bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

Relationship between biased mutagenesis and H3K4me1-targeted DNA repair in plants

Daniela Quiroz[1], Kehan Zhao[1], Pablo Carbonell-Bejerano[2], Grey Monroe[1,#]

[1] Department of Plant Sciences, University of California Davis, Davis, CA, USA 95616
[2] Institute for Grape and Wine Sciences (ICVV, CSIC-CAR-UR), 26007 Logroño, La Rioja, Spain
[#] Correspondence: gmonroe@ucdavis.edu

Abstract

Mutation is the ultimate source of genetic variation in natural populations and crops. To study mechanisms determining mutation rate variation within plant genomes, we analyzed 43,483 *de novo* germline single base substitutions in 1,504 fast neutron mutation lines of the model rice cultivar Kitaake (*Oryza sativa* ssp *japonica*) (from Li et al. 2017). We find that, like previously observed for *de novo* germline mutations in *Arabidopsis thaliana*, mutation rates are significantly lower in genomic regions marked by H3K4me1, a histone modification found in the gene bodies of actively expressed genes in plants. We also 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. By examining existing ChIP-seq data for PDS5C in *A. thaliana*, we find that it localizes to genome regions marked by H3K4me1: regions of low mutation rates, coding regions, essential genes, constitutively expressed genes, and genes under stronger purifying selection, mirroring mutation biases observed in rice as well. We searched the *A. thaliana* proteome for genes containing similar Tudor domains and found that they are significantly enriched for DNA repair functions ($p<1x10^{-11}$), including the mismatch repair *MSH6* gene (in both rice and *A. thaliana*), suggesting the potential for multiple DNA repair pathways to specifically target gene bodies and essential genes through H3K4me1 reading. These findings inspire further research to characterize mechanisms that localize DNA repair via histone interactions, leading to hypomutation in functionally constrained regions and potentially tuning the evolutionary trajectories of plant genomes.

Introduction

Mutation occurs when DNA damage or replication error goes unrepaired. Mechanisms that localize DNA repair proteins to certain genome regions can therefore affect local rates of mutation. Interactions between DNA repair and histone modifications are predicted to evolve if they promote repair in regions 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 such mutation biases can be driven by the localization of DNA repair proteins has been established empirically, especially in studies of mutation rate variation within human genomes (Supek and Lehner 2015, 2017, 2019; Katju et al. 2022b; Foster et al. 2015). In vertebrates, H3K36me3 is targeted by PWWP domains in genes contributing to homology-directed and mismatch repair. H3K36me3 marks the gene bodies and exons of active genes in humans (F. Li et al. 2013; Huang, Gu, and Li 2018; Fang et al. 2021; Aymard et al. 2014; Sun et al. 2020), and reduced mutation rates in active genes, regions, and gene bodies, as a consequence, have been observed in humans and other animals (Moore et al. 2021; Akdemir et al. 2020; R. Li et al. 2021; Katju et al. 2022a; Supek and Lehner 2017).

Reduced mutation rates in gene bodies and active genes have also been observed in studies of mutation rates in algae and land plants (Belfield et al. 2021; Z. 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). However, the precise mechanisms underlying these patterns in plants are unclear, as plants lack the enrichment of H3K36me3 in gene bodies that is characteristic of vertebrate genomes. Knockout lines of *msh2*, the mismatch repair gene that dimerizes with MSH6 to form MutSα, experience elevated relative mutation rates in gene bodies in *A. thaliana*, indicating that mismatch repair

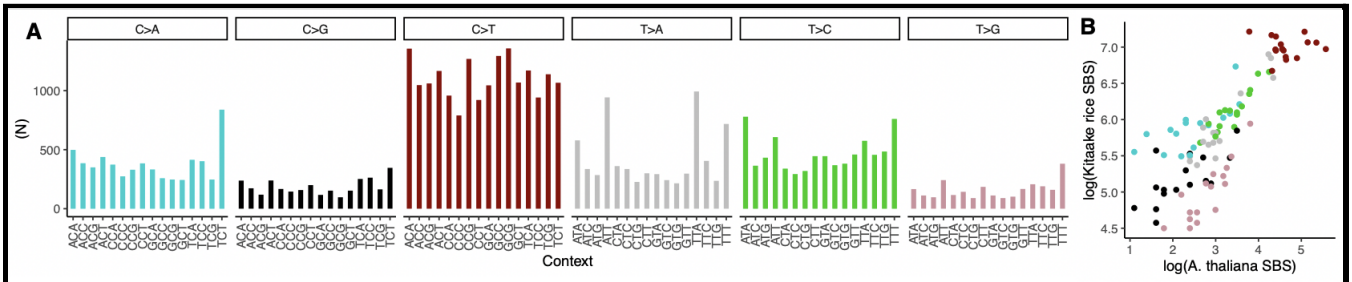

Figure 1. SBS mutations in fast neutron rice. A, Trinucleotide contexts of *de novo* mutations in rice were correlated with **B**, *de novo* germline mutations in *A. thaliana* ($r>0.8$, $p<2x10^{-16}$) (Weng et al. 2019, Monroe et al. 2022).

can preferentially target gene bodies, yet a specific mechanism of such targeting has been unknown (Belfield et al. 2018).

Unlike in humans, in plants, H3K4me1 marks the gene bodies of active genes. Recent work has demonstrated that this is mediated by a combination of transcription-coupled (ATXR7) and epigenome-encoded (ATX1, ATX2) methyltransferases (Oya et al. 2021). Once placed, H3K4me1 reading can then occur by proteins containing histone-reader domains such as Tudor "Royal family" domains, which bind methylated lysine residues on H3 histone tails (Kim et al. 2006; R. 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 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; G. 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 (G. Li et al. 2017) and ask whether mutation rates are predicted by epigenomic features that could function as targets for DNA repair pathways.

Results and Discussion

We analyzed *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 (G. 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. The ratio of non-synonymous to synonymous mutations in mutation accumulation lines was 2.33 (N/S=5,370/2,155), a 1.9X 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 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 undectability 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

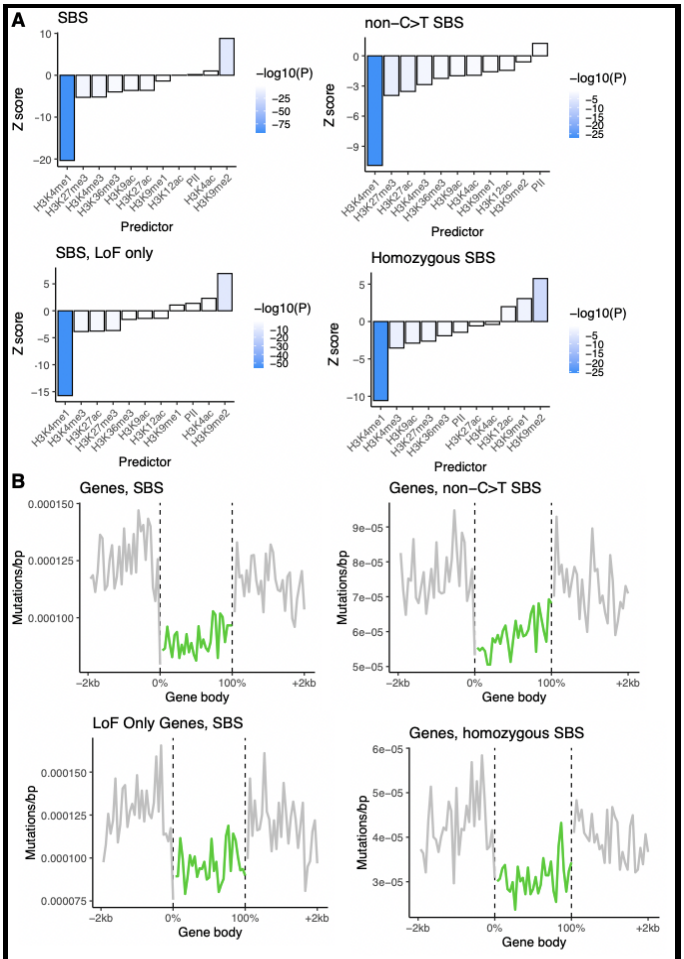

Figure 2. Relationships between epigenomic features, mutation and gene bodies A, Results from logistic regression modeling mutation probability in 100bp windows as a function of overlap with epigenomic marks around genes in rice. Four panels reflect various subsets of data to test for effects of cytosine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

intragenomic mutation rate heterogeneity (G. Li et al. 2017; 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 ($r=0.8$, $p<2x10^{-16}$). We cannot know how much of the residual difference in SBS spectra is due to 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 (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 mutational spectra in plants.

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. 2). 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). To account for the possibility that some 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. 2). 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. 2). H3K4me1-associated hypomutation was also the same when analyses were restricted to only homozygous mutations (Fig. 2). We calculated mutation rates in genes and their neighboring sequences and observed a significant reduction in mutation rates in gene bodies (Fig. 2).

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. 3). 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. 2).

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, potentially as a consequence of enrichment for H3K4me1, 22% lower in expressed genes (X^2 = 63.7, p = $1x10^{-15}$) while 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). 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 < $2x10^{-16}$), experienced 19% lower mutation rates (X^2 = 188.5, p < $2x10^{-16}$), but did not have a lower ratio of non-synonymous to synonymous in *de novo* mutations (X^2 = 0.22, p=0.63) (Fig. 3). As such, we find no evidence that these patterns could be explained by selection in the mutation accumulation lines.

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 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. 4). Here we also find that PDS5C is enriched in regions of lower germline mutation rates in *A. thaliana*, consistent with its function in facilitating DNA repair (Fig. 4).

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, and genes constitutively expressed across tissues, and genes under stronger purifying selection in natural populations of *A. thaliana* (Fig. 4).

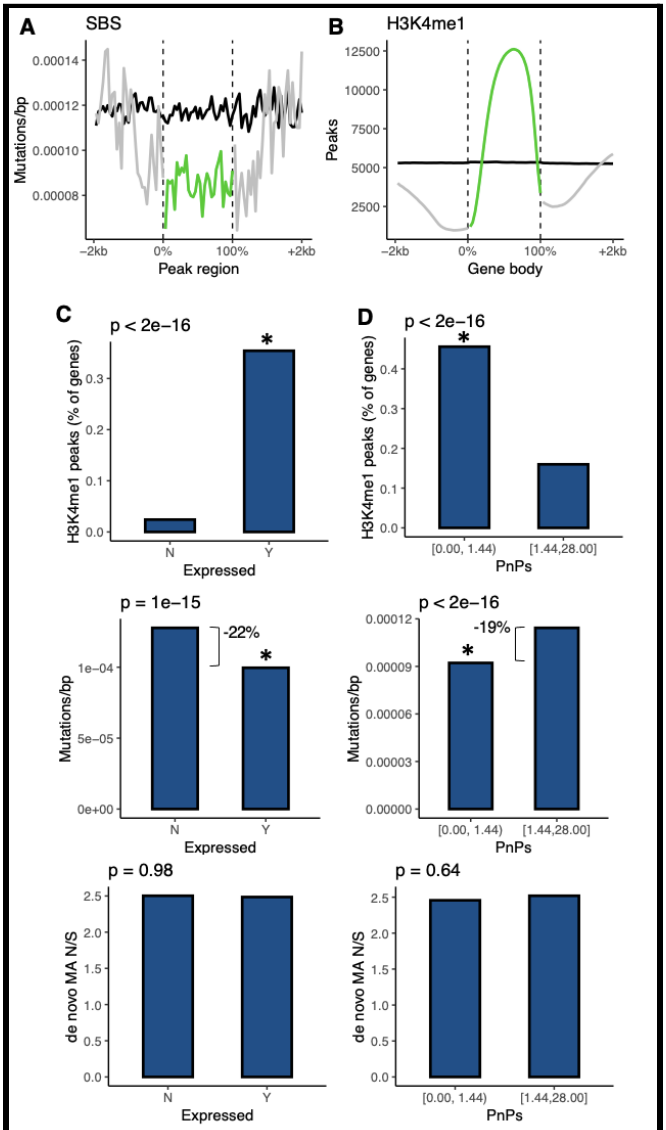

Figure 3. Relationships between H3K4me1, mutation and gene bodies, and function in rice A, Mutation rates are lower within and around H3K4me1 peaks in the rice genome. The black line shows results from randomly generated peaks matching lengths of H3K4me1 peak sizes. **B,** H3K4me1 is enriched in gene 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).

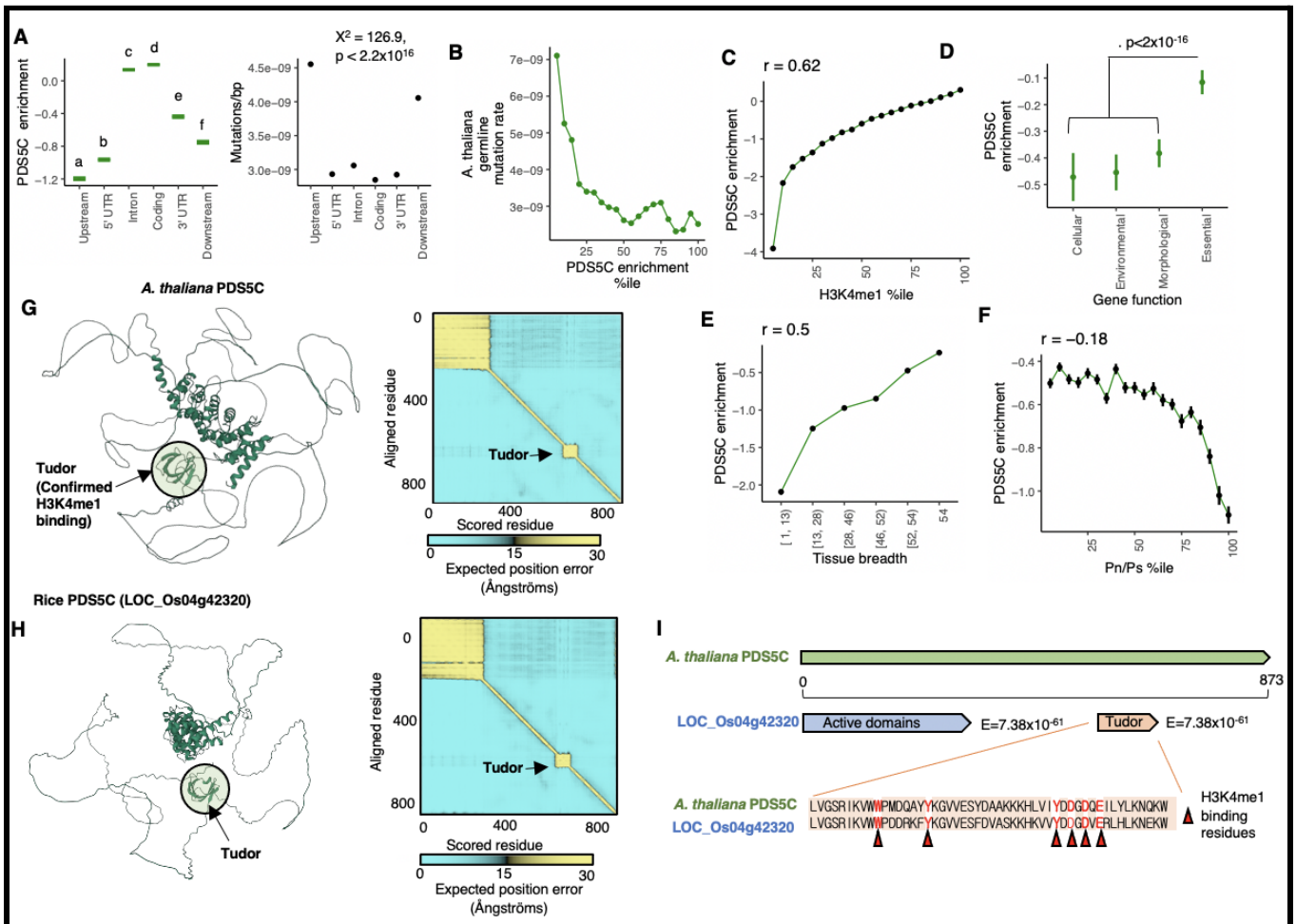

Figure 4. Example of potential H3K4me1 targeted repair mechanism. A, ChIP-seq analysis of PDS5C in *A thaliana*, cohesion cofactor involved in HDR and possibly NER confirms enrichment in features in gene bodies. Raw data from Niu et al. 2021. Letters represent TukeyHSD contrast at p=0. Upstream/downstream = 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.

Visualizing its structure generated by Alphafold2 (Jumper et al. 2021) reveals that the PDS5C active domain is separated from the Tudor domain by several hundred unstructured amino acids, 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 contribute to lower mutation rates in regions adjoining PDS5C, such as in UTRs next to enriched coding regions and introns. PDS5C is a cohesion cofactor linked to multiple DNA repair pathways. In its role in cohesion between sister chromatids, it has been reported to promote HDR (Pradillo et al. 2015). This is consistent with the known contribution of cohesion and PDS orthologs in other species to the HDR pathway (Morales et al. 2020; Phipps and Dubrana 2022; Hill, Kim, and Waldman 2016; Ren et al. 2005). PDS5C has also been found recently to interact with MED17 *in vivo*, which may be involved in the NER pathway (Giustozzi et al. 2022). The observation that mutation rates are reduced at H3K4me1 peak regions (Fig. 3) supports the hypothesis that Tudor domain-mediated targeting in PDS5C or other repair-related

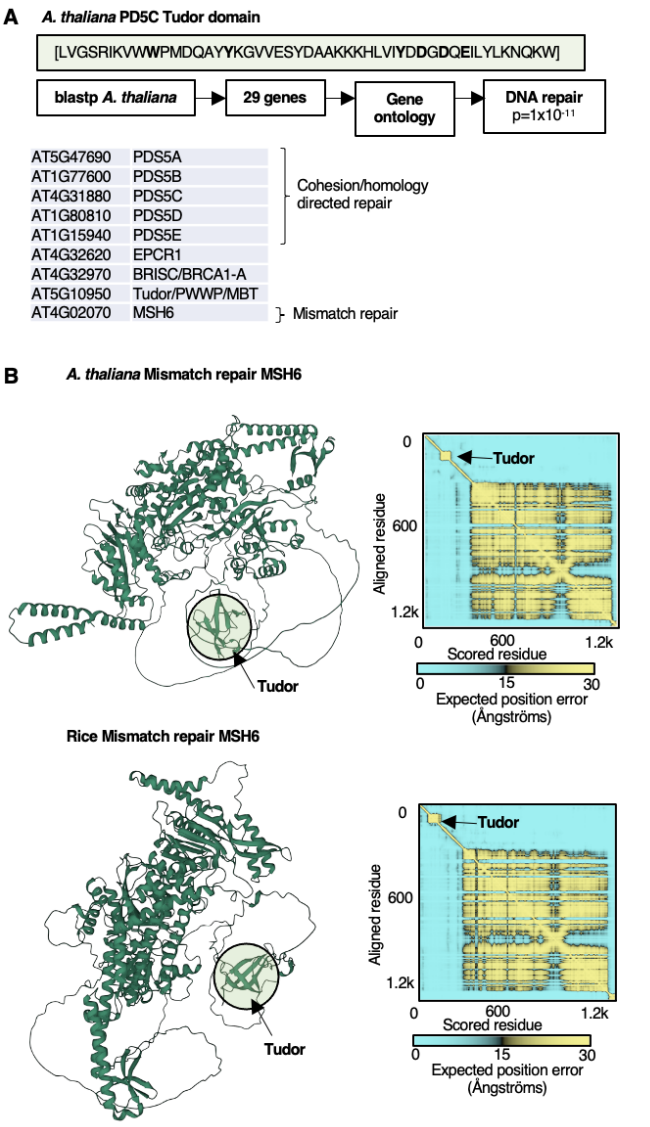

Figure 5. Evidence of multiple repair pathways with H3K4me1 targeting potential. A, Results from blastp of PCDS5C Tudor domain, which has confirmed H3K4me1 binding specificity (Niu et al. 2021) identified 29 genes. Nine of these are annotated as having DNA repair function (significant enrichment by gene 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).

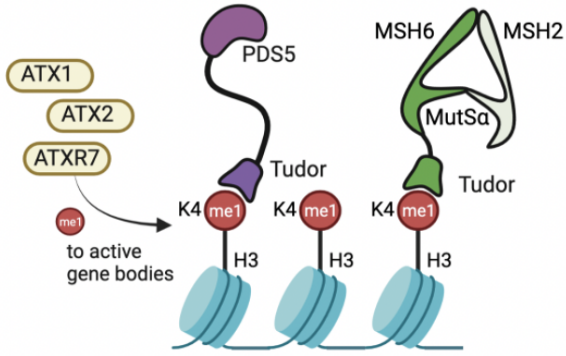

Figure 6. Potential mechanistic model of H3K4me1-targeted DNA repair. ATX1, ATX2, ATXR7 are responsible for the enrichment of H3K4me1 in the gene bodies of active genes (Oya et al. 2021). Tudor domain of PDS5C binds H3K4me1 (Niu et al. 2021). Other repair genes with Tudor domains, such as PDS paralogs and MSH6 may also target H3K4me1 via Tudor domain binding. Graphic made in Biorender.

genes contributes 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 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. 4), suggesting a potential role of PDS5C in the mutation biases observed here in rice (Fig. 2, Fig. 3).

The discovery of the PDS5C Tudor domain as an H3K4me1 targeting domain (Niu et al. 2021) provides an opportunity to identify other genes with potential for H3K4me1-mediated gene body recruitment. We used *blastp* to search the *A. thaliana* proteome for other genes containing Tudor domains similar to that of PDS5C (Fig. 5). These revealed 29 genes with 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 PDS5 orthologs. We also found that mismatch repair MSH6 contains a Tudor domain similar to that of PDS5C, which is an obvious candidate for further consideration. This 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. The structure of MSH6 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. 5). 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 a body of work showing that mutation rates can be lower in gene bodies of active and conserved genes. 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 other histone states via Tudor domains or other histone readers (Fig. 6).

We found evidence of mutation bias caused by H3K4me1-mediated DNA repair in rice, and examined potential mechanisms. Our findings here, derived from reanalyses of data generated by independent research groups (G. Li et al. 2017; Niu et al. 2021; Xie et al. 2021), are consistent with our previous observations of mutation biases in *A. thaliana* (Monroe et al. 2022). They are aligned with evolutionary models of evolved mutation bias and provide a higher resolution mechanistic model of gene body and essential genes hypomutation in plants, motivating experimental investigations into the several DNA repair pathways with potential to target active and constrained genes.

Acknowledgements

Research was conducted at the University of California Davis on land which was the home of Patwin people for thousands of years. We are honored and grateful to be on their traditional lands.

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 (G. 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 effect on 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 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 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).

Protein 3D structure and *blastp*

We used Alphafold2 predicted structures (alphafold.ebi.ac.uk/) to visualize PDS5C and MSH6 in *A. thaliana (Jumper et al. 2021)*. We also generated predicted structures for rice PDS5C and MSH6 with Alphafold2's Google Colab tool (AlphaFold.ipynb). We confirmed the structures from Alphafold's own database for *A. thaliana* genes.

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, Jung, and Yao 2019).

Code and data

Figures, code, and data are 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 (2): 481–91.

Akdemir, Kadir C., Victoria T. Le, Justin M. Kim, Sarah Killcoyne, Devin A. King, Ya-Ping Lin, Yanyan Tian, et al. 2020. "Somatic Mutation Distributions in Cancer Genomes Vary with Three-Dimensional Chromatin Structure." *Nature Genetics* 52 (11): 1178–88.

Aymard, François, Beatrix Bugler, Christine K. Schmidt, Emmanuelle Guillou, Pierre Caron, Sébastien Briois, Jason S. Iacovoni, et al. 2014. "Transcriptionally Active Chromatin Recruits Homologous Recombination at DNA Double-Strand Breaks." *Nature Structural & Molecular Biology* 21 (4): 366–74.

Belfield, Eric J., Carly Brown, Zhong Jie Ding, Lottie Chapman, Mengqian Luo, Eleanor Hinde, Sam W. van Es, et al. 2021. "Thermal Stress Accelerates Arabidopsis Thaliana Mutation Rate." *Genome Research* 31 (1): 40–50.

Belfield, Eric J., Zhong Jie Ding, Fiona J. C. Jamieson, Anne M. Visscher, Shao Jian Zheng, Aziz Mithani, and Nicholas P. Harberd. 2018. "DNA Mismatch Repair Preferentially Protects Genes from Mutation." *Genome Research* 28 (1): 66–74.

Cagan, Alex, Adrian Baez-Ortega, Natalia Brzozowska, Federico Abascal, Tim H. H. Coorens, Mathijs A. Sanders, Andrew R. J. Lawson, et al. 2022. "Somatic Mutation Rates Scale with Lifespan across Mammals." *Nature* 604 (7906): 517–24.

Fang, Hu, Xiaoqiang Zhu, Haocheng Yang, Jieun Oh, Jayne A. Barbour, and Jason W. H. Wong. 2021. "Deficiency of Replication-Independent DNA Mismatch Repair Drives a 5-Methylcytosine Deamination Mutational Signature in Cancer." *Science Advances* 7 (45): eabg4398.

Foster, Patricia L., Heewook Lee, Ellen Popodi, Jesse P. Townes, and Haixu Tang. 2015. "Determinants of Spontaneous Mutation in the Bacterium Escherichia Coli as Revealed by Whole-Genome Sequencing." *Proceedings of the National Academy of Sciences of the United States of America* 112 (44): E5990–99.

Ge, Steven Xijin, Dongmin Jung, and Runan Yao. 2019. "ShinyGO: A Graphical Gene-Set Enrichment Tool for Animals and Plants." *Bioinformatics* 36 (8): 2628–29.

Giustozzi, Marisol, Santiago Nicolás Freytes, Aime Jaskolowski, Micaela Lichy, Julieta Mateos, María Lorena Falcone Ferreyra, Germán L. Rosano, Pablo Cerdán, and Paula Casati. 2022. "Arabidopsis Mediator Subunit 17 Connects Transcription with DNA Repair after UV-B Exposure." *The Plant Journal: For Cell and Molecular Biology*, February. https://doi.org/10.1111/tpj.15722.

Goodstein, David M., Shengqiang Shu, Russell Howson, Rochak Neupane, Richard D. Hayes, Joni Fazo, Therese Mitros, et al. 2012. "Phytozome: A Comparative Platform for Green Plant Genomics." *Nucleic Acids Research* 40 (Database issue): D1178–86.

Habig, Michael, Cecile Lorrain, Alice Feurtey, Jovan Komluski, and Eva H. Stukenbrock. 2021. "Epigenetic Modifications Affect the Rate of Spontaneous Mutations in a Pathogenic Fungus." *Nature*

Communications 12 (1): 5869.
Hill, Victoria K., Jung-Sik Kim, and Todd Waldman. 2016. "Cohesin Mutations in Human Cancer." *Biochimica et Biophysica Acta* 1866 (1): 1–11.
Huang, Yaping, Liya Gu, and Guo-Min Li. 2018. "H3K36me3-Mediated Mismatch Repair Preferentially Protects Actively Transcribed Genes from Mutation." *The Journal of Biological Chemistry* 293 (20): 7811–23.
Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* 596 (7873): 583–89.
Katju, Vaishali, Anke Konrad, Thaddeus C. Deiss, and Ulfar Bergthorsson. 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 (1): jkab364.
———. 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 (1). https://doi.org/10.1093/g3journal/jkab364.
Kim, Jeesun, Jeremy Daniel, Alexsandra Espejo, Aimee Lake, Murli Krishna, Li Xia, Yi Zhang, and Mark T. Bedford. 2006. "Tudor, MBT and Chromo Domains Gauge the Degree of Lysine Methylation." *EMBO Reports* 7 (4): 397–403.
Krasovec, Marc, Adam Eyre-Walker, Sophie Sanchez-Ferandin, and Gwenael Piganeau. 2017. "Spontaneous Mutation Rate in the Smallest Photosynthetic Eukaryotes." *Molecular Biology and Evolution* 34 (7): 1770–79.
Li, Feng, Guogen Mao, Dan Tong, Jian Huang, Liya Gu, Wei Yang, and Guo-Min Li. 2013. "The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutSα." *Cell* 153 (3): 590–600.
Li, Guotian, Rashmi Jain, Mawsheng Chern, Nikki T. Pham, Joel A. Martin, Tong Wei, Wendy S. Schackwitz, et al. 2017. "The Sequences of 1504 Mutants in the Model Rice Variety Kitaake Facilitate Rapid Functional Genomic Studies." *The Plant Cell* 29 (6): 1218–31.
Li, Ruoyan, Lin Di, Jie Li, Wenyi Fan, Yachen Liu, Wenjia Guo, Weiling Liu, et al. 2021. "A Body Map of Somatic Mutagenesis in Morphologically Normal Human Tissues." *Nature* 597 (7876): 398–403.
Liu, Yue, Tian Tian, Kang Zhang, Qi You, Hengyu Yan, Nannan Zhao, Xin Yi, Wenying Xu, and Zhen Su. 2018. "PCSD: A Plant Chromatin State Database." *Nucleic Acids Research* 46 (D1): D1157–67.
Lloyd, Johnny, and David Meinke. 2012. "A Comprehensive Dataset of Genes with a Loss-of-Function Mutant Phenotype in Arabidopsis." *Plant Physiology* 158 (3): 1115–29.
López-Cortegano, Eugenio, Rory J. Craig, Jobran Chebib, Toby Samuels, Andrew D. Morgan, Susanne A. Kraemer, Katharina B. Böndel, Rob W. Ness, Nick Colegrave, and Peter D. Keightley. 2021. "De Novo Mutation Rate Variation and Its Determinants in Chlamydomonas." *Molecular Biology and Evolution* 38 (9): 3709–23.
Lu, Rui, and Gang Greg Wang. 2013. "Tudor: A Versatile Family of Histone Methylation 'Readers.'" *Trends in Biochemical Sciences* 38 (11): 546–55.
Lu, Zhaogeng, Jiawen Cui, Li Wang, Nianjun Teng, Shoudong Zhang, Hon-Ming Lam, Yingfang Zhu, et al. 2021. "Genome-Wide DNA Mutations in Arabidopsis Plants after Multigenerational Exposure to High Temperatures." *Genome Biology* 22 (1): 160.
Lynch, Michael. 2010. "Evolution of the Mutation Rate." *Trends in Genetics: TIG* 26 (8): 345–52.
Lynch, Michael, Matthew S. Ackerman, Jean-Francois Gout, Hongan Long, Way Sung, W. Kelley Thomas, and Patricia L. Foster. 2016. "Genetic Drift, Selection and the Evolution of the Mutation Rate." *Nature Reviews. Genetics* 17 (11): 704–14.
Makova, Kateryna D., and Ross C. Hardison. 2015. "The Effects of Chromatin Organization on Variation in Mutation Rates in the Genome." *Nature Reviews. Genetics* 16 (4): 213–23.
Martincorena, Iñigo, and Nicholas M. Luscombe. 2013. "Non-Random Mutation: The Evolution of Targeted Hypermutation and Hypomutation." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 35 (2): 123–30.
Maurer-Stroh, Sebastian, Nicholas J. Dickens, Luke Hughes-Davies, Tony Kouzarides, Frank Eisenhaber, and Chris P. Ponting. 2003. "The Tudor Domain 'Royal Family': Tudor, Plant Agenet, Chromo, PWWP and MBT Domains." *Trends in Biochemical Sciences* 28 (2): 69–74.
Mergner, Julia, Martin Frejno, Markus List, Michael Papacek, Xia Chen, Ajeet Chaudhary, Patroklos Samaras, et al. 2020. "Mass-Spectrometry-Based Draft of the Arabidopsis Proteome." *Nature* 579 (7799): 409–14.
Monroe, J. Grey, Thanvi Srikant, Pablo Carbonell-Bejerano, Claude Becker, Mariele Lensink, Moises Exposito-Alonso, Marie Klein, et al. 2022. "Mutation Bias Reflects Natural Selection in Arabidopsis Thaliana." *Nature*, January. https://doi.org/10.1038/s41586-021-04269-6.
Moore, Luiza, Alex Cagan, Tim H. H. Coorens, Matthew D. C. Neville, Rashesh Sanghvi, Mathijs A. Sanders, Thomas R. W. Oliver, et al. 2021. "The Mutational Landscape of Human Somatic and Germline Cells." *Nature*, August. https://doi.org/10.1038/s41586-021-03822-7.
Morales, Carmen, Miguel Ruiz-Torres, Sara Rodríguez-Acebes, Vanesa Lafarga, Miriam Rodríguez-Corsino, Diego Megías, David A. Cisneros,

Jan-Michael Peters, Juan Méndez, and Ana Losada. 2020. “PDS5 Proteins Are Required for Proper Cohesin Dynamics and Participate in Replication Fork Protection.” *The Journal of Biological Chemistry* 295 (1): 146–57.

Niu, Qingfeng, Zhe Song, Kai Tang, Lixian Chen, Lisi Wang, Ting Ban, Zhongxin Guo, 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.” *Nature Communications* 12 (1): 3367.

Oya, Satoyo, Mayumi Takahashi, Kazuya Takashima, Tetsuji Kakutani, and Soichi Inagaki. 2021. “Transcription-Coupled and Epigenome-Encoded Mechanisms Direct H3K4 Methylation.” *bioRxiv*. https://doi.org/10.1101/2021.06.03.446702.

Peña, Mariana Villalba de la, Pauliina A. M. Summanen, Martta Liukkonen, and Ilkka Kronholm. 2022. “Variation in Spontaneous Mutation Rate and Spectrum across the Genome of Neurospora Crassa.” *bioRxiv*. https://doi.org/10.1101/2022.03.13.484164.

Phipps, Jamie, and Karine Dubrana. 2022. “DNA Repair in Space and Time: Safeguarding the Genome with the Cohesin Complex.” *Genes* 13 (2). https://doi.org/10.3390/genes13020198.

Pradillo, Mónica, Alexander Knoll, Cecilia Oliver, Javier Varas, Eduardo Corredor, Holger Puchta, and Juan L. Santos. 2015. “Involvement of the Cohesin Cofactor PDS5 (SPO76) During Meiosis and DNA Repair in Arabidopsis Thaliana.” *Frontiers in Plant Science* 6 (December): 1034.

Ren, Qun, Hui Yang, Matthew Rosinski, Michael N. Conrad, Michael E. Dresser, Vincent Guacci, and Zhaojie Zhang. 2005. “Mutation of the Cohesin Related Gene PDS5 Causes Cell Death with Predominant Apoptotic Features in Saccharomyces Cerevisiae during Early Meiosis.” *Mutation Research* 570 (2): 163–73.

Sasani, Thomas A., David G. Ashbrook, Annabel C. Beichman, Lu Lu, Abraham A. Palmer, Robert W. Williams, Jonathan K. Pritchard, and Kelley Harris. 2022. “A Natural Mutator Allele Shapes Mutation Spectrum Variation in Mice.” *Nature* 605 (7910): 497–502.

Schep, Ruben, Eva K. Brinkman, Christ Leemans, Xabier Vergara, Robin H. van der Weide, Ben Morris, Tom van Schaik, et al. 2021. “Impact of Chromatin Context on Cas9-Induced DNA Double-Strand Break Repair Pathway Balance.” *Molecular Cell* 81 (10): 2216–30.e10.

Schuster-Böckler, Benjamin, and Ben Lehner. 2012. “Chromatin Organization Is a Major Influence on Regional Mutation Rates in Human Cancer Cells.” *Nature* 488 (7412): 504–7.

Sun, Zhongxing, Yanjun Zhang, Junqi Jia, Yuan Fang, Yin Tang, Hongfei Wu, and Dong Fang. 2020. “H3K36me3, Message from Chromatin to DNA Damage Repair.” *Cell & Bioscience* 10 (January): 9.

Supek, Fran, and Ben Lehner. 2015. “Differential DNA Mismatch Repair Underlies Mutation Rate Variation across the Human Genome.” *Nature* 521 (7550): 81–84.

———. 2017. “Clustered Mutation Signatures Reveal That Error-Prone DNA Repair Targets Mutations to Active Genes.” *Cell* 170 (3): 534–47.e23.

———. 2019. “Scales and Mechanisms of Somatic Mutation Rate Variation across the Human Genome.” *DNA Repair* 81 (September): 102647.

Wang, Wensheng, Ramil Mauleon, Zhiqiang Hu, Dmytro Chebotarov, Shuaishuai Tai, Zhichao Wu, Min Li, et al. 2018. “Genomic Variation in 3,010 Diverse Accessions of Asian Cultivated Rice.” *Nature* 557 (7703): 43–49.

Weiss, Trevor, Peter A. Crisp, Krishan M. Rai, Meredith Song, Nathan M. Springer, and Feng Zhang. 2022. “Drastic Differential CRISPR-Cas9 Induced Mutagenesis Influenced by DNA Methylation and Chromatin Features.” *bioRxiv*. https://doi.org/10.1101/2022.02.28.482333.

Weng, Mao-Lun, Claude Becker, Julia Hildebrandt, Manuela Neumann, Matthew T. Rutter, Ruth G. Shaw, Detlef Weigel, and Charles B. Fenster. 2019. “Fine-Grained Analysis of Spontaneous Mutation Spectrum and Frequency in Arabidopsis Thaliana.” *Genetics* 211 (2): 703–14.

Wyant, Skylar R., M. Fernanda Rodriguez, Corey K. Carter, Wayne A. Parrott, Scott A. Jackson, Robert M. Stupar, and Peter L. Morrell. 2022. “Fast Neutron Mutagenesis in Soybean Enriches for Small Indels and Creates Frameshift Mutations.” *G3* 12 (2). https://doi.org/10.1093/g3journal/jkab431.

Xie, Liang, Minghao Liu, Lun Zhao, Kai Cao, Peng Wang, Wenhao Xu, Wing-Kin Sung, Xingwang Li, and Guoliang Li. 2021. “RiceENCODE: A Comprehensive Epigenomic Database as a Rice Encyclopedia of DNA Elements.” *Molecular Plant* 14 (10): 1604–6.

Yang, Xiaoxu, Martin W. Breuss, Xin Xu, Danny Antaki, Kiely N. James, Valentina Stanley, Laurel L. Ball, et al. 2021. “Developmental and Temporal Characteristics of Clonal Sperm Mosaicism.” *Cell* 184 (18): 4772–83.e15.

Yan, Wei, Xing Wang Deng, Chengwei Yang, and Xiaoyan Tang. 2021. “The Genome-Wide EMS Mutagenesis Bias Correlates With Sequence Context and Chromatin Structure in Rice.” *Frontiers in Plant Science* 12 (March): 579675.

Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum, et al. 2008. “Model-Based Analysis of ChIP-Seq (MACS).” *Genome Biology* 9 (9): R137.

Zhu, Xiaohong, Shaojun Xie, Kai Tang, Rajwant K. Kalia, Na Liu, Jinbiao Ma, Ray A. Bressan, and Jian-Kang Zhu. 2021. “Non-CG DNA Methylation-Deficiency Mutations Enhance Mutagenesis Rates during Salt Adaptation in Cultured Arabidopsis Cells.” *Stress Biology* 1 (1): 12.

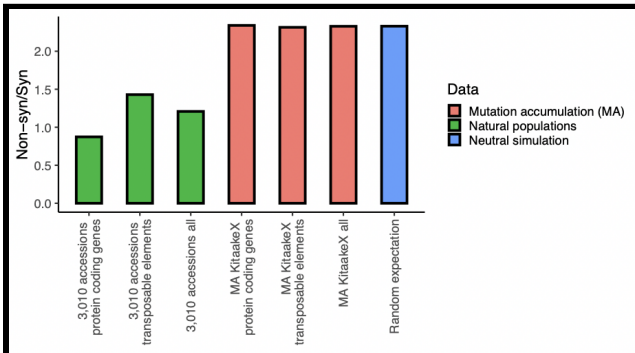

Figure S1. Non-synonymous to synonymous ratios. The non-synonymous to synonymous ratio was significantly lower ($p<2.x10^{-16}$) in natural populations compared to mutation accumulation lines and neutral expectation. The non-synonymous to synonymous ratio was similar in mutation accumulation lines to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