

# Dual activities of an X-family DNA polymerase regulate CRISPR-induced insertional mutagenesis across species

Received: 10 January 2024

Accepted: 18 July 2024

Published online: 26 July 2024



Trevor Weiss<sup>1,2,3,4,7</sup>, Jitesh Kumar<sup>1,2,3,4,7</sup>, Chuan Chen<sup>5</sup>, Shengsong Guo<sup>1,2</sup>, Oliver Schlegel<sup>6</sup>, John Lutterman<sup>6</sup>, Kun Ling<sup>5</sup> & Feng Zhang<sup>1,2,3,4</sup>✉

The canonical non-homologous end joining (c-NHEJ) repair pathway, generally viewed as stochastic, has recently been shown to produce predictable outcomes in CRISPR-Cas9 mutagenesis. This predictability, mainly in 1-bp insertions and small deletions, has led to the development of in-silico prediction programs for various animal species. However, the predictability of CRISPR-induced mutation profiles across species remained elusive. Comparing CRISPR-Cas9 repair outcomes between human and plant species reveals significant differences in 1-bp insertion profiles. The high predictability observed in human cells links to the template-dependent activity of human Polλ. Yet plant Polλ exhibits dual activities, generating 1-bp insertions through both templated and non-templated manners. Polλ knockout in plants leads to deletion-only mutations, while its overexpression enhances 1-bp insertion rates. Two conserved motifs are identified to modulate plant Polλ's dual activities. These findings unveil the mechanism behind species-specific CRISPR-Cas9-induced insertion profiles and offer strategies for predictable, precise genome editing through c-NHEJ.

Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas-mediated genome editing, a potent tool for introducing intended genetic modifications in both fundamental and translational research, operates through the sequential steps of searching, binding, and cleaving the targeted sequences, guided by programmable CRISPR RNA<sup>1–3</sup>. Subsequent repair of double-stranded DNA cleaved breaks can be accomplished via either end-joining or homology-directed pathways. Generally, the end joining pathways, including canonical or alternative non-homologous end joining (c-NHEJ or a-NHEJ), are considered more efficient but imprecise, resulting in stochastic insertion or deletion (indel) events. In contrast, the homology-directed repair (HDR) pathway is commonly employed to achieve precise editing

outcomes, such as base substitutions or precise insertions/deletions, albeit with much less efficiency<sup>4,5</sup>. Despite substantial efforts to enhance its efficiency, achieving precise genome editing through the HDR pathway remains challenging in many species<sup>6</sup>.

An alternative strategy to achieve predictable and precise genome editing has been explored by harnessing the end-joining pathways. Although generally considered imprecise, the end-joining pathways have been utilized to facilitate the knock-in of DNA templates or deletion of intervening sequences between two cleavage sites<sup>7–10</sup>. Furthermore, recent studies indicated that the indel outcomes induced by the CRISPR-*SpCas9* (i.e. *Streptococcus pyogenes* Cas9; hereafter referred to as Cas9) system exhibited predictable

<sup>1</sup>Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN 55108, USA. <sup>2</sup>Center for Precision Plant Genomics, University of Minnesota, Saint Paul, MN 55108, USA. <sup>3</sup>Microbial and Plant Genomics Institute, University of Minnesota, Minneapolis, MN 55108, USA. <sup>4</sup>Center for Genome Engineering, University of Minnesota, Minneapolis, MN 55108, USA. <sup>5</sup>Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN 55905, USA. <sup>6</sup>Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN 55455, USA. <sup>7</sup>These authors contributed equally: Trevor Weiss, Jitesh Kumar. ✉ e-mail: [zhangumn@umn.edu](mailto:zhangumn@umn.edu)

patterns<sup>8,11,12</sup>. By targeting a large set of CRISPR-Cas9 sites in human and mouse cells, it was observed that up to 47% of the sites produced a single indel type<sup>12</sup>. These findings led to the development of in-silico programs, such as FORECasT and InDelphi, with remarkable accuracy in predicting the end joining-mediated indel mutations in various species, including yeast, zebrafish, *Xenopus*, mouse, and human<sup>11–14</sup>.

Among the predominant indel mutations induced by CRISPR-Cas9, 1-bp insertions were identified as the most commonly occurring events<sup>11,12</sup>. The templated insertion model has been proposed to explain the 1-bp insertions observed in yeast and animal species<sup>13</sup>. In this model, Cas9 generates a 1-nt 5' overhang by cleaving the –3rd position on the CRISPR RNA complementary strand and the –4th position on the non-complementary strand upstream of the Protospacer Adjacent Motif (PAM) sequence. Subsequently, the overhangs can be filled in by an X-family DNA polymerase, DNA Pol 4 or  $\lambda$ , and ligated through the c-NHEJ pathway resulting in the duplication of the nucleotide at the –4th position<sup>13,15</sup>.

Recent efforts to extend CRISPR-Cas9 genome editing into plants have confirmed the prevalence of 1-bp insertions in various plant species<sup>9,16–18</sup>. While these findings suggested a potential extension of predictability observed in animals to plants, recent studies have also indicated that 1-bp insertion profiles may not be consistently predictable as proposed by the templated insertion model<sup>9,16,17</sup>. The existence of distinctive 1-bp insertion profiles and the underlying mechanism remains unclear. Addressing these questions is crucial for refining predictability and enhancing the end-joining mediated precise genome editing across species.

In this study, we systematically compare the CRISPR-Cas9-induced 1-bp insertion profiles among two plant species, *Arabidopsis thaliana*, representing a dicot model plant, and *Setaria viridis*, representing a monocot model plant, and human cells. Our observations reveal the presence of distinctive 1-bp insertion profiles induced by CRISPR-Cas9 in plants. To delve deeper into the underlying mechanism, we identify the X-family DNA Polymerase, Pol $\lambda$ , responsible for plant-specific 1-bp insertion profiles and demonstrate that the efficiency and profiles of 1-bp insertions can be modulated by dual activities of Pol $\lambda$ . Finally, we propose an updated model to account for distinctive CRISPR-induced 1-bp insertions. Overall, our findings provide a thorough evaluation of the predictability of CRISPR-Cas9 mutagenesis and shed light on the intricate interplay between CRISPR-Cas9 and DNA repair pathways across species. This research underscores that the prediction tools for CRISPR mutagenesis need to be tailored and optimized for each individual species, accounting for the inherent differences in the DNA repair machinery. Our work will also pave the way for the development of more precise and predictable genome editing strategies using the c-NHEJ pathway.

## Results

### Low predictive power of CRISPR-Cas9 mutagenesis prediction programs for plants

To assess the predictability of the CRISPR-Cas9-induced mutations in plants, we examined the performance of two widely used CRISPR mutagenesis prediction programs, FORECasT and InDelphi<sup>11,12</sup>. We generated CRISPR-induced mutations at 59 sites, including 26 from *Arabidopsis* and 33 from *Setaria*, by introducing the corresponding CRISPR-Cas9 constructs into each species (Supplementary Data 1). In *Arabidopsis*, each CRISPR-Cas9 construct was introduced using the floral dip-based stable transgenic approach. Individual seedlings of each T1 transgenic plant were collected for the CRISPR mutation analysis at each target site. In *Setaria*, individual CRISPR-Cas9 constructs were transformed via transient protoplasts transfection. Transformed protoplast cells were collected after 48 h for the mutation assay. Subsequently, the mutation profile of each site was obtained by using the next-generation sequencing (NGS) based assay. The indel mutagenesis rates averaged at 8.9% and 28.4% at the

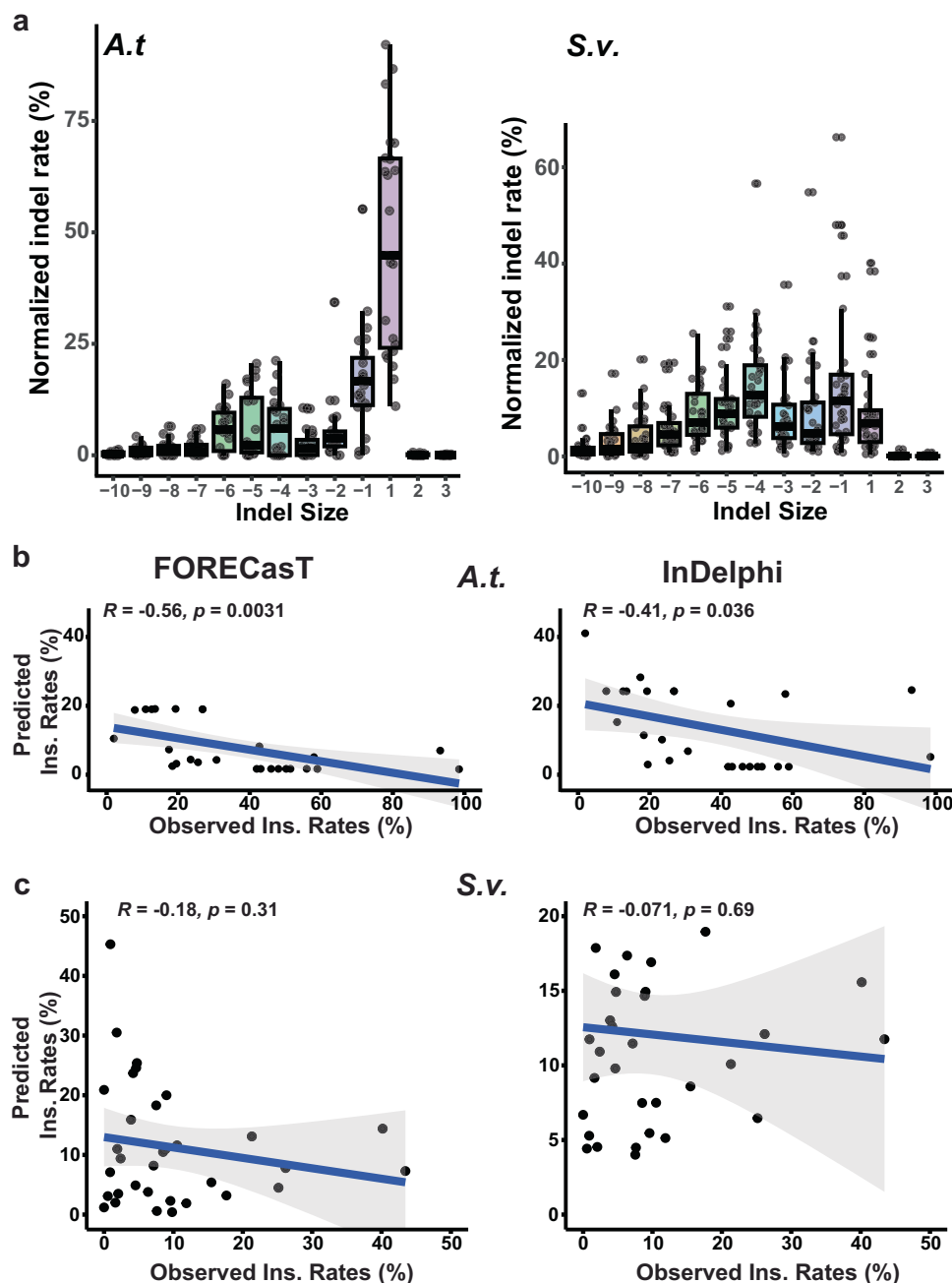
sites from *Arabidopsis* and *Setaria*, respectively (Supplementary Fig. 1a). Additionally, the indel profile from each site was further characterized into individual insertion and deletion types for each species (Fig. 1a). Notably, the 1-bp insertions represent one of the most common occurring mutation types, as previously observed in human cell lines. In *Arabidopsis*, 1-bp insertions were the most prevalent mutation types, accounting for an average of 44.6% of all mutations across 26 CRISPR sites (Fig. 1a). For *Setaria viridis*, the average 1-bp insertion rate appeared to be the 4th highest at 9.6% across 33 CRISPR sites (Fig. 1a).

Simultaneously, the predicted mutation profile was generated for each target site using FORECasT and InDelphi. In this study, we chose to focus on the insertion rates for correlation analyses on the predicted versus observed values for the following reasons: (1) CRISPR-induced insertions appeared to exhibit less stochastic patterns than deletions; and (2) previous studies have suggested that these prediction tools demonstrate greater predictive power for insertions compared to other indel types<sup>11,12,14</sup>. As a result, we observed no positive correlations using either FORECasT or InDelphi for both *Arabidopsis* and *Setaria* datasets (Fig. 1b, c). Weak negative correlations were observed in the *Arabidopsis* dataset ( $r = -0.56$ ,  $p < 0.0031$  and  $r = -0.4$ ,  $p < 0.036$ ; Fig. 1b), while no correlation was found in the *Setaria* dataset ( $r = -0.18$ ,  $p < 0.31$  and  $r = -0.07$ ,  $p < 0.69$ ; Fig. 1c). Thus, our data suggested that both prediction programs developed with human datasets exhibited low predictive power for the CRISPR-Cas9-induced mutation profile in plants.

### Distinctive templated versus non-templated 1-bp insertion patterns across species

The limited predictive power from the human cell-based indel prediction tools prompted us to further examine CRISPR-Cas9-induced insertion profiles in plants. Both FORECasT and InDelphi predicted CRISPR-induced insertions primarily as 1-bp insertion events occurred at the –4th position upstream of the PAM; and most of these insertions were derived from templated insertions by duplicating the –4th nucleotide. When we analyzed the observed insertions from the *Arabidopsis* and *Setaria* target sites, 1-bp insertions were consistently predominant, accounting for averagely 95.9% of insertions across all sites (Supplementary Fig. 1b). However, when the 1-bp insertion patterns were plotted according to the –4th nucleotide, the observed insertions did not consistently exhibit characteristics of templated insertions in plants (Fig. 2a). When the –4th nucleotide was T, the inserted nucleotide appeared to follow the templated insertion model with 78.8% and 75.7% of insertions as T in *Arabidopsis* and *Setaria*, respectively (Fig. 2a). With the –4th nucleotide as A, while A remained the predominant inserted nucleotide (58.5% and 58.7% in *Arabidopsis* and *Setaria*), the fractions of other types of insertions, termed as non-templated insertions, increased substantially (Fig. 2a). In cases where the –4th nucleotide was either C or G, non-templated insertions became predominant by increasing to 61.4% and 66.0% for the –4th nucleotide C, and 98.4% and 99.5% for the –4th nucleotide G in *Arabidopsis* and *Setaria*, respectively.

To compare with 1-bp insertion patterns in human cells, we analyzed the insertion profiles of 150 target sites previously reported from the human cell lines<sup>13</sup>. The results were largely consistent with the templated insertion model, showing the 1-bp insertion pattern with the –4th nucleotide duplications, while low levels of non-templated 1-bp insertions were observed at the target sites with –4th nucleotide as C or G (Fig. 2a). Taken together, our observations revealed distinct 1-bp insertion patterns between plants and human cells. The 1-bp insertion profiles from plant species exhibited a higher incidence of non-templated insertions, deviating from the templated insertion model. Notably, the rates of non-templated insertions appeared to vary depending on the –4th nucleotide, increasing in the order of T, A, C, and G.



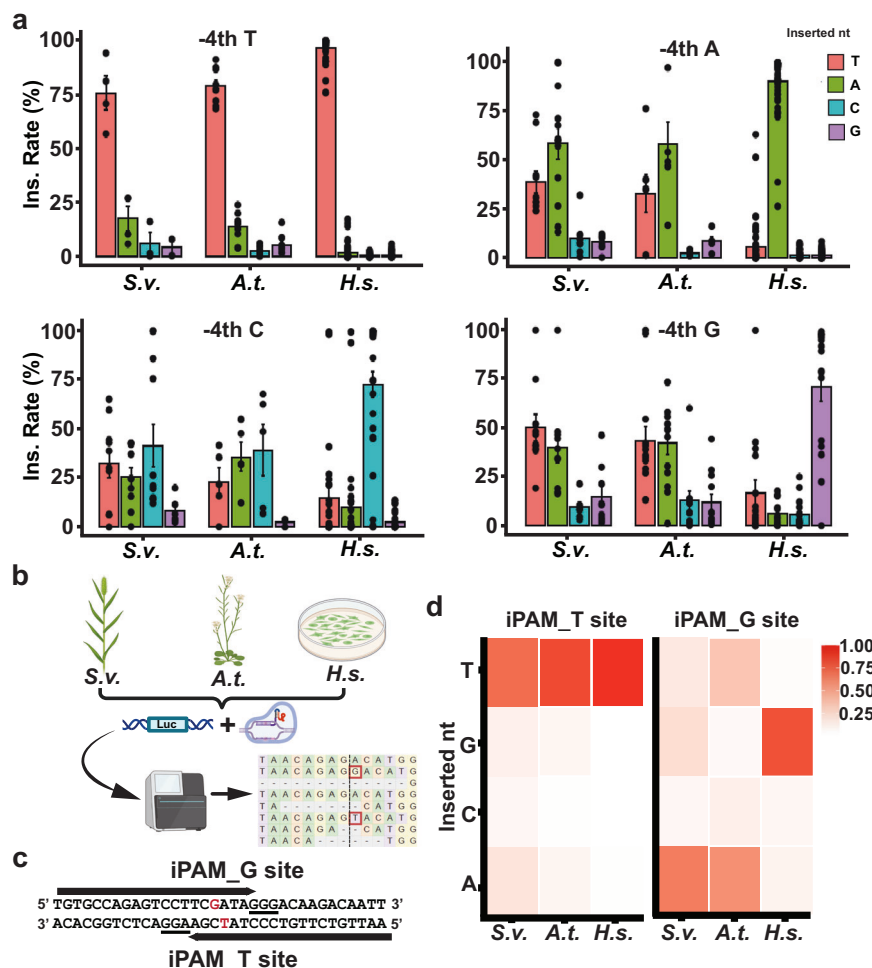
**Fig. 1 | Predictability of FORECasT and InDelphi tools for CRISPR-Cas9 induced insertions in plants.** **a** CRISPR-Cas9 induced mutation profiles across 59 target sites in *Arabidopsis* ( $n = 26$ ) and *Setaria* ( $n = 33$ ). X-axis represents individual indel sizes. The normalized mutation rates (Y-axis) were determined by dividing the number of reads containing mutations within each indel size by the total number of reads containing all types of mutations. The horizontal bars within boxes represent medians. The top and bottom edges of the boxes represent the 75th and 25th

percentiles, respectively. The upper and lower whiskers extend to data no more than 1.5x the interquartile range from the upper and lower edges of the box, respectively. **b, c** Two-sided Pearson correlation analysis were performed using scatter plots to compare predicted versus experimentally observed insertion (ins.) rates for each CRISPR gRNA in *Arabidopsis* ( $n = 26$ ) and *Setaria* ( $n = 33$ ). The 95% confidence interval (CI) were indicated with gray color. The source data are provided in the Source Data file.

### Plant-specific 1-bp insertion profiles dependent on the –4th nucleotide context

To further explore the distinctive 1-bp insertion profiles across species, we conducted direct comparisons by targeting identical CRISPR sites in *Arabidopsis*, *Setaria*, and human cell line, HEK293. This involved initially integrating the firefly luciferase gene and subsequently expressing the CRISPR-Cas9 expression cassette in the genomes of these three species (Fig. 2b). We designed two CRISPR guide RNAs (gRNAs) to target overlapping sites located on opposite strands, referred to as inverted PAM (iPAM) targets, as described in previous research<sup>13</sup> (Fig. 2c). These

two gRNAs, with one –4th position as T (iPAM\_T) and the other as G (iPAM\_G), represented the sequence contexts for the highest and lowest templated insertion rates observed in plants (Fig. 2c). In *Arabidopsis*, CRISPR-Cas9 constructs were assembled with the firefly luciferase reporter gene in T-DNA. The resulting constructs were transformed using the floral dip-based stable transgenic approach. Three seedlings from each T1 transgenic group were collected for CRISPR mutation analysis at each target site. For *Setaria viridis*, a homozygous *Setaria* line with the firefly luciferase reporter gene integrated into the genome was obtained from previous research<sup>19</sup>. Individual CRISPR-Cas9 constructs



**Fig. 2 | Cross-species analyses of templated versus non-templated 1-bp insertions according to the -4th nucleotide.** **a** Cross-species 1-bp insertion patterns to the -4th nucleotide. The 1-bp insertions were divided into 4 groups according to the inserted nucleotide for each CRISPR gRNA. The normalized 1-bp insertion (ins.) rates were calculated by dividing the number of reads containing each type of 1-bp insertions by the number of reads with all types of 1-bp insertions and were plotted to the -4th nucleotides (T, A, C, and G) for *Setaria viridis* (*S.v.*;  $n = 33$  biologically independent samples), *Arabidopsis thaliana* (*A.t.*;  $n = 26$  biologically independent samples), and the human cell line (*H.s.*;  $n = 150$  biologically independent samples).

Data are presented as mean values  $\pm$  SEM. **b** The schematic workflow to compare 1-bp insertion patterns across *S.v.*, *A.t.*, and *H.s.* line through the next-generation sequencing assay. **c** The CRISPR targeted sequences of iPAM\_T and G. The -4th nucleotide was highlighted in red with the PAM sequence underlined. **d** Heatmap analyses of the proportion of each inserted nucleotide type (nt) at the -4th position of the iPAM\_T and G sites across *S.v.* ( $n = 3$ ), *A.t.* ( $n = 3$ ), and *H.s.* ( $n = 3$ ). The source data are provided in the Source Data file. **b** Created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

were then transformed into protoplast cells isolated from the luciferase gene-containing plants. Transformed protoplasts were collected after 48 h for the mutation assay with 3 replications for each target site. When insertion rates were examined, both CRISPR gRNAs induced substantial 1-bp insertions ranging from 33.8% to 89.0% for the iPAM\_T site and from 33.4% to 74.4% for the iPAM\_G site in three species (Supplementary Fig. 2).

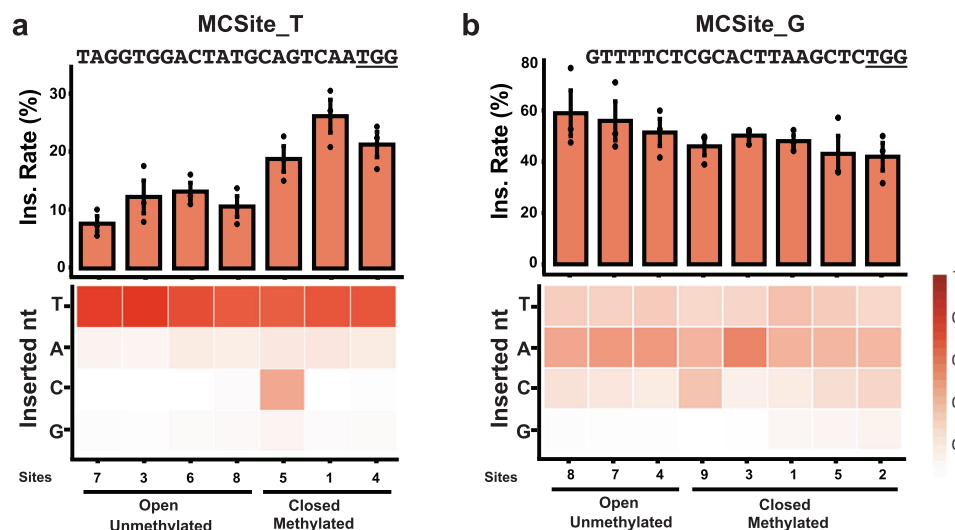
Next, we analyzed templated versus non-templated insertion patterns at each target site. In the HEK293 cells, consistent with the templated insertion model, templated insertions were predominantly presented at both target sites with rates of 97.0% and 84.8%, respectively (Fig. 2d). However, in *Arabidopsis* and *Setaria*, predominant templated insertions were primarily observed at the iPAM\_T site, ranging from 73.3% to 87%. At the iPAM\_G site, non-templated insertions were predominant, accounting for 72.6% to 95.8% of 1-bp insertions in both plant species (Fig. 2d). Taken together, these findings corroborated the observations from 59 individual target sites, revealing distinct plant-specific 1-bp insertion profiles. These profiles exhibited either templated or non-templated dominant patterns associated by the -4th nucleotide upstream of PAM.

### Limited influence of chromatin states on 1-bp insertion profiles in plants

As indicated by prior studies, both epigenetic and genetic factors could influence CRISPR-Cas9 induced mutation profiles<sup>15,17,20</sup>. To explore the mechanism underlying the distinctive 1-bp insertion profiles in plants, we investigated the impact of the chromatin states on these insertions. We used the multi-copy CRISPR target site (MCSite) system previously developed in *Arabidopsis*<sup>17</sup>. Two sets of MCSites, designated as MCSite\_T and MCSite\_G based on their -4th nucleotide, are located in diverse epigenetic contexts as described previously<sup>17</sup>. Individual sites within each MCSite family can be categorized into two major groups as either open and unmethylated or closed and methylated chromatin (Fig. 3a, b).

When the 1-bp insertion rates of individual MCSites were examined, variations were found across different chromatin states as previously indicated<sup>17</sup>. For MCSite\_T sites, insertion rates ranged from 7.9% to 26.8%, and for MCSite\_G sites, they ranged from 41.9% to 58.9% (Fig. 3a, b). In contrast, heatmap analysis of the 1-bp insertion profiles revealed a consistent pattern within each MCSite family. Specifically, for MCSite\_T sites, templated insertions were predominantly observed





**Fig. 3 | Effect of chromatin states on the 1-bp insertion rate and profile in *Arabidopsis*.** Normalized 1-bp insertion rates were plotted for individual MCSite\_T (a) and MCSite\_G (b) sites (X-axis). The 20-bp targeted sequences with 3-bp underlined PAM sequences were shown with each plot. The normalized 1-bp insertion (ins.) rates were determined by dividing the number of reads containing 1-bp insertions by the total number of reads containing all types of indel mutations.

Data are presented as mean values  $\pm$  SEM from three independent plants. Heatmaps in the lower panel illustrated the proportion of each inserted nucleotide type (T, A, C, G) at the -4th position of individual MCSite\_T (a) and MCSite\_G (b) sites. Chromatin states of individual sites were categorized into Open and Unmethylated or Closed and Methylated groups. The source data are provided in the Source Data file.

across individual sites, regardless of their chromatin states (Fig. 3a). On the other hand, all individual sites within the MCSite\_G family exhibited a predominant 1-bp non-templated insertion pattern across different epigenetic contexts, ranging from 94.0% to 99.8% (Fig. 3b). These results suggested that chromatin states may have limited impacts on CRISPR-Cas9 induced 1-bp insertion profile.

### Pol $\lambda$ homolog responsible for both templated and non-templated 1-bp insertions in plants

We then investigated the genetic factors contributing to the distinct 1-bp insertion profiles in plants. Previous studies have pointed to the X-family DNA polymerase, Pol $\lambda$ , and its homolog as pivotal players in mediating 1-bp templated insertions in human and yeast cells<sup>13,15</sup>. A single copy of the Pol $\lambda$  homolog was identified in both *Arabidopsis* and *Setaria* genomes through sequence homology searches<sup>21</sup>. No other X-family DNA polymerases were found in plants from the homology search. Phylogenetic analyses confirmed that this plant X-family DNA polymerase exhibited a close evolutionary relationship with Pol $\lambda$  as opposed to other members, such as DNA Pol  $\mu$  and Terminal deoxynucleotidyl Transferase (TdT) (Supplementary Fig. 3 and Supplementary Data 2).

To explore the involvement of the plant Pol $\lambda$  homolog in CRISPR-Cas9 induced 1-bp insertions, we obtained an *Arabidopsis* T-DNA knock-out mutant line (*atpol1-1*), previously characterized with no notable growth or physiological defects<sup>22,23</sup>. Using the wild type and the homozygous *atpol1-1* mutant *Arabidopsis* plants, we generated stable transgenic plants with the CRISPR-Cas9 T-DNA construct to target three distinct sites: the single-copy site in the *Arabidopsis Chelatease 12* gene (*AtCHL12*), as well as the MCSite\_T and MCSite\_G sites. Three T1 CRISPR-Cas9 transgenic plants from each genotype were used to survey CRISPR-induced mutations for each target site. The single-copy *CHL12* site would allow for a rapid assessment of the involvement of Pol $\lambda$  in 1-bp insertions, while the two MCSites provided additional insights in different epigenetic contexts.

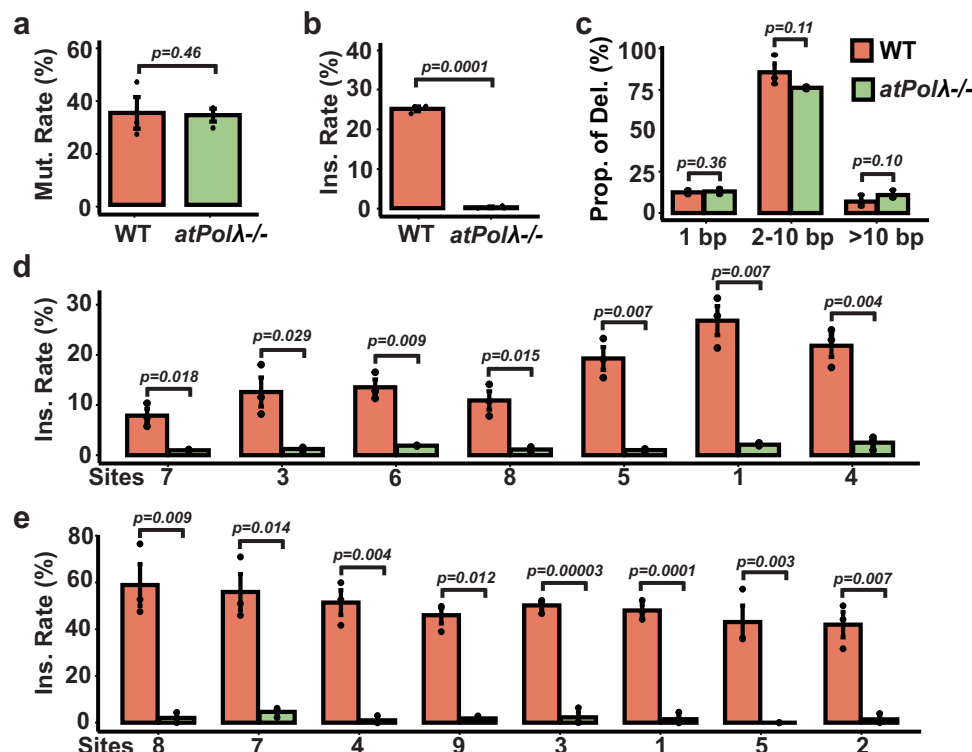
When we examined CRISPR-Cas9 mutagenesis at the *CHL12* site, both wild-type and mutant CRISPR-Cas9 plants displayed comparable overall mutagenesis rates, averaging 38.9% and 37.9%, respectively (Fig. 4a). In wild-type plants, approximately 25.3% of indel mutations were identified as 1-bp insertions at the -4th position, with non-templated insertions being predominant at a rate of 65.2%, attributable

to the G nucleotide at the -4th position in the *CHL12* site (Fig. 4b; Supplementary Fig. 4a). In contrast, in Pol $\lambda$  mutant plants, the 1-bp insertion rates, encompassing both non-templated and templated insertions, were reduced to undetectable levels (0.2%; Fig. 4b; Supplementary Fig. 4b). Additionally, we explored the potential involvement of this Pol $\lambda$  homolog in CRISPR-Cas9 induced deletions. As a result, we observed similar levels of deletions within three different deletion groups, 1-bp, 2 to 10-bp and more than 10-bp, between the wild-type and mutant plants (Fig. 4c). Thus, the plant Pol $\lambda$  homolog appeared to be the pivotal gene for CRISPR-Cas9 induced 1-bp insertions, operating in both templated and non-templated manners, with limited involvement in deletions.

Furthermore, we investigated the role of this Pol $\lambda$  homolog at additional CRISPR target sites within diverse epigenetic contexts. When examining the 1-bp insertion rates at the MCSite\_T and G sites, we observed significant reductions of 1-bp insertions, both templated and non-templated, across all sites, irrespective of their chromatin states. In the MCSite\_T sites, the 1-bp insertion rates decreased from an average of 19.5% in wild-type plants to 1.6% in the mutant plants, while in the MCSite\_G sites, the rates were reduced from an average of 49.4% to 1.8% (Fig. 4d, e). These results substantiated that the plant Pol $\lambda$  homolog is responsible for both templated and non-templated 1-bp insertions regardless of chromatin states.

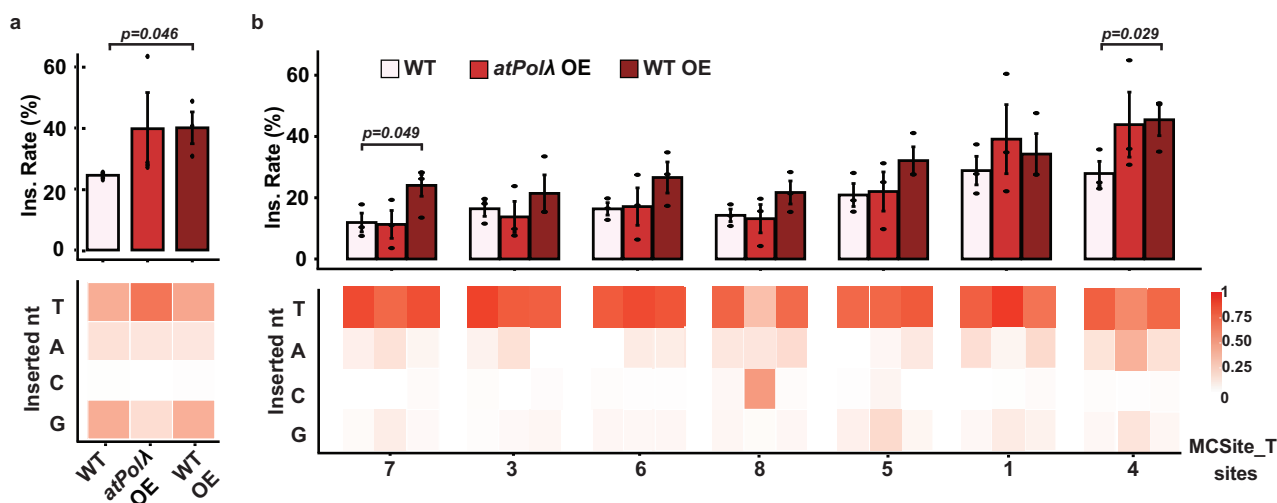
### Overexpression of AtPol $\lambda$ restores or enhances templated and non-templated 1-bp insertions

Next, we hypothesized that overexpression of *Atpol1* could restore or even enhance the 1-bp insertion rates. To test this hypothesis, we generated stable transgenic plants by overexpressing the *AtPol $\lambda$*  gene in the *atpol1-1* mutant plants. The *AtPol $\lambda$*  coding sequence was driven under the constitutive *Arabidopsis Ubiquitin-10* promoter and cloned into the final construct with a CRISPR-Cas9 expression cassette to target the *CHL12* and MCSite\_T sites. Three T1 CRISPR-Cas9 transgenic plants with the *atpol1-1* mutant genotype were used to survey CRISPR-induced mutations for each target site. When 1-bp insertions were examined at the *CHL12* site, the *AtPol $\lambda$*  overexpression plants exhibited a 1.6-fold increase compared to wild-type plants, with an average rate of 39.8% (Fig. 5a). The 1-bp insertion profiles appeared similar between the *AtPol $\lambda$*  overexpression and the wild-type plants, with non-



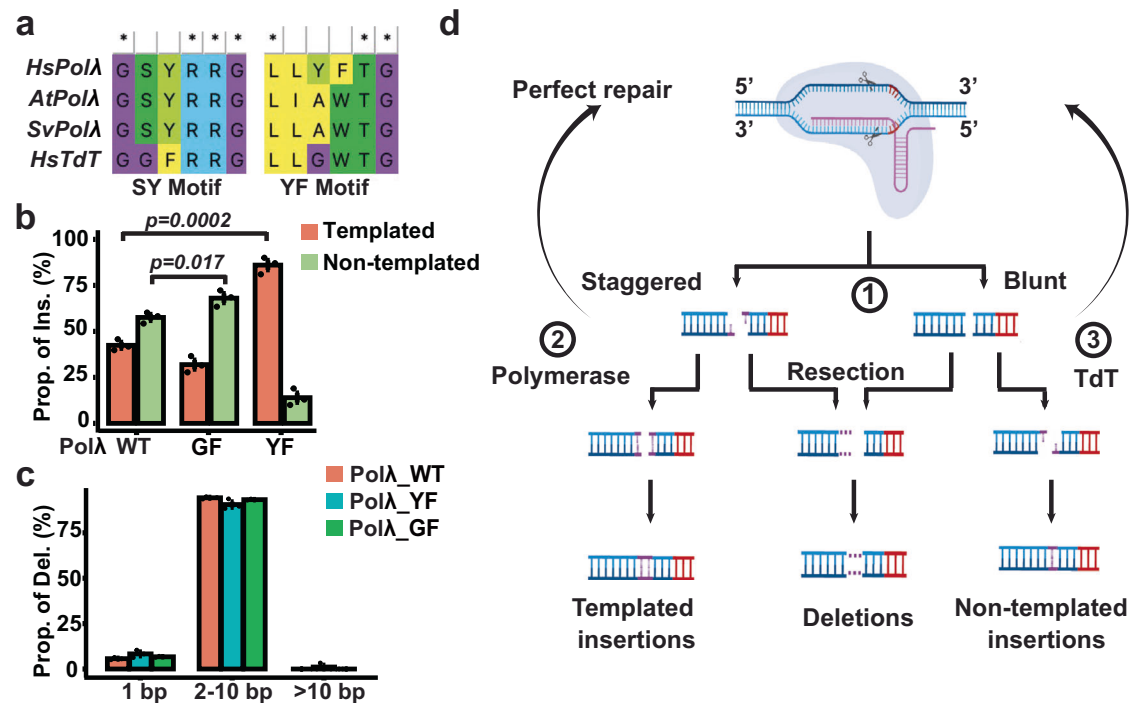
**Fig. 4 | Polλ is responsible for 1-bp insertions in *Arabidopsis*.** **a** CRISPR-Cas9 mutation (mut.) rates between the wild type and *atpolλ*-1 mutant plants. The mutation rates (Y-axis) were determined by dividing the number of reads containing indel mutations by the total number of NGS reads. **b** Normalized 1-bp insertion rates between the wild type and *atpolλ*-1 mutant plants at the CHLI2 site. The normalized 1-bp insertion (ins.) rates were determined by dividing the number of reads containing 1-bp insertions by the total number of reads containing all types of indel mutations. **c** Normalized deletion rates between the wild type (WT) and *atpolλ*-1 mutant plants at the CHLI2 site. The normalized proportion of deletion

(Prop. of Del. as Y-axis) were determined by dividing the number of reads containing deletions within each category (1-bp, 2-10 bp, or >10 bp) by the total number of reads containing all types of deletions. **d, e** Normalized 1-bp insertion (ins.) rates between the wild type and *atpolλ*-1 mutant plants at the MCSite\_T (**d**) and MCSite\_G (**e**) sites. Heatmaps under the bar plots illustrate the proportion of each inserted nucleotide type (T, A, C, G) at the -4th position of individual MCSite\_T (**d**) and MCSite\_G (**e**) sites. Data are presented as mean values ± SEM from 3 independent plants. *P*-values were derived from unpaired one-tailed Student's *t* test. The source data are provided in the Source Data file.



**Fig. 5 | Overexpression of Polλ restores or enhances 1-bp insertions in *Arabidopsis*.** Normalized 1-bp insertion rates at CHLI2 (**a**) and MCSite\_T (**b**) among three lines: wild-type plants (WT), Polλ overexpression plants in the *atpolλ*-1 mutant (*atpolλ* OE), and Polλ overexpression plants in the wild-type backgrounds (WT OE). The normalized 1-bp insertion rates (Y-axis) were determined by dividing the number of reads containing 1-bp insertions by the total number of

reads containing all types of indel mutations. Heatmaps under each plot illustrated the proportion of each inserted nucleotide type (T, A, C, G) at the -4th position. Data are presented as mean values ± SEM from 3 independent plants. *P*-values were derived from unpaired one-tailed Student's *t* test. The source data are provided in the Source Data file.



**Fig. 6 | Dual activities of AtPolλ are modulated by two conserved motifs.** **a** Sequence alignment of two conserved motifs, SY and YF, across Human Polλ, AtPolλ, SvPolλ, and human TdT. **b** Comparisons of templated versus non-templated insertion rates between the wild type AtPolλ and two variants, Polλ<sup>S366G/Y367F</sup> and Polλ<sup>A459Y/W460F</sup> at the CHLI2 site. The templated (indicated by orange) or non-templated insertion (indicated by green) rates (Y-axis) were determined by dividing the number of reads containing each type of 1-bp insertions by the total number of reads containing 1-bp insertions in each sample.  $P$ -values were derived from unpaired one-tailed Student's  $t$  test. The source data are provided in the Source Data file. **d** The proposed

model for the dual activities of Polλ in generating templated and non-templated 1-bp insertions. **Step 1:** CRISPR-Cas9 generates a blunt or staggered cut at the targeted site. Blunt-ended cleavages occur at the -3rd position upstream of the PAM (indicated by the red vertical lines) on both strands, while staggered cleavages take place with one cut at the -4th position on the non-targeted strand and the other cut at the -3rd position on the targeted strand, producing 5' 1-nt overhangs. **Step 2:** The staggered product can be filled in by Polλ with template-dependent activity. **Step 3:** The blunt-ended product can be processed by Polλ with template-independent activity to extend 1-nt at the 3' end of each strand. After ligation and correction by c-NHEJ and mismatch repair, non-templated 1-bp insertions occur at the -4th position. Additionally, cleavage products could be processed through either perfect ligation, indicated by the curved arrowheads, or through resection to generate deletion, indicated by the purple dash lines.

templated insertions still being predominant at an average rate of 74.8% (Fig. 5a). When examining the 1-bp insertions at the MCSite\_T sites, overexpression of the AtPolλ transgene in the mutant plant appeared to restore 1-bp insertion rates to the levels observed in wild-type plants at five of seven MCSite\_T sites. At the other 2 sites, sites 1 and 4, the 1-bp insertion rates exhibited substantial increases by 1.4 to 1.6 folds, respectively (Fig. 5b). When comparing the 1-bp insertion profiles, similar insertion patterns were observed between the overexpression and wild-type plants with predominant templated insertions across nearly all the sites except for one site, site 8 (Fig. 5b). These results confirmed that overexpression of AtPolλ could restore or may enhance CRISPR-Cas9 induced templated and non-templated 1-bp insertions in the knockout mutant plants, further validating its pivotal role in generating 1-bp insertions.

We further hypothesized that overexpression of this gene should have the potential to enhance 1-bp insertions in wild-type plants. To test this idea, we introduced the same overexpression construct to wild-type plants. Three T1 CRISPR-Cas9 transgenic plants with the wild-type background were used to survey CRISPR-induced mutations for each target site. At the CHLI2 site, we observed a similar increase in the 1-bp insertion rate between the overexpression wild-type plants and the overexpression mutant plants compared to the wild-type control plants (Fig. 5a). At the MCSite\_T sites, when comparing the 1-bp insertion rates between the overexpression wild-type plants and the wild-type control plants, we observed substantial increases in all seven

sites by 1.2 to 2.0 folds (Fig. 5b). When comparing the 1-bp insertion profiles, similar insertion patterns were observed between the overexpression wild-type plants, the overexpression mutant plants, and the wild-type control across all the sites, irrespective of their epigenetic states (Fig. 5b). Taken together, these observations corroborated that overexpressing the Polλ homolog in wild-type plants could further increase 1-bp insertions.

### Conserved motifs modulate the dual activities of AtPolλ

To gain insights into the mechanism(s) underpinning the distinct properties of Polλ across species, we conducted protein sequence analyses by aligning AtPolλ with X-family DNA Polymerases in humans (Supplementary Fig. 5). Previous studies have indicated two conserved motifs in human X-family DNA Polymerase that contribute to template dependency<sup>24</sup>. The first motif, identified as GSYRRG in template-dependent human DNA polymerases λ, features two amino acids, serine and tyrosine (SY), which are replaced by glycine and phenylalanine (GF) in the template-independent human TdT (Fig. 6a and Supplementary Fig. 5)<sup>24,25</sup>. The second motif, known as the YF motif, contains tyrosine and phenylalanine at the catalytically active sites of the DNA polymerases λ. In contrast, these two residues are changed to glycine and tryptophan (GW) in TdT (Supplementary Fig. 5)<sup>24,25</sup>. When analyzing these motifs in DNA polymerase λ homologs from *Arabidopsis*, *Setaria*, Tobacco, and rice, the first motif was identical to the sequences in human Polλ, while the second motif, characterized by

alanine and tryptophan (AW), showed a closer resemblance to the GW motif found in human TdT (Fig. 6a and Supplementary Fig. 5). Thus, the plant Pol $\lambda$  homologs appear to combine characteristic motifs from human Pol $\lambda$  and TdT.

The presence of both human Pol $\lambda$  and TdT motifs could potentially contribute to the observed dual templated-dependent and independent activities in AtPol $\lambda$ . We then hypothesized that the dual activities of AtPol $\lambda$  could be modulated by modifying each motif individually. To test this hypothesis, we generated two variants of AtPol $\lambda$  through site-directed mutagenesis on the respective motifs. The first variant, AtPol $\lambda^{\text{YF}}$ , was engineered by substituting Alanine and Tryptophan (AW) with Tyrosine and Phenylalanine (YF) at the second motif to mimic human Pol $\lambda$  (Fig. 6a). Similarly, the second variant, AtPol $\lambda^{\text{GF}}$ , was created to mimic human TdT by replacing Serine and Tyrosine (SY) with Glycine and Phenylalanine (YF) at the first motif (Fig. 6a).

The coding sequence of each AtPol $\lambda$  variant was cloned into the T-DNA vector described above, with the constitutive *Arabidopsis Ubiquitin-10* promoter and a CRISPR-Cas9 expression cassette to target the CHL12 site. We used an *agrobacterium*-mediated transient expression approach to transform individual T-DNA constructs into young seedlings of the *atpol\lambda* knock-out mutant, and then examined the CRISPR-Cas9 mutation profile at the CHL12 site using the NGS assay (Supplementary Fig. 6a). The average mutation rates from these samples are 17.3% (AtPol $\lambda^{\text{WT}}$ ), 12.7% (AtPol $\lambda^{\text{GF}}$ ) and 10.3% (AtPol $\lambda^{\text{YF}}$ ), respectively (Supplementary 6b). When analyzing templated versus non-templated 1-bp insertion patterns, the samples expressing the wild type AtPol $\lambda$  gene exhibited higher proportions of non-templated insertions compared to those of templated insertions (57.6% non-templated insertions versus 42.4% templated insertions) consistent with the observations from the stable transgenic plants (Fig. 5a, b). In contrast, the samples transformed with the AtPol $\lambda^{\text{YF}}$  variant demonstrated altered 1-bp insertion profiles with templated insertion proportions being significantly higher than those from the overexpression of the wildtype AtPol $\lambda$  by 100% (86.0% versus 42.4%; Fig. 6b and Supplementary Fig. 6c, d). Conversely, the samples transformed with the AtPol $\lambda^{\text{GF}}$  variant displayed significantly higher proportions of non-templated insertions compared to the wild-type AtPol $\lambda$  overexpression lines by 18% (67.9 % versus 57.6%; Fig. 6b and Supplementary Fig. 6c, d). Regarding the deletion profiles, no evident differences were observed within three different deletion groups, 1-bp, 2 to 10-bp and more than 10-bp, among AtPol $\lambda^{\text{WT}}$  and the two variants (Fig. 6c).

Notably, the overall 1-bp insertion rates from the samples with each variant reduced to 4.4% and 5.5% compared to 31.7% in the wild-type AtPol $\lambda$  overexpression control, suggesting the involvement of additional amino acids in regulating enzymatic activity. (Supplementary Fig. 6e). Collectively, these observations align with our hypothesis that these two conserved motifs play crucial roles in modulating the dual template-dependent and independent activities of AtPol $\lambda$ . Further investigation is required to refine the enzymatic activities of these variants.

## Discussion

In this study, we initially investigated CRISPR-Cas9-induced insertion profiles across different species and subsequently delved into the mechanism responsible for plant-specific 1-bp insertion patterns. Our findings revealed that prediction tools developed with human datasets demonstrated limited accuracy in predicting insertional mutations in plants. This discrepancy seemingly stemmed from plant unique 1-bp insertion patterns. Specifically, in human cells, templated 1-bp insertions were predominantly found at most CRISPR-Cas9 target sites, while both templated insertions and non-templated 1-bp insertions could occur at substantial levels in plants, inconsistent with the templated insertion model. Our observations highlighted the significance of the sequence context, especially -4th nucleotide, in determining

the 1-bp insertion profile. The rates of non-templated insertions exhibited substantial variations based on the type of -4th nucleotide, increasing in the order of T, A, C, to G. In some cases, however, variations could be observed even with the same -4th nucleotide, suggesting the involvement of other factors in determining 1-bp insertion patterns (Fig. 2a). Further investigation will be needed to fully address this question.

We identified the crucial gene responsible for both templated and non-templated 1-bp insertions in *Arabidopsis* as the homolog of a specialized X-family DNA polymerase, Pol $\lambda$ . Disrupting this Pol $\lambda$  homolog resulted in a nearly complete reduction of 1-bp insertions while overexpressing the gene led to the restoration or enhancement of 1-bp insertions. Previous research suggested that plant Pol $\lambda$  homologs may play a role in repairing oxidative base damage, UV-induced DNA repair or general double-strand DNA break repair, akin to their human counterpart<sup>24,26,27</sup>. However, the specific function of plant Pol $\lambda$  has remained elusive<sup>22,23,28</sup>. Our study provided genetic evidence indicating the direct involvement of plant Pol $\lambda$  in the end processing during the c-NHEJ repair. However, unlike human and yeast Pol $\lambda$  genes, which predominantly filled in gaps in a template-based manner, the *Arabidopsis* Pol $\lambda$  homolog seemed capable of processing broken ends through both templated and non-templated mechanisms.

Different levels of template dependence have been observed in X-family DNA polymerases. In human and most vertebrates, the X-family DNA polymerases consist of four members, Pol $\alpha$ ,  $\beta$ ,  $\mu$ , and Terminal deoxynucleotidyl Transferase (TdT)<sup>29</sup>. Human Pol $\lambda$  shows a high level of template dependence, while TdT appears to be entirely template-independent<sup>30</sup>. This aligns with the observation that Pol $\lambda$ -mediated 1-bp insertions are mostly template-based in human cells. However, in most plants, only one X-family DNA polymerase, Pol $\lambda$ , is present, with no other X-family paralogs in their genome. One possible explanation for increased levels of non-templated 1-bp insertions in plants could be that the plant Pol $\lambda$  may possess dual activities, catalyzing both template-dependent and -independent reactions. Interestingly, an in vitro biochemical study has suggested that the rice Pol $\lambda$  homolog exhibited both template-based polymerase and non-templated-based TdT activities<sup>28,31</sup>. Furthermore, protein sequence analysis revealed that plant Pol $\lambda$  homologs possess characteristic motifs from both human Pol $\lambda$  and TdT, which could elucidate the observed dual activities of AtPol $\lambda$ . Our in vivo assay with the AtPol $\lambda$  variants, generated by modifying these motifs, further supports this hypothesis. The AtPol $\lambda$  variant resembling human Pol $\lambda$  significantly increased the proportion of templated insertions, while the human TdT-like variant significantly increased the proportion of non-templated insertions. These findings suggest that the dual activities of plant Pol $\lambda$  stem from the combination of conserved motifs from both template-dependent and independent polymerases.

Moreover, it is worth noting that the frequencies of a specific inserted nucleotide can vary for non-templated insertions. For example, lower frequencies of cytosine (C) insertions compared to adenine (A), thymine (T), and guanine (G) were observed in the CHL12 and MCSite\_T sites (Fig. 3b and Supplementary Fig. 4). One possible explanation could be that the AtPol $\lambda$  exhibits preferences for certain types of nucleotides when installing non-templated insertions. Substrate nucleotide preferences have been previously reported in human Terminal deoxynucleotidyl transferase (TdT), which catalyzes non-templated nucleotide additions. In vitro studies have shown that human TdT exhibits a preference for incorporating nucleotides in the order of G>T=A>C<sup>32,33</sup>. It is plausible that AtPol $\lambda$  also displays nucleotide preferences, potentially favoring G, T, and A over C during non-templated insertions in vivo. To explore this hypothesis and elucidate the AtPol $\lambda$  activity during non-templated synthesis, further experiments would be required.

Taking these observations together, we propose that dual activities of the X-family DNA polymerase, Pol $\lambda$ , plays a pivotal role in



determining templated versus non-templated 1-bp insertion patterns (Fig. 6d). According to this model, CRISPR-Cas9 can cleave its target sites in two modes: blunt-ended cleavage at the -3rd position upstream of the PAM and staggered cleavages with one cut at the -4th position on the non-targeted strand and the other cut at the -3rd position on the targeted strand<sup>8,9,13,15,34</sup>. Following cleavage, Pol $\lambda$  homologs with dual activities could use either staggered or blunt products as substrates to generate a spectrum of templated and non-templated insertions. In support of this view, previous studies have indicated that template-dependent DNA polymerase favors sticky DNA ends with overhangs as substrates, while the template-independent TdT prefers blunt ends<sup>30</sup>. Moreover, the ratio between staggered and blunt cleavage can vary in different CRISPR sites. Recent studies in human cell lines have suggested that the sequence context, particularly the nucleotide at the -4 position, could play a crucial role in influencing the ratio of blunt and staggered cleavages. For example, target sites with a thymine (T) at the -4 position appeared to exhibit higher staggered cleavage compared to those with other nucleotides<sup>11,35</sup>. In our model, a higher frequency of staggered cleavages would result in more templated insertions, and vice versa. While the nucleotide at the -4 position has been identified as a key contributing factor, it is likely that additional sequence features or other factors could influence the ratio of blunt and staggered cleavages. To systematically explore this question further, experiments employing approaches to capture double-strand break (DSB) end structures *in vivo* could be employed<sup>36</sup>. Furthermore, noticeable levels of non-templated 1-bp insertions were also observed from CRISPR sites in human cells particularly when the -4th nucleotides were C or G. It has been suggested that other non-template dependent X-family DNA polymerases, such as Pol or TdT, could contribute to this phenomenon<sup>35</sup>. Further research will be required to systematically address these questions.

It is noteworthy that, in this study, we compared 1-bp insertion profiles across different cell states from three different species, including non-dividing *Setaria* protoplast cells, dividing cells from *Arabidopsis* seedlings, and human dividing cell lines. It has been known that cell states, especially cell cycle stages, could influence DNA repair pathway choices and expression of DNA repair genes<sup>37</sup>. To address the potential differences in 1-bp insertion profiles between dividing and non-dividing cells, we conducted an experiment using *Arabidopsis* protoplasts to target three CRISPR sites: CHLI2, MCSite\_T, and MCSite\_G. When comparing the 1-bp insertion profiles of these sites between the stable transgenic lines and protoplasts, we observed highly similar 1-bp mutation outcomes between the two groups (Supplementary Fig. 6f, Supplementary Data 4). While our data indicate that the cell cycle state may not be a major determinant of the 1-bp insertion patterns observed in this study, it has been reported that environmental factors can regulate the expression level and protein stability of AtPol $\lambda$ <sup>38,39</sup>. Further investigation would be required to examine the CRISPR mutagenesis profiles under different growth conditions.

To rigorously test this dual activity model, it will be imperative to further dissect the properties of Pol $\lambda$  homologs through both genetic and biochemical approaches. Nevertheless, our findings unveiled a unique role of a specialized DNA polymerase in regulating DNA repair outcomes, holding several implications for enhancing the predictability and precision of CRISPR-Cas9 mediated genome editing. The varying rates and profiles of 1-bp insertions observed across species could stem from distinct properties of Pol $\lambda$  activity or its differential involvement in repairing CRISPR-Cas9-induced double-strand breaks. This underscores how inherent variability in DNA repair pathways and the enzymatic properties of key repair proteins, such as Pol $\lambda$ , contribute to the species-specific CRISPR-Cas9 editing profiles. Consequently, enhancing the predictability of CRISPR-Cas9 mutagenesis prediction tools necessitates tailoring and optimizing these tools for individual species. Moreover, manipulating the activities of Pol $\lambda$  and its

homologs could serve to fine-tune the frequency and profiles of CRISPR-Cas9 mutagenesis. For instance, disrupting Pol $\lambda$  may result in deletion-only CRISPR-Cas9 mutagenesis profiles. Conversely, in scenarios where precise or predictable 1-bp insertions are desirable, strategies involving the overexpression or targeted recruitment of Pol $\lambda$  at specific target sites can be explored. In *Arabidopsis* plants overexpressing Pol $\lambda$ , we have demonstrated substantial increases in the frequency of 1-bp insertions at CRISPR target sites. Further investigation is warranted on whether directly recruiting Pol $\lambda$  to targeted genomic loci could further augment 1-bp insertion frequencies, thereby improving the predictability of mutagenesis outcomes. This strategy could pave the way for more predictable CRISPR-Cas9 mutagenesis across diverse species by circumventing the intrinsic variability in DNA repair pathway machinery among organisms. As such, it could also present a powerful tool for precisely introducing 1-bp insertions at specific sites in gene therapy, especially considering that over 8000 human disease-related pathogenic alleles result from 1-bp deletions<sup>40</sup>.

## Methods

### Plant materials and growth conditions

The *Arabidopsis thaliana* Columbia ecotype (Col-0) and *Setaria viridis* (ME034 ecotype) were used in these experiments. *Arabidopsis* plants were grown in a growth chamber with the following conditions: 16-/8-h light/dark cycle, 22 °C, and 55% humidity. *Setaria* growth conditions for protoplasting were performed as previously described<sup>19</sup>.

Homozygous *Arabidopsis* Pol $\lambda$  (AT1G10520) knock-out line (*atPol $\lambda$ -1*) was obtained from *Arabidopsis* Biological Resource Center (ABRC) with the stock number SALK\_075391C<sup>22</sup>. The mutant plants were grown under the same conditions as the wild-type Col-0 plants.

### Vector construction

Plasmids for *Arabidopsis* and *Setaria* transformation and CRISPR-Cas9 mutagenesis experiments were created using the method described previously<sup>41</sup>. In brief, the gRNAs, listed in Supplementary Data 1, were first cloned into pMOD\_B2301 (for *Arabidopsis*; Addgene #91067) or pMOD\_B2303 (for *Setaria*; Addgene #91068) to create gRNA expression plasmids. The *Setaria* protoplast transfection constructs were generated by using the Golden Gate assembly method with pMOD\_A1110 (the Cas9 expressing plasmid; Addgene #91031), the gRNA plasmids, pMOD\_C3001 (the GFP reporter plasmid; Addgene #91094), and pTRANS100 (the destination plasmid; Addgene #91198). The *Arabidopsis* T-DNA transformation constructs were generated using the Golden Gate assembly method with pMOD\_A0101 (the Cas9 expressing plasmid; Addgene #90998), the gRNA plasmids, pMZ105 (the luciferase reporter plasmid), and pTRANS230d (the T-DNA destination plasmid with the BASTA selection gene; Addgene #91113).

The Plasmid for luciferase expression in HEK293T cells was created by inserting the firefly luciferase gene into the lentiviral cloning vector pCDH-Neo via *Xba*I and *Eco*RI sites. The CRISPR-Cas9 plasmids (pTW224 and pTW225) targeting the luciferase gene were created by inserting CRISPR gRNA into pX458-GFP (pTW184) via *Bbs*I digestion and ligating annealed oligos containing complementary overhangs (Supplementary Data 3).

### Plant transformation for CRISPR-Cas9 expression

All *Arabidopsis* transgenic plants were generated using the Columbia ecotype (Col-0). T-DNA-mediated floral dip transformation was performed according to the protocol as previously outlined<sup>42</sup>. Transgenic T1 seeds were sown on soil and treated with BASTA selection to select transgenic plants as previously described<sup>42</sup>. In brief, developing *Arabidopsis* inflorescences were simply dipped for a few seconds into a 5% sucrose solution containing 0.05% Silwet L-77 and resuspended *Agrobacterium* cells carrying the genes to be transferred. Whole plants were used for genotyping at 3 weeks post-germination using the amplicon-based next-generation sequencing assay.

*Arabidopsis* and *Setaria* protoplast transfection was performed by following the protocol described previously<sup>19</sup>. In brief, leaves from 14-day-old seedlings were sliced and digested in the enzyme solution (1.5% Cellulase, 0.75% Macerozyme, Kanematsu USA Inc.) for 4–5 h on a shaker at 40 rpm. Digested tissues were filtered through a 70 μm nylon filter (Fisher Scientific LLC) into W5 buffer (2 mM MES with pH 5.7, 154 mM NaCl, 125 mM CaCl<sub>2</sub>, 5 mM KCl). Protoplasts were collected and resuspended in W5 buffer with a gentle centrifuge at 100g for 5 min. Approximately 200,000 protoplasts were mixed with DNA plasmids (15 μg per construct) in 20% PEG buffer, incubated at room temperature in the dark for 48 h, then collected for the amplicon-based next-generation sequencing.

We adopted an Agrobacterium-mediated transient transformation approach, AGROBEST (Agrobacterium-mediated enhanced seedling transformation), for testing AtPolλ and its variants in the atpolλ-1 mutant. T-DNA constructs were assembled using the Golden Gate assembly method with pMOD\_A0101 (the Cas9 expressing plasmid, Addgene #90998), the gRNA plasmids, the *AtUbi10* promoter driving *AtPolλ* coding sequences (either wild type or with amino acid variants), the AmCyan fluorescence reporter gene (Addgene #197731), and pTRANS230d (the T-DNA destination plasmid with the BASTA selection gene; Addgene #91113). The transient T-DNA transformation protocol was performed as described by Wu et al.<sup>43</sup> with the following modifications. The *Arabidopsis* seeds were germinated in 6-well plates in 1 mL of ½ MS liquid media with 0.5% sucrose (pH 5.5) in a growth incubator set to a 16-/8-h light/dark cycle, 22 °C, and 55% humidity. Thirteen-day-old seedlings were co-cultivated with the Agrobacterium strain EHA105 carrying each T-DNA construct at O.D. 0.02 for 3 days. The transformed plants were then transferred to 1 mL of ½ MS medium containing 250 μg/mL Timentin for 3 days under the same growth condition. Individual transformed plants underwent genomic DNA extraction followed by Next Generation Sequencing assay.

#### Human cell culture, transfection, and lentivirus transduction

HEK293 and 293T cells (from ATCC with Cat# CRL-1573 and CRL-3216) were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. For lentivirus production, 293T cells were co-transfected with pCDH carrying the luciferase gene and the packaging plasmids pCMV-ΔR8.91 and pMDG-VSV-G at a ratio of 3:1:3 using PEI STAR™ (Tocris Bioscience, Minneapolis, Minnesota, USA), a polyethylenimine (PEI) transfection reagent, following the manufacturer's instruction. Medium-containing virus particles were collected 72 h after transfection and applied to HEK293 cells for 48 h. Transduced HEK293 cells were then selected using G418 (800 μg/ml) for 5 days to establish a cell line stably expressing luciferase. HEK293-luciferase cells were transfected with pX458 carrying CRISPR gRNA using TransIT-X2 (Mirus Bio) based on the manufacturer's instruction. 48 h after the transfection, cells were harvested to extract genomic DNA for the amplicon-based next-generation sequencing.

#### Mutation genotyping and characterization of mutation profiles

Total genomic DNA was extracted from the *Setaria* protoplast or *Arabidopsis* whole plant samples using the Cetyl trimethylammonium bromide (CTAB) method as described earlier<sup>9,16,17</sup>. Polymerase chain reaction (PCR) was performed from the genomic DNA with the oligonucleotides listed in Supplementary Data 3 by using GoTaq Green Mastermix (Promega Corp., Madison, Wisconsin, USA). The PCR conditions were followed as per the manufacturer's instruction with an annealing temperature at 55 °C (CHL12 and MCsite\_G) or 60 °C (MCsite\_T) and an extension time of 1 min. Individual PCR amplicons were subjected to Illumina paired-end read sequencing (Genewiz Inc., South Plainfield, New Jersey, USA). The paired-end NGS reads were

analyzed using CRISPResso2 (version 2.1.1) to identify NHEJ-mediated indel mutations<sup>44</sup> with the default settings. The overall mutagenesis rates for each site were summarized in the Source Data file and Supplementary Data 4. The corresponding output files were loaded in R studio (Version 2023.09.1) for data plotting.

#### Statistics and reproducibility

An unpaired one-tailed Student's *t* test was conducted using R studio (Version 2023.09.1) to assess the significance of the mean value differences between the Polλ mutant or overexpression plants and the wild-type control groups. No statistical method was used to pre-determine the sample size. No data were excluded from the analyses. For each experiment, all samples were randomly chosen. The treatments were compared to a control treatment without any prior knowledge of whether the experimental variables being altered would have a positive or negative impact on the results.

#### Phylogenetic analyses

All X-family DNA polymerase protein sequences were obtained from NCBI GenBank (Supplementary Data 2). The neighbor-joining tree was generated using MEGA (version 11) with the default setting using the yeast DNA Pol4 as the outgroup<sup>45</sup>.

#### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

#### Data availability

The plasmids used in the Golden Gate assembly are available from Addgene with the ID number listed in the Methods. The Next-Generation sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under accession code [PRJNA1062985](#). Sequence data used from previously published article can be found in NCBI SRA under accession code [PRJNA795172](#). The processed CRISPResso2 output files for each sample are available from github [[https://github.com/ZhangLab-UMN/Weiss\\_Kumar\\_ibp\\_manuscript.git](https://github.com/ZhangLab-UMN/Weiss_Kumar_ibp_manuscript.git)]. Other data generated in this study are provided in the Supplementary Information/Source Data file. Source data are provided with this paper.

#### References

- Wang, J. Y. & Doudna, J. A. CRISPR technology: a decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).
- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Liu, G., Lin, Q., Jin, S. & Gao, C. The CRISPR-Cas toolbox and gene editing technologies. *Mol. Cell* **82**, 333–347 (2022).
- Nambiar, T. S., Baudrier, L., Billon, P. & Ciccio, A. CRISPR-based genome editing through the lens of DNA repair. *Mol. Cell* **82**, 348–388 (2022).
- Chen, K., Wang, Y., Zhang, R., Zhang, H. & Gao, C. CRISPR/cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* **70**, 667–697 (2019).
- Yamamoto, Y. & Gerbi, S. A. Making ends meet: targeted integration of DNA fragments by genome editing. *Chromosoma* **127**, 405–420 (2018).
- Shou, J., Li, J., Liu, Y. & Wu, Q. Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol. Cell* **71**, 498–509.e4 (2018).
- Kumar, J. et al. Efficient protein tagging and cis-regulatory element engineering via precise and directional oligonucleotide-based

- targeted insertion in plants. *Plant Cell* <https://doi.org/10.1093/plcell/koad139> (2023).
10. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
  11. Allen, F. et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* <https://doi.org/10.1038/nbt.4317>. (2018).
  12. Shen, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
  13. Lemos, B. R. et al. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc. Natl Acad. Sci. USA* **115**, E2040–E2047 (2018).
  14. Naert, T. et al. Maximizing CRISPR/Cas9 phenotype penetrance applying predictive modeling of editing outcomes in *Xenopus* and zebrafish embryos. *Sci. Rep.* **10**, 14662 (2020).
  15. Hussmann, J. A. et al. Mapping the genetic landscape of DNA double-strand break repair. *Cell* **184**, 5653–5669.e25 (2021).
  16. Molla, K. A., Shih, J., Wheatley, M. S. & Yang, Y. Predictable NHEJ insertion and assessment of HDR editing strategies in plants. *Front. Genome Ed.* **4**, 825236 (2022).
  17. Weiss, T. et al. Epigenetic features drastically impact CRISPR–Cas9 efficacy in plants. *Plant Physiol.* **190**, 1153–1164 (2022).
  18. Slaman, E., Lammers, M., Angenent, G. C. & de Maagd, R. A. High-throughput sgRNA testing reveals rules for Cas9 specificity and DNA repair in tomato cells. *Front. Genome Ed.* **5**, 1196763 (2023).
  19. Weiss, T. et al. Optimization of multiplexed CRISPR/Cas9 system for highly efficient genome editing in *Setaria viridis*. *Plant J.* **104**, 828–838 (2020).
  20. Schep, R. et al. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol. Cell* **81**, 2216–2230.e10 (2021).
  21. Roy, S., Choudhury, S. R., Singh, S. K. & Das, K. P. AtPol $\lambda$ , a homolog of mammalian DNA polymerase  $\lambda$  in *Arabidopsis thaliana*, is involved in the repair of UV-B induced DNA damage through the dark repair pathway. *Plant Cell Physiol.* **52**, 448–467 (2011).
  22. Roy, S., Choudhury, S. R., Sengupta, D. N. & Das, K. P. Involvement of AtPol $\lambda$  in the repair of high salt- and DNA cross-linking agent-induced double strand breaks in *Arabidopsis*. *Plant Physiol.* **162**, 1195–1210 (2013).
  23. Furukawa, T., Angelis, K. J., Britt, A. B. & Arabidopsis, D. N. A. polymerase lambda mutant is mildly sensitive to DNA double strand breaks but defective in integration of a transgene. *Front. Plant Sci.* **6**, 357 (2015).
  24. Moon, A. F. et al. The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair* **6**, 1709–1725 (2007).
  25. Delarue, M. et al. Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *EMBO J.* **21**, 427–439 (2002).
  26. Pedroza-Garcia, J.-A., De Veylder, L. & Raynaud, C. Plant DNA Polymerases. *Int. J. Mol. Sci.* **20** (2019).
  27. Amoroso, A. et al. Oxidative DNA damage bypass in *Arabidopsis thaliana* requires DNA polymerase  $\lambda$  and proliferating cell nuclear antigen 2. *Plant Cell* **23**, 806–822 (2011).
  28. Uchiyama, Y., Kimura, S., Yamamoto, T., Ishibashi, T. & Sakaguchi, K. Plant DNA polymerase lambda, a DNA repair enzyme that functions in plant meristematic and meiotic tissues. *Eur. J. Biochem.* **271**, 2799–2807 (2004).
  29. Uchiyama, Y., Takeuchi, R., Kodera, H. & Sakaguchi, K. Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie* **91**, 165–170 (2009).
  30. Nick McElhinny, S. A. et al. A gradient of template dependence defines distinct biological roles for family X polymerases in non-homologous end joining. *Mol. Cell* **19**, 357–366 (2005).
  31. Ramadan, K. et al. Human DNA polymerase  $\lambda$  possesses terminal deoxyribonucleotidyl transferase activity and can elongate RNA primers: Implications for Novel. *Funct. J. Mol. Biol.* **328**, 63–72 (2003).
  32. Kuznetsova, A. A. et al. substrate specificity diversity of human terminal deoxynucleotidyltransferase may be a naturally programmed feature facilitating its biological function. *Int. J. Mol. Sci.* **25** (2024).
  33. Kuznetsova, A. A. et al. Insight into the mechanism of DNA synthesis by human terminal deoxynucleotidyltransferase. *Life Sci. Alliance* **5**, e202201428 (2022).
  34. Zuo, Z. & Liu, J. Cas9-catalyzed D. N. A. cleavage generates staggered ends: evidence from molecular dynamics simulations. *Sci. Rep.* **5**, 37584 (2016).
  35. Gisler, S. et al. Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nat. Commun.* **10**, 1598 (2019).
  36. Ben-Tov, D. et al. Uncovering the dynamics of precise repair at CRISPR/Cas9-induced double-strand breaks. *bioRxiv* <https://doi.org/10.1101/2023.01.10.523377> (2023).
  37. Hustedt, N. & Durocher, D. The control of DNA repair by the cell cycle. *Nat. Cell Biol.* **19**, 1–9 (2016).
  38. Roy, S., Banerjee, V. & Das, K. P. Understanding the physical and molecular basis of stability of *Arabidopsis* Pol $\lambda$  under UV-B and high NaCl stress. *PLoS ONE* **10**, e0133843 (2015).
  39. Roy, S., Roy Choudhury, S. & Das, K. The interplay of DNA polymerase  $\lambda$  in diverse DNA damage repair pathways in higher plant genome in response to environmental and genotoxic stress factors. *Plant Signal. Behav.* **8**, e22715 (2013).
  40. Bermudez-Cabrera, H. C. et al. Small molecule inhibition of ATM kinase increases CRISPR-Cas9 1-bp insertion frequency. *Nat. Commun.* **12**, 5111 (2021).
  41. Čermák, T. et al. A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* **29**, 1196–1217 (2017).
  42. Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W. & Chua, N.-H. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* **1**, 641–646 (2006).
  43. Wu, H.-Y. & Lai, E.-M. AGROBEST: a highly efficient agrobacterium-mediated transient expression system in *Arabidopsis* seedlings. *Methods Mol. Biol.* **2379**, 113–123 (2022).
  44. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
  45. Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary genetics analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).

## Acknowledgements

J.K. and F.Z. are partially supported by the National Science Foundation (IOS-2040218 and IOS-2206920) awards. T.W. was supported by a Bernard and Jean Phinney Graduate Fellowship in Plant Biology and a Doctoral Dissertation Fellowship from the University of Minnesota. C.C. and K.L. are supported by the Department of Defense CDMRP research award (W81XWH2010214). We thank all the Zhang lab members for their inputs.

## Author contributions

T.W., J.K., K.L., and F.Z. conceived and planned the study. T.W., J.K., C.C., S.G., J.L., O.S., and K.L. performed the experiments. F.Z., T.W., S.G., and J.K. analyzed the data. F.Z., T.W., J.K., and K.L. wrote the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50676-4>.

**Correspondence** and requests for materials should be addressed to Feng Zhang.

**Peer review information** *Nature Communications* thanks Yiping Qi, Cécile Raynaud, and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024