH2 that confer mismatch repair defects but do not affect nonhomologous-tail removal during recombination. Mol Cell Biol 19: 7558-7567.
20. Alani E, Sokolsky T, Studamire B, Miret JJ, Lahue RS (1997) Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. Mol Cell Biol 17: 2436-2447.
21. Wu TH, Marinus MG (1994) Dominant negative mutator mutations in the mutS gene of Escherichia coli. J Bacteriol 176: 5393-5400.
22. Rédei GP, Libbenga K, Torrey J, Kamra OP, Kesavan PC, et al. (1975) Arabidopsis as a genetic tool. Annu Rev Genet 9: 111-127.
23. Initiative TAG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815.
24. Page DR, Grossniklaus U (2002) The art and design of genetic screens: Arabidopsis thaliana. Nat Rev Genet 3: 124-136.
25. Zhang X, Henriques R, Lin SS, Niu QW, Chua NH (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat Protoc 1: 641-646.
26. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40: D1202-1210.
27. Reddy GV, Meyerowitz EM (2005) Stem-cell homeostasis and growth dynamics can be uncoupled in the Arabidopsis shoot apex. Science 310: 663-667.
28. Yadav RK, Girke T, Pasala S, Xie M, Reddy GV (2009) Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. Proc Natl Acad Sci U S A 106: 4941-4946.
29. Li SL, Rédei GP (1969) Estimation of mutation rate in autogamous diploids. Radiation Botany 9: 125-131.

30. Tran HT, Keen JD, Kricker M, Resnick MA, Gordenin DA (1997) Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. Mol Cell Biol 17: 2859-2865.
31. Baker SM, Bronner CE, Zhang L, Plug AW, Robatzek M, et al. (1995) Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. Cell 82: 309-319.
32. Gurtu VE, Verma S, Grossmann AH, Liskay RM, Skarnes WC, et al. (2002) Maternal effect for DNA mismatch repair in the mouse. Genetics 160: 271-277.
33. Cinalli RM, Rangan P, Lehmann R (2008) Germ cells are forever. Cell 132: 559-562.
34. Twell D (2011) Male gametogenesis and germline specification in flowering plants. Sex Plant Reprod 24: 149-160.
35. Yadegari R, Drews GN (2004) Female gametophyte development. Plant Cell 16 Suppl: S133-141.
36. Tautz D, Schlotterer (1994) Simple sequences. Curr Opin Genet Dev 4: 832-837.
37. Moxon R, Bayliss C, Hood D (2006) Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet 40: 307-333.
38. Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 18: 1161-1167.
39. Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol 7: R14.
40. Sia EA, Jinks-Robertson S, Petes TD (1997) Genetic control of microsatellite stability. Mutat Res 383: 61-70.
41. Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274-276.
42. Loeb LA (1994) Microsatellite instability: marker of a mutator phenotype in cancer. Cancer Res 54: 5059-5063.
43. Azaiez A, Bouchard EF, Jean M, Belzile FJ (2006) Length, orientation, and plant host influence the mutation frequency in microsatellites. Genome 49: 1366-1373.
44. Vogelstein B, Kinzler KW (1999) Digital PCR. Proc Natl Acad Sci U S A 96: 9236-9241.
45. Coolbaugh-Murphy M, Maleki A, Ramagli L, Frazier M, Lichtiger B, et al. (2004) Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR. Genomics 84: 419-430.
46. Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res 19: 1349.

47. Pomraning KR, Smith KM, Freitag M (2011) Bulk segregant analysis followed by high-throughput sequencing reveals the Neurospora cell cycle gene, ndc-1, to be allelic with the gene for ornithine decarboxylase, spe-1. Eukaryot Cell 10: 724-733.
48. Li R, Yu C, Li Y, Lam TW, Yiu SM, et al. (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966-1967.
49. White RA, 3rd, Blainey PC, Fan HC, Quake SR (2009) Digital PCR provides sensitive and absolute calibration for high throughput sequencing. BMC Genomics 10: 116.
50. Bargmann BO, Birnbaum KD (2010) Fluorescence activated cell sorting of plant protoplasts. J Vis Exp.
51. Lee LY, Gelvin SB (2008) T-DNA binary vectors and systems. Plant Physiol 146: 325-332.
52. Min Ni DC, Jane Einstein, Soma Narasimhulu, Claudia Vergara Stanton B Gelvin (1995) Strength and tissue specificity of chimeric promoters derived from the octopine and manopine synthase genes. The plant journal 7: 661-676.
53. Lee LY, Kononov ME, Bassuner B, Frame BR, Wang K, et al. (2007) Novel plant transformation vectors containing the superpromoter. Plant Physiol 145: 1294-1300.
54. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev 18: 2368-2379.
55. Li R, Li Y, Fang X, Yang H, Wang J, et al. (2009) SNP detection for massively parallel whole-genome resequencing. Genome Res 19: 1124-1132.
56. Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327: 92-94.
57. Kunkel TA (2009) Evolving views of DNA replication (in)fidelity. Cold Spring Harb Symp Quant Biol 74: 91-101.
58. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, et al. (2010) The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature 465: 473-477.
59. Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, et al. (2009) A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. Proc Natl Acad Sci U S A 106: 16310-16314.
60. Lynch M, Sung W, Morris K, Coffey N, Landry CR, et al. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci U S A 105: 9272-9277.
61. Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, et al. (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res 19: 1195-1201.

62. Lynch M (2007) The origin of genome architecture. Sunderland, MA: Sinauer Associates, Inc.
63. Denver DR (2003) Phylogenetics in *Caenorhabditis elegans*: An Analysis of Divergence and Outcrossing. Molecular Biology and Evolution 20: 393-400.
64. Denver DR, Feinberg S, Estes S, Thomas WK, Lynch M (2005) Mutation rates, spectra and hotspots in mismatch repair-deficient *Caenorhabditis elegans*. Genetics 170: 107-113.
65. Kunkel TA (2009) Evolving views of DNA replication (in)fidelity. Cold Spring Harbor symposia on quantitative biology 74: 91-101.
66. Leonhardt H, Rahn HP, Weinzierl P, Sporbert A, Cremer T, et al. (2000) Dynamics of DNA replication factories in living cells. J Cell Biol 149: 271-280.
67. Kitamura E, Blow JJ, Tanaka TU (2006) Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. Cell 125: 1297-1308.
68. Lee TJ, Pascuzzi PE, Settlage SB, Shultz RW, Tanurdzic M, et al. (2010) Arabidopsis thaliana chromosome 4 replicates in two phases that correlate with chromatin state. PLoS Genet 6: e1000982.
69. De S, Michor F (2011) DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. Nat Biotechnol 29: 1103-1108.
70. Barton MK, Poethig RS (1993) Formation of the shoot apical meristem in *Arabidopsis thaliana*: an analysis of development in the wild type and in the shoot meristemless mutant. Development 119: 823-831.
71. Jürgens G (1994) Pattern Formation in the Embryo. In: Meyerowitz EM, Somerville CR, editors. Arabidopsis. Plainview, NY: Cold Spring Harbor Laboratory Press. pp. 297-312.
72. Buermeyer AB, Deschenes SM, Baker SM, Liskay RM (1999) Mammalian DNA mismatch repair. Annu Rev Genet 33: 533-564.
73. Larkins BA, Dilkes BP, Dante RA, Coelho CM, Woo YM, et al. (2001) Investigating the hows and whys of DNA endoreduplication. J Exp Bot 52: 183-192.
74. Furukawa T, Curtis MJ, Tominey CM, Duong YH, Wilcox BW, et al. (2010) A shared DNA-damage-response pathway for induction of stem-cell death by UVB and by gamma irradiation. DNA Repair (Amst) 9: 940-948.
75. Curtis MJ, Hays JB (2011) Cooperative responses of DNA-damage-activated protein kinases ATR and ATM and DNA translesion polymerases to replication-blocking DNA damage in a stem-cell niche. DNA Repair (Amst) 10: 1272-1281.
76. Fulcher N, Sablowski R (2009) Hypersensitivity to DNA damage in plant stem cell niches. Proc Natl Acad Sci U S A 106: 20984-20988.
77. Medford JI, Callos JD, Behringer FJ, Link BM (1994) Development of the Vegetative Shoot Apical Meristem. In: Meyerowitz EM, Somerville CR,

editors. Arabidopsis. Plainview, NY: Cold Spring Harbor Laboratory Press. pp. 355-378.
78. Reddy GV, Heisler MG, Ehrhardt DW, Meyerowitz EM (2004) Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of Arabidopsis thaliana. Development 131: 4225-4237.
79. Gordon SP, Heisler MG, Reddy GV, Ohno C, Das P, et al. (2007) Pattern formation during de novo assembly of the Arabidopsis shoot meristem. Development 134: 3539-3548.
80. Sugimoto K, Jiao Y, Meyerowitz EM (2010) Arabidopsis regeneration from multiple tissues occurs via a root development pathway. Dev Cell 18: 463-471.

Appendix

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	114001	A→T	intergenic
1	192361	G→A	intergenic
1	355307	C→T	intergenic
1	478566	C→T	CDS
1	484625	G→A	intergenic
1	578660	C→A	intron
1	701148	C→A	intergenic
1	1211717	G→A	CDS
1	1299563	G→A	intergenic
1	1494652	A→C	intergenic
1	1712951	C→A	CDS
1	1833244	G→T	intron
1	1863727	T→C	CDS
1	2037328	G→A	3'UTR
1	2503338	C→T	intron
1	2516028	A→G	CDS
1	2571268	G→A	CDS
1	2579526	G→A	CDS
1	2639183	G→A	intron

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	2803472	A→G	intron
1	2849952	G→A	CDS
1	3088959	T→C	intergenic
1	3145629	C→T	CDS
1	3187449	T→C	intergenic
1	3819732	C→T	CDS
1	3845942	A→G	intron
1	3909938	A→G	transposable element
1	4733059	C→T	CDS
1	4838900	C→T	intron
1	5708481	C→T	CDS
1	6075473	C→T	transposable element
1	6220157	T→G	intergenic
1	6246252	C→A	intron
1	6461455	T→A	CDS
1	6599437	A→T	intron
1	6961131	C→T	intergenic
1	7420546	T→G	intron
1	7530734	C→T	intron

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	7690032	C→T	intergenic
1	7745806	C→T	intron
1	7798869	T→C	intron
1	7812035	C→T	transposable element
1	8002862	C→T	5'UTR
1	8012266	C→A	intergenic
1	8181950	C→T	CDS
1	8580845	T→C	CDS
1	8692447	G→T	intergenic
1	9088803	C→T	intron
1	9461724	G→A	intergenic
1	9821735	A→T	intergenic
1	10118698	C→T	CDS
1	10243042	G→A	intergenic
1	10247111	T→G	intergenic
1	10530173	A→G	transposable element
1	10590154	G→A	intergenic
1	10598599	G→A	transposable element
1	10600582	G→C	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	10693422	T→C	intergenic
1	10752972	C→T	intron
1	10791377	A→G	intron
1	10853569	C→A	intergenic
1	10957272	T→C	transposable element
1	11108504	C→T	intergenic
1	11180593	G→A	intergenic
1	11307904	G→A	transposable element
1	11325861	A→G	intergenic
1	11335891	C→T	intergenic
1	11416741	C→T	CDS
1	11525430	G→A	transposable element
1	11714125	G→A	intron
1	11932649	C→G	CDS
1	12127275	C→T	intergenic
1	12230345	G→A	transposable element
1	12272777	T→C	transposable element
1	12301851	T→C	CDS
1	12310920	G→A	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	12390591	G→A	CDS
1	12513365	C→T	transposable element
1	12668755	G→A	transposable element
1	12864502	C→T	intron
1	13139209	C→T	transposable element
1	13205204	C→T	transposable element
1	13206809	G→A	CDS
1	13219957	G→A	intergenic
1	13283327	C→A	intergenic
1	13284440	C→T	intergenic
1	13399694	G→A	transposable element
1	13786607	T→C	transposable element
1	13939466	T→C	transposable element
1	14057244	C→T	transposable element
1	14430153	C→A	transposable element
1	14432966	T→C	transposable element
1	15535371	A→T	transposable element
1	15542229	T→C	transposable element
1	15757884	C→T	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	15909729	C→T	transposable element
1	15985391	G→A	transposable element
1	16032986	G→A	transposable element
1	16108325	C→T	transposable element
1	16497876	G→A	intergenic
1	16655292	C→T	intron
1	16794660	C→A	intergenic
1	16802842	G→A	transposable element
1	16867852	C→T	transposable element
1	17455192	A→G	transposable element
1	17653766	G→A	transposable element
1	18296469	G→A	intergenic
1	18729282	C→T	intron
1	19025408	T→C	transposable element
1	19142718	C→T	CDS
1	19212673	G→A	transposable element
1	19238175	C→T	CDS
1	19240334	A→T	CDS
1	19246890	C→T	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	19627462	T→G	transposable element
1	20293659	G→A	intron
1	20660615	G→A	CDS
1	20709663	G→A	transposable element
1	20723882	A→C	CDS
1	21439034	T→G	intergenic
1	22139521	A→G	intron
1	22453935	G→A	CDS
1	22936705	C→T	intron
1	23236937	C→T	intergenic
1	23698139	C→T	CDS
1	23702486	G→A	CDS
1	23931207	C→T	transposable element
1	24173223	G→A	CDS
1	24289142	C→T	3'UTR
1	24696999	G→T	intron
1	24774681	G→T	CDS
1	25110836	C→T	CDS
1	25254094	C→T	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	25363078	A→G	intergenic
1	25371575	C→T	CDS
1	25394658	C→T	CDS
1	25889495	T→G	CDS
1	26023157	C→T	intron
1	26348481	C→T	transposable element
1	26784742	C→T	CDS
1	26960879	C→T	intergenic
1	27141323	G→T	transposable element
1	27228827	G→A	intron
1	27407805	C→T	5'UTR
1	27426508	A→G	intergenic
1	27570347	T→C	transposable element
1	27695727	C→T	CDS
1	27704304	G→A	CDS
1	27841030	C→A	CDS
1	28453254	C→T	CDS
1	28473599	T→C	intergenic
1	28818937	C→T	CDS

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
1	28883451	G→A	CDS
1	28916738	C→T	intergenic
1	28982954	A→T	pseudogene
1	29398812	G→A	intergenic
1	29427430	A→G	intergenic
1	29443326	T→C	intergenic
1	30052606	C→T	CDS
2	139011	A→G	intergenic
2	197400	C→T	intron
2	685400	G→A	intergenic
2	824515	C→T	intergenic
2	946139	A→G	CDS
2	1109916	A→G	transposable element
2	2295468	G→A	intergenic
2	2422714	C→T	transposable element
2	2540502	G→A	transposable element
2	2816247	G→A	transposable element
2	3132948	G→A	intergenic
2	3155616	C→T	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
2	3179419	G→A	transposable element
2	3198848	G→A	intergenic
2	3219739	C→T	intergenic
2	3535105	C→T	transposable element
2	3595458	A→G	intergenic
2	3657466	C→T	transposable element
2	3690428	G→A	transposable element
2	3727113	A→C	transposable element
2	3753653	G→A	transposable element
2	4008976	C→T	pseudogene
2	4362701	C→A	intron
2	4405001	C→T	transposable element
2	4469560	C→T	transposable element
2	4579729	C→T	transposable element
2	4849961	G→A	transposable element
2	5021494	T→C	transposable element
2	5448660	T→C	transposable element
2	5464403	C→T	transposable element
2	5471374	T→C	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
2	5694248	C→T	intergenic
2	5723631	G→A	transposable element
2	5726676	G→A	transposable element
2	5784487	G→C	transposable element
2	5805021	C→T	intergenic
2	5979595	G→A	intron
2	6055842	T→G	intergenic
2	6511333	A→T	CDS
2	6531302	G→A	transposable element
2	6542961	G→T	CDS
2	6566029	G→A	intergenic
2	7028450	A→G	intron
2	7065224	C→T	pseudogene
2	7125822	G→A	intergenic
2	7238666	G→A	intergenic
2	7343134	C→A	CDS
2	7379550	G→A	intron
2	7584692	C→A	CDS
2	7595351	G→A	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
2	7640744	C→T	CDS
2	7701346	C→T	intergenic
2	7789801	A→G	intron
2	7838676	C→A	intergenic
2	8068824	C→A	transposable element
2	8211307	A→G	intergenic
2	8301834	G→A	intergenic
2	8428844	C→T	intron
2	8481255	A→G	intron
2	8652484	C→T	intron
2	8790978	G→A	intergenic
2	9146865	T→C	intron
2	9238631	G→A	5'UTR
2	9317586	G→T	CDS
2	9626531	C→T	intron
2	9863959	C→T	3'UTR
2	9876721	C→T	intergenic
2	10476720	C→T	intron
2	10532224	A→T	ncRNA

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
2	10591090	T→C	intergenic
2	10949577	G→T	5'UTR
2	11416274	G→A	intron
2	11448571	C→T	CDS
2	11498369	G→A	intergenic
2	11681569	A→T	CDS
2	11721845	G→A	ncRNA
2	12186829	A→G	CDS
2	12486394	G→A	CDS
2	12530565	C→T	intergenic
2	12555317	G→A	transposable element
2	12877551	A→T	transposable element
2	13178759	C→T	intergenic
2	13312370	A→G	intergenic
2	13315873	C→T	intergenic
2	13337131	C→T	transposable element
2	13559630	T→A	CDS
2	13904602	T→C	intergenic
2	14226746	T→C	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
2	14336445	A→G	intergenic
2	14515974	C→T	intergenic
2	14572728	C→T	transposable element
2	14643361	T→C	intron
2	14646213	A→G	intergenic
2	14965399	G→A	5'UTR
2	15102179	C→T	intergenic
2	15150274	G→A	intron
2	15172767	T→C	intron
2	15733497	C→T	intergenic
2	15745372	C→T	CDS
2	16140179	G→A	intron
2	16150820	A→G	transposable element
2	16285503	A→G	intergenic
2	17044886	T→C	intergenic
2	17053996	C→T	intron
2	17309886	T→C	intergenic
2	17372014	T→C	intergenic
2	17587554	G→T	5'UTR

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
2	17712922	A→G	intergenic
2	17892780	G→A	CDS
2	17907289	T→C	intergenic
2	18040508	C→T	CDS
2	18081052	A→G	intergenic
2	18357914	G→A	intron
2	18383964	G→A	CDS
2	18701313	G→A	5'UTR
2	18814224	G→A	intergenic
2	18837531	C→T	intron
2	18929756	C→T	CDS
2	19051669	C→T	intergenic
2	19080636	C→T	CDS
2	19141644	C→T	intergenic
2	19167505	C→T	intergenic
2	19405935	T→G	intron
2	19471313	C→T	CDS
2	19512205	G→A	CDS
2	19598359	C→T	3'UTR

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
2	19624407	A→G	CDS
3	13278	C→A	5'UTR
3	205101	T→C	intergenic
3	236495	C→T	intron
3	334932	C→T	intergenic
3	437394	G→A	CDS
3	471427	T→G	intron
3	1029875	C→T	CDS
3	1569064	C→T	intron
3	1882946	G→A	CDS
3	1916352	C→T	intergenic
3	2020974	C→T	intron
3	2521240	C→T	CDS
3	2701450	C→T	CDS
3	3017716	C→T	CDS
3	3039057	C→T	pseudogene
3	3096092	G→A	intergenic
3	3107801	G→A	CDS
3	3168336	G→A	intron

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
3	3401812	T→A	intron
3	3415875	T→C	intergenic
3	3539520	C→A	intron
3	3549071	C→T	intergenic
3	4118639	G→A	transposable element
3	4161993	C→A	5'UTR
3	4199438	C→T	CDS
3	4753700	A→G	intergenic
3	4828231	C→T	intergenic
3	4835910	T→C	intergenic
3	4869716	T→C	pseudogene
3	4958254	C→T	CDS
3	5141493	G→A	intron
3	5241629	G→A	intergenic
3	5800728	T→C	intron
3	5881834	C→T	intergenic
3	6166085	C→T	CDS
3	6236161	T→C	intergenic
3	6283408	G→A	CDS

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
3	6334913	G→T	3'UTR
3	6350418	A→C	CDS
3	6746414	G→T	CDS
3	7334424	T→C	intergenic
3	7390332	C→T	intergenic
3	7509172	T→A	3'UTR
3	7809807	T→G	intergenic
3	7975103	G→A	intron
3	8112691	C→T	intergenic
3	8126607	T→C	transposable element
3	8147996	G→A	transposable element
3	8307694	G→A	intergenic
3	8327076	G→A	CDS
3	8369552	G→A	transposable element
3	8731184	G→T	intron
3	9080152	G→A	CDS
3	9146257	G→A	intergenic
3	9225038	T→C	CDS
3	9715238	G→A	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
3	9816331	A→T	intergenic
3	9874086	C→T	intergenic
3	10114502	C→T	CDS
3	10388590	C→T	intergenic
3	10559637	G→A	intergenic
3	10686142	C→T	CDS
3	10729768	C→T	CDS
3	10764994	T→C	transposable element
3	11560663	C→T	transposable element
3	11658050	G→A	transposable element
3	11658051	A→G	transposable element
3	11694216	G→A	transposable element
3	11810979	C→T	CDS
3	11856469	G→A	transposable element
3	12158817	A→G	intergenic
3	12304155	C→A	transposable element
3	12328429	C→T	transposable element
3	12344415	C→T	transposable element
3	12575810	C→T	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
3	12613229	G→A	intergenic
3	12901087	G→A	transposable element
3	13082086	G→A	transposable element
3	13082862	G→A	transposable element
3	13123482	C→T	transposable element
3	13196177	C→T	transposable element
3	13425126	G→T	intron
3	13575538	C→T	intergenic
3	13887971	C→T	transposable element
3	13948244	C→T	transposable element
3	14012224	G→T	transposable element
3	14262024	A→G	transposable element
3	14372168	C→T	transposable element
3	14448798	G→A	transposable element
3	14534279	G→A	transposable element
3	14667603	C→T	intron
3	14841894	G→A	pseudogene
3	15479926	G→A	CDS
3	15536726	T→C	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
3	15770690	G→A	transposable element
3	16178674	C→T	transposable element
3	16186119	G→A	transposable element
3	16187772	G→A	intergenic
3	16266748	T→A	transposable element
3	16360791	G→A	intergenic
3	16576393	G→A	intergenic
3	16696671	G→A	transposable element
3	17145903	G→T	transposable element
3	17239565	C→T	CDS
3	17523877	A→G	CDS
3	17903090	A→G	intergenic
3	17926465	G→T	CDS
3	17944955	A→G	intergenic
3	18048483	C→A	CDS
3	18132420	C→T	CDS
3	18493646	C→A	intron
3	18681345	G→A	CDS
3	18938190	A→G	CDS

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
3	18984859	G→A	CDS
3	19204410	T→C	5'UTR
3	19391724	C→T	intergenic
3	19481069	G→A	intergenic
3	19671738	T→C	intergenic
3	20107493	G→A	intergenic
3	20141740	C→T	intron
3	20212189	C→T	CDS
3	20499969	G→A	3'UTR
3	21350877	T→C	intron
3	21564335	T→C	intergenic
3	21771820	C→A	intron
3	21925647	T→C	intron
3	22093436	G→A	CDS
3	22268604	G→A	intergenic
3	22718471	G→A	intergenic
3	22722120	G→A	CDS
3	22770815	C→T	CDS
3	22791650	G→A	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
3	22881935	T→C	intergenic
3	22965588	A→T	intron
3	23153771	A→C	intergenic
3	23162220	G→A	intergenic
3	23360323	C→T	CDS
4	361057	C→T	intergenic
4	388845	G→T	CDS
4	618978	T→C	transposable element
4	650678	T→C	intergenic
4	692483	C→T	intergenic
4	780486	C→T	intergenic
4	1162347	A→G	intron
4	1189515	A→G	intron
4	1525706	G→A	CDS
4	1605184	C→A	CDS
4	1634390	C→T	transposable element
4	1822226	C→A	intergenic
4	1910735	G→A	transposable element
4	2235505	G→A	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
4	2250803	T→C	transposable element
4	2658923	A→G	intergenic
4	2701834	C→T	transposable element
4	2979449	C→T	intergenic
4	2995346	A→G	transposable element
4	3101074	C→T	transposable element
4	3396579	G→A	transposable element
4	3398926	C→T	transposable element
4	3421613	C→T	transposable element
4	3434470	G→A	transposable element
4	3755170	C→T	transposable element
4	3785322	G→C	transposable element
4	3903749	G→A	transposable element
4	4020829	G→A	transposable element
4	4082019	A→C	intergenic
4	4284925	C→T	transposable element
4	4589347	C→T	intergenic
4	4974679	G→A	transposable element
4	5019374	T→G	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
4	5093804	T→G	transposable element
4	5452830	G→A	CDS
4	5739338	T→C	transposable element
4	5801609	T→G	transposable element
4	5896203	A→G	transposable element
4	5942323	C→T	CDS
4	6104080	T→A	intergenic
4	6257836	A→T	transposable element
4	6267539	G→A	intergenic
4	6308310	A→T	transposable element
4	6404912	C→T	transposable element
4	6423913	C→A	CDS
4	6468501	C→T	transposable element
4	6484027	C→A	CDS
4	6588357	G→A	intergenic
4	6749983	G→A	intron
4	6828372	C→T	transposable element
4	6887130	A→G	intergenic
4	6940029	G→A	CDS

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
4	7217001	C→T	transposable element
4	7745442	C→T	intron
4	7833373	G→A	intergenic
4	7898355	C→T	intergenic
4	8042225	C→T	intergenic
4	8079082	C→T	CDS
4	8128088	G→A	CDS
4	8280518	C→A	CDS
4	8305041	G→A	CDS
4	8313485	G→T	intergenic
4	8731643	G→A	CDS
4	8875969	C→T	CDS
4	9194404	A→C	transposable element
4	9277607	T→C	intergenic
4	10042345	T→C	CDS
4	10171542	C→T	transposable element
4	10343312	G→T	intergenic
4	10494099	G→A	CDS
4	10626687	G→T	CDS

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
4	10644678	T→C	CDS
4	10662438	A→G	intron
4	10727641	G→A	CDS
4	10912809	A→G	transposable element
4	10922212	G→A	intron
4	11202092	C→T	intergenic
4	11213191	C→T	CDS
4	11292161	A→G	CDS
4	11632679	C→T	intergenic
4	11757380	C→T	intergenic
4	12227829	C→G	CDS
4	12290157	A→G	intergenic
4	12357371	C→G	intergenic
4	12549351	C→A	intergenic
4	12574045	G→A	transposable element
4	12636631	G→A	intron
4	12720597	C→T	CDS
4	12809889	A→G	intron
4	12972023	C→T	CDS

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
4	13219460	G→A	3'UTR
4	13522608	T→C	intron
4	13649370	G→A	intergenic
4	13781638	C→T	intergenic
4	13783349	A→G	intron
4	14036902	C→T	intron
4	14132373	C→T	CDS
4	14177337	C→T	CDS
4	14271549	G→A	CDS
4	14353472	G→T	intron
4	14391042	A→G	transposable element
4	14860225	T→C	intergenic
4	14885713	C→A	intergenic
4	14956698	A→T	5'UTR
4	15417902	C→T	intergenic
4	15447361	G→A	intergenic
4	15483138	G→A	CDS
4	15534837	G→A	CDS
4	15569802	G→A	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
4	16156204	C→T	intron
4	16284662	G→A	intron
4	16355737	A→G	intron
4	16381944	T→A	CDS
4	16525772	C→T	intergenic
4	16551414	C→T	intergenic
4	16816334	T→A	intron
4	16938541	G→A	intron
4	17193448	C→T	intergenic
4	17323539	C→T	intergenic
4	17346840	T→C	5'UTR
4	17419904	T→C	3'UTR
4	17579968	C→T	CDS
4	17658609	C→A	intergenic
4	17695343	C→T	intergenic
4	17905584	G→T	CDS
4	18236655	C→T	intron
5	216606	G→A	intergenic
5	267336	G→A	pseudogene

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	317804	G→T	intergenic
5	465802	A→G	intergenic
5	473293	T→C	transposable element
5	552144	C→T	intron
5	581003	T→C	intergenic
5	937801	A→G	intergenic
5	952969	A→G	intergenic
5	1468722	G→A	CDS
5	1560177	G→A	CDS
5	1611466	A→T	transposable element
5	1815412	C→T	intron
5	1887527	C→T	intergenic
5	2033438	A→G	CDS
5	2052690	C→A	intron
5	2537420	T→C	intron
5	2703895	C→A	intron
5	2711619	A→G	intron
5	2859667	C→G	CDS
5	3279367	T→C	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	3465489	G→A	3'UTR
5	3483351	C→A	intergenic
5	3485529	C→T	CDS
5	3627822	G→A	intergenic
5	3665069	G→A	CDS
5	3675316	C→T	CDS
5	3903348	T→C	transposable element
5	3991465	A→G	intergenic
5	4951530	G→A	CDS
5	5016820	G→A	CDS
5	5196673	T→C	CDS
5	5248113	C→T	intron
5	5443494	C→T	intergenic
5	5507278	G→A	intron
5	5591102	C→T	intron
5	5595989	G→T	CDS
5	6011829	C→T	intron
5	6666645	A→G	intergenic
5	6868548	C→T	intron

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	7041293	T→G	transposable element
5	7110284	C→T	CDS
5	7478816	T→C	intergenic
5	7495522	C→T	transposable element
5	7887985	C→T	intergenic
5	8244950	C→A	intergenic
5	8378290	C→A	intergenic
5	9152039	T→C	intron
5	9420319	A→G	intron
5	9543361	T→C	intron
5	9575629	G→A	intergenic
5	9660554	T→A	intron
5	9678197	C→T	CDS
5	9818989	G→A	intergenic
5	9937975	G→T	intergenic
5	9973869	T→C	transposable element
5	10055326	C→A	transposable element
5	10067189	C→T	transposable element
5	10180845	T→C	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	10354939	G→A	transposable element
5	10457369	T→A	transposable element
5	10500687	G→A	transposable element
5	10539918	G→C	intergenic
5	10576677	G→A	transposable element
5	10606784	C→T	transposable element
5	10960388	G→A	transposable element
5	11029991	A→G	transposable element
5	11111570	C→A	intergenic
5	11218723	C→A	transposable element
5	11320368	G→A	transposable element
5	11530068	G→T	transposable element
5	12023878	A→G	transposable element
5	12294251	C→G	transposable element
5	12573466	G→A	transposable element
5	12615272	G→C	transposable element
5	12658768	G→A	intergenic
5	12722001	G→A	transposable element
5	12752757	G→A	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	12767299	C→T	transposable element
5	12966711	C→T	transposable element
5	13044022	C→T	transposable element
5	13307427	G→T	transposable element
5	13376583	C→T	transposable element
5	13477609	G→A	CDS
5	13511913	G→A	transposable element
5	13562731	A→T	transposable element
5	13784202	C→A	transposable element
5	13787359	C→T	intergenic
5	13896807	C→T	CDS
5	14004195	G→A	transposable element
5	14062168	T→A	transposable element
5	14275479	A→G	intron
5	14546288	G→T	intergenic
5	14818833	G→A	intron
5	15237003	G→A	intron
5	15264420	C→T	CDS
5	15328157	G→A	transposable element

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	15457044	A→G	intron
5	15749382	C→A	intron
5	15824668	G→T	intron
5	16021723	C→T	CDS
5	16468698	C→T	5'UTR
5	16680202	C→T	intergenic
5	16697079	A→G	CDS
5	16718634	G→T	CDS
5	16783775	G→A	intron
5	16958374	C→T	CDS
5	17067969	C→A	CDS
5	17095768	G→A	transposable element
5	17280843	C→T	transposable element
5	17430934	C→T	ncRNA
5	17442850	G→T	transposable element
5	17705665	G→A	intergenic
5	17736846	G→A	CDS
5	17907661	T→C	intergenic
5	18004930	G→T	CDS

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	18188986	A→G	CDS
5	18601926	G→A	intergenic
5	18813988	G→T	intron
5	19056845	G→A	CDS
5	19227922	G→A	intron
5	19322032	C→A	intron
5	19325965	C→A	intron
5	19674224	A→T	transposable element
5	19925404	C→T	CDS
5	20098024	G→A	intergenic
5	20243079	G→A	intron
5	20316106	G→A	intron
5	20553128	C→T	intergenic
5	20846900	A→G	intron
5	20945863	C→T	CDS
5	21034388	A→G	intergenic
5	21440345	C→A	transposable element
5	21512583	C→G	intron
5	21990482	G→A	intergenic

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	22090087	T→C	intergenic
5	22128716	G→A	intergenic
5	22257941	A→G	intergenic
5	22333909	C→T	intergenic
5	22486133	T→C	intergenic
5	22488393	C→A	CDS
5	22507685	G→A	intron
5	22519686	C→T	intergenic
5	22520537	G→C	intron
5	22602603	T→C	intergenic
5	23012321	C→T	CDS
5	23123134	C→A	CDS
5	23382704	C→T	intergenic
5	24055937	C→T	intergenic
5	24140628	C→T	CDS
5	24385887	T→C	intergenic
5	24626408	C→G	intron
5	25180589	C→T	intergenic
5	25249079	C→A	CDS

Base substitutions in *msh2*$^{-/-}$ after 17 generations

chr.	position	mutation	context
5	25348845	T→C	intergenic
5	25445450	T→C	intron
5	25481715	C→A	transposable element
5	25659060	G→T	intergenic
5	25893509	G→A	CDS
5	26180452	A→C	intergenic
5	26230698	G→A	intergenic
5	26371855	G→A	CDS
5	26393554	C→A	intron
5	26418057	C→T	CDS
5	26643548	C→T	CDS
5	26645130	G→T	intergenic
5	26743193	C→T	intergenic
5	26747138	G→T	CDS
5	26764813	A→G	intergenic