Programmable Single and Multiplex Base-Editing in *Bombyx mori* Using RNA-Guided Cytidine Deaminases

Yufeng Li,*,[1] Sanyuan Ma,*,†,[1] Le Sun,* Tong Zhang,* Jiasong Chang,* Wei Lu,* Xiaoxu Chen,* Yue Liu,* Xiaogang Wang,* Run Shi,* Ping Zhao,*,† and Qingyou Xia*,†,[2]

*State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing, 400716 P. R. China and †Chongqing Engineering and Technology Research Center for Novel Silk Materials, Southwest University, Chongqing 400716, China

ABSTRACT Genome editing using standard tools (ZFN, TALEN, and CRISPR/Cas9) rely on double strand breaks to edit the genome. A series of new CRISPR tools that convert cytidine to thymine (C to T) without the requirement for DNA double-strand breaks was developed recently and quickly applied in a variety of organisms. Here, we demonstrate that CRISPR/Cas9-dependent base editor (BE3) converts C to T with a high frequency in the invertebrate *Bombyx mori* silkworm. Using BE3 as a knock-out tool, we inactivated exogenous and endogenous genes through base-editing-induced nonsense mutations with an efficiency of up to 66.2%. Furthermore, genome-scale analysis showed that 96.5% of *B. mori* genes have one or more targetable sites that can be edited by BE3 for inactivation, with a median of 11 sites per gene. The editing window of BE3 reached up to 13 bases (from C1 to C13 in the range of gRNA) in *B. mori*. Notably, up to 14 bases were substituted simultaneously in a single DNA molecule, with a low indel frequency of 0.6%, when 32 gRNAs were co-transfected. Collectively, our data show for the first time that RNA-guided cytidine deaminases are capable of programmable single and multiplex base editing in an invertebrate model.

The clustered regularly interspaced short palindromic repeat (CRISPR) system has been widely used for site-specific genome editing in various organisms and cell lines (Sander and Joung 2014). Under the direction of a guide RNA (gRNA), the Cas9 nuclease binds to an opened DNA strand that pairs with the gRNA and induces a double-strand break (DSB). At this point, the intrinsic cellular DNA-repair mechanism will generally repair the DSB via non-homologous end joining (NHEJ), resulting in deletions or insertions (indels) in most cases. When present, the cell can use a homologous DNA fragment as a template to repair the DSB by another mechanism, known as homology-directed repair (HDR). The efficiency of NHEJ-induced indel formation is always higher than that of HDR-mediated gene correction, although various efforts have been made to increase the frequency of HDR (Chu *et al.* 2015; Maruyama *et al.* 2015; Paquet *et al.* 2016; Richardson *et al.* 2016; Sakuma *et al.* 2016). ZFN, TALEN, and CRISPR/Cas9 all rely on the generation of a DSB, which can lead to many potential defects such as unexpected indels, off-target cleavage, and decreased cell proliferation when targeted to copy number-amplified (CNA) genomic regions (Shen and Ideker 2017). Thus, approaches to precisely edit the genome while avoiding DSBs are needed.

Several modified systems have been developed to overcome the drawbacks of DSB-based genome editing. Komor *et al.* showed that the rat cytidine deaminase rAPOBEC1 could be linked to nCas9 (Cas9 nickase) to efficiently substitute C with T at target sites without generating DSBs (Komor *et al.* 2016). After three generations of modification, they developed the final version of base editor (BE), named BE3, which showed up to a 74.9% mutation efficiency in mammalian cells and was composed of rAPOBEC1, nCas9 (A840H), and uracil DNA glycosylase inhibitor (UGI). Kondo *et al.* engineered CRISPR/Cas9 and the activation-induced cytidine deaminase (AID) ortholog PmCDA1 to assemble a complex (Target-AID) that performed highly efficient

specific point mutations (Nishida *et al.* 2016). In a parallel study, Ma *et al.* fused UGI with dcas9-AIDx (AICDA-P182X) to convert targeted cytidine specifically to thymine, in a process referred to as targeted AID-mediated mutagenesis (Ma *et al.* 2016). Additionally, to further optimize BE3, a series of studies were conducted to expand the number of target sites, narrow the width of the editing window (Kim *et al.* 2017a), improve the efficiency and product purity (Komor *et al.* 2017), and reduce off-target effects (Kim *et al.* 2017b). To show that BE3 converts C:G base pairs to T:A base pairs with high efficiency, several groups have used BE3 to inactivate genes through base-editing-induced nonsense mutations (Billon *et al.* 2017; Kim *et al.* 2017c; Kuscu *et al.* 2017).

The development of base-editing systems has both improved the scope and effectiveness of genome editing. Site-directed mutagenesis in the genome can be achieved by design, rather than randomly achieved through wild-type Cas9. To date, several organisms have been subjected to base editing, mostly using BE, including mammalian cells (Komor *et al.* 2016; Ma *et al.* 2016), rice (Lu and Zhu 2017; Zong *et al.* 2017), wheat (Zong *et al.* 2017), tomatoes (Shimatani *et al.* 2017), mice (Kim *et al.* 2017b), and zebrafish (Zhang *et al.* 2017). Komor *et al.* corrected two disease-relevant mutations that cause Alzheimer's disease in mammalian cells by BE3. Zhang *et al.* converted a proline residue at position 302 in the *tyr* gene to serine, threonine, or alanine to mimic the oculocutaneous albinism (OCA) mutation in zebrafish by BE3. However, no reports have described the performance of any base-editing system in invertebrates.

Bombyx mori (silkworm) is a model organism for studying invertebrate biology and an economically important lepidopteran insect (Goldsmith *et al.* 2005). The genome-editing tools ZFN, TALEN, and CRISPR/Cas9 have all been applied in *B. mori* (Ma *et al.* 2012; Liu *et al.* 2014; Ma *et al.* 2014a; Ma *et al.* 2014b; Ma *et al.* 2014c; Liu *et al.* 2017). Here, we demonstrated that BE3 performed base editing by targeting the *Blos2* and *Yellow-e* genes with efficiencies of 25% and 51.2% in *B. mori*, respectively. Then, we demonstrated that BE3 could efficiently knock out exogenous and endogenous genes through base-editing-induced nonsense mutations with efficiencies of up to 66.2%. Furthermore, we found 151,551 targetable knockout sites within the *B. mori* genome. The deamination window of BE3 in *B. mori* spans from C1 to C13 within the protospacers. In addition, by co-transfecting 32 gRNAs and BE3 simultaneously, we generated up to 14 C:G base-pair substitutions in one DNA molecule, with few indels observed.

MATERIALS AND METHODS

Design and construction of the BE3- and gRNA-expression vectors

DNA encoding rAPOBEC1, the XTEN linker, partial nCas9 with the A840H mutation, and UGI was *B. mori* codon-optimized, synthesized by GenScript service, and then inserted into the pUC57-T-simple plasmid. The complete BE3 vector was constructed by inserting the same fragment of Cas9 and nCas(A840H) into the synthesized plasmid after digestion of *Nde*I and *Bam*HI (New England BioLabs). The gRNA-expression vector pUC57-gRNA was described previously (Liu *et al.* 2014). gRNA sequences (Table S1 in File S1) were synthesized as two complementary oligonucleotides, annealed, and inserted into the pUC57-gRNA plasmid after *Bbs*I digestion.

Cell culture and transfection

The *B. mori* embryo cell lines BmE, BmE-mCherry, and BmE-EGFP were established in our laboratory and maintained at 27° in Grace's insect medium (Gibco, Thermo Fisher) containing 10% fetal bovine serum (Gibco, Thermo Fisher). The cells were seeded in 12-well plates (Corning) and transfected at approximately 80% confluency. Plasmids (1.8 μg total) were transfected at a 1:1 ratio into cells using the X-tremeGENE HP DNA Transfection Reagent (Roche), following the manufacturer's recommended protocol.

Flow cytometry and fluorescence imaging

Cells were harvested at 12 days post-transfection. A MoFlo XDP flow cytometer (Beckman) was used to measure mCherry fluorescence, and Summit software was used to analyze the data. In addition, light and fluorescence microscopy images were captured using an EVOS FL Auto microscope (Life Technologies).

Protein extraction and western blot analysis

Cellular proteins were extracted using NP-40 lysis buffer (Beyotime) on day 12 post-transfection. Proteins were quantified using a BCA Kit (Beyotime). Equal amounts of proteins were resolved by 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred to a polyvinylidene fluoride (PVDF) membrane. Next, 5% milk was used to block the PVDF membranes, which were then incubated with primary antibodies (diluted 1: 1,000 in 1% milk) for 2 h. Subsequently, the PVDF membranes were incubated with secondary anti-mouse or anti-rabbit antibodies (diluted at 1: 10,000 in 1% milk) for 1 h. Membranes were developed with Thermo Fisher ECL reagent and then imaged with a western blot processor.

Genome-wide analysis of knock-out sites

To identify knock-out sites on a genome-wide scale, we first downloaded all CDSs for *B. mori* from the Silkworm Genome Database. gRNAs and the nucleotide motifs CAG, CGA, CAA, and TGG were searched using the fuzznuc EMBOSS explorer. Based on the editing window of BE3 for *B. mori*, nucleotide motifs (with a multiple of three for the last base) at position 1–13 within each gRNA were selected using the intersectBed and shell script programs. These nucleotide motifs were defined as potential knock-out sites.

Sanger sequencing

Genomic DNA extracted using the E.Z.N.A. Tissue DNA Kit (Omega) following the manufacturer's recommended protocols after cells were transfected for 3 days. The genomic regions of interest were amplified using site-specific primers (Table S2 in File S1). PCR products were generated with PrimerSTAR Max DNA Polymerase (Takara) according to the manufacturer's instructions, then separated on a 2% agarose gel. The Gel Extraction Kit (Omega) and pEASY-T5 vector (Transgene) were used for PCR product purification and T-A cloning, respectively. The PCR and T clone products were then subjected to Sanger sequencing.

NGS experiments

Thirty-two gRNA vectors targeting EGFP were co-transfected individually with the BE3 vector. Cellular genomic DNA was extracted using the E.Z.N.A. Tissue DNA Kit (Omega) following the manufacturer's protocols at 3 days post-transfection. The target regions were amplified using site-specific primers with added barcodes. Each sample corresponded to a unique pair of barcodes. All 32 PCR products were mixed well at a 1:1 molar ratio and subjected to NGS on an Illumina HiSequation 2500 PE250 instrument. The 32 gRNA vectors were also mixed well at a 1:1 molar ratio and transfected together with BE3 vector. The targeted genomic regions were amplified using site-specific primers with added barcodes and subjected to NGS on an Illumina MiSeq PE300 instrument. Control samples were sequenced in the same way, except that they were only transfected with the BE3 vector.
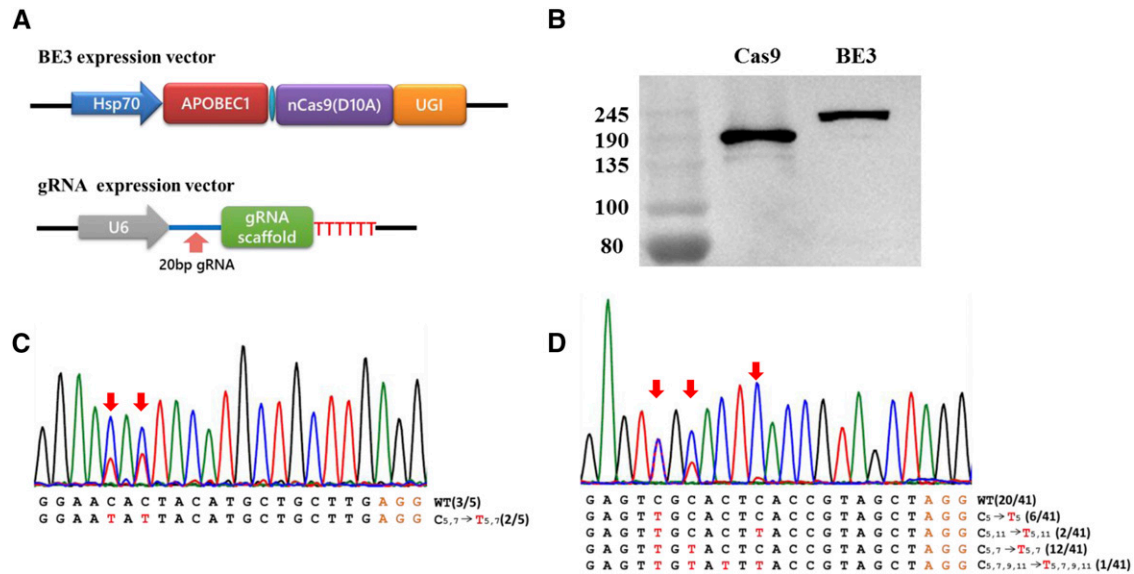
Figure 1 Base editing in *B. mori*. (A) Schematic representation of the BE3- and gRNA-expression vectors. (B) Detection of BE3 expression in the BmE cell line by western blotting. (C, D) Sanger sequencing of targeted genomic regions of *Blos2* and *Yellow-e*. *Red arrows* point to overlapping peaks. The orange and red letters mark PAM sites and converted bases, respectively. The numbers in trumpet font indicate the position of C or T.

NGS data analysis

The sequencing reads generated for each sample were first filtered using the Trimmomatic tool with the parameters "LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50" to remove low-quality bases at the ends of each read and to truncate reads containing consecutive bases with an average quality score below 15. Using different barcodes, we extracted the sequencing reads for each sample using shell script. We removed unedited reads at gRNA-target sites (20 bp) and collected reads with C:G to T:A conversions at gRNA-target sites (20 bp). The efficiencies of all 32 gRNAs were analyzed by counting the proportion of reads with conversions among all clean reads for each sample. All the above data-processing steps were performed using shell script. We also analyzed the efficiency of each C:G substitution for each gRNA in the same manner. A heatmap was prepared using R software. To determine the indel frequencies, sequence-alignment files in SAM format were generated using sequencing reads and reference sequences via bowtie2. Reads that contained an insertion or a deletion (according to the SAM files) were considered to represent indels.

Data availability

NGS data were deposited in the NCBI Sequence Read Archive under accession number PRJNA434087. The plasmids and cell lines developed in this study are available upon request. The authors state that all data necessary for confirming the conclusions presented are represented fully in this article.

RESULTS

Establishing a base-editing system for *B. mori*

To confirm whether base editing could be achieved by BE3 (rAPOBEC1-XTEN-nCas9-UGI) in *B. mori*, we first constructed the BE3 vector. DNA encoding rAPOBEC1, the XTEN linker, a partial nCas9 sequence with an A840H mutation, and UGI was commercially synthesized, using *B. mori* codon-optimized sequences, and then was inserted into the pUC57-T-simple plasmid. The Hsp70 promoter and SV40 terminator sequence (Ma *et al.* 2017) were used to initiate and terminate BE3 expression, respectively. The complete BE3 vector was constructed by inserting a

fragment encoding either Cas9 or nCas9 (A840H) into the synthesized plasmid after the digestion with *Nde*I and *Bam*HI (Figure 1A). Then, the BE3 vector was transfected into the *B. mori* embryonic cell line, BmE. Cellular proteins were extracted at 3 days post-transfection. Expression of the BE3 protein was confirmed by western blot analysis (Figure 1B). To evaluate the efficacy of BE3-based genome editing in invertebrate cells, two previously constructed gRNAs targeting *Blos2* and *Yellow-e* in *B. mori* cells were used (Liu *et al.* 2014). Based on the typical editing window *in vitro* (Komor *et al.* 2016), ranging from position 4 to 8 in the proto-spacer counting from the distal end to the protospacer-adjacent motif (PAM), both gRNAs have a C located in position 5 (C5) and 7 (C7). We co-transfected the gRNA and BE3 vector into BmE cells, and cellular genomic DNA was extracted after 3 days without selection. Genomic regions spanning the target sites were amplified by PCR for sequencing. Sanger sequencing chromatograms for both the *Blos2* and *Yellow-e* PCR products showed a set of C/T peaks in the target sites, indicating that the intended base substitutions did occur (Figure 1C, D). To confirm the base substitutions, the PCR products were cloned into the pEASY-T5 vector by T-A cloning. Sanger sequencing of individual clones showed different types of mutations in the DNA sequences. For *Blos2*, two out of five clones (40%) had C–T substitutions. The efficiency of C–T substitutions for *Yellow-e* was 51.2% in 41 examined clones. Among the studied positions (C5, C7, C9, and ^{11}C), C5 showed the highest mutation rate. The observation of base substitutions in C9 and C13 was beyond our expectation because C9 and C11 laid outside of the typical editing window observed *in vitro* and in plants (Komor *et al.* 2016; Zong *et al.* 2017). Taken together, these results demonstrated that the BE3–gRNA nuclease complex effectively and site-specifically substituted C with T in invertebrate *B. mori* cells. Notably, the base-editing window of BE3 in *B. mori* may be wider than that *in vitro* and plants.

Using BE3 to knock out the exogenous genes *mCherry* and *Puromycin* in *B. mori* cells

To expand the application of BE3 in *B. mori*, we evaluated BE3 as a knock-out tool through the induction of premature stop codons (TAG, TGA, or TAA) by converting C:G base pairs to T:A base pairs for four
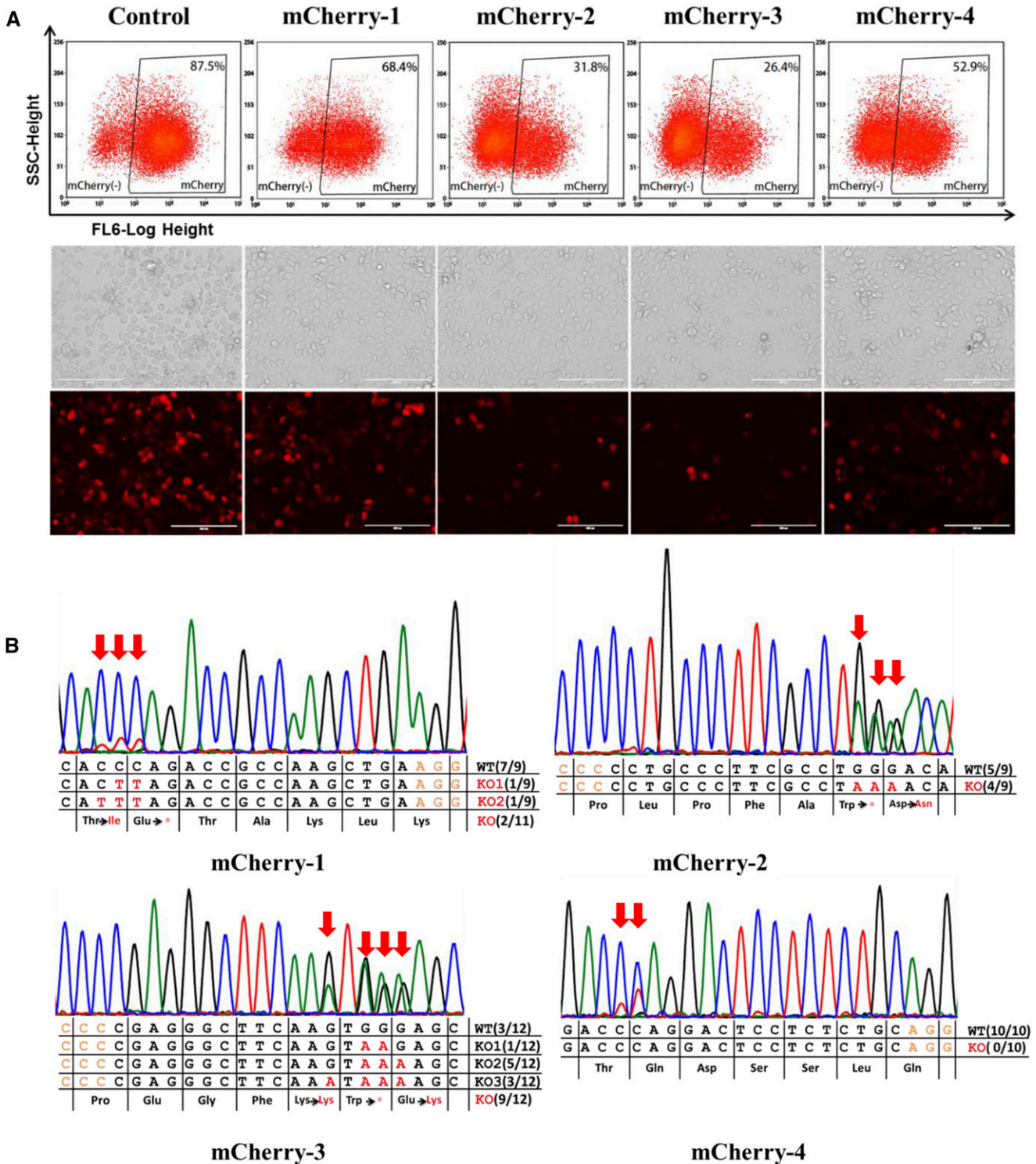
Figure 2 Using BE3 to knock out the *mCherry* reporter gene by introducing a premature stop codon. (A) Flow cytometry-based quantification of mCherry expression at 12 days post-transfection (upper panel). Light and fluorescence images show decreased mCherry-positive cells (lower panel). (B) Sanger sequencing results show the base conversions in targeted genomic regions of four *mCherry* gRNAs. The encoded amino acids are shown below. Asterisks in red represent stop codons.

codons (CAA, CAG, CGA, TGG) in coding strands. Four gRNAs with the potential of generating stop codons in the exogenous *mCherry* reporter gene were designed (Table S1 in File S1). We co-transfected these four gRNAs (mCherry-1, 2, 3, or 4) together with BE3 individually into

a transgenic cell line (BmE-mCherry), which harbors the *mCherry* expression cassette (unpublished). At 12 days post-transfection, we evaluated the knock-out efficiency by measuring the average fluorescence intensity of mCherry via flow cytometry. Significantly decreased mCherry signal was

found with all four gRNAs compared with control cells (Figure 2A, Figure S1 in File S1). We also observed markedly decreased mCherry-positive cells in cells harboring the BE3–gRNA complex by fluorescence microscopy (Figure 2A), in agreement with the flow cytometry results. However, it should be noted that the most effective gRNAs were mCherry-2 and mCherry-3 (59.1 and 66.2%, respectively), rather than mCherry-1, which targeted the 5′-most mCherry sequence among the four gRNAs. These results may have been observed because mCherry-2/3 targeted TGG with two potentialities of inducing stop codons. To further confirm that mCherry was knocked out through a substitution-induced nonsense mutation, we amplified and sequenced the targeted genome regions. Sanger sequencing of both PCR products and individual clones showed several base substitutions, resulting in stop codon mutations (Figure 2B). Although no mutations were found with mCherry-2 among 10 clones, the Sanger sequencing chromatogram for the PCR product did show an overlapping peak at the target site, where CAG was converted to the stop codon TAG (Figure 2B), resulting in *mCherry* being functionally knocked out. To test the universality of BE3 in *B. mori*, we then designed two gRNAs (Puromycin-1, 2) targeting another exogenous gene *Puromycin*. Cellular proteins were extracted at 12 days post-transfection. Western blot results indicated that *Puromycin* protein levels dramatically decreased after transfection with either gRNA (Figure 3A). Sanger sequencing of PCR products and clones also showed C:G-to-T:A conversions. The stop codons TAA and TGA generated by BE3 caused *Puromycin* to be knocked out with efficiencies of 58.3% (7/12) and 27.2% (3/11), respectively (Figure 3B). Taken together, our data indicated that we effectively knocked out the exogenous genes *mCherry* and *Puromycin* via substitution-induced nonsense mutations mediated by BE3 in *B. mori*.

Using BE3 to knock out the endogenous genes *BmGAPDH* and *BmV-ATPase B*, and predicting knocked-out loci on a genome-wide scale

To determine whether BE3 could generate effective substitution-induced nonsense mutations in endogenous genes, four gRNAs were designed to target *BmGAPDH* (GAPDH-1, 2) and *BmV-ATPase B* (ATPase-1, 2). We extracted cellular proteins at 12 days post-transfection. Western blot analyses indicated that protein levels noticeably decreased for GAPDH-1 and ATPase-1, compared with control cells (Figure 4A). Sanger sequencing chromatograms for four PCR products showed overlapping peaks in the target sites, where nonsense mutations were generated (Figure 4B). The Sanger sequencing results were consistent with the western blot results. Collectively, these findings showed that BE3 could mediate effective substitution-induced nonsense mutations for endogenous genes in *B. mori* cells. However, not all gRNAs work exceedingly well with BE3 in knocking out genes (as also found with Cas9), such as GAPDH-2 and ATPase-2 (Figure 4A, B).

To determine the targetable sites available for knockout by BE3 at the genome scale, we identified all candidate codons (CAA, CAG, CGA, and TGG) that can potentially be converted to stop codons (TAA, TAG, TGA) by BE3. This genome-scale analysis revealed 151,551 potentially targetable knockout sites in 14,106 CDSs, with a median of 11 sites per gene, and 96.5% of all genes (14,623) were targetable (Figure 4C). Furthermore, the distributions of these codons were well-distributed within the CDSs, suggesting that approximately half of these sites could be targeted to stop mRNA translation within the first 50% of their encoded protein sequences (Figure 4D).

The editing window of BE3 mostly resides Between C4 and C7, although effective C substitutions can occur within C1–C13

Sanger sequencing for T clones showed that base substitutions could be generated at C3–C9 and ^{11}C. However, these results may not fully
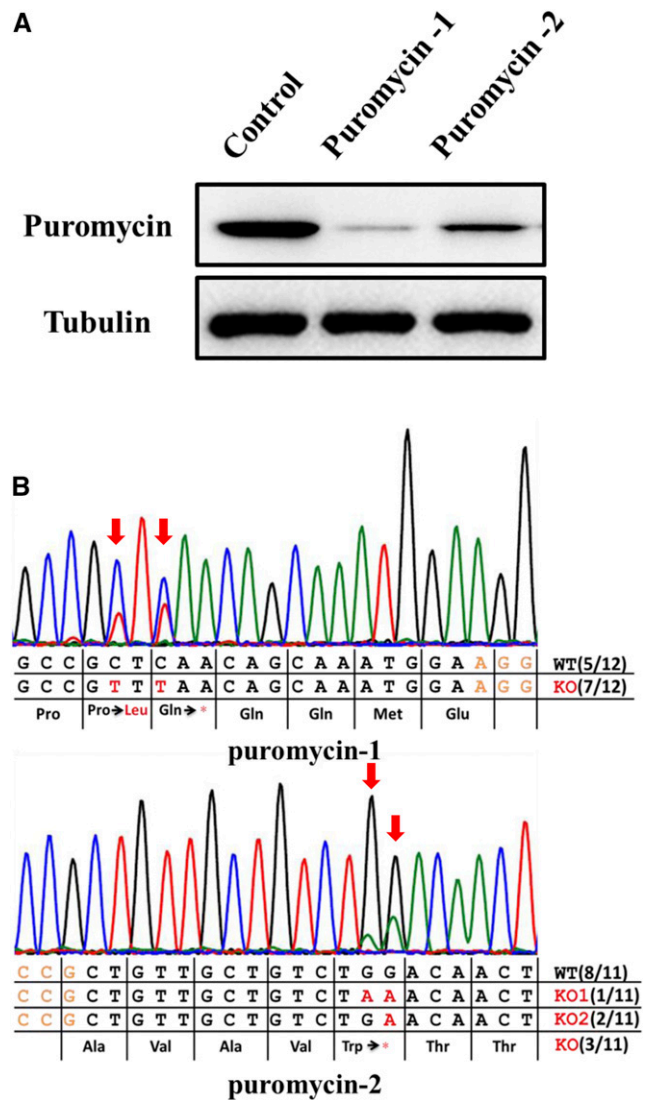

Figure 3 Using BE3 to knock out the exogenous *Puromycin* gene by introducing a premature stop codon. (A) Sanger sequencing results showing base conversions in the targeted genomic regions of two *Puromycin* gRNAs. (B) Protein levels were analyzed by western blotting at 12 days post-transfection. Alpha-tubulin was detected as a control protein.

represent the editing window of BE3 in *B. mori*. Next, we designed 32 gRNAs to target the enhanced green fluorescent protein (EGFP) gene within the 5′ 400 base pairs (bp) (Figure 5A). Each gRNA was co-transfected with the BE3 vector into the BmE-EGFP cell line (a transgenic cell line harboring an EGFP-expression cassette, unpublished). The PCR products for targeted genomic regions were sequenced by next-generation sequencing (NGS). We obtained an average of 96,782 DNA reads for each cell line transfected with a different gRNA. The efficiencies of 32 gRNAs, ranging from 3.4 to 25.3%, were calculated by counting every DNA read with one or more C:G-to-T:A substitutions within the 20-bp gRNA target sites (Figure 5B). The indel frequencies of the 32 gRNAs were ≤5.0%, except for gRNA21 (8.1%), while the control showed a 1.2% indel frequency (Table S3 in File S1). Then, we analyzed the efficiency of each base substitution for each gRNA in detail (Figure 5C). The different positions of the targeted C:G base pairs in the 32 gRNAs probably affected
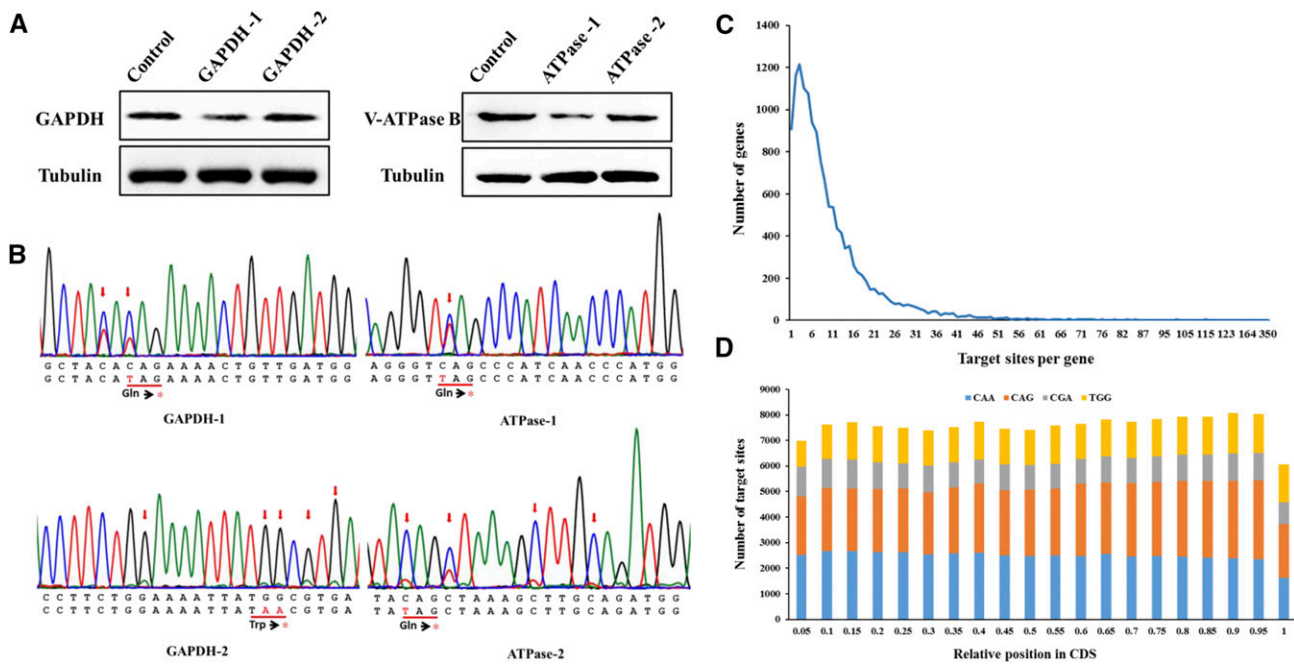
Figure 4 Using BE3 to knock out the endogenous *GAPDH* and *V-ATPase B* genes by introducing a premature stop codon. (A) Protein levels were analyzed by western blotting at 12 days post-transfection. Alpha-tubulin was detected as a control protein. (B) Sanger sequencing of genomic regions targeted by *GAPDH* and *V-ATPase B* gRNAs. The encoded amino acids are shown below. Asterisks in red represent stop codons. (C) The number of genes with different numbers of targetable knockout sites per gene. (D) Relative positions of knockout sites in CDSs. Four targetable codons are shown in different colors.

the editing efficiencies (Table S1 in File S1). Not every C:G base pair in the 32 designed gRNAs were substituted. For example, there were 16 gRNAs with a C1 residing in, but only 2 of them had a C1 substitution. We counted the number of designed gRNAs with Cn and the number of mutational gRNAs with corresponding Cn. The ratio of these two numbers indicated the probability of each C:G substitution within 20-bp gRNA target site (Figure 5D). These data provide a clear understanding of the editing window, which was mostly contained within C4–C7, although effective C substitutions occurred within the range of C1–C13 (Figure 5C).

Multiple base-editing With an extremely low indel frequency by BE3

Previous studies have indicated that DSBs generated by Cas9 might lead to excessive DNA damage that could cause cell death (Shen and Ideker 2017) and targeted chromosome elimination (Zuo *et al.* 2017) when targeting several genomic loci simultaneously. However, BE3, which is much more temperate than Cas9, with respect to the genome, demonstrated base editing without generating DSBs. Thus, BE3 should be more suitable for performing multiple genome edits, without causing excessive DNA damage and indels (Komor *et al.* 2016; Kim *et al.* 2017a; Zong *et al.* 2017). All 32 gRNAs that target EGFP were co-transfected together with the BE3 vector into cells. The PCR products for the targeted genomic regions were sequenced by NGS. We found a range of mutation types (1–14 base substitutions) with an extremely low indel frequency of approximately 0.6%, which was the equal to the control (Figure 6A, Table S3 in File S1). The more base substitutions were generated by BE3 simultaneously, the fewer reads that were observed (Figure 6A). We further found that several C:G base pairs located in different positions were mutated to T:A base pairs simultaneously in a single DNA read (Figure 6B). Although only a few reads showed 10 or more base-substitutions, they did not exist in the control. Together, our data indicate that BE3 represents an improvement over the Cas9 nuclease in inducing multiple base mutations with an extremely low indel frequency and no DSBs.

DISCUSSION

In recent years, CRISPR/Cas9 genome-editing technology has been rapidly and widely adapted for cutting-edge applications in stem cell biology, genomic biology, developmental biology, and cancer research because of its high efficiency and simplicity (Hsu *et al.* 2014). However, the existing CRISPR/Cas9 system does not sufficiently meet the requirement for gene therapy-based treatment of human genetic diseases caused by point mutations. Advances and developments in base-editing tools show great potential for maintaining the power of genome editing, while reducing its uncertainty and risk. By fusing nCas9 with cytidine deaminases, scientists have used BE3 to generate site-specific base conversions from C:G base pairs to T:A base pairs without inducing DSBs, while effectively avoiding indels. In addition, a series of subsequent reports described improvements in the performance of BE3 in generating nonsense mutations (Billon *et al.* 2017; Kim *et al.* 2017b; Kuscu *et al.* 2017).

Recently, BE3 was confirmed to function stably and efficiently in a variety of organisms. The efficiencies were up to 74.9%, 43.48%, and 28% in mammalian cells (Komor *et al.* 2016), plants (Zong *et al.* 2017) and zebrafish (Zhang *et al.* 2017), respectively. Here we report that BE3 can mediate the conversion of cytidine to thymine with an efficiency of up to 66.2% in *B. mori*, which to our knowledge, is the first report demonstrating BE3-mediated base editing in an invertebrate species. Previous reports have shown the successful use of BE3 as a knock-out tool
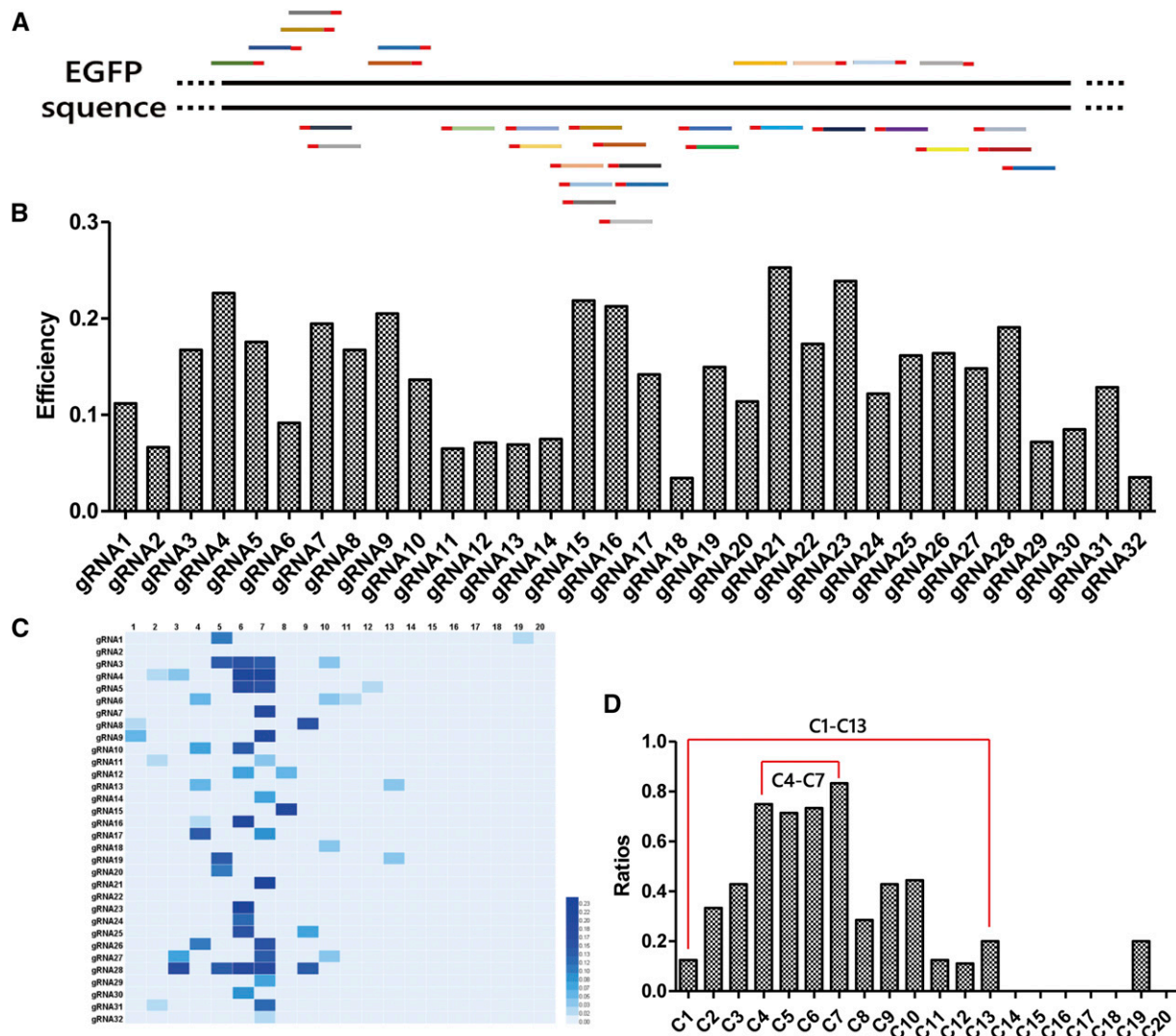
Figure 5 The efficiencies of 32 EGFP gRNAs and the editing window of BE3 in *B. mori*. (A) Schematic representation of 32 gRNAs targeting EGFP. (B) Bar graphs show the efficiencies of the 32 gRNAs designed to target EGFP. (C) The heatmap shows the efficiency of each base substitution in each gRNA. The darker the colors are, the more efficient the base substitutions were at the indicated position. (D) Bar graphs showing that the editing window of BE3 in *B. mori* ranges from C1 to ¹³C.

with high efficiency in mouse embryos and mammalian cells (Billon *et al.* 2017; Kim *et al.* 2017b; Kuscu *et al.* 2017). Here, we also used BE3 to efficiently inactivate exogenous and endogenous genes through substitution-induced nonsense mutations in *B. mori*. A recent report showed that the plant base editor (nCas9-PBE) could induce C-to-T substitutions with deamination windows covering seven bases from C3–C9 in the protospacers (Zong *et al.* 2017), whereas we detected a broader editing window with effective C substitutions spreading from C1 to C13.

Genome editing without generating DSBs is the biggest advantage of the base-editing system over Cas9. Such a capacity will guarantee the integrity of the genome to a much larger extent. In this study, we only detected a few indels generated by 32 gRNAs individually targeting EGFP with an average indel frequency of 2.2%, compared with a 1.2% indel frequency in control samples (which might have been induced by PCR or sequencing) (Table S3 in File S1). Previous data showed that the indel frequency for BE3 was also very low in mammalian cells and

plants (Komor *et al.* 2016; Zong *et al.* 2017), but that is different in *B. mori*. We suspect that the frequency of indels may be diverse because the efficiency of BE3 and the mechanisms of DNA-damage repair differ between various organisms. However, when we co-transfected all 32 gRNAs together with the BE3 vector, the indel frequency in the experimental group was almost the same as that in control cells. This result indicated that multiple bases can be edited simultaneously with almost no indels occurring around the target sites.

Using the CRISPR/Cas9 system and a pool of gRNAs, two groups (Katherine 2016; Garst *et al.* 2017) mapped the functional domains of proteins and promoters through error-prone NHEJ or correction-prone HDR. However, the unavoidable defects of both approaches were that indels were caused by both NHEJ and HDR, resulting in genome damage. Due to its DSB-free nature, BE3 is more useful for identifying alterations conferring a gain of function to modified proteins. Our findings showed that, with 32 gRNAs together, 14 C:G-to-T:A conversions occurred in EGFP with a indel frequency that was comparable to
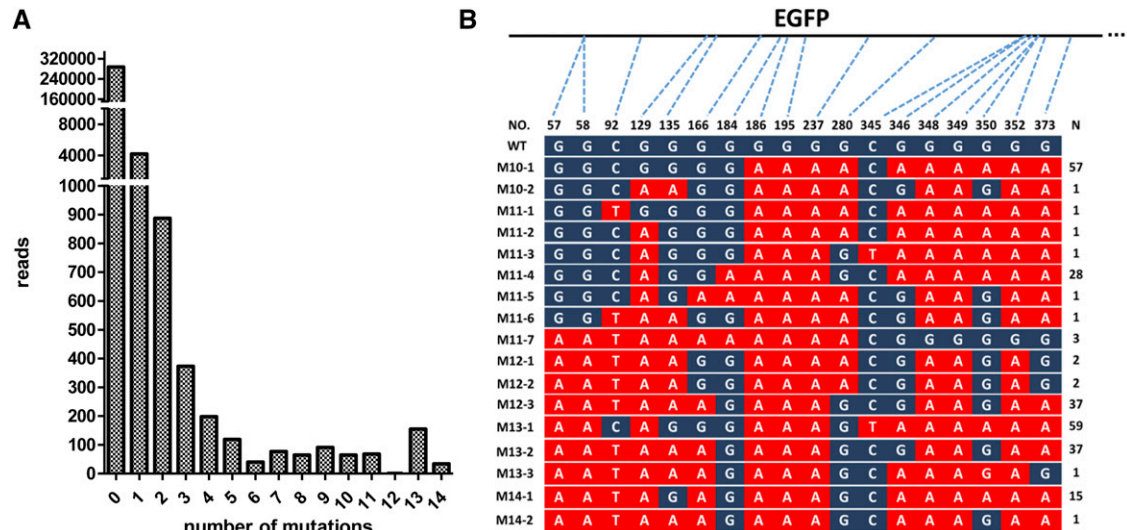
Figure 6 Multiple base substitutions generated by BE3. (A) Bar graphs showing the reads of multiple mutations after co-transfection of 32 gRNA vectors and the BE3 vector. (B) Schematic representation of 10 or more base substitutions in a single DNA read. The converted bases are marked in red. n1: number of base substitutions. n2: number of DNA reads with multiple base substitutions.

the control. Using the C:G base editor BE3 and the recently described A:T base editor ABE (Gaudelli *et al.* 2017), all four transition mutations can be programed with a library of gRNAs to enable mapping of thousands of parallel amino acids and promoter mutations simultaneously. Additionally, the limits associated with the PAM sites can be significantly relieved by the development of various Cas9 enzymes that recognize PAM sites other than NGG (Kleinstiver *et al.* 2015a; Kleinstiver *et al.* 2015b; Ran *et al.* 2015).

In summary, we demonstrated site-specific base editing by BE3 with high efficiency in an invertebrate organism, for the first time. This study may help in generating efficient base-editing systems in other organisms. The editing window, ranging from C1 to C13, is wider than that in plants (Zong *et al.* 2017), which provides greater potential and feasibility for modifying the genome at target sites with a larger scope. We also used BE3 to knock out exogenous and endogenous genes with high efficiency through substitution-induced nonsense mutations in *B. mori*. The frequency of introducing a stop codon was up to 66.2%, which was much more efficient than that found with SpCas9, SaCas9, and Cpf1 in *B. mori* cells (Liu *et al.* 2014; Ma *et al.* 2017). From this point of view, BE3 is more suitable for knock-out studies in invertebrates. Moreover, genome-wide bioinformatics analysis revealed 151,551 targetable knockout sites located in 96.5% of all *B. mori* genes, with approximately 11 sites per gene. BE3 provides an alternative for functional-genomics studies and other knock-out experiments and represents an improvement over Cas9 in terms of its ability to perform multiple base substitutions, without generating DSBs. Using 32 gRNAs simultaneously, we found up to 14 base mutations occurred in EGFP, with scarcely any indels. This result illustrates a novel strategy that using a pool of gRNAs and BE3 offers an enormous potential for mapping functional protein domains by generating diverse variants via BE3-mediated mutagenesis. Hence, BE3 is more applicable for identifying gain-of-function mutations in proteins. This is also good news because specific modifications via Cas9-mediated HDR have been difficult to achieve in many organisms. This new CRISPR tool for base editing should have a substantial impact on basic research and gene therapy for clinical treatments.

LITERATURE CITED

Billon, P., E. E. Bryant, S. A. Joseph, T. S. Nambiar, S. B. Hayward *et al.*, 2017 CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. Mol Cell 67: 1068–1079 e1064. https://doi.org/10.1016/j.molcel.2017.08.008

Chu, V. T., T. Weber, B. Wefers, W. Wurst, S. Sander *et al.*, 2015 Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. 33(5): 543–548. https://doi.org/10.1038/nbt.3198

Garst, A. D., M. C. Bassalo, G. Pines, S. A. Lynch, A. L. Halweg-Edwards *et al.*, 2017 Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat. Biotechnol. 35(1): 48–55. https://doi.org/10.1038/nbt.3718

Gaudelli, N. M., A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran *et al.*, 2017 Programmable base editing of A∗T to G∗C in genomic DNA without DNA cleavage. Nature 551(7681): 464–471. https://doi.org/10.1038/nature24644

Goldsmith, M. R., T. Shimada, and H. Abe, 2005 The genetics and genomics of the silkworm, Bombyx mori. Annu. Rev. Entomol. 50(1): 71–100. https://doi.org/10.1146/annurev.ento.50.071803.130456

Hsu, P. D., E. S. Lander, and F. Zhang, 2014 Development and applications of CRISPR-Cas9 for genome engineering. Cell 157(6): 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010

Katherine, M. H., Meagan Sullender, Emma W Vaimberg, Cory M Johannessen, David E Root, John G Doench, 2016 Mapping protein function with CRISPR/Cas9-mediated mutagenesis. bioRxiv.

Kim, D., K. Lim, S. T. Kim, S. H. Yoon, K. Kim *et al.*, 2017a Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nat. Biotechnol. 35(5): 475–480. https://doi.org/10.1038/nbt.3852

Kim, K., S. M. Ryu, S. T. Kim, G. Baek, D. Kim *et al.*, 2017b Highly efficient RNA-guided base editing in mouse embryos. Nat. Biotechnol. 35(5): 435–437. https://doi.org/10.1038/nbt.3816

Kim, Y. B., A. C. Komor, J. M. Levy, M. S. Packer, K. T. Zhao *et al.*, 2017c Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. 35(4): 371–376. https://doi.org/10.1038/nbt.3803

Kleinstiver, B. P., M. S. Prew, S. Q. Tsai, N. T. Nguyen, V. V. Topkar *et al.*, 2015a   Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. 33(12): 1293–1298. https://doi.org/10.1038/nbt.3404

Kleinstiver, B. P., M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen *et al.*, 2015b   Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523(7561): 481–485. https://doi.org/10.1038/nature14592

Komor, A. C., Y. B. Kim, M. S. Packer, J. A. Zuris, and D. R. Liu, 2016   Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533(7603): 420–424. https://doi.org/10.1038/nature17946

Komor, A. C., K. T. Zhao, M. S. Packer, N. M. Gaudelli, A. L. Waterbury *et al.*, 2017   Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. Sci Adv 3: eaao4774. https://doi.org/10.1126/sciadv.aao4774

Kuscu, C., M. Parlak, T. Tufan, J. Yang, K. Szlachta *et al.*, 2017   CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. Nat. Methods 14(7): 710–712. https://doi.org/10.1038/nmeth.4327

Liu, Y., S. Ma, J. Chang, T. Zhang, X. Wang *et al.*, 2017   Tissue-specific genome editing of laminA/C in the posterior silk glands of Bombyx mori. J. Genet. Genomics 44(9): 451–459. https://doi.org/10.1016/j.jgg.2017.09.003

Liu, Y., S. Ma, X. Wang, J. Chang, J. Gao *et al.*, 2014   Highly efficient multiplex targeted mutagenesis and genomic structure variation in Bombyx mori cells using CRISPR/Cas9. Insect Biochem. Mol. Biol. 49: 35–42. https://doi.org/10.1016/j.ibmb.2014.03.010

Lu, Y., and J. K. Zhu, 2017   Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. Mol. Plant 10(3): 523–525. https://doi.org/10.1016/j.molp.2016.11.013

Ma, S., J. Chang, X. Wang, Y. Liu, J. Zhang *et al.*, 2014a   CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in Bombyx mori. Sci. Rep. 4(1): 4489. https://doi.org/10.1038/srep04489

Ma, S., Y. Liu, Y. Liu, J. Chang, T. Zhang *et al.*, 2017   An integrated CRISPR Bombyx mori genome editing system with improved efficiency and expanded target sites. Insect Biochem. Mol. Biol. 83: 13–20. https://doi.org/10.1016/j.ibmb.2017.02.003

Ma, S., R. Shi, X. Wang, Y. Liu, J. Chang *et al.*, 2014b   Genome editing of BmFib-H gene provides an empty Bombyx mori silk gland for a highly efficient bioreactor. Sci. Rep. 4(1): 6867. https://doi.org/10.1038/srep06867

Ma, S., X. Wang, Y. Liu, J. Gao, S. Zhang *et al.*, 2014c   Multiplex genomic structure variation mediated by TALEN and ssODN. BMC Genomics 15(1): 41. https://doi.org/10.1186/1471-2164-15-41

Ma, S., S. Zhang, F. Wang, Y. Liu, Y. Liu *et al.*, 2012   Highly efficient and specific genome editing in silkworm using custom TALENS. PLoS One 7(9): e45035. https://doi.org/10.1371/journal.pone.0045035

Ma, Y., J. Zhang, W. Yin, Z. Zhang, Y. Song *et al.*, 2016   Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat. Methods 13(12): 1029–1035. https://doi.org/10.1038/nmeth.4027

Maruyama, T., S. K. Dougan, M. C. Truttmann, A. M. Bilate, J. R. Ingram *et al.*, 2015   Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. 33(5): 538–542. https://doi.org/10.1038/nbt.3190

Nishida, K., T. Arazoe, N. Yachie, S. Banno, M. Kakimoto *et al.*, 2016   Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353 aaf8729. https://doi.org/10.1126/science.aaf8729

Paquet, D., D. Kwart, A. Chen, A. Sproul, S. Jacob *et al.*, 2016   Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature 533(7601): 125–129. https://doi.org/10.1038/nature17664

Ran, F. A., L. Cong, W. X. Yan, D. A. Scott, J. S. Gootenberg *et al.*, 2015   In vivo genome editing using Staphylococcus aureus Cas9. Nature 520(7546): 186–191. https://doi.org/10.1038/nature14299

Richardson, C. D., G. J. Ray, M. A. DeWitt, G. L. Curie, and J. E. Corn, 2016   Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat. Biotechnol. 34(3): 339–344. https://doi.org/10.1038/nbt.3481

Sakuma, T., S. Nakade, Y. Sakane, K. T. Suzuki, and T. Yamamoto, 2016   MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. Nat. Protoc. 11(1): 118–133. https://doi.org/10.1038/nprot.2015.140

Sander, J. D., and J. K. Joung, 2014   CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. 32(4): 347–355. https://doi.org/10.1038/nbt.2842

Shen, J. P., and T. Ideker, 2017   Correcting CRISPR for copy number. Nat. Genet. 49(12): 1674–1675. https://doi.org/10.1038/ng.3994

Shimatani, Z., S. Kashojiya, M. Takayama, R. Terada, T. Arazoe *et al.*, 2017   Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat. Biotechnol. 35(5): 441–443. https://doi.org/10.1038/nbt.3833

Zhang, Y., W. Qin, X. Lu, J. Xu, H. Huang *et al.*, 2017   Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. Nat. Commun. 8(1): 118. https://doi.org/10.1038/s41467-017-00175-6

Zong, Y., Y. Wang, C. Li, R. Zhang, K. Chen *et al.*, 2017   Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat. Biotechnol. 35(5): 438–440. https://doi.org/10.1038/nbt.3811

Zuo, E., X. Huo, X. Yao, X. Hu, Y. Sun *et al.*, 2017   CRISPR/Cas9-mediated targeted chromosome elimination. Genome Biol. 18(1): 224. https://doi.org/10.1186/s13059-017-1354-4

Communicating editor: B. Andrews