Supporting Information Appendix

Germline Replications and Somatic Mutation Accumulation are Independent of Vegetative Lifespan in Arabidopsis

J. Matthew Watson[a], Alexander Platzer[a], Anita Kazda[a], Svetlana Akimcheva[a], Sona Valuchova[b], Viktoria Nizhynska[a], Magnus Nordborg[a], Karel Riha[a,b,1]

SI methods

Sequence analysis

We first mapped all reads to a slightly modified reference genome (TAIR10 + T-DNA) using BWA (version 0.5.9-r16) (1) with default parameters. The aligned reads were then filtered for quality ≥ 30 and uniqueness. Samtools (version 0.1.18) (2) was used for conversions and to remove duplicated alignments. Alignments were then combined with GATK (version 1.6.5) (3) using the UnifiedGenotyper and IndelRealigner functions to minimize single-alignment artifacts. Identification of large structural variants, particularly the T-DNA insertion site, was performed by breakdancer (version 0.0.1r81) (4). A mask was generated for each individual with the following filtering rules:

- Coverage ≥ 20 and ≤ 1000
- At least 5 bp calls with a quality ≤ 30
- No pericentromeric regions (S4 Table)
- No repetitive region (defined as not more than 4 repeats of a word length up to 3 bp)

The final mask is the intersection of all individual masks, and only SNPs not masked were considered in further analysis. As we were specifically interested in novel SNPs arising during MMR-deficient growth, we subtracted all SNPs present in the F1 MSH2+/- founder line.

Calling ratios

Our experimental design necessitated to reliably identify heterozygous SNPs. As our founding plant was also an F1 hybrid of a wild type Col0 and an MMR mutant, we binned SNPs identified in this plant based on their allele ratios. Bins were set at 50±5, and then with further steps of 10, and ten SNPs from each bin were randomly selected for validation by Sanger sequencing. Additionally, as this founder line was sequenced twice, we validated the 23 SNPs that were called in one sequencing run but not the other. Despite the high

sequence coverage, we found that only the first bin, covering allele ratios of 45-55%, provided a low false error rate (SI Appendix, Fig. S7). Further analysis of the actual allele ratios demonstrated that the false positive rate rose from roughly 10% to over 50% outside of the range 50±10%. We attribute the large error observed below 40% to hidden duplications (see below). We therefore set a boundary of allele ratios between 0.4 and 0.6, inside which SNPs would be considered heterozygous. Allele ratios above 0.9 were considered homozygous.

Following the completion of all sequences, we re-analyzed the calling ratio profile of all samples (SI Appendix, Fig. S8). We observed that the MSH2 samples had a roughly bimodal distribution, with peaks around 0.5 and 0.25. The lower peak likely represents duplication events for the following reasons. First, the majority of the SNPs called with these allele ratios are located close to our masked pericentromeric regions. Second, nearly all of these SNPs (with the exceptions presented in SI Appendix, Table S5) are present in more than one sample. Third, these SNPs are often found in clusters ranging in size from hundreds of base pairs up to 15 kb. We have however not specifically tested whether these SNPs do represent duplications. Importantly, we queried the SNPs from all samples that fell outside of our pre-determined range, and found that, even when setting the range of allele ratios to 0.15-0.85, there was no dramatic change in the number of SNPs between LD and SD (SI Appendix, Table S5) and the difference between the two growth conditions remains insignificant (Fisher's exact test, P=0.4624).

Sequencing error

In order to estimate the technical error in sequencing, we used the same filtering steps that produced our combined mask, and additionally filtered for loci where the same base is called in all samples, and no sample contains more than 20% alternate calls. With this set, we calculate an average technical error of 0.21%, which can be further subdivided into roughly 4 classes (SI Appendix, Fig. S9). This is a lower bound of our error rate, as it

only includes base calling errors during sequencing. Several other error sources exist, particularly PCR amplification during library generation.

Mutation rate

The average number of new heterozygous SNPs in our population is 31.8, which would yield a simple mutation rate of 6.5x10-7/site/generation. However, this mutation rate should minimally be adjusted for our technical error rates. Our samples have an average coverage of 120, with a range from 10 to 230, our SNPs were called between a range of allele ratios from 0.4-0.6 (which is also the range used for error estimation), and the total number of sites is 97,165,755.

To estimate the false positive rate due to technical sequencing error, the error frequency is divided by 3, as the same base must be called in each case. The expected error is then 3 times the error of a binomial distribution between the calling ratios for heterozygous SNPs. Even if our coverage had been as low as 10 at all loci, this error rate would predict one miscalled base across the 97 MB genome. Our true false positive rate is much higher however as our measured error rate does not include any errors derived from upstream steps during library generation. In validation of 65 SNPs across all samples, Sanger sequencing identified the reference base in one case, for a false positive rate of 1.5%.

In contrast, the false negative rate is combinatorial, as we have considered heterozygous SNPs those with a calling ratio between 0.4 and 0.6. In this case, the error rate is the area of a binomial distribution outside our calling ratio. This error rate ranges from 34% for a coverage of 10, to 0.2% with a coverage of 230. Over our total coverage range, we estimate the false negative rate of be 8%. The technical error rate is ignored in this case, as it is insignificantly small. Secondly, an expanded range of calling ratios would be expected to reduce this error, while at the same time increase the false positive error rate. In conclusion, we estimate our mutation rate as follows. The experimentally determined mutation rate of 31.8, times the false negative rate of 1.08, times 2 due to outcrossing, divided by the total number of sites, resulting in a mutation rate of 7x10-7/site/generation.

SI References

1. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
2. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
3. DePristo MA*, et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491-498.
4. Chen K*, et al.* (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 6(9):677-681.
5. Huffman KE, Levene SD, Tesmer VM, Shay JW, & Wright WE (2000) Telomere shortening is proportional to the size of the G-rich telomeric 3 '-overhang. *Journal of Biological Chemistry* 275(26):19719-19722.
6. Kazda A*, et al.* (2012) Chromosome end protection by blunt-ended telomeres. *Genes Dev* 26(15):1703-1713.
7. Riha K, McKnight TD, Fajkus J, Vyskot B, & Shippen DE (2000) Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. *Plant Journal* 23(5):633-641.
8. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, & Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152(1-2):352-364.
9. Ossowski S*, et al.* (2010) The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science* 327(5961):92-94.
10. Watson JM & Shippen DE (2007) Telomere rapid deletion regulates telomere length in Arabidopsis thaliana. *Mol Cell Biol* 27(5):1706-1715.

	Leaves			Flowering		
	average	SD	n	average	SD	n
LD	10,3	1,3	35	35,5	2,1	17
SD	74,5	18,3	24	91,3	8,2	21

Table S1. Growth characteristics of plants grown in long day (LD) and short day (SD) conditions. Number of leaves produced and days before bolting.

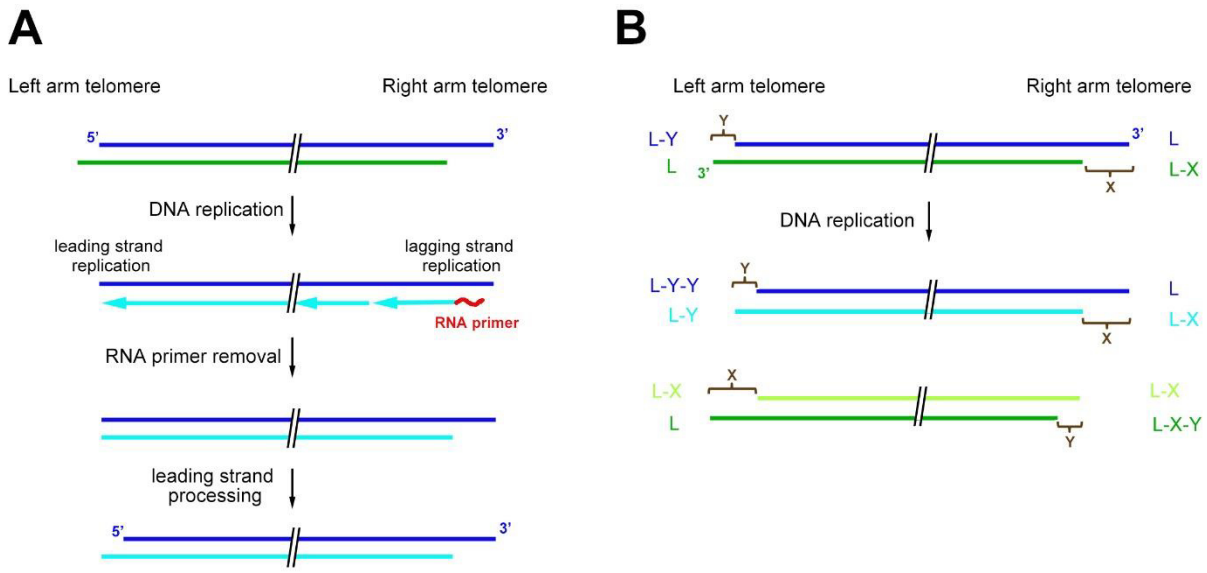

Fig. S1. Relationship between telomere shortening and DNA replication. (*A*) The diagram illustrates the end replication problem by depicting replication of a hypothetical linear chromosome with two telomeres (replication of only a single chromosomal DNA strand is shown for simplicity). DNA polymerases synthesize DNA in a 5' to 3' direction and require a primer for initiation of synthesis. Removal of the RNA primer from the last Okazaki fragment on the telomere replicated by the lagging strand mechanism leaves a gap that cannot be filled, producing a single stranded G-overhang. The telomere replicated by the leading strand mechanism is synthesized to the end of the template strand, forming a blunt end. This blunt end is subject to further 5' to 3' post-replicative nucleolytic processing, forming a smaller G-overhang also on this telomere. Thus, in the absence of telomerase, which normally extends the G-overhangs, telomeres are subjected to a replication dependent shortening that is proportional to the gaps left after DNA processing and removal of the last Okazaki RNA primer. (*B*) Calculation of telomere loss (∆L) after one round of DNA replication based on G-overhang structure. This estimate is based on an existing model of telomeric DNA loss to fit known parameters in Arabidopsis (5). Post-replicative resection of leading telomeres in Arabidopsis is greatly reduced (Y= 0-4 nt) (6). Estimates of G-overhangs at lagging telomeres (X) are in the range of 20-30 nt (7). The following equations can be used to calculate the rates of telomere shortening on the left and right chromosomal arms:

$$\Delta L_{\text{left arm}} = \left(\frac{(L\text{-}Y)+L}{2}\right) - \left(\frac{(L\text{-}Y)+(L\text{-}Y\text{-}Y)+(L\text{-}X)+L}{4}\right) = (L\text{-}0.5Y)\text{-}(L\text{-}0.75Y\text{-}0.25X) = 0.25Y+0.25X$$

$$\Delta L_{\text{right arm}} = \left(\frac{L+(L\text{-}X)}{2}\right) - \left(\frac{L+(L\text{-}X)+(L\text{-}X\text{-}Y)+(L\text{-}X)}{4}\right) = (L\text{-}0.5X)\text{-}(L\text{-}0.75X\text{-}0.25Y) = 0.25X+0.25Y$$

$$\Delta L_{\text{all telomeres}} = 0.25X+0.25Y$$

At the low end of G-overhang estimates, with a Y of 0 and X of 20, we calculate an average loss of 5 bp/division. For high end values, we use a Y of 4 and an X of 30, for a loss of 8.5 bp/division, providing a rough average of 7 bp/division telomeric DNA loss.

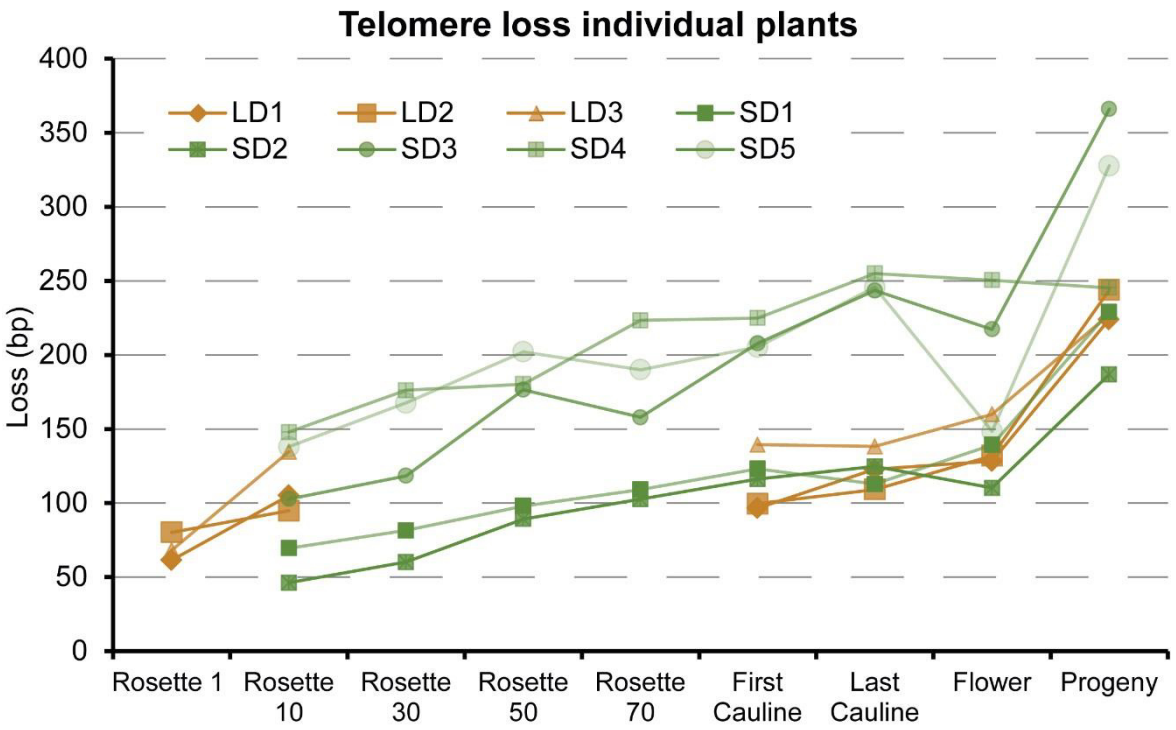

Fig. S2. Loss of telomeric DNA represented as individual plants used in this study. Two of the short day plants (SD1 and SD2) followed very closely the telomere loss curves of the long day plants. One plant (SD5) showed additional telomeric DNA loss during early vegetative growth but telomeres in the progeny had lost roughly the same amount of telomeric DNA as LD plants. SD3 and SD4 both displayed excessive loss during vegetative growth and the progeny had also lost roughly 100 bp more telomeric DNA than the LD plants and the three other SD plants. When tested on an individual basis, the difference between LD and SD plants is not significant (Student's t-test, P = 0.14).

Table S2. List of SNPs identified in this study. Methylation data is from the average of three sequence tracks from (8).

chr.	position	line	SNP	context	Dipyrimidine	Methylation	
1	11479041	LD3	C→T	intergenic	Yes	CHH	0,00
1	12012192	SD3	G→A	nonsynonymous	Yes	CHH	0,00
1	12403875	SD3	G→A	transposon	No	CG	0,86
1	12595130	SD3	C→A	synonymous	Yes	CHH	0,37
1	12666340	SD1	C→T	transposon	No	CG	0,88
1	12865870	LD1	T→C	intron	No	NA	
1	1566472	SD2	A→G	synonymous	No	NA	
1	16044944	LD3	C→T	transposon	Yes	CHH	0,01
1	16573200	LD1	C→A	transposon	Yes	CHH	UN
1	16601010	SD2	G→A	intergenic	Yes	CHH	0,11
1	17609272	SD2	G→A	transposon	Yes	CHG	0,71
1	18296004	LD1	C→T	intergenic	No	CHH	0,00
1	18535761	LD3	C→T	intron	No	CHH	0,00
1	18667171	SD1	G→A	intron	Yes	CHH	0,00
1	18865366	SD1	G→A	synonymous	No	CG	0,00
1	19141842	LD1	C→T	intergenic	Yes	CHH	0,00
1	19165202	SD3	C→T	UTR	Yes	CHH	0,00
1	19828811	SD2	C→T	intergenic	No	CHH	0,00
1	2085475	LD1	C→T	nonsynonymous	Yes	CHH	0,00
1	209897	SD1	A→G	intron	No	NA	
1	22025854	LD3	T→A	intergenic	No	NA	
1	22642730	LD2	G→A	synonymous	No	CHG	0,00
1	23111421	SD2	C→T	intergenic	No	CG	0,00
1	23238419	LD3	A→G	nonsynonymous	No	NA	
1	23884811	LD1	G→A	intergenic	Yes	CHH	0,00
1	24658522	SD1	C→T	intergenic	Yes	CHH	0,00
1	26005667	SD3	T→C	nonsynonymous	No	NA	
1	26106866	SD2	A→G	intergenic	No	NA	
1	26585116	SD3	T→C	synonymous	No	NA	
1	2755151	SD2	A→G	intergenic	No	NA	
1	27996336	LD1	A→G	intron	No	NA	
1	28180650	LD2	C→T	nonsynonymous	Yes	CHH	0,07
1	2844443	SD2	C→T	synonymous	Yes	CHH	0,00

chr.	position	line	SNP	context	Dipyrimidine	Methylation	
1	29938742	LD1	T→C	intergenic	No	NA	
1	3047660	SD1	C→T	intergenic	No	CHH	0,00
1	3580403	SD1	T→C	intergenic	No	NA	
1	3663090	LD2	C→G	intron	Yes	CHG	0,00
1	3897010	LD2	C→T	intergenic	Yes	CHH	0,00
1	4737626	SD3	G→A	UTR	Yes	CHH	0,00
1	4989502	SD2	A→G	intergenic	No	NA	
1	51718	LD1	T→C	intergenic	No	NA	
1	5461348	LD1	G→A	UTR	Yes	CHG	0,00
1	5897118	SD1	G→A	nonsynonymous	Yes	CHH	0,00
1	6414170	LD3	G→A	synonymous	Yes	CHH	UN
1	6467544	SD3	G→A	intron	No	CHH	0,00
1	6911142	SD3	A→G	intergenic	No	NA	
1	7590386	SD1	C→T	nonsynonymous	Yes	CHH	0,00
1	77769	SD2	C→A	intergenic	Yes	CG	UN
1	789661	LD2	G→A	intergenic	Yes	CHH	0,00
1	7949923	SD1	G→A	nonsynonymous	Yes	CHH	0,00
1	8267483	LD2	T→C	intergenic	No	NA	
1	8392778	SD3	T→G	intergenic	No	NA	
1	8458519	SD1	G→A	intergenic	Yes	CHH	UN
1	8643844	LD3	C→T	transposon	No	CG	0,93
1	8886892	SD2	A→C	intergenic	No	NA	
1	9507061	SD1	C→T	UTR	Yes	CHH	0,00
1	9888590	LD3	G→A	intergenic	No	CHH	0,00
2	10091196	LD3	G→A	transposon	Yes	CHH	0,02
2	10331375	SD1	C→T	intergenic	No	CHH	0,00
2	11500049	SD1	C→T	intergenic	Yes	CHH	0,00
2	14186087	SD2	C→T	intron	No	CHH	0,00
2	15355717	SD3	G→A	nonsynonymous	Yes	CG	0,00
2	15474298	LD2	C→T	nonsynonymous	Yes	CHG	0,00
2	15553073	LD2	C→T	intergenic	Yes	CHG	0,00
2	15922246	LD3	A→G	UTR	No	NA	
2	16899807	SD2	C→T	intron	Yes	CHH	0,00

chr.	position	line	SNP	context	Dipyrimidine	Methylation	
2	18195181	LD2	T→G	pseudogene	No	NA	
2	19511588	LD1	G→A	nonsynonymous	Yes	CHH	0,00
2	2108660	SD1	C→T	transposon	No	CHH	0,35
2	2473978	SD2	C→T	intergenic	No	CHH	0,03
2	341664	LD3	A→C	intron	No	NA	
2	5476133	SD2	G→A	transposon	Yes	CG	0,95
2	6107543	LD2	G→A	transposon	No	CHH	0,04
2	6165982	LD3	C→T	pseudogene	Yes	CHH	UN
2	6663320	LD3	C→T	intergenic	Yes	CHH	UN
2	6769074	SD2	T→C	transposon	No	NA	
2	6953490	LD3	A→G	transposon	No	NA	
2	9439566	SD3	C→G	intergenic	Yes	CHH	0,00
2	9588326	LD1	G→A	nonsynonymous	No	CHH	0,00
2	984233	SD3	G→A	intron	Yes	CHH	0,00
2	9907078	LD2	A→G	intergenic	No	NA	
3	10222865	LD3	G→A	intergenic	No	CHH	0,00
3	10293328	LD1	C→T	intergenic	Yes	CHH	UN
3	11077991	LD1	G→A	intergenic	Yes	CHH	0,06
3	11415721	SD2	C→T	intergenic	Yes	CHG	0,91
3	1229393	SD2	C→T	UTR	Yes	CHG	0,01
3	15932961	LD3	C→T	intergenic	No	CHH	0,00
3	15986137	LD3	A→G	intergenic	No	NA	
3	16287427	LD2	G→A	synonymous	Yes	CHH	0,00
3	16456284	LD2	C→T	intergenic	No	CHH	0,00
3	1711932	LD1	A→G	intron	No	NA	
3	17266527	LD3	C→T	nonsynonymous	Yes	CHH	0,00
3	17495282	SD3	T→C	intergenic	No	NA	
3	17539860	SD2	C→T	nonsynonymous	Yes	CHG	0,07
3	17621907	LD1	G→A	intron	Yes	CHH	0,00
3	17971812	SD3	T→G	intergenic	No	NA	
3	18469662	LD1	T→A	UTR	No	NA	
3	1881350	LD1	T→G	nonsynonymous	No	NA	
3	18961970	LD3	G→T	intergenic	Yes	CHH	0,00

chr.	position	line	SNP	context	Dipyrimidine	Methylation	
3	19797574	SD1	C→A	intergenic	Yes	CG	0,00
3	20785263	SD1	C→T	nonsynonymous	Yes	CHH	0,00
3	2160324	LD3	G→A	nonsynonymous	Yes	CHH	UN
3	21611382	LD1	C→A	intron	Yes	CHH	0,00
3	22073422	LD2	C→T	intron	Yes	CHG	0,00
3	22234336	SD1	G→A	transposon	No	CG	0,91
3	22280963	LD1	C→T	intergenic	No	CHH	0,00
3	22303453	LD3	G→A	nonsynonymous	No	CHG	UN
3	22834599	SD3	T→C	intergenic	No	NA	
3	3757108	SD2	G→A	synonymous	No	CG	0,88
3	3923406	SD1	C→T	intergenic	No	CG	0,00
3	4148075	SD1	G→A	nonsynonymous	No	CG	0,89
3	4204488	SD1	C→A	nonsynonymous	Yes	CHG	0,00
3	4750955	LD2	C→T	intergenic	Yes	CHH	0,00
3	5278109	LD1	C→G	intergenic	Yes	CHH	0,00
3	5279719	LD1	T→C	transposon	No	NA	
3	5676422	LD1	A→G	intergenic	No	NA	
3	6068879	LD2	G→T	intron	Yes	CHH	0,00
3	6472896	LD1	T→A	intergenic	No	NA	
3	7097070	SD1	G→T	intron	Yes	CHH	0,00
3	8279239	LD1	G→A	intergenic	No	CHH	0,00
3	854482	LD3	T→C	nonsynonymous	No	NA	
3	8630902	LD3	G→A	synonymous	Yes	CHH	UN
3	9610586	LD1	G→A	intergenic	No	CHH	UN
3	9826048	SD3	T→C	intron	No	NA	
4	10142544	SD2	T→C	intergenic	No	NA	
4	10226348	LD3	A→C	intron	No	NA	
4	10756521	SD2	C→T	transposon	Yes	CHH	0,16
4	10824392	LD2	C→T	nonsynonymous	No	CHH	0,00
4	10982640	LD2	C→T	UTR	Yes	CHH	0,00
4	11791681	SD1	A→C	intron	No	NA	
4	12656121	SD1	C→T	intron	No	CHH	0,02
4	12682333	LD3	A→G	intron	No	NA	

chr.	position	line	SNP	context	Dipyrimidine	Methylation	
4	12703571	LD3	A→G	nonsynonymous	No	NA	
4	13511696	LD3	G→A	intergenic	No	CHH	0,00
4	14223070	LD1	T→C	intron	No	NA	
4	14994897	LD2	C→T	intron	Yes	CHH	0,00
4	15225862	SD3	T→C	intergenic	No	NA	
4	15924026	LD2	T→A	intergenic	No	NA	
4	17471277	LD3	G→A	intergenic	No	CHH	0,01
4	17589018	LD3	C→A	intergenic	Yes	CHH	0,01
4	17889263	LD3	C→T	intron	No	CHH	0,00
4	18217498	LD2	T→C	intergenic	No	NA	
4	1903176	SD1	C→T	transposon	No	CG	0,79
4	2275147	LD1	G→A	intergenic	Yes	CHG	0,73
4	2277918	LD2	G→T	transposon	Yes	CHH	0,06
4	499046	SD2	G→A	intergenic	No	CHH	0,00
4	524069	SD2	C→T	nonsynonymous	Yes	CHG	0,00
4	6957710	SD1	C→T	intron	No	CG	0,84
4	7065310	LD2	C→T	intron	Yes	CHH	0,00
4	7269274	SD1	C→T	intron	No	CHH	0,02
4	736340	LD2	G→A	nonsynonymous	Yes	CHG	0,00
4	7819067	LD3	C→T	intergenic	Yes	CHG	0,28
4	8465299	SD2	A→G	intergenic	No	NA	
4	9434636	LD3	G→T	UTR	Yes	CG	0,00
4	9579055	LD3	G→A	nonsynonymous	No	CG	0,96
4	9846809	SD1	G→A	nonsynonymous	No	CHH	0,00
5	1134190	LD1	C→T	intron	No	CHG	0,00
5	1318873	LD3	A→G	intergenic	No	NA	
5	14535155	SD2	G→T	intron	Yes	CHH	0,00
5	14673601	SD1	G→A	transposon	Yes	CHG	0,08
5	14822539	SD3	C→T	intergenic	No	CG	0,00
5	14860463	SD3	A→G	intergenic	No	NA	
5	16686076	SD2	T→C	nonsynonymous	No	NA	
5	1709058	LD3	T→C	intergenic	No	NA	
5	17624822	SD1	A→C	intergenic	No	NA	

chr.	position	line	SNP	context	Dipyrimidine	Methylation	
5	18506441	SD1	C→T	intergenic	No	CG	0,00
5	19299664	LD2	A→C	intergenic	No	NA	
5	19632391	SD2	C→T	synonymous	Yes	CHH	0,00
5	20306035	LD1	A→G	nonsynonymous	No	NA	
5	20437690	SD1	A→G	UTR	No	NA	
5	20915295	SD1	G→A	synonymous	No	CHH	0,00
5	22359699	SD2	G→A	intergenic	Yes	CHH	0,00
5	23010894	LD3	G→A	synonymous	No	CHH	0,00
5	23013027	LD1	C→T	nonsynonymous	No	CHG	0,00
5	23323616	SD1	G→T	intergenic	Yes	CHH	0,00
5	24230048	SD3	G→A	nonsynonymous	Yes	CHG	0,00
5	24493489	LD2	A→G	intergenic	No	NA	
5	24850995	LD3	C→T	nonsynonymous	No	CHH	0,00
5	26436540	SD1	A→C	nonsynonymous	No	NA	
5	2658178	SD2	C→T	intron	Yes	CHH	0,00
5	3564292	LD1	G→C	intergenic	Yes	CHH	0,00
5	4003685	SD3	C→T	intron	No	CG	0,42
5	4289529	LD1	T→C	UTR	No	NA	
5	4470145	LD3	C→A	intron	Yes	CHH	0,00
5	4763414	LD2	T→C	intergenic	No	NA	
5	5118721	LD2	C→T	intron	No	CHH	0,00
5	5985542	SD1	G→T	nonsynonymous	Yes	CHG	0,00
5	655601	SD2	C→T	nonsynonymous	Yes	CHG	0,00
5	6782423	SD1	C→G	UTR	No	CG	0,00
5	7849596	LD2	G→A	nonsynonymous	Yes	CHG	0,00
5	9655068	SD1	G→T	intergenic	Yes	CHH	0,00

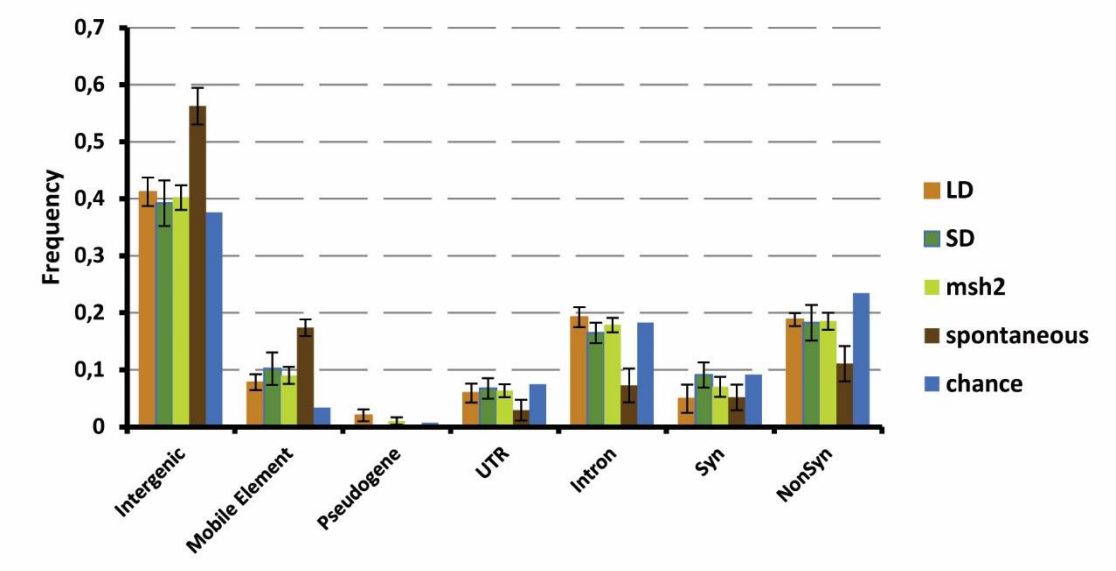

Fig. S3. Ratio of mutations in different sequence contexts. The msh2 category represents a combination of LD and SD data, spontaneous is from the spontaneous mutation rate determined by (9) and chance refers to the fraction of bases in each category from our masked genome.

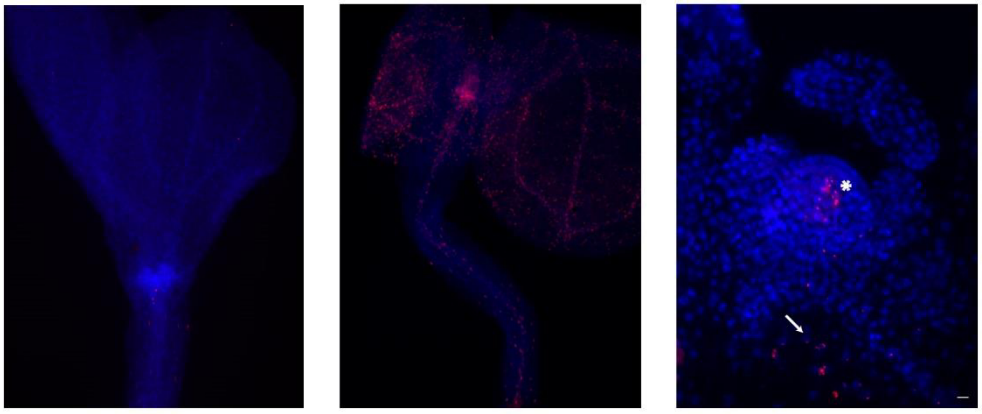

Fig. S4. EdU Labeling of aerial portions of plants.

Left, Middle: While labeling of roots was consistently strong, labeling of aerial portions of seedling was highly variable, as demonstrated by these two seedlings (left and middle panels) which were pulsed under the same conditions and fluorescently labeled on the same slide 24 hr after the initial pulse. Right: An 8d old SAM with labeling in the lower stem (indicated by arrow) and in the L3 layer of the SAM (asterisk). Strong staining was observed in the lower stem at all time-points, and was usually an indicator of whether EdU uptake had occurred during the EdU pulse. Plants that lacked EdU staining were counted as unstained in our analysis.

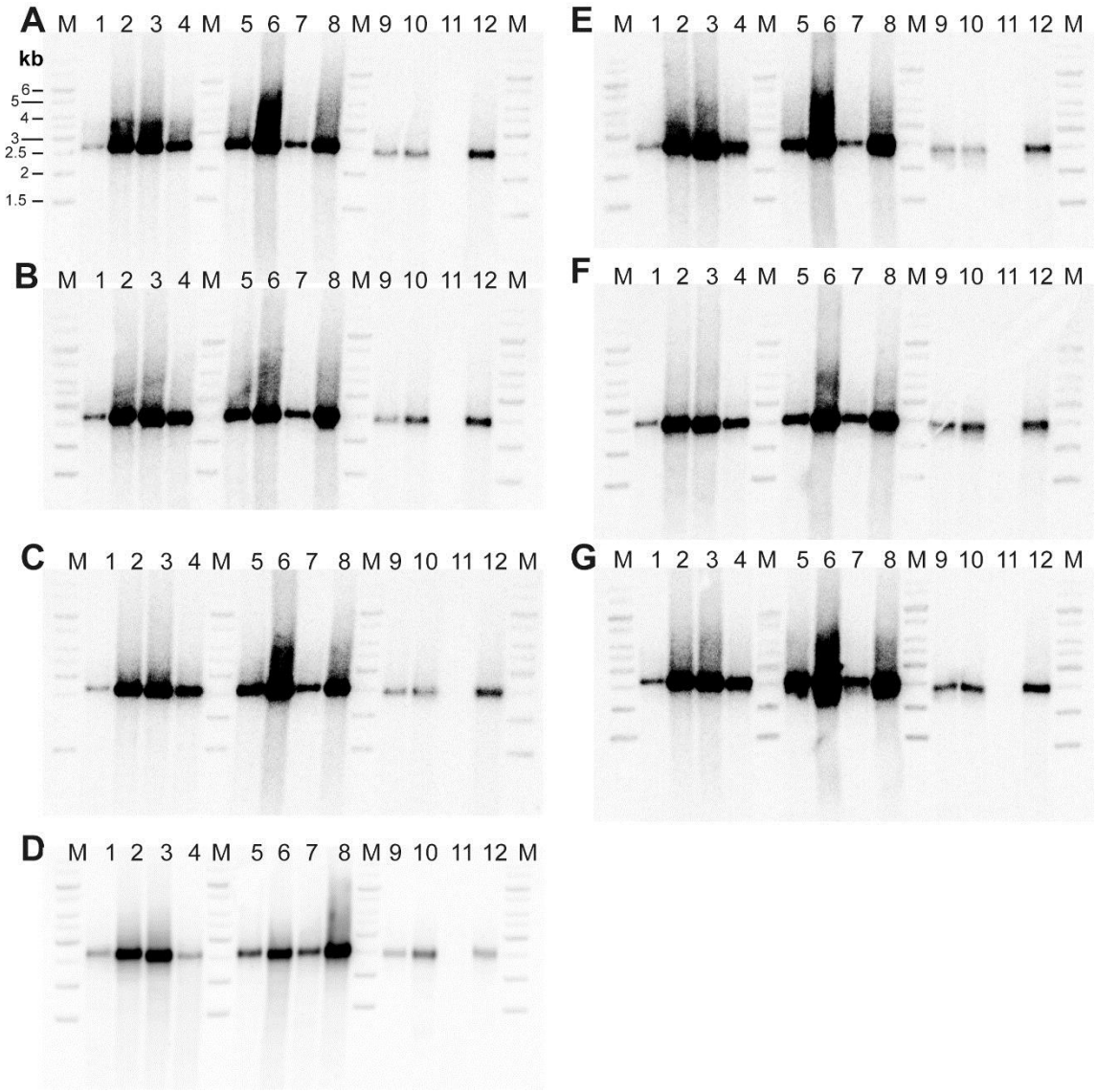

Fig. S5. Analysis of PETRA accuracy. (*A*) represents the original analysis used to measuring telomere shortening, (*B-D*) are the same sample run on three gels. (*E-G*) the products are from the same primer extension but different PCR reactions. Calculated lengths of each product and the standard deviation are given in Table S3. The average overall standard deviation is 31.7 bp. To test for specific errors arising due to individual steps of the procedure, we tested standard deviations within the samples after each of the three steps with Fisher's exact tests and nested ANOVAs, all of which produced insignificant results. Therefore, we cannot point to a single experimental step as being primarily responsible for the differences we observe. However, we conclude from the low overall standard deviation, that PETRA is an accurate and precise measure of telomere lengths of individual chromosome arms.

	A	B	C	D	E	F	G	average	standard deviation
cotyledon	2691	2671	2650	2655	2679	2653	2652	2664	16,11
rosette 1	2651	2651	2586	2626	2624	2596	2618	2622	24,72
rosette 2	2599	2575	2524	2586	2526	2549	2572	2562	29,32
caluine 1	2571	2565	2506	2582	2546	2546	2563	2554	24,84
cauline 2	2588	2594	2544	2584	2557	2585	2599	2579	20,32
flower	2544	2579	2545	2591	2526	2554	2557	2557	22,04
off-shoot cauline	2557	2601	2556	2598	2540	2599	2570	2574	24,89
off shoot flower	2570	2552	2469	2563	2488	2522	2508	2525	38,72
progeny 1	2478	2437	2462	2480	2369	2375	2439	2434	45,73
progeny 2	2476	2443	2444	2500	2386	2323	2382	2422	61,34
progeny 3	2537	2484	2456	2481	2431	2427	2427	2463	40,55

Table S3. Quantification of the data in Fig. S5. Calculated sizes for each tissue across all seven gels, the average, and standard error.

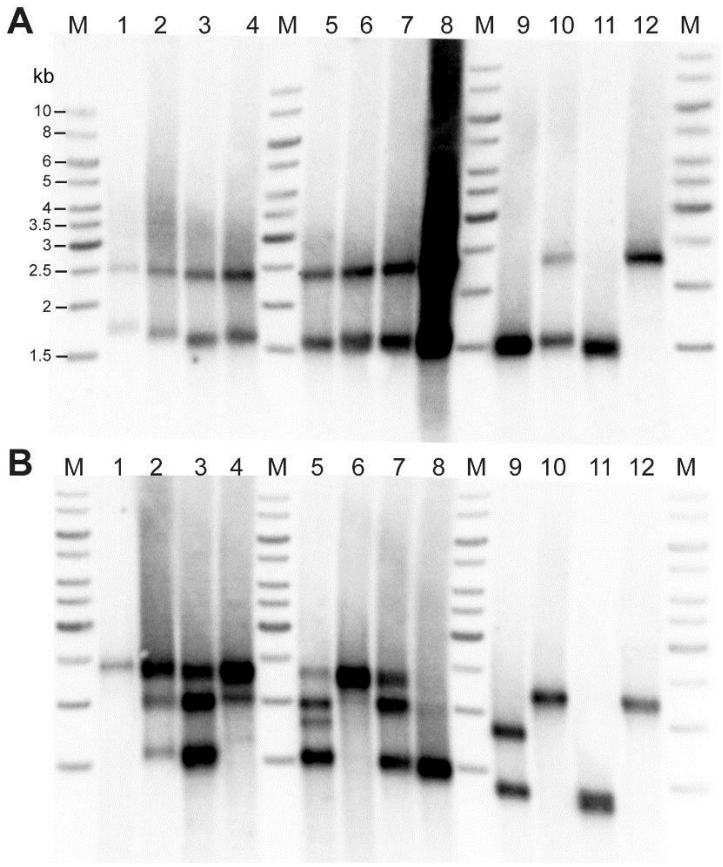

Fig. S6. PETRA analysis of complicated telomere patterns. (*A*) Samples where telomeres at homologous chromosome arms exhibit different length. In this case the large (~2.5 kb) and small (~1.7kb) PCR products are treated as two independent sets. (*B*) Samples in which telomeres produce multiple bands that exhibit sudden changes in their size in comparison to a neighboring sample indicate rapid telomere deletion events (10), which are caused by homologous recombination and not by the end replication problem. As previously reported, TRD occurred more often in telomeres longer than 2 kb (10). Telomere shortening occurred primarily as a gradual loss of telomeric DNA over time, which is consistent with the loss of telomeric DNA via the end replication problem. Occasional TRD events appeared as sudden loss of large amounts (more than 500 bp) of telomeric DNA in a short developmental window, usually between two vegetative developmental tissues. Further, in all cases, TRD affected only a subset of telomeres within the tissue – as telomere shortening should be roughly equal in all telomeres, TRD events identified in this way were excluded.

	Pericentromere	Centromere
Chr. 1	13-16	14.4-15.7
Chr. 2	2.8-5.1	3.1-4.1
Chr. 3	11.8-15.4	13.6-13.7
Chr. 4	2.5-5.8	2.9-4.0
Chr. 5	10.4-14.3	11.7-12.1

Table S4. Locations of centromeric and pericentromeric areas excluded in our analysis. The coordinates are given in MB and refer to TAIR10.

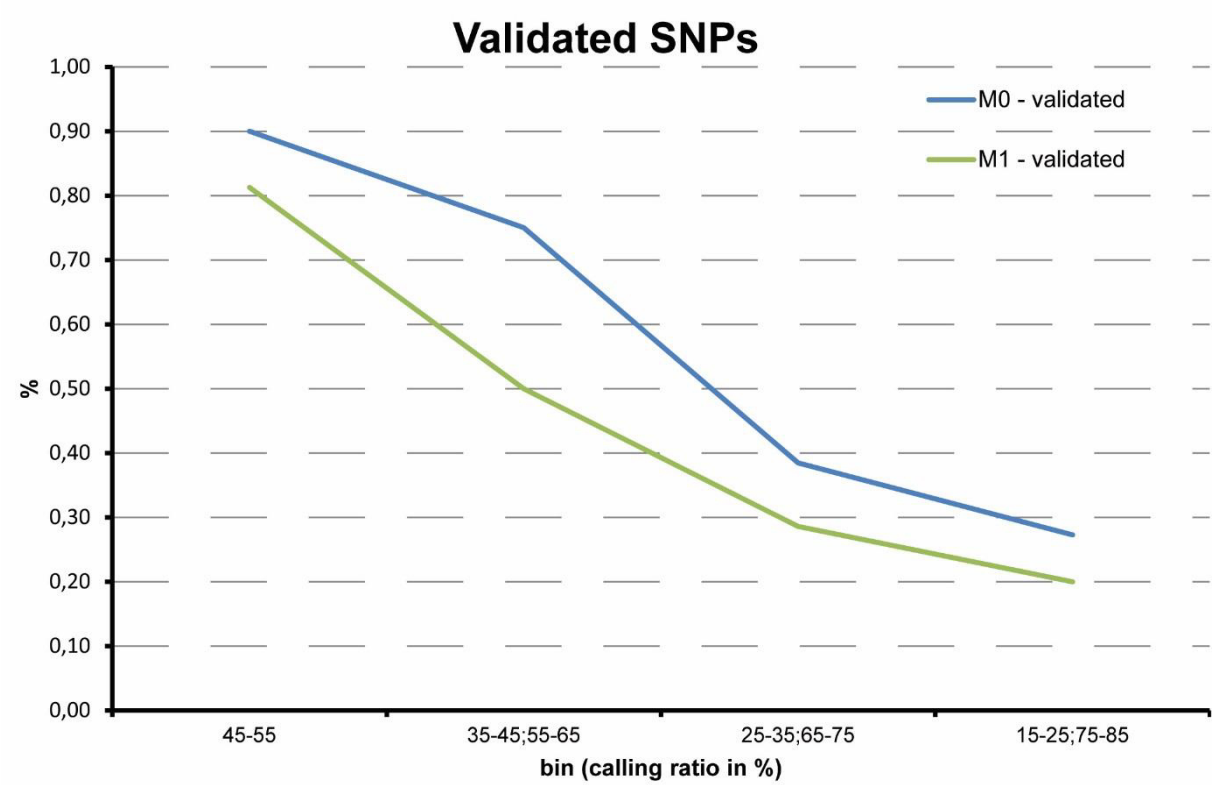

Fig. S7. Validation of heterozygous SNPs in different calling ratio bins. SNPs were randomly selected from each bin (based on the M0 calls) and verified by Sanger sequencing. The y-axis represents the % of SNPs called by our pipeline that were verified as heterozygous SNPs. M0 and M1 refer to two independent sequencing dataset of the *MSH2*$^{+/-}$ founder plant.

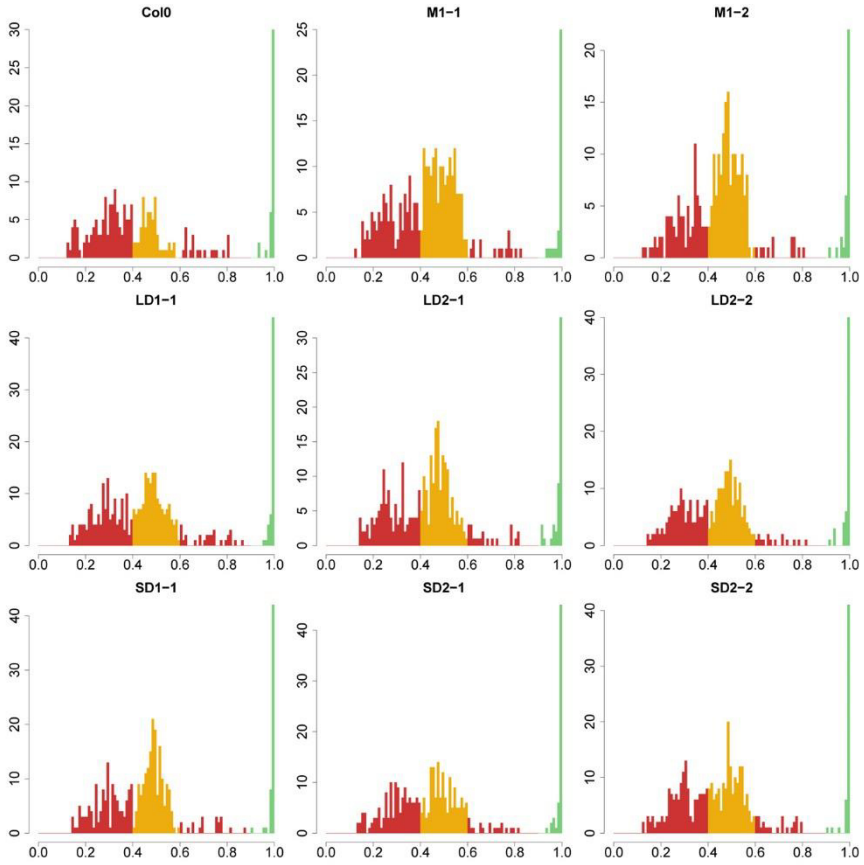

Fig. S8. Histogram of calling ratios for mutations called by our pipeline across all samples. Green represents calls we considered homozygous, orange heterozygous, and red mutations were ignored in our analysis.

SD1 (**43**)	SD2 (**35**)	SD3 (**28**)	LD1 (**44**)	LD2 (**35**)	LD3 (**39**)
0,165975104	0,278106509	0,267605634	0,1875	0,155339806	0,380952381
0,174757282	0,293193717	0,365384615	0,291666667	0,2	
0,330708661	0,303664921	0,373626374	0,356164384	0,235955056	
0,387096774	0,376811594	0,383333333	0,370967742	0,253012048	
0,397849462		0,396226415	0,372881356	0,390410959	
		0,611940299	0,373333333	0,4	
			0,395721925		
			0,614457831		
			0,684039088		
			0,717105263		
			0,740412979		

Table S5. Calling ratios of additional SNPs identified with extended thresholds (0.15-0.85). The bolded number in parentheses next to the sample name indicates the total number of mutations identified by extending the calling ratio threshold.

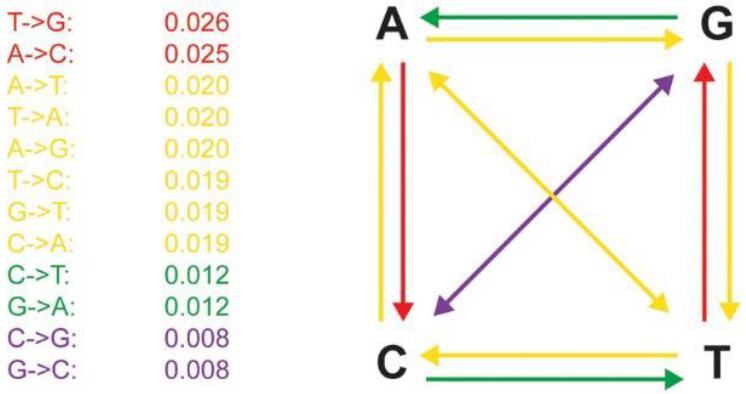

Fig. S9. Technical sequencing error frequency by erroneous base-calling errors. At each base-calling error, the base on the left side is the most likely correct base and the base on the right is the erroneous base; this can be seen as directional error. The error frequency per transition is relative to all bases where the missing fraction to the total technical error frequency are the indels. These frequencies can be grouped into certain ranges, and this pattern is shown on the right side.