Biased mutagenesis and H3K4me1-targeted DNA repair in plants

Daniela Quiroz[1,2], Diego Lopez-Mateos[3,4], Kehan Zhao[1], Pablo Carbonell-Bejerano[5], Vladimir Yarov-Yarovoy[3,4], J. Grey Monroe[1,2,#]

[1] Department of Plant Sciences, University of California Davis, Davis, CA, USA 95616
[2] Integrative Genetics and Genomics, University of California Davis, Davis, CA, USA 95616
[3] Department of Physiology and Membrane Biology, University of California Davis, Davis, CA, USA 95616
[4] Biophysics Graduate Group, University of California Davis, Davis, California
[5] Institute for Grape and Wine Sciences (ICVV, CSIC-CAR-UR), 26007 Logroño, La Rioja, Spain
[#] Correspondence: gmonroe@ucdavis.edu

Abstract

Mutations are the ultimate source of genetic variation. To study mechanisms determining intragenomic mutation rate variability, we reanalyzed 43,483 *de novo* germline single base substitutions in 1,504 fast neutron irradiated mutation accumulation lines in Kitaake rice. Mutation rates were significantly lower in genomic regions marked by H3K4me1, a histone modification found in the gene bodies of actively expressed and evolutionarily conserved genes in plants. We observed conservation in rice for PDS5C, a cohesion cofactor involved in the homology-directed repair pathway that in *A. thaliana* binds to H3K4me1 via its Tudor domain and localizes to regions exhibiting reduced mutation rates: coding regions, essential genes, constitutively expressed genes, and genes under stronger purifying selection, mirroring mutation biases observed in rice as well. We find that Tudor domains are significantly enriched in DNA repair proteins (p<1e-11). These include the mismatch repair *MSH6* protein, suggesting that plants have evolved multiple DNA repair pathways that target gene bodies and essential genes through H3K4me1 binding, which is supported by models of protein-peptide docking. These findings inspire further research to characterize mechanisms localizing DNA repair, potentially tuning the evolutionary trajectories of plant genomes.

Introduction

Mutations occur when DNA damage or replication error goes unrepaired. Mechanisms that localize DNA repair proteins to certain genome regions can therefore reduce local mutation rates. Interactions between DNA repair and histone modifications are predicted to evolve if they promote repair in regions that are prone to deleterious mutations, such as coding regions of essential genes (Lynch, 2010; Martincorena and Luscombe, 2013; Lynch et al., 2016).

Associations between histone modifications and mutation rates have been observed across diverse organisms (Habig et al., 2021; de la Peña et al., 2022; Yang et al., 2021; Yan et al., 2021; Monroe et al., 2022; Makova and Hardison, 2015; Schuster-Böckler and Lehner, 2012). That the localization of DNA repair proteins can drive such mutation biases has been well-established in humans (Supek and Lehner, 2015, 2017, 2019; Katju et al., 2022a; Foster et al., 2015). In vertebrates, H3K36me3 is targeted by PWWP domains in proteins contributing to homology-directed and mismatch repair, with H3K36me3 marking the gene bodies and exons of active genes (Li et al., 2013; Huang et al., 2018; Fang et al., 2021; Aymard et al., 2014; Sun et al., 2020). As predicted, reduced mutation rates in active genes, regions, and gene bodies have been observed in humans and other animals (Moore et al., 2021; Akdemir et al., 2020; Li et al., 2021; Katju et al., 2022b; Supek and Lehner, 2017).

Reduced mutation rates in gene bodies and active genes have also been observed in algae and land plants, but the mechanism has been unclear (Belfield et al., 2021; Lu et al., 2021; Zhu et al., 2021; Monroe et al., 2022; Belfield et al., 2018; López-Cortegano et al., 2021; Yan et al., 2021; Krasovec et al., 2017). Knockout lines of *msh2*, the mismatch repair protein that dimerizes with MSH6 to form MutSα, indicate that mismatch repair can preferentially target gene bodies in plants, yet a specific mechanism of such targeting is unresolved (Belfield et al., 2018). The precise mechanisms underlying these patterns in plants are unclear in part because plants lack gene body enrichment of H3K36me3 which functions as the target of DNA repair in vertebrate genomes.

Unlike in humans, in plants, H3K4me1 (rather than H3K36me3) marks the gene bodies of active genes. Recent work has demonstrated that H3K4me1 enrichment is

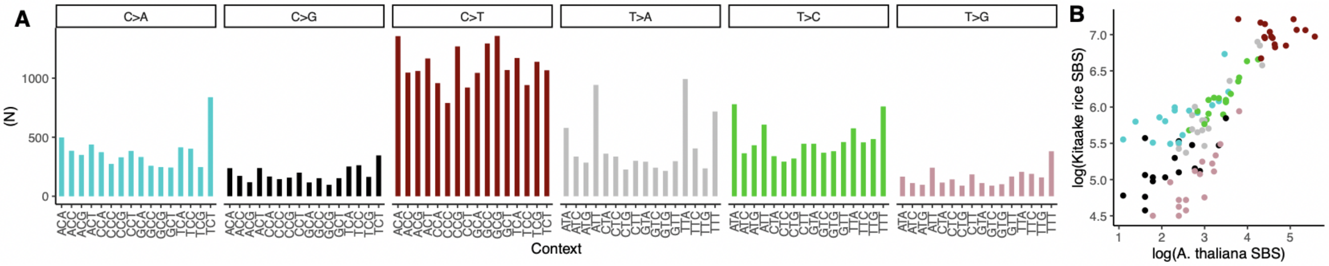


Figure 1. SBS mutations in fast neutron rice. A, Trinucleotide contexts of *de novo* mutations in rice were correlated with **B**, *de novo* germline mutations in *A. thaliana* ($r>0.8$, $p<2x10^{-16}$) (Weng et al. 2019, Monroe et al. 2022).

mediated by a combination of transcription-coupled (ATXR7) and epigenome-encoded (ATX1, ATX2) methyltransferases (Oya et al., 2021). Once established, H3K4me1 reading can then occur by proteins containing histone-reader domains such as “Royal family” Tudor domains, which bind methylated lysine residues on H3 histone tails (Kim et al., 2006; Lu and Wang, 2013; Maurer-Stroh et al., 2003).

Recently, the Tudor domain of PDS5C(RDM15) was shown to specifically bind H3K4me1 in *A. thaliana* (Niu et al., 2021). This gene is also shown to be a cohesion cofactor that facilitates homology-directed repair (Pradillo et al., 2015; Phipps and Dubrana, 2022; Morales et al., 2020). Recent studies of CRISPR-mediated mutation efficiency show that H3K4me1 is associated with lower mutation efficacy (R = -0.64), supporting more efficient repair (Weiss et al., 2022; Schep et al., 2021; Zhu et al., 2021). These findings are consistent with analyses of mutation accumulation lines in *A. thaliana* which indicate H3K4me1 to be associated with lower mutation rates (Monroe et al., 2022).

Mutagenesis has been used extensively in the generation and study of mutation in plants. With single base substitutions (SBS) in fast-neutron mutation accumulation lines largely reflecting native mutational processes (Wyant et al., 2022; Li et al., 2017), analyses of the distribution of these *de novo* mutations could provide insights into mechanisms underlying intragenomic heterogeneity in mutation rate. Here we analyzed *de novo* mutations from whole-genome-sequenced fast neutron mutation accumulation lines in Kitaake rice (Li et al., 2017) and then ask whether mutation rates are predicted by epigenomic features that could function as targets for DNA repair pathways.

Results and Discussion

Mutations in rice FN-irradiated lines reflect no selection

We first reanalyzed *de novo* mutations in a population of 1,504 rice lines that accumulated mutations upon fast neutron radiation. These data were previously described and single base-pair substitutions (SBS) were validated with a >99% true positive rate (Li et al., 2017). In total, these data included 43,483 SBS, reflecting a combination of fast neutron-related and “spontaneous” mutations (Fig. 1) This population was generated with minimal selection, evidenced by the existence of loss-of-function mutations detected in 28,419 genes, making it an exceptional resource for both functional genomics and mutation analysis (kitbase.ucdavis.edu). The ratio of non-synonymous to synonymous mutations in mutation accumulation lines was 2.33 (N/S=5,370/2,155), a 190% increase over this ratio (Pn/Ps = 1.21) observed in polymorphisms of 3,010 sequenced rice accessions (Wang et al., 2018) (X^2 = 670.63, $p<2x10^{-16}$). The ratio of non-synonymous to synonymous *de novo* mutations was not higher in transposable elements (TE) (N/S=2.31) than in non-TE protein-coding genes (N/S=2.34) (X^2 = 0.035, p = 0.85) nor was it less in coding genes than neutral expectations (N/S=2.33) based on mutation spectra and nucleotide composition of coding regions in the rice genome (X^2 = 0.029, df = 1, p = 0.86) (Fig. S1). Thus, the effects of selection appear to have been minimized to the point of undetectability in the generation of these mutation accumulation lines. Nevertheless, some selection (e.g. on loss-of-function hemizygous lethal sterility mutations) could, in principle, have occurred so we attempted to account for any such cryptic selection by restricting analyses to genes in which loss-of-function was found in this population and therefore apparently tolerated by whatever, if any, level of selection did occur (ie. the 28,419 genes where loss-of-function mutations were observed, N/S = 2.26 in these genes).

Compared with EMS-induced mutagenesis, the SBS spectra of fast neutron mutation lines more closely mirror spontaneous mutational patterns providing an opportunity to investigate the mechanisms governing intragenomic mutation rate heterogeneity (Li et al., 2017;

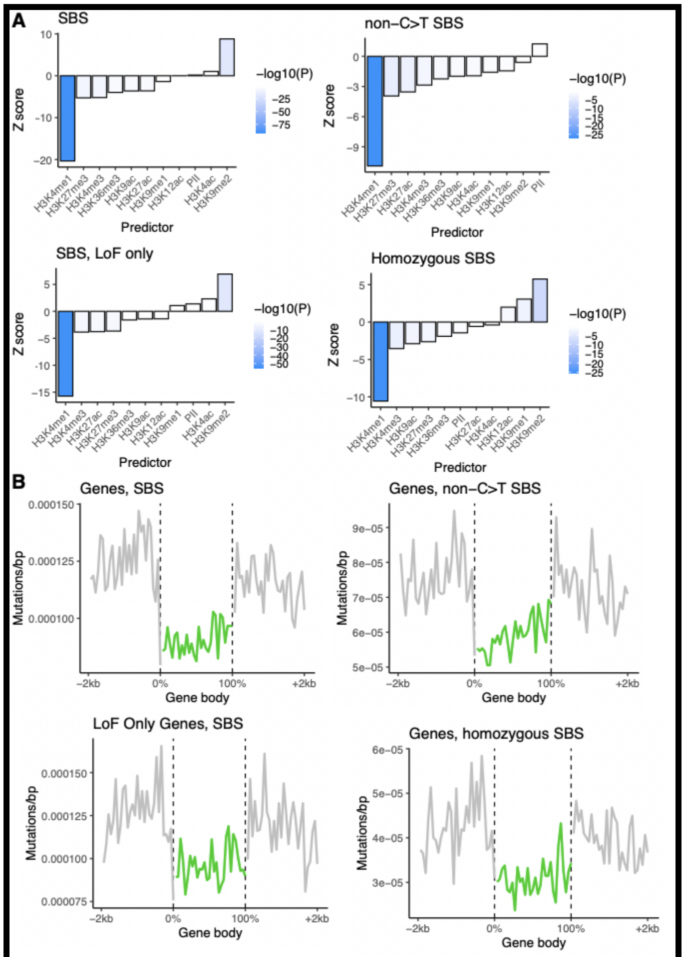


Figure 2. Relationships between epigenomic features, mutation and gene bodies A, Results from logistic regression modeling mutation probability in 100bp windows as a function of overlap with epigenomic marks around genes in rice. Four panels reflect various subsets of data to test for effects of cytosine deamination in transposable elements as alternative explanation, selection, or mapping. **B,** Mutation rates in relation to gene bodies in rice.

Wyant et al., 2022). We compared SBS spectra in trinucleotide contexts from these lines with *de novo* germline mutations in *A. thaliana (Weng et al., 2019; Monroe et al., 2022)* and found that they are significantly correlated (Fig.1, r=0.8, p<2x10^{-16}). We cannot know how much of the residual difference in SBS spectra is due to the effects of fast neutron mutagenesis versus inherent differences between rice and *A. thaliana*, as differences in the spectra of SBS have been reported between related species and even different genotypes of the same species (Jiang et al., 2021; Sasani et al., 2022; Cagan et al., 2022). Future work will benefit from the evaluation of *de novo* germline mutations arising in different species under diverse conditions to gain a complete understanding of the environmental and genetic controls of SBS mutational spectra in plants.

H3K4me1 marks low mutation rate regions in rice, gene bodies, transcriptionally active and evolutionarily conserved genes

To test whether the genome-wide distribution of mutations in rice is associated with histone modifications, we used data from the riceENCODE epigenomic database which includes H3K4me1, H3K9me1, H3K4me3, H3K36me3, H3K9me2, H3K27me3, H3K27ac, H3K4ac, H3K12ac, H3K9ac, and RNA polymerase II (PII) measured by chromatin immunoprecipitation sequencing (ChIP-seq) (Xie et al., 2021). We then tested whether mutation probabilities in 100bp windows in genic regions (the probability of observing a mutation) were predicted by epigenomic features and found a significant reduction in mutation probabilities in windows that overlapped with H3K4me1 peaks (Fig. 2A). These data are consistent with observations in other plant species where lower mutation rates have been observed in regions marked by H3K4me1 (Monroe et al., 2022; Weiss et al., 2022). While no evidence of selection was found in these data (Fig. S1), to account for the possibility that some undetectable selection occurred in the generation of this mutant population (ie. removing loss-of-function mutations in essential genes that would cause sterility or lethality), we restricted our analyses to only those genes in which loss-of-function mutations were found in the population and observed similar results (Fig. 2A). We considered the possibility that mutation rate heterogeneity was caused by GC>AT mutations in transposable elements with elevated cytosine methylation in non-genic sequences rather than histone-mediated mutation reduction, and therefore restricted our analyses to exclude all such GC>AT mutations and observed similar results, though H3K9me2 associated hypermutation was no longer detected (Fig. 2A). H3K4me1-associated hypomutation was also the same when analyses were restricted to only homozygous mutations (Fig. 2A). We calculated mutation rates in genes and their neighboring sequences and observed a significant reduction in mutation rates in gene bodies (Fig. 2B).

Mutation rates were lower both in and around H3K4me1 peaks, which could indicate the action of local recruitment and targeting of DNA repair to H3K4me1 marked sequences (Fig. 3A). That mutation rates were also lower in sequences immediately neighboring H3K4me1 peaks could indicate a spatially distributed effect on mutation in relation to H3K4me1 positioning, or the effect of conservative peak calling. Only 8.9% of H3K4me1 peaks were found outside of non-TE protein-coding genes. Nevertheless, we could use these instances of non-genic H3K4me1 to test whether the reduction in mutation rates in

H3K4me1 peaks was due simply to selection against coding region mutations having affected our results. When considering all H3K4me1 peaks, we observed a 20.1% reduction in mutation rates compared to regions within 2kb outside of peaks (X^2 = 124.38, p < 2.2e-16 -). For non-genic H3K4me1 peaks, we observed the same reduction: -20.2% (X^2 = 9.88, p = 0.00167). Together, these results suggested a role of H3K4me1 in localized hypomutation which could explain the reduced gene body mutation rates observed since gene bodies are enriched for H3K4me1 (Fig. 3B, Fig. 2B).

Mutation biases in rice were consistent with the expected effects of increased DNA repair in functionally constrained genes as well, which could be caused by H3K4me1-localized repair. H3K4me1 peaks were enriched in genes annotated as expressed compared with those not expressed (X^2 = 2550961, p < $2x10^{-16}$). Mutation rates were, as predicted by their enrichment for H3K4me1, 22% lower in expressed genes (X^2 = 63.7, p = $1x10^{-15}$). The ratio of non-synonymous to synonymous *de novo* mutations in the data was not different between expressed and non-expressed genes (X^2 = 0.0007, p = 0.98)(Fig. 3C). Comparing genes that exhibit different degrees of selection in natural accessions of rice, those under elevated purifying selection with low Pn/Ps (non-synonymous/synonymous polymorphisms), were enriched for H3K4me1 peaks (X^2 = 8045711, p < $2x10^{-16}$) and experienced 19% lower mutation rates (X^2 = 188.5, p < $2x10^{-16}$). These genes did not have a lower ratio of non-synonymous to synonymous in *de novo* mutations (X^2 = 0.22, p=0.63) (Fig. 3D). As such, we find no evidence that these patterns could be explained by selection in the mutation accumulation lines.

PDS5C targets H3K4me1 and is associated with lower mutation rates

Our findings are consistent with reports of reduced mutation rates in gene bodies of expressed and constrained genes in *A. thaliana* and other species (Krasovec et al., 2017; Moore et al., 2021; Monroe et al., 2022). While in humans, this is known to be mediated by H3K36me3 targeting by DNA repair genes, our results suggest that H3K4me1 may be a target of DNA repair in plants. To examine this further, we considered genes with known H3K4me1 targeting. *PDS5C*, a gene belonging to a family of cohesion cofactors that facilitate homology-directed repair (HDR) and possibly interact with nucleotide excision DNA repair (NER) repair pathway, contains a Tudor domain that was recently discovered to specifically bind H3K4me1 (Niu et al., 2021). Analyses of ChIP-seq data of PDS5C-Flag from *A. thaliana* show PDS5C is targeted to gene bodies (which are enriched for H3K4me1 in both rice and *A. thaliana*) (Fig. 4A). We also find that PDS5C is enriched in regions of lower

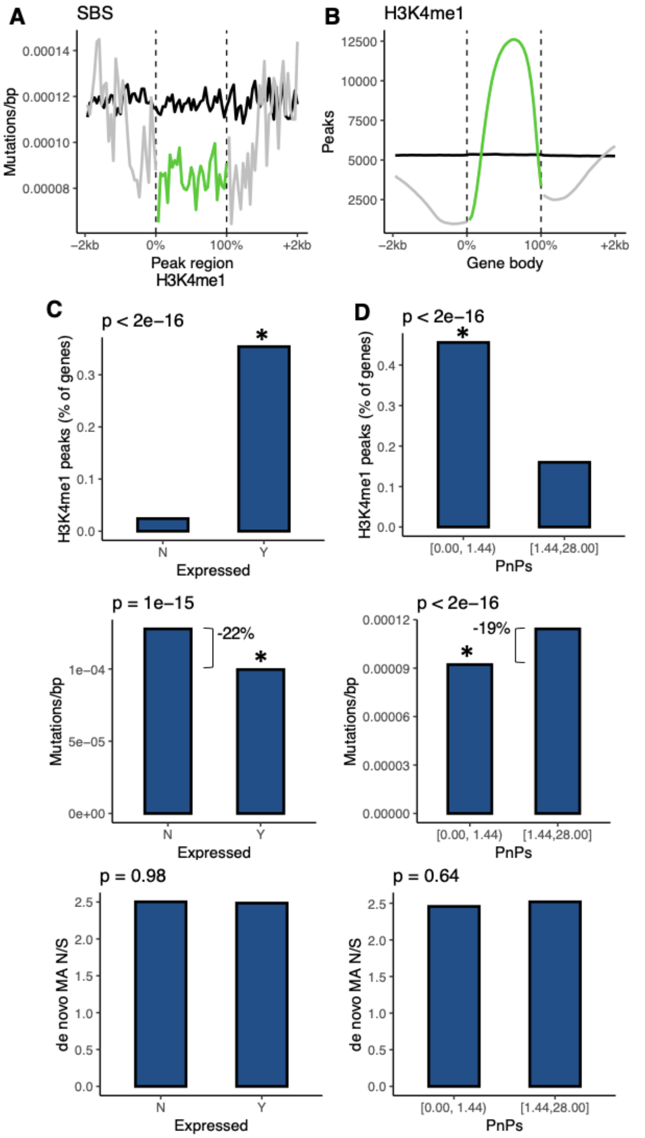


Figure 3. Relationships between H3K4me1, mutation and gene bodies, and function in rice A, Mutation rates are lower within and around H3K4me1 peaks in the rice genome. The black line shows results from randomly generated peaks matching lengths of H3K4me1 peak sizes. **B,** H3K4me1 is enriched in gene bodies of rice, consistent with observations in *A. thaliana* and other plants. The black line shows results from overlaps with randomly generated peaks matching lengths of H3K4me1 peak sizes. **C,** H3K4me1, gene body mutation rates, non-synonymous to synonymous *de novo* mutations in mutation accumulation lines (MA N/S) in expressed vs non-expressed genes and **D,** in genes under different selection in natural accessions; strong purifying selection (low Pn/Ps) vs more relaxed purifying selection or possibly balancing/positive selection (high Pn/Ps).

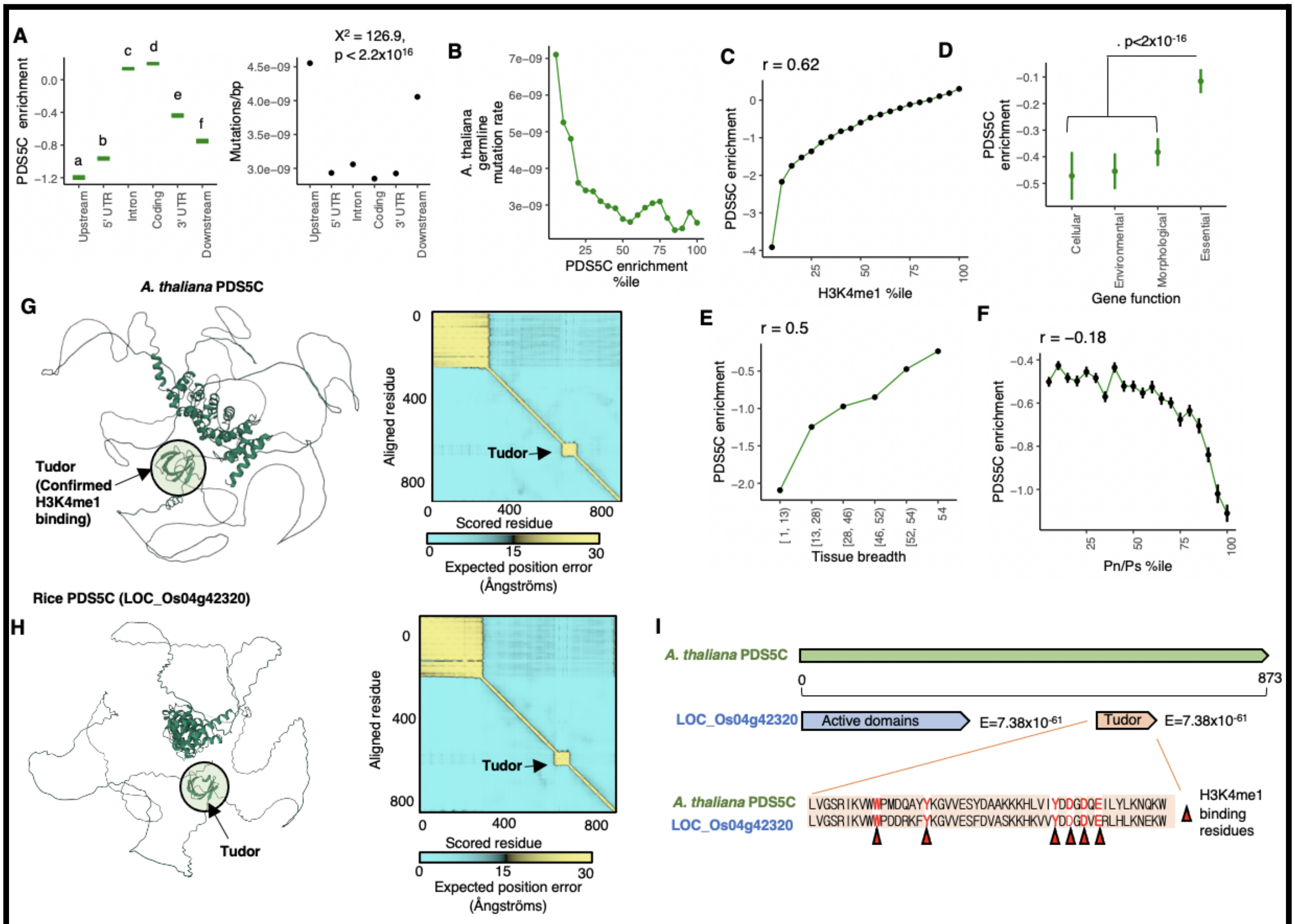
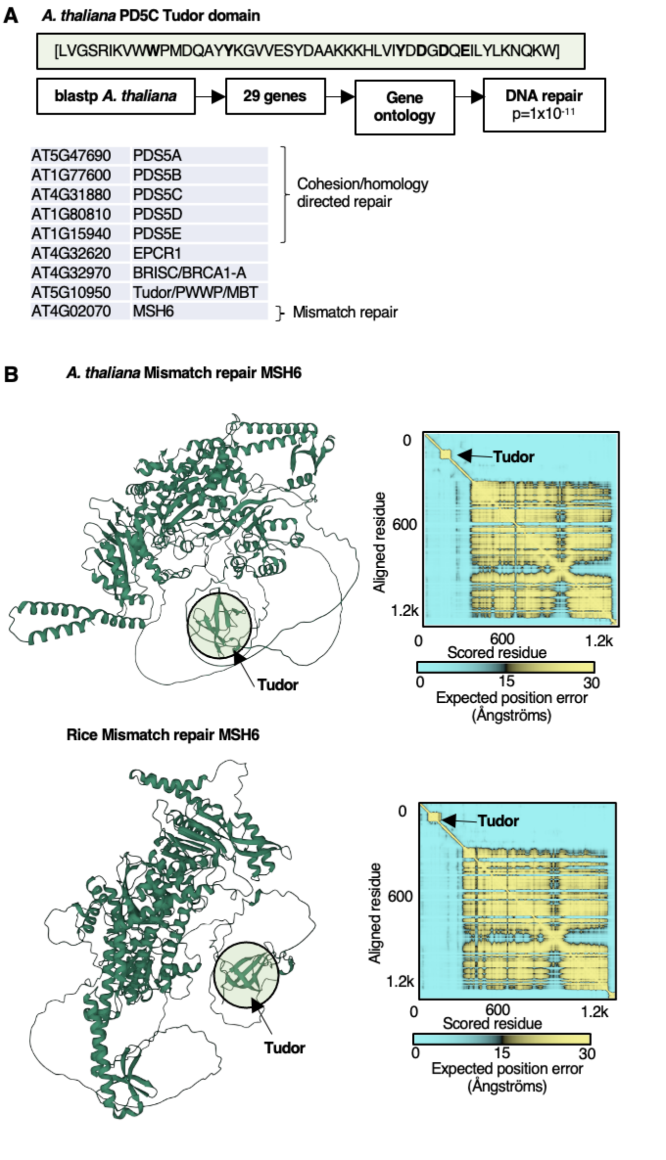


Figure 4. Example of potential H3K4me1 targeted repair mechanism. A, ChIP-seq analysis of PDS5C in *A thaliana*, cohesion cofactor involved in HDR and possibly NER confirms enrichment in features in gene bodies. Raw data from Niu et al. 2021. Letters represent TukeyHSD contrast at p=0. Upstream/downstream = 1000bp. **B,** Features marked by PDS5C experience reduced germline mutation rates in *A. thaliana*. **C,** Gene bodies of genes with elevated H3K4me1 have greater levels of PDS5C, as do **D**, essential genes, **E,** genes expressed in many tissues, and **F**, genes under stronger purifying selection. Pn/Ps = Non-synonymous /synonymous polymorphisms in *A. thaliana* populations. Mean +- SE for percentiles shown. Pearson correlation (r) reflects entire dataset. P<2x10^{-16} for all. **G**, Structure of PDS5C protein from Alphafold2 indicates Tudor domain is tethered to active domains. **H**, similar structure observed in putative rice ortholog, **I**, Comparison of *A. thaliana* PDS5C and putative rice ortholog LOC_Os4g43220. Results from *blastp* of A. thaliana PDS5C against rice proteome confirm conservation of active domain and Tudor domain. Sequence comparison of Tudor domain confirms conservation of amino acids experimentally validated by Niu et al. 2021 to be responsible for H3K4me1 binding specificity.

germline mutation rates in *A. thaliana*, consistent with its function in facilitating DNA repair (Fig. 4B).

Evolutionary models predict that histone-mediated repair mechanisms should evolve if they facilitate lower mutation rates in sequences under purifying selection. As predicted by this theory, we find PDS5C targeting (ChIP-seq) is enriched in coding sequences, essential genes (determined by experiments of knockout lines), and genes constitutively expressed (detected in 100% of tissues sampled), and genes under stronger purifying selection in natural populations of *A. thaliana* (lower Pn/Ps) (Fig. 4C-F).

Visualizing the PDS5C full-length model generated by Alphafold (Jumper et al., 2021) reveals that the PDS5C active domain is separated from the Tudor domain by long unstructured and flexible segments (Fig. 4G), suggesting that the Tudor domain operates as an anchor, localizing PDS5C to H3K4me1 and gene bodies of active genes. One interesting possibility is that the unstructured tether may enable elevated repair and thus lower mutation rates in regions adjoining PDS5C bound chromatin, such as in UTRs next to PDS5C enriched coding regions and introns. PDS5C is a cohesion cofactor linked to multiple DNA repair


A
A. thaliana PD5C Tudor domain
[LVGSRIKVWWPMDQAYYKGVVESYDAAKKKHLVIYDDGDQEILYLKNQKW]
blastp *A. thaliana*
29 genes
Gene ontology
DNA repair
p=1x10-11
Cohesion/homology directed repair
Mismatch repair
B
A. thaliana Mismatch repair MSH6
Tudor
Aligned residue
0
600
1.2k
Scored residue
Expected position error (Ångströms)
15
30
Rice Mismatch repair MSH6

AT5G47690	PDS5A
AT1G77600	PDS5B
AT4G31880	PDS5C
AT1G80810	PDS5D
AT1G15940	PDS5E
AT4G32620	EPCR1
AT4G32970	BRISC/BRCA1-A
AT5G10950	Tudor/PWWP/MBT
AT4G02070	MSH6

Figure 5. Evidence of multiple repair pathways with H3K4me1 targeting potential. A, Results from blastp of PCDS5C Tudor domain, which has confirmed H3K4me1 binding specificity (Niu et al. 2021) identified 29 genes. Nine of these are annotated as having DNA repair function (significant enrichment by gene ontology analysis). **B,** Mismatch repair gene MSH6 which contains a Tudor domain functions as a dimer with MSH2 to form MutSα, which could help explain how mismatch repair preferentially targets gene bodies, as others have described (Belfield et al. 2018).

pathways. In its role in cohesion between sister chromatids, it has been reported to promote HDR (Pradillo et al., 2015). This is consistent with its known interaction with repair proteins, and the contribution of cohesion and PDS orthologs to the HDR pathway (Morales et al., 2020; Phipps and Dubrana, 2022; Hill et al., 2016; Ren et al., 2005; Bolaños-Villegas et al., 2013; Schubert et al., 2009). PDS5C has also been found recently to interact with MED17 *in vivo*, which is involved in the NER pathway (Giustozzi et al., 2022). The observation that mutation rates are reduced at H3K4me1 peak regions (Fig. 2, Fig. 3) supports the hypothesis that Tudor domain-mediated targeting in PDS5C, its orthologs (PDS5A, B, D, and E), or other repair-related proteins contribute to targeted hypomutation in the functionally important regions of the genome. Still, additional experiments are needed to quantify the precise local effect of PDS5C on mutation rate. We compared the PDS5C Tudor domain sequence between *A. thaliana* and rice, and find that the critical amino acids constituting the aromatic cage, where H3K4me1 binding specificity is determined, are conserved (Fig. 4I), suggesting a potential role of PDS5C in the mutation biases observed here in rice (Fig. 2, Fig. 3).

Multiple DNA repair mechanisms could be influencing the mutation biases

The discovery of the PDS5C Tudor domain as an H3K4me1 targeting domain (Niu et al., 2021) provides an opportunity to identify other proteins with potential for H3K4me1-mediated gene body recruitment. We used *blastp* to search the *A. thaliana* proteome for other proteins containing Tudor domains similar to that of PDS5C (Fig. 5A). These revealed 29 genes encoding amino acid sequence regions similar to the PDS5C Tudor domain. An analysis of gene ontologies indicated that this gene set is highly enriched for genes with DNA repair functions (9/29 genes, $p=1x10^{-11}$). Five of these were PDS5C homologs. We also found that MSH6, a DNA mismatch repair protein, contains a Tudor domain similar to that of PDS5C, which was an obvious candidate for further consideration.

Structural modeling prediction of MSH6 structure using AlphaFold (Jumper et al., 2021) indicates that its Tudor domain may, like that in PDS5C, function as an anchor, tethering it to H3K4me1 leading to local increases in DNA repair (Fig. 5B). Sequence comparison suggests that MSH6 Tudor domain may present a similar binding preference for H3K4me1 as PDS5C Tudor domain, since key residues forming monomethylated lysine binding site are conserved between both homologous domains (Fig. 6A). In order to support this hypothesis, we modeled *A. thaliana* and *O. sativa* MSH6 and *O. sativa* PDS5C Tudor domains with AlphaFold and compared them with the experimental structure of *A. thaliana* PDS5C Tudor domain (PDB:7DE9) (Niu et al., 2021). Superimposition of the three modeled Tudor domains onto the PDS5C Tudor domain showed a remarkably similar fold and backbone root mean square deviation (RMSD) values below 1 Å (Fig. 6D-G, Fig. S2).

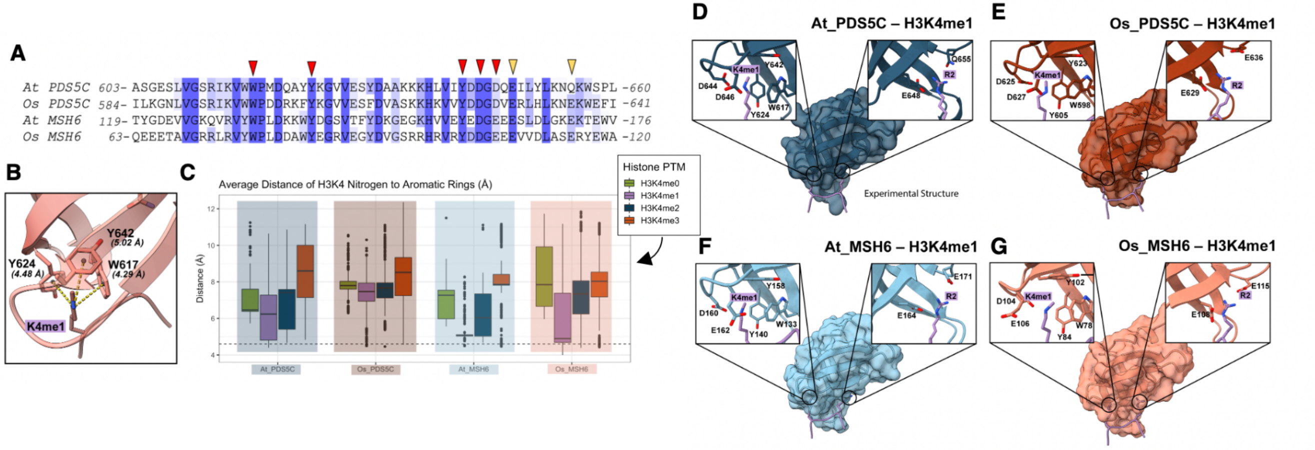


Figure 6. *In silico* prediction of *Arabidopsis thaliana* and *Oryza sativa* PDS5C and MSH6 Tudor domains affinity to H3K4 methylation. A, ClustalW alignment of Tudor domains. Red arrows mark residues shown to be important for H3K4me1 specific binding, and yellow arrows mark residues important to interact with H3R2 (Niu et al. 2021) **B,** Aromatic cage and H3K4me1in experimental structure of At_PDS5C (PDB: 7DE9) (Niu et al. 2021) highlighting in yellow the distances to the center of the three aromatic rings. **C,** Average distance from the side chain nitrogen of H3K4 to the center of aromatic rings forming the aromatic cage (see B) of top 1,000 models (10%) based on Rosetta total score after docking H3 tail peptides in different methylation states to the Tudor domains of PDS5C and MSH6 from both A. thaliana and O. sativa. The dashed line marks the average distance in the experimental structure as a reference. **D-G,** FlexPepDock models of the H3K4me1 peptide in the aromatic cage of Tudor domains of At_PDS5C, At_MSH6, Os_PDS5C and OS_MSH6. Residues interacting with H3K4me1 are amplified in the left boxes. Residues interacting with H3R2 are highlighted in the right boxes.

Subsequently, we modified the K4me1 from the H3 tail peptide bound to the PDS5C Tudor domain in ChimeraX (Goddard et al., 2018) to obtain H3K4, H3K4me2 and H3K4me3 H3 tail peptides. Using Rosetta FlexPepDock (Raveh et al., 2010), we simulated docking of the different methylation states of H3K4 to the Tudor domains of PDS5C and MSH6 from *A. thaliana* and *O. sativa*. We analyzed the interface scores and geometries of top 10% of models based on Rosetta total score. Interface scores shows a trend for preferential affinity for the H3K4me1 followed by H3K4me2 (Fig. S3). This is consistent with ITC experimental results: *A. thaliana* PDS5C Tudor domain binds H3K4me2 peptide with the second-best affinity (K_D: 30.9 uM) after H3K4me1 (K_D: 1.47 uM) (Niu et al., 2021). As the main molecular interactions involved in H3K4me1 binding are cation-π types, which are difficult to capture in computational energy functions, we reasoned that there might be some energy contributions that are not being accounted for and that might be important to determine the affinity difference between monomethylated and dimethylated states (Daze and Hof, 2013). We therefore analyzed the binding site geometry of the top docking models, measuring the average distance of H4K4 nitrogen to the center of the 3 aromatic rings forming the aromatic cage in the Tudor domain. The experimental structure of the PDS5C Tudor domain bound to H3K4me1 provided a reference to compare *in silico* models (PDB:7DE9) (Niu et al., 2021) where the average distance between H3K4 nitrogen and the aromatic rings is 4.6 Å (Fig. 6B). For all Tudor domains, FlexPepDock sampled geometries close to the experimental more often for H3K4me1 peptide than for the other peptides (Fig. 6C). This result suggests a lower energetic barrier to access the aromatic cage for H3K4 when it is monomethylated than for other methylation states. Interestingly, lack of methylation resulted in no sampling of geometries within the aromatic cage, which aligns with ITC experimental results that show no detectable binding of H3 tail peptide by the PDS5C Tudor domain when K4 was not methylated (Niu et al., 2021). In summary, based on previous experimental data and our computational analysis, we conclude that MSH6 and PDS5C Tudor domains bind to H3K4 preferentially when it is in the monomethylated state (H3K4me1).

The similar binding affinity of PDS5C and MSH6 Tudor domains suggests that multiple repair pathways could have evolved H3K4me1-targeting potential in plants, motivating additional experiments and investigations into the evolutionary origins of these mechanisms. Because MSH6 operates as a dimer with MSH2 to form the MutSα complex which recognizes and repairs small mismatches, its Tudor domain could explain the previous observation that MSH2 preferentially targets gene bodies to reduce mutation rates therein (Belfield et al., 2018). These findings are consistent with extensive work showing that mutation rates can be lower in gene bodies of active and conserved genes, but suggest that these mechanisms are independent of and

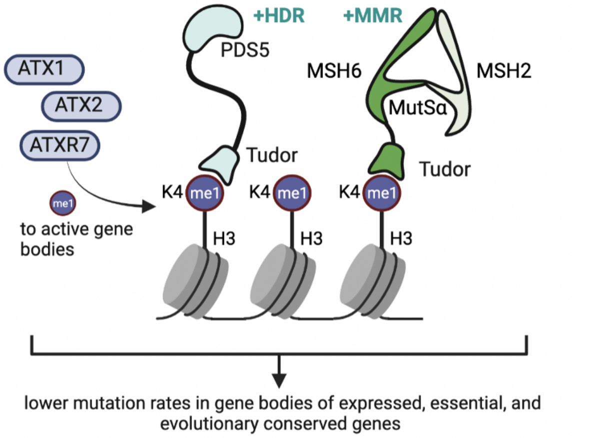


Figure 7. Emerging mechanistic model of H3K4me1-targeted DNA repair. ATX1, ATX2, ATXR7 are responsible for the enrichment of H3K4me1 in the gene bodies of active genes (Oya et al. 2021). Tudor domain of PDS5C binds H3K4me1 and facilitates homology directed repair (HDR) (Niu et al. 2021). Other repair genes including PDS5 paralogs and MSH6 (responsible for mismatch repair – MMR) also contain functionally similar Tudor domains, with H3K4me1 targeting potential.

functionally analogous to similar mechanisms known in vertebrates (H3K36me3 targetted repair described in introduction). Further experiments are needed - the reduced mutation rates in gene bodies in active genes in plants may be explained by multiple mechanisms collectively targeting H3K4me1 or additional histone states via Tudor domains as well as other histone modifications and readers (Davarinejad et al., 2022; Liu et al., 2022).

Conclusions

We found evidence of mutation bias associated with H3K4me1-mediated DNA repair in rice and examined potential mechanisms conserved in plants (Fig. 7). Our observations here, are derived from reanalyses of data generated by independent research groups (Li et al., 2017; Niu et al., 2021; Xie et al., 2021) and consistent with previous reports of mutation biases in *A. thaliana* (Monroe et al., 2022). The mechanisms revealed are aligned with evolutionary models of evolved mutation bias, indicating targetting of mismatch repair and homology-directed repair pathways to regions of the genome functioanlly sensitive to mutation: coding regions and genes under stronger evolutionary constraints. These findings provide a plant-specific and higher resolution mechanistic model of hypomutation in gene bodies and essential genes, motivating experimental investigations to further elucidate the extent and evolutionary origins of targetted DNA repair.

Acknowledgments

The research was conducted at the University of California Davis, which is located on land which was the home of the Patwin people for thousands of years. We thank Satoyo Oya, Tetsuji Kakutani, and Soichi Inagaki for feedback and insights on H3K4me1.

Methods

Mutation dataset in rice

Germline *de novo* mutations in 1,504 fast neutron mutagenesis lines were downloaded from Kitbase at kitbase.ucdavis.edu. These were independently called and validated as previously described (Li et al., 2017). We focused specifically on single base substitutions (SBS) which were validated with a >99% accuracy by Li et al (2017). We annotated each SBS in coding regions as being a synonymous or non-synonymous mutation based on the effect on the amino acid sequence. We compared non-synonymous and synonymous ratios with values from genomes of 3,010 natural accessions (Wang et al., 2018) and neutral expectations based on mutation spectra, coding region nucleotide composition, and codon table with the *Null_ns_s* function from the *polymorphology* package in R.

Epigenomic data collection

Epigenome features were accessed from the RiceENCODE database (glab.hzau.edu.cn/RiceENCODE/) which has been previously described (Xie et al., 2021). In brief, peaks were called from ChIP-seq data with MACS2 (Zhang et al., 2008) narrow-peak calling settings. We analyzed peak distributions for H3K4me1, H3K9me1, H3K4me3, H3K36me3, H3K9me2, H3K27me3, H3K27ac, H3K4ac, H3K12ac, H3K9ac, and RNA polymerase II (PII) measured in Nipponbare rice plant seedlings, which constituted the most complete set of histone modifications available. We repeated analyses but with H3K4me1 measurements derived from panicles and leaves (rather than seedlings) and found essentially the same results.

Estimation of the relationship between mutation rates and rice epigenomic features

We divided the genome into 100 bp windows surrounding genes (+- 3000 bp of genes). This allowed us to, in later steps, restrict our analyses to only genes known to have accumulated loss-of-function mutations, and thus be less likely to be affected by selection. We also divided the genome into 100bp windows and repeated analyses, to confirm that results were generally the same. We calculated the number of single base-pair substitutions and peaks for each epigenomic feature overlapping within each window. We then estimated the relationships between epigenomic features and mutation rates with a binomial generalized linear model where the response was a binary state defined as whether a substitution occurred in that window, predicted by all features, with predictors defined as whether that window overlapped with an epigenome peak. We also repeated the analyses with a linear regression model where the response was the number of mutations in a window and found essentially the same results, so we show the binomial regression results. To test whether findings were driven simply by GC>AT mutations in transposable elements, we removed all GC>AT and repeated analyses. To further control for any residual selection in the mutation accumulation experiment, we also restricted our analyses to genes harboring loss-of-function mutations in the population and the repeated analyses. Finally, we restricted analyses to homozygous SBS and repeated the analyses. Mutation frequencies were plotted around genes in 100 bp windows. Since gene bodies are different lengths, the position of the window was converted into a percent of gene length. H3K4me1 peaks around gene bodies were plotted similarly. We also visualized mutation frequencies relative to H3K4me1 peaks in the same manner.

Analysis of ChIP-seq data of AtPDS5C

To study the distribution of PDS5C, we used ChIP-seq data as described by Niu et al. (2021). PDS5C enrichment was calculated as described by Niu et al. (2021) among regions as log2[(1 + n_ChIP)/N_ChIP] − log2[(1 + n_Input)/N_Input)], where n_ChIP and n_Input represent the total depth of mapped ChIP and Input fragments in a region, and N_ChIP and N_Input are the numbers total depths of mapped unique fragments. We calculated PDS5C enrichment in genic features (1000 bp upstream and downstream of genes, UTRs, introns, coding regions) and gene bodies (TSS to TTS) across the TAIR10 *A. thaliana* genome (arabidopsis.org).

Relationship between AtPDS5C and functional constraint

We analyzed the enrichment of the PDS5C ChIP-seq peaks in *A. thaliana* in genetic features and estimated the relationships between those regions and mutation rates, H4K4me1, Pn/Ps, tissue expression depth. Tissue expression data are from (Mergner et al., 2020). H3K4me1 in Arabidopsis is from the Plant Chromatin State Database (Liu et al., 2018). Synonymous (Ps) and non-synonymous polymorphism (Pn) data are from the 1001 Genomes project (1001 Genomes Consortium, 2016). Essential genes were based on findings from (Lloyd and Meinke, 2012). Germline mutation rates are from (Weng et al., 2019; Monroe et al., 2022).

Blastp and protein structure prediction and visualization

We used *blastp* on Phytozome(Goodstein et al., 2012) to search the rice proteome for PDS5C and MHS6 orthologs, and to search the *A. thaliana* proteome for genes containing Tudor domains similar to that of PDS5C, which was validated experimentally to bind H3K4me1(Niu et al., 2021). We submitted the resulting list of 29 genes with putative Tudor domains to gene ontology analysis with ShinyGO (bioinformatics.sdstate.edu/go/)(Ge et al., 2019).

Protein structure predictions were performed using AlphaFold(Jumper et al., 2021) in Google Colab (Mirdita et al., 2022) in no-template mode. All structures were visualized, processed, and analyzed using UCSF ChimeraX (Goddard et al., 2018).

Peptide docking

H3 tail peptides comprising 5 amino acids with the different methylation states for K4 (none, mono, di or trimethylated) were docked to the experimental structure of *A. thaliana* PDS5C Tudor domain and to the models of *A. thaliana* and *O. sativa* MSH6 and *O. sativa* PDS5C Tudor domains using Rosetta FlexPepDock tool (Raveh et al., 2010) in refinement mode. We generated 10,000 docked models per case and analyzed the top 10% based on Rosetta total score. Analysis of the outputs was conducted using PyRosetta home-made scripts (Chaudhury et al., 2010).

Code and data

Figures, code, and data are also located on: https://github.com/greymonroe/rice_mutation_project.

References

1001 Genomes Consortium (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell **166**: 481–491.

Akdemir, K.C. et al. (2020). Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. Nat. Genet. **52**: 1178–1188.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P., and Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at

DNA double-strand breaks. Nat. Struct. Mol. Biol. **21**: 366–374.

Belfield, E.J., Brown, C., Ding, Z.J., Chapman, L., Luo, M., Hinde, E., van Es, S.W., Johnson, S., Ning, Y., Zheng, S.J., Mithani, A., and Harberd, N.P. (2021). Thermal stress accelerates Arabidopsis thaliana mutation rate. Genome Res. **31**: 40–50.

Belfield, E.J., Ding, Z.J., Jamieson, F.J.C., Visscher, A.M., Zheng, S.J., Mithani, A., and Harberd, N.P. (2018). DNA mismatch repair preferentially protects genes from mutation. Genome Res. **28**: 66–74.

Bolaños-Villegas, P., Yang, X., Wang, H.-J., Juan, C.-T., Chuang, M.-H., Makaroff, C.A., and Jauh, G.-Y. (2013). Arabidopsis CHROMOSOME TRANSMISSION FIDELITY 7 (AtCTF7/ECO1) is required for DNA repair, mitosis and meiosis. Plant J. **75**: 927–940.

Cagan, A. et al. (2022). Somatic mutation rates scale with lifespan across mammals. Nature **604**: 517–524.

Chaudhury, S., Lyskov, S., and Gray, J.J. (2010). PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics **26**: 689–691.

Davarinejad, H. et al. (2022). The histone H3.1 variant regulates TONSOKU-mediated DNA repair during replication. Science **375**: 1281–1286.

Daze, K.D. and Hof, F. (2013). The cation-π interaction at protein-protein interaction interfaces: developing and learning from synthetic mimics of proteins that bind methylated lysines. Acc. Chem. Res. **46**: 937–945.

Fang, H., Zhu, X., Yang, H., Oh, J., Barbour, J.A., and Wong, J.W.H. (2021). Deficiency of replication-independent DNA mismatch repair drives a 5-methylcytosine deamination mutational signature in cancer. Sci Adv **7**: eabg4398.

Foster, P.L., Lee, H., Popodi, E., Townes, J.P., and Tang, H. (2015). Determinants of spontaneous mutation in the bacterium Escherichia coli as revealed by whole-genome sequencing. Proc. Natl. Acad. Sci. U. S. A. **112**: E5990–9.

Ge, S.X., Jung, D., and Yao, R. (2019). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics **36**: 2628–2629.

Giustozzi, M., Freytes, S.N., Jaskolowski, A., Lichy, M., Mateos, J., Falcone Ferreyra, M.L., Rosano, G.L., Cerdán, P., and Casati, P. (2022). Arabidopsis mediator subunit 17 connects transcription with DNA repair after UV-B exposure. Plant J.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. **27**: 14–25.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. **40**: D1178–86.

Habig, M., Lorrain, C., Feurtey, A., Komluski, J., and Stukenbrock, E.H. (2021). Epigenetic modifications affect the rate of spontaneous mutations in a pathogenic fungus. Nat. Commun. **12**: 5869.

Hill, V.K., Kim, J.-S., and Waldman, T. (2016). Cohesin mutations in human cancer. Biochim. Biophys. Acta **1866**: 1–11.

Huang, Y., Gu, L., and Li, G.-M. (2018). H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. J. Biol. Chem. **293**: 7811–7823.

Jiang, P., Ollodart, A.R., Sudhesh, V., Herr, A.J., Dunham, M.J., and Harris, K. (2021). A modified fluctuation assay reveals a natural mutator phenotype that drives mutation spectrum variation within Saccharomyces cerevisiae. Elife **10**.

Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature **596**: 583–589.

Katju, V., Konrad, A., Deiss, T.C., and Bergthorsson, U. (2022a). Mutation rate and spectrum in obligately outcrossing Caenorhabditis elegans mutation accumulation lines subjected to RNAi-induced knockdown of the mismatch repair gene msh-2. G3 **12**.

Katju, V., Konrad, A., Deiss, T.C., and Bergthorsson, U. (2022b). Mutation rate and spectrum in obligately outcrossing Caenorhabditis elegans mutation accumulation lines subjected to RNAi-induced knockdown of the mismatch repair gene msh-2. G3 **12**: jkab364.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep. **7**: 397–403.

Krasovec, M., Eyre-Walker, A., Sanchez-Ferandin, S., and Piganeau, G. (2017). Spontaneous Mutation Rate in the Smallest Photosynthetic Eukaryotes. Mol. Biol. Evol. **34**: 1770–1779.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.-M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. Cell **153**: 590–600.

Li, G. et al. (2017). The Sequences of 1504 Mutants in the Model Rice Variety Kitaake Facilitate Rapid Functional Genomic Studies. Plant Cell **29**: 1218–1231.

Li, R. et al. (2021). A body map of somatic mutagenesis in morphologically normal human tissues. Nature **597**: 398–403.

Liu, Q., Liu, P., Ji, T., Zheng, L., Shen, C., Ran, S., Liu, J., Zhao, Y., Niu, Y., Wang, T., and Dong, J. (2022). The histone methyltransferase SUVR2 promotes DSB repair via chromatin remodeling and liquid-liquid phase separation. Mol. Plant.

Liu, Y., Tian, T., Zhang, K., You, Q., Yan, H., Zhao, N., Yi, X., Xu, W., and Su, Z. (2018). PCSD: a plant chromatin state database. Nucleic Acids Res. **46**: D1157–D1167.

Lloyd, J. and Meinke, D. (2012). A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. Plant Physiol. **158**: 1115–1129.

López-Cortegano, E., Craig, R.J., Chebib, J., Samuels, T., Morgan, A.D., Kraemer, S.A., Böndel, K.B., Ness, R.W., Colegrave, N., and Keightley, P.D. (2021). De Novo Mutation Rate Variation and Its Determinants in Chlamydomonas. Mol. Biol. Evol. **38**: 3709–3723.

Lu, R. and Wang, G.G. (2013). Tudor: a versatile family of histone methylation "readers." Trends Biochem. Sci. **38**: 546–555.

Lu, Z., Cui, J., Wang, L., Teng, N., Zhang, S., Lam, H.-M., Zhu, Y., Xiao, S., Ke, W., Lin, J., Xu, C., and Jin, B. (2021). Genome-wide DNA mutations in Arabidopsis plants after multigenerational exposure to high temperatures. Genome Biol. **22**: 160.

Lynch, M. (2010). Evolution of the mutation rate. Trends Genet. **26**: 345–352.

Lynch, M., Ackerman, M.S., Gout, J.-F., Long, H., Sung, W., Thomas, W.K., and Foster, P.L. (2016). Genetic drift, selection and the evolution of the mutation rate. Nat. Rev. Genet. **17**: 704–714.

Makova, K.D. and Hardison, R.C. (2015). The effects of chromatin organization on variation in mutation rates in the genome. Nat. Rev. Genet. **16**: 213–223.

Martincorena, I. and Luscombe, N.M. (2013). Non-random mutation: the evolution of targeted hypermutation and hypomutation. Bioessays **35**: 123–130.

Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem. Sci. **28**: 69–74.

Mergner, J. et al. (2020). Mass-spectrometry-based draft of the Arabidopsis proteome. Nature **579**: 409–414.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods **19**: 679–682.

Monroe, J.G. et al. (2022). Mutation bias reflects natural selection in Arabidopsis thaliana. Nature.

Moore, L. et al. (2021). The mutational landscape of human somatic and germline cells. Nature.

Morales, C., Ruiz-Torres, M., Rodríguez-Acebes, S., Lafarga, V., Rodríguez-Corsino, M., Megías, D., Cisneros, D.A., Peters, J.-M., Méndez, J., and Losada, A. (2020). PDS5 proteins are required for proper cohesin dynamics and participate in replication fork protection. J. Biol. Chem. **295**: 146–157.

Niu, Q. et al. (2021). A histone H3K4me1-specific binding protein is required for siRNA accumulation and DNA methylation at a subset of loci targeted by RNA-directed DNA methylation. Nat. Commun. **12**: 3367.

Oya, S., Takahashi, M., Takashima, K., Kakutani, T., and Inagaki, S. (2021). Transcription-coupled and epigenome-encoded mechanisms direct H3K4 methylation. bioRxiv: 2021.06.03.446702.

de la Peña, M.V., Summanen, P.A.M., Liukkonen, M., and Kronholm, I. (2022). Variation in spontaneous mutation rate and spectrum across the genome of Neurospora crassa. bioRxiv: 2022.03.13.484164.

Phipps, J. and Dubrana, K. (2022). DNA Repair in Space and Time: Safeguarding the Genome with the Cohesin Complex. Genes **13**.

Pradillo, M., Knoll, A., Oliver, C., Varas, J., Corredor, E., Puchta, H., and Santos, J.L. (2015). Involvement of the Cohesin Cofactor PDS5 (SPO76) During Meiosis and DNA Repair in Arabidopsis thaliana. Front. Plant Sci. **6**: 1034.

Raveh, B., London, N., and Schueler-Furman, O. (2010). Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins **78**: 2029–2040.

Ren, Q., Yang, H., Rosinski, M., Conrad, M.N., Dresser, M.E., Guacci, V., and Zhang, Z. (2005). Mutation of the cohesin related gene PDS5 causes cell death with predominant apoptotic features in Saccharomyces cerevisiae during early meiosis. Mutat. Res. **570**: 163–173.

Sasani, T.A., Ashbrook, D.G., Beichman, A.C., Lu, L., Palmer, A.A., Williams, R.W., Pritchard, J.K., and Harris, K. (2022). A natural mutator allele shapes mutation spectrum variation in mice. Nature **605**: 497–502.

Schep, R. et al. (2021). Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. Mol. Cell **81**: 2216–2230.e10.

Schubert, V., Weissleder, A., Ali, H., Fuchs, J., Lermontova, I., Meister, A., and Schubert, I. (2009). Cohesin gene defects may impair sister chromatid alignment and genome stability in Arabidopsis thaliana. Chromosoma **118**: 591–605.

Schuster-Böckler, B. and Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature **488**: 504–507.

Sun, Z., Zhang, Y., Jia, J., Fang, Y., Tang, Y., Wu, H., and Fang, D. (2020). H3K36me3, message from chromatin to DNA damage repair. Cell Biosci. **10**: 9.

Supek, F. and Lehner, B. (2017). Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. Cell **170**: 534–547.e23.

Supek, F. and Lehner, B. (2015). Differential DNA mismatch repair underlies mutation rate variation across the human genome. Nature **521**: 81–84.

Supek, F. and Lehner, B. (2019). Scales and mechanisms of somatic mutation rate variation across the human genome. DNA Repair **81**: 102647.

Wang, W. et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature **557**: 43–49.

Weiss, T., Crisp, P.A., Rai, K.M., Song, M., Springer, N.M., and Zhang, F. (2022). Drastic differential CRISPR-Cas9 induced mutagenesis influenced by DNA methylation and chromatin features. bioRxiv: 2022.02.28.482333.

Weng, M.-L., Becker, C., Hildebrandt, J., Neumann, M., Rutter, M.T., Shaw, R.G., Weigel, D., and Fenster, C.B. (2019). Fine-Grained Analysis of Spontaneous Mutation Spectrum and Frequency in Arabidopsis thaliana. Genetics **211**: 703–714.

Wyant, S.R., Rodriguez, M.F., Carter, C.K., Parrott, W.A., Jackson, S.A., Stupar, R.M., and Morrell, P.L. (2022). Fast neutron mutagenesis in soybean enriches for small indels and creates frameshift mutations. G3 **12**.

Xie, L., Liu, M., Zhao, L., Cao, K., Wang, P., Xu, W., Sung, W.-K., Li, X., and Li, G. (2021). RiceENCODE: A comprehensive epigenomic database as a rice Encyclopedia of DNA Elements. Mol. Plant **14**: 1604–1606.

Yang, X. et al. (2021). Developmental and temporal characteristics of clonal sperm mosaicism. Cell **184**: 4772–4783.e15.

Yan, W., Deng, X.W., Yang, C., and Tang, X. (2021). The Genome-Wide EMS Mutagenesis Bias Correlates With Sequence Context and Chromatin Structure in Rice. Front. Plant Sci. **12**: 579675.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. **9**: R137.

Zhu, X., Xie, S., Tang, K., Kalia, R.K., Liu, N., Ma, J., Bressan, R.A., and Zhu, J.-K. (2021). Non-CG DNA methylation-deficiency mutations enhance mutagenesis rates during salt adaptation in cultured Arabidopsis cells. Stress Biology **1**: 12.

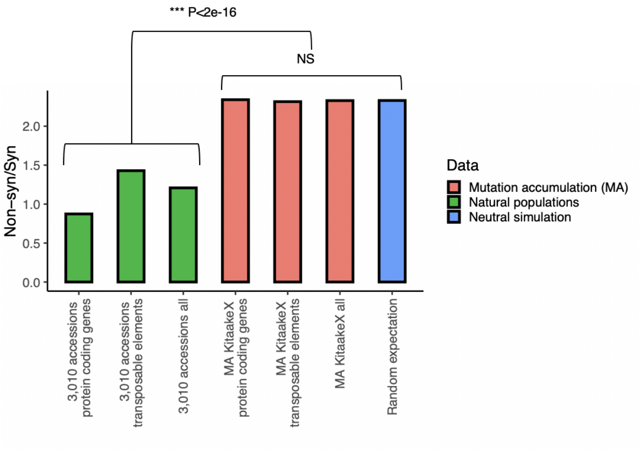


Figure S1. Non-synonymous to synonymous ratios. The non-synonymous to synonymous ratio was significantly lower ($p<2.x10^{-16}$) in natural populations compared to mutation accumulation lines and neutral expectation. The non-synonymous to synonymous ratio was similar in mutation accumulation lines to neutral expectation.

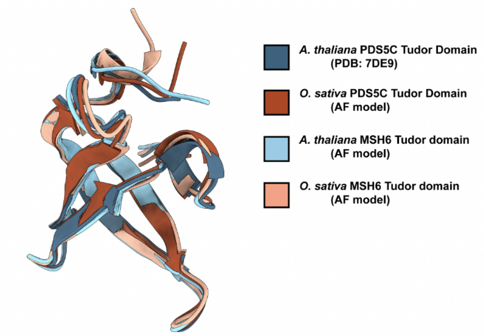

A. thaliana PDS5C Tudor Domain (PDB: 7DE9)
O. sativa PDS5C Tudor Domain (AF model)
A. thaliana MSH6 Tudor domain (AF model)
O. sativa MSH6 Tudor domain (AF model)

Figure S2. AlphaFold predictions of Tudor domains. AlphaFold models of *A. thaliana* and *O. sativa* MSH6 and *O. sativa* PDS5C Tudor are shown superimposed onto the experimental structure of A. thaliana PDS5C Tudor domain (PDB:7DE9) (Niu et al. 2021).

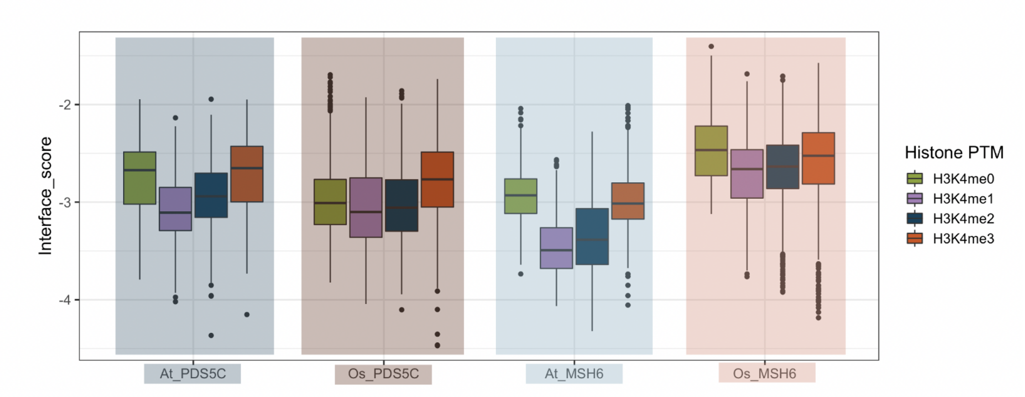


Figure S3. Interface score analysis of peptide docking simulations. Boxplots show the average interface score of top 1,000 models (10%) based on Rosetta total score after docking H3 tail peptides with different K3 modifications to the four Tudor domains.