REVIEW ARTICLE

Beyond CRISPR: single base editors for human health and crop improvement

P. K. Gupta

Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut 250 004, India

During 2016–2018, CRISPR/Cas9 technology was modified using disabled Cas9 with nickase activity in combination with cytosine/adenine deaminases for the development of four generations of cytosine base editors (BE1–BE4) for C → U conversion and at least seven generations of adenine base editors (ABE1–ABE7) for A → I conversion. These base editors exhibited improved efficiency and reduced frequency of deletions among the products. Further improvement in the form of enhanced base editors and high-fidelity base editors was achieved through the use of 1–3 copies of uracil N-glycosylase inhibitors and phage Mu-Gam protein. The technology will bring precision to gene editing technology for human healthcare and crop improvement.

Keywords: AID/APOBEC, base editing, CRISPR/Cas9, cytidine/adenine deaminases, target AID.

THE last five years have witnessed enormous excitement with the discovery of genome editing approaches involving sequence-specific designer nucleases (ZFN, TALEN, CRISPR/Cas), which create double strand breaks (DSB) in DNA (for a review, see ref. 1). However, of the three nucleases, CRISPR/Cas9 attracted the maximum attention for developing several plant and animal products with desired genetic modifications through genome editing. Soon an alternative for Cas9 in the form of Cpf1 became available paving way to a superior system in the form of CRISPR/Cpf1, which has several advantages over CRISPR/Cas[2,3].

ZFN/TALEN/CRISPR-mediated genome editing has been an approach that is preferred over transgenics, since no foreign gene is being introduced, and only an existing gene is altered, using cell's own machinery. Therefore, it has been argued that products of genome editing technologies like CRISPR/Cas9 should not be subjected to the regulatory system, which is typically used in case of genetically modified organisms (GMOs). This has made commercialization of genome-edited products easier at least in some countries[4]. As an example, a strain of 'mushroom' with white buttons, which will not turn brown (when stored) was developed using CRISPR and commercialized in USA without being subjected to regulations that are commonly applied to GMOs[5,6]. In these edited mushrooms, a gene for PPO (polyphenol oxidase) that causes browning of mushrooms was altered, thus reducing the quantity of PPO to 30%. A mutant waxy corn that gave higher yield under drought conditions has also been developed through genome editing by DuPont; this genome-edited waxy corn was also approved in USA for commercial cultivation, and may become available to the farmers for commercial cultivation within the next few years[7].

The popular CRISPR/Cas9 system for genome editing involves use of the nuclease Cas9 for creating a double strand break (DSB) at the target site, which is recognized with the help of an associated synthetic guide-RNA (sgRNA), which is programmable and is designed using the target sequence that is intended to be edited. The sgRNA consists of a scaffold sequence, which facilitates DNA binding of Cas9, and a ~20 nucleotide protospacer that is complementary to the sequence to be edited. This 20-nucleotide protospacer needs to be so designed that it should lie upstream of a characteristic PAM (protospacer adjacent motif), which differs in different microbes. The canonical PAM is 5′-NGG-3′ (Figure 1), which is associated with the endogenous Cas9 nuclease of *Streptococcus pyogenes*, so that the corresponding Cas9 is designated as SpCas9. In the absence of PAM, genome editing may not take place.

Genome editing due to CRISPR/Cas9 is known to have low efficiency, since NHEJ competes with the preferred HDR-dependent genome editing, and creates high frequency of indels and off-site alterations during genome editing. Also, genome editing does not allow an alteration of a specific existing base pair in a DNA or

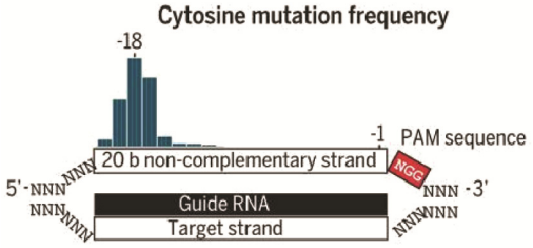

Figure 1. The occurrence of PAM sequence on the non-target strand 18 bases downstream of the target site for base alteration; the presence of downstream PAM is essential for base editing to occur in a range of few bases near 18th position upstream of PAM sequence.

e-mail: pkgupta36@gmail.com

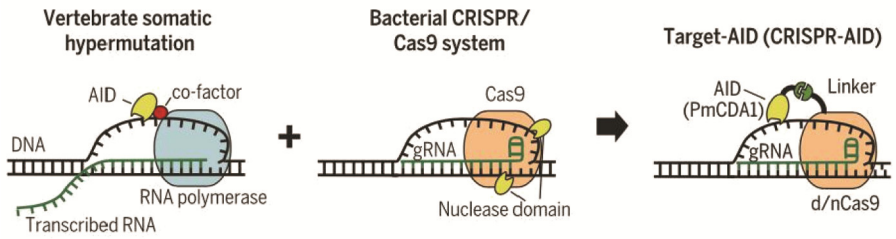

Figure 2. Use of vertebrate AID system with nuclease deficient CRISPR/Cas for the development of a hybrid 'Target-AID' complex that was later used for development of a variety of base editors, which create site specific single base C → T mutations with great precision[8]. AID, Activator induced deaminase.

RNA molecule in a predictable manner. In actual practice, it has been noticed that for a variety of products, selection needs to be exercised to obtain the desired product, which is generally available at a frequency of not more than 5%. CRISPR-Cas9 also introduces random insertions, deletions, translocations and other base-to-base conversions, which is another limitation associated with CRISPR/Cas9 system.

Keeping the above limitations of CRISPR/Cas in view, a hybrid system has been developed by combining the nuclease deficient bacterial CRISPR/Cas system with the vertebrate AID (activation-induced deaminase) system that causes deamination of cytidine and protects the vertebrate cells from invaders by causing alterations in the genome of these invaders (like prokaryotic CRISPR/Cas). The vertebrate AID system also causes hypermutation (C → U = C → T) in the variable region of the immunoglobulin locus that produces antibodies, thus generating antibody diversity in vertebrates. In the hybrid system, CRISPR/Cas9 could generate a suitable target for cytidine deamination by AID. The system was described as Target-AID, CRISPR-AID or 'CRISPR Nickase System' (Figure 2) and was first used in yeast, demonstrating that this can really be used for predictable gene editing in a eukaryotic system[8,9].

Starting in 2016, the above hybrid 'Target-AID' or 'CRISPR Nickase System' of cytosine deaminases was extensively used for developing a large number of plasmid vectors[10], which were described as base editors (BEs). These BEs were successfully used for generating precise CG → TA mutations in mouse/*Xenopus*/human living cells and embryos (Figure 3). In the case of mouse and *Xenopus*, the modified embryos were also used for transplantation in pseudopregnant surrogate mothers for generating offspring exhibiting altered phenotypes (e.g., black to albino body colour). In plant systems also, protoplasts/calli were successfully used for creating mutations in specific genes, and then regenerated into whole plants exhibiting desirable altered traits. The progress made in the development and use of BEs till early months of 2017 was also summarized in five articles that appeared in a special issue of *Nature Biotechnology* in May 2017 (see later). These five papers included papers on base editing in mouse embryos as well as in crops including rice, wheat, maize and tomato. The work on C → T base editing was followed by two important reports that were simultaneously published in October 2017. In one report, DNA adenine deaminases, which did not occur in nature were developed in the laboratory using naturally occurring RNA adenine deaminases (used for tRNA processing), so that a series of adenine base editors (ABEs) became available for conversion of adenine into inosine in a DNA molecule. Thus it also became possible to replace A : T base pair by G : C (A : T → I : T → G : C; Figure 3)[11]. In the second report, a technology for base editing of RNA transcripts was described[12,13]. In this manner, CRISPR/Cas was modified and used for editing DNA/RNA molecules in the form of BEs and ABEs. This brought many-fold efficiency and precision to gene-editing technology.

The technique of CRISPR/Cas9 was pioneered independently by Fang Zhang of Massachusetts Institute of Technology (MIT) in USA and by a team led by Jennifer Doudna of UC Berkeley (USA) and Emmanuelle Charpentier of Max Planck Institute for Infection Biology in Germany. Using this technology, the pioneering research of single base editing was however initiated and nurtured by a team of workers led by David Liu (Figure 4) of Harvard University in USA. During 2016–2017, the BEs brought about a revolution in gene-editing technology, so that David Liu was listed as the first of ten people who mattered in 2017 according to Nature10 (ref. 14). Several reports on BEs and their use for human health and crop improvement (particularly for rice), also appeared in 2018, suggesting that BEs will be increasingly used in future both for health care and crop improvement. In the present article, an attempt has been made to describe this new approach of single base editing and to review the progress made in this subject during 2016–2018. The author hopes that the article will raise awareness among readers about this innovative and exciting area of research.

Cytosine base editors (BE1, BE2, BE3, BE4)

For the purpose of precise C → U base editing, a number of plasmids called BEs were developed during 2016–2017. These plasmids facilitate base editing (a

'transition') involving conversion of cytosine into uracil (Figure 5 *a*), eventually leading to replacement of cytosine/guanine (C : G) base pair by thymine/adenine (T : A) base pair (Figure 3). Since these BEs were meant for alteration of cytosine only, these could be better named as cytosine base editors (CBEs) as against ABEs that were developed for A → I(G) (I = insone) conversion and were so named later in 2017.

The first-generation of C → U BEs was developed by attaching a cytidine deaminase AID/APOBEC1 to a nuclease deficient Cas (dCas) through a XTEN linker[10]; XTEN used in this study is a biodegradable unstructured peptide that increases the half-life of the fusion protein; AIDs/APOBECs used in this study represent a family of naturally occurring cytidine deaminases in vertebrates, providing protection against invaders and also providing antibody diversity (AID = activation induced deaminases; APOBEC = apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like). These deaminases make use of single-stranded DNA/RNA as a substrate[15]. Thus the original BEs consisted of the following two components: (i) a disabled Cas9 (dCas9) fused to a cytidine deaminase; (ii) a sgRNA for delivery of dCas9 with associated deaminase to the target locus carrying C within a window of 4–8 nucleotides and associated with a downstream PAM (protospacer adjacent motif) sequence available ~18–20 base pairs away. These first generation base editors (BE1) were further improved, so that a series of 2nd, 3rd and 4th generation BEs (BE2, BE3, BE4) became available (Table 1). In each case, high-throughput DNA sequencing (HTS) was used to quantify base editing efficiency. Digenome seq (sequencing of digested DNA) was also used for assessment of off-target effects in human cells[16].

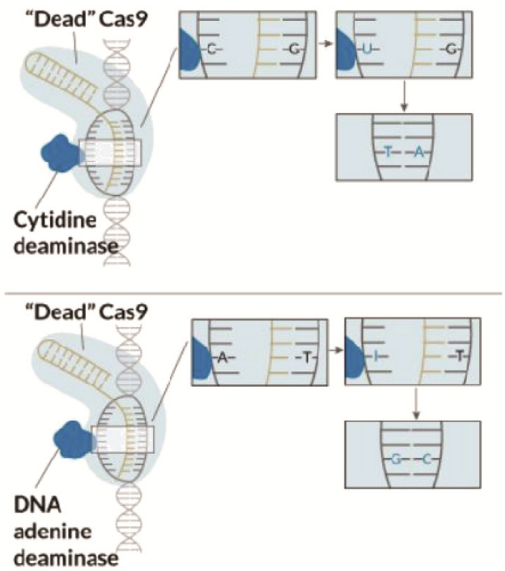

Figure 3. Coversion of C : G base pair into T : A base pair and that of A : T base pair into G : C base pair through single base editing[13].

Figure 4. David Lui, who pioneered base editing.

Improvement of BEs using Uracil N-Glycosylase Inhibitors

The major problem with the BE1 included the formation of undesired products due to the following two reasons: (i) frequent removal of uracil by cellular N-glycosylase (UNG) and (ii) possible occurrence of more than one cytosines (Cs) within the target window of 4–8 bases, which allows targeting of even off-site cytosines that were not intended to be edited. The enzyme UNG works during Base Excision Repair (BER) and therefore, treats transitional edited base pair U : G (derived from C : G) as a case of DNA damage, so that in the edited U : G base pair, U is excised, and U : G is not able to produce T : A. This leads to a failure of desired conversion of C : G into T : A. In order to improve *in vivo* editing efficiency, BE2 were developed, which carried a gene encoding uracil glycosylase inhibitor (UGI) fused with the gene encoding dCas9, so that the enzyme UNG is inhibited and is not able to excise U from the U : G base pair, which is intended to be converted to T : A during DNA replication. These BE2 had an improved editing efficiency (reaching up to ~20%) and the formation of indels was reduced significantly (<0.1%) over that obtained in CRISPR-mediated genome editing. The second problem of the occurrence of more than one cytosines (Cs) in the editing window was partly resolved by reducing the size of editing window from 4–8 base pairs to 1 or 2 base pairs (see later).

The next stage for improvement of BEs was achieved by converting dCas9 to a nickase (then designated as nCas9) through replacement of either amino acid aspartate (D) by alanine (A) at position 10 (D10A) or replacement of amino acid histidine (H) by alanine at position 840 (H840A). The modified forms of Cas9 in the form of nCas9 produce nicks in opposite strands and have been suitably utilized in single base gene editing. For instance D10A mutant of Cas9 retains a domain that generates a single strand DNA nick in the non-target strand instead of creating double strand breaks at the desired site;

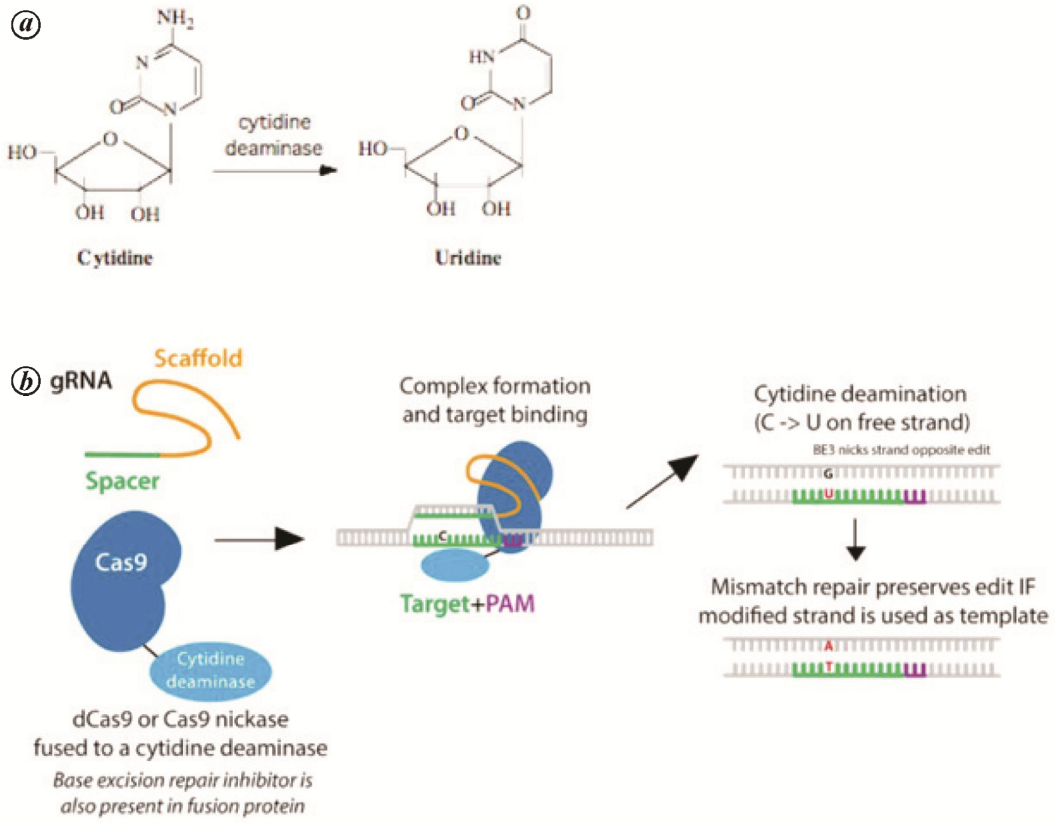

Figure 5. ***a***, Conversion of cytidine into uridine using cytidine deaminase; ***b***, An improved method of cytosine base editing, where cytosine deamination occurs in one strand with no DSBs, and nCas nicks the opposite strand helping mismatch repair, thus retaining the edited base[36].

Table 1. Four generations of base editors (BE1–BE4) that convert cytosine to uracil

Base Editors	Characteristics	Reference
BE1	Cas9 altered to dCas9; rat APOBEC* cytidine deaminase to convert C to U without cutting DNA	10
BE2	BER inhibitor UGI fused to dCas9, raising editing efficiency three-fold to a maximum of ~20%	10
BE3	dCas9 altered to a nCas9 to simulate mismatch repair. BE3 nicks the unmodified DNA strand so that it appears 'newly synthesized' to the cell; the cell repairs the DNA using the U-containing strand as a template, copying the base edit	10
BE4	2nd copy of UGI increased product purity; decreased C → G or C → A products and also decreased indel formation	17
BE4-Gam	Adding Gam to BE4 or SaBE4 decreased indel frequency 1.5–2.0 fold	17

*APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) are a family of cytidine deaminase that prefer single-stranded nucleic acids as substrates.

this simulates 'mismatch repair', so that the unmodified opposite DNA strand mimics a DNA strand undergoing synthesis, where the strand containing the edited base is used as a template (C → U), taking U as T (Figure 5 *b*). Therefore, BE3 had the following three components in addition to sgRNA, which guides the editor to the target site: (i) an AID/APOBEC1 deaminase, that was fused to a (ii) nickase (nCas9) that was deficient for nuclease activity [nCas9(D10A)], and (iii) a UGI that was linked to nCas9 through a 4 amino acid linker.

The importance of UGI in base editing was demonstrated by showing that the UGI-deleted BE3 (BE3-ΔUGI) was less competent in base editing compared to original BE3, and produced not only lower frequency of desired C → T editing, but also produced a higher frequency of unwanted indels. A number of improved BE3 variants were also developed (Table 2), which resulted in much more efficient conversion of the G : U intermediate to desired A : U and A : T products[17].

Another problem associated with BE1 and BE2 was the occurrence of more than one Cs within the base-editing window, so that the cytosine deaminase converts even a non-target C into U. This problem was overcome by the development of a number of BE3 variants with

modified SpCas9, where even the non-NGG PAM sequence could be used for base editing (Table 2; also see later).

It was also shown that an addition of another copy of UGI to BE3 further reduced the frequency of indels, so that BEs with more than one UGI were developed and were described as BE4, which were found to be more efficient[18]. BE4 or SaBE4 were further improved by adding Gam to the cassette, so that the use of BE4-Gam resulted in a further 1.5–2.0 fold decrease in indel frequency (Table 1).

Two limitations of gene editing: PAM site and size of targeting window

As mentioned earlier, the occurrence of a downstream protospacer adjacent motif (PAM) sequence in association with target sequence is necessary for dCas9 or nCas9 to bind (with the help of gRNA) and cleave the target DNA sequence for base editing. In the absence of PAM, targeted base editing will not take place. However, the PAM sequence associated with Cas9 varies across bacterial species, although the canonical PAM sequence 5′-NGG-3′ (Figure 1), is associated with SpCas9 nuclease from *Streptococcus pyogenes*. This limits the number of target genome sites, which can be efficiently targeted by BE3 and its variants, because many sites would lack NGG PAM at the desired location. Therefore, attempts have been made to improve BE3 variants through engineering Cas9, enabling CRISPR/Cas9 mediated gene editing at any desired genome location (using non-canonical-PAM); Cas9 was further modified to reduce the editing window to 1–2 nucleotides through introducing appropriate mutations, so that off-site editing could be avoided (Table 2). For instance, SpCas9 was replaced by SaCas9 (from *Staphylococcus aureus*)[19], which is considerably smaller and is associated with NNGRRT PAM; SaCas9 has been successfully used in several studies. The editing efficiency in several of these BE3 and BE4 editors was as high as ~50–100%.

Table 2. BE3 variants, developed with different Cas9 variants (including those for non-canonical PAM)[25,39]

BE3 Plasmid name	Cas9 (PAM)
pJL-SaBE3	SaCas9 (NNGRRT)
pJL-SaKKH-BE3	SaCas9 (NNNRRT)
pBK-VQR-BE3	VQR-Cas9 (NGA)
pBK-EQR-BE3	EQR-Cas9 (NGAG)
pBK-VRER-BE3	VRER-Cas9 (NGCG)
pBK-YE1-BE3	SpCas9 (NGG)
pBK-EE-BE3	SpCas9 (NGG)
pBK-YE2-BE3	SpCas9 (NGG)
pBK-YEE-BE3	SpCas9 (NGG)
pET42-HF-BE3	HF-Cas9
pCMV-HF-BF3	HF-Cas9

Earlier studies have established SpCas9 variants, which can function with a broader range of PAMs. In one report[20] a variant named SpCas9-NG was used, which was shown to mediate the C-to-T conversion at target sites with NG (rather than NGG) PAMs in human cells. In the other report[21], a technique called 'phage assisted continuous evolution' (PACE) was used to obtain a SpCas9 variant called xCas9, which could recognize a very wide range of PAM sequences including NG, GAA and GAT. The xCas9 variant was also shown to exhibit a very low off-target activity at all NGG and non-NGG PAM targets with a higher level of DNA specificity thus overcoming the restriction due to PAM compatibility.

Enhanced BEs (eBE-S1, eBE-S3) and high fidelity BEs (HF-BE)

Two additional classes of improved BEs included enhanced base editors (eBEs) and high fidelity BEs (HF-BEs). The enhanced BE (eBE) involved designing of the following two systems, and their expression was tested in FT292 and HeLa cells: (i) In one system, three independent vectors were used, one carrying sgRNA, the second carrying BE3 and the third carrying UGI gene (Figure 6 *a*); this would allow co-expression of BE3 and sgRNA with free UGI in trans; expression level of UGI was also manipulated to study the relationship between the level of UGI and precision in base editing; it was observed that the ratio of the frequency of C → T conversion to indel frequency was negatively correlated and that at high level of UGI, there was six fold increase in the proportion of C → T conversion relative to indels; unwanted C → A and C → G substitutions relative to C → T conversion also declined. (ii) In the other system, sgRNA was allowed to coexpress with BE3 carrying one (eBE-S1) or three (eBE-S3) additional copies of 2A-UGI sequence (Figure 6 *b*). In one study, five different sgRNAs were independently used for targeting five different loci in FT293 cells; both enhanced BE3s (eBE3) exhibited a decline in indel frequency and an improvement in the frequency of targeted C → T, when compared with the frequency obtained with original BE3 carrying only one copy of UGI[18] (Figure 6 *a*). It has been recommended that the approach should be combined with altered PAM and nCas9 to get a very high level of precision[18]. The approach has since been deployed in a variety of organisms, ranging from cereals (rice wheat and maize) to tomato among crops and zebrafish to mice among model animal systems.

Fusion of Mu Gam protein for improved base editors

Further improvement in BE4 was achieved through fusion of Gam protein derived from Phage Mu. The Gam

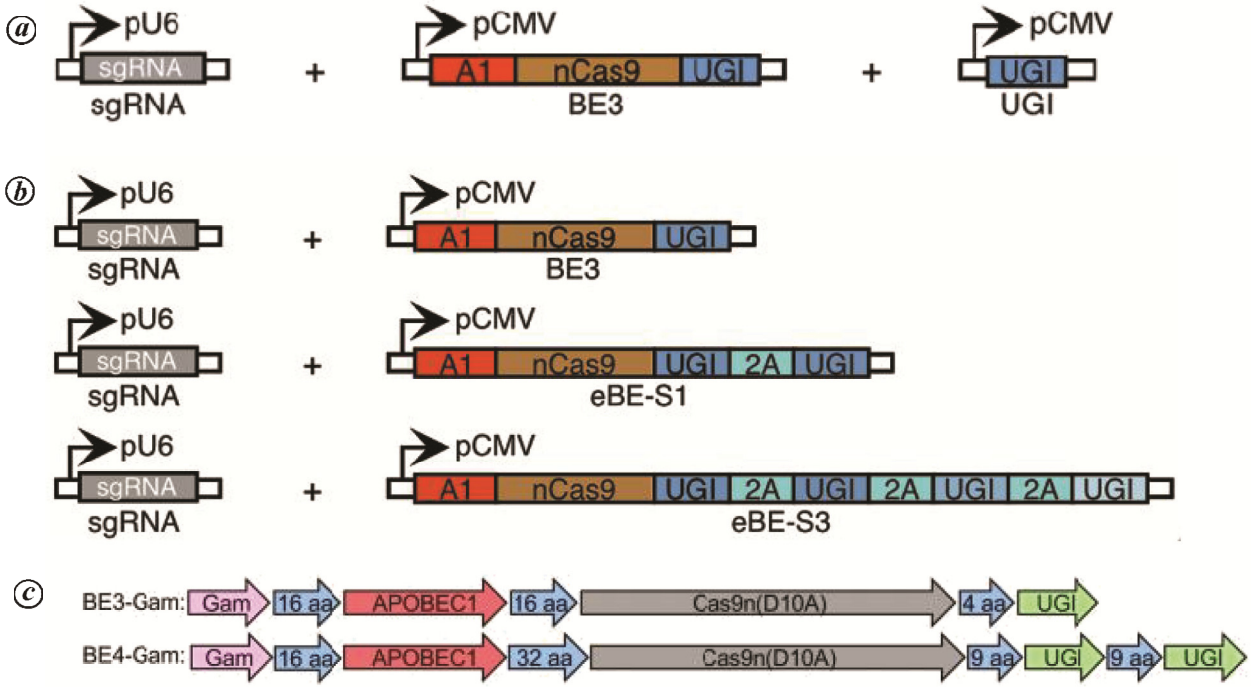

Figure 6. Composition of enhanced base editors (BEs): ***a***, sgRNA with BE3 and free UGI; ***b***, BE3 and two versions of enhanced BE3 (eBE-S1 and eBE-S3)[39]. ***c***, Components of plasmids BE3-GAM (with one copy of UGI) and BE4-Gam (with two copies of UGI)[16].

protein fusion products in the form of BE3-Gam and BE4-Gam vectors help in binding with DSBs leading to reduction in the frequency of indels and an improvement in product purity. Thus fusion of Mu Gam protein to BE3 and BE4 gave one of the best base editors for conversion of C : G to T : A[15] (Figure 6 *c*).

High fidelity base editors

High-fidelity base editors (HF-BE) represents another class of BEs, which contained high fidelity Cas9 variants (HF-Cas9) including Cas9-HF1 with four mutations (N497A, R661A, Q695A and Q926A) and Cas9-HF2 with one additional mutation (D1135A), thus conferring higher level of specificity[22,23] (Figure 7). High fidelity HF-BE2 and HF-BE3 were also developed and utilized for base-editing in *tyrosinase* gene (*tyr*) in mouse embryos to examine the efficiency of these BEs. It was observed that the desired base editing occurred and off-target activity was low, although proximal off-target deamination was also observed, up to 30–40 bp away from the gRNA binding site and also away from the PAM sequence. The base edited mice were also obtained through transplantation of base edited embryos in surrogate mothers, suggesting no embryonic toxicity during base editing. The mutant albino pups were obtained in a frequency of 18.2% for gRNA1 and 63.6% for gRNA2[24]. Often, with HF-BEs, C → T conversion in mouse embryos occurred with 100% efficiency, associated with 37-fold reduction in off-target editing in case of HF-BE3[24].

Delivery of base editors as RNPs

Recently it was argued that DNA delivery method using plasmid vectors results in uncontrolled Cas9 and sgRNA expression even after the on-target locus has been edited, thus providing opportunity for off-target editing[25]. Therefore, delivery of exogenous DNA plasmid vectors was replaced by DNA-free lipid mediated delivery of ribonucleoproteins (RNPs) prepared through association of BE proteins with gRNA. Delivery of RNPs when tried in mice inner ear and zebrafish embryos exhibited much higher target specificity and reduced off-target editing[25].

Adenine base editors

Initially (in 2016 and early 2017), base editing was restricted to C → T conversion involving the use of naturally occurring DNA cytosine deaminase. However, for A → G conversion, no DNA adenine deaminase occurred in nature, although RNA adenine deaminases occurred for modification of tRNAs. Therefore, synthetic DNA adenine deaminases had to be developed from RNA adenine deaminase through protein engineering using directed evolution. Once DNA adenine deaminase was successfully

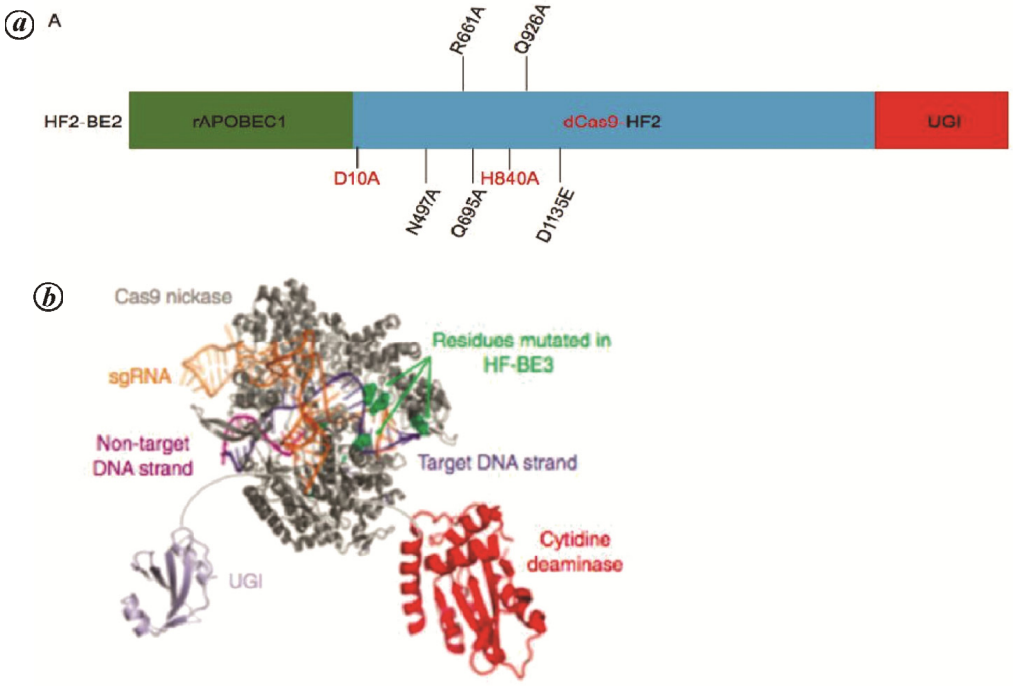

Figure 7. High fidelity base editors: ***a***, Structure of HF2-BE2 carrying dCas9HF2 with seven mutations, two mutations (in red) for disabling the nuclease activity and five mutations (in black) for converting dCas9 into dCas9-HF2 (ref. 22). ***b***, A representation of different components HF-BE3 (HF-Cas9 nickase) with UGI, a cytidine deaminase and sgRNA) bound to target and non-target strands of DNA[19].

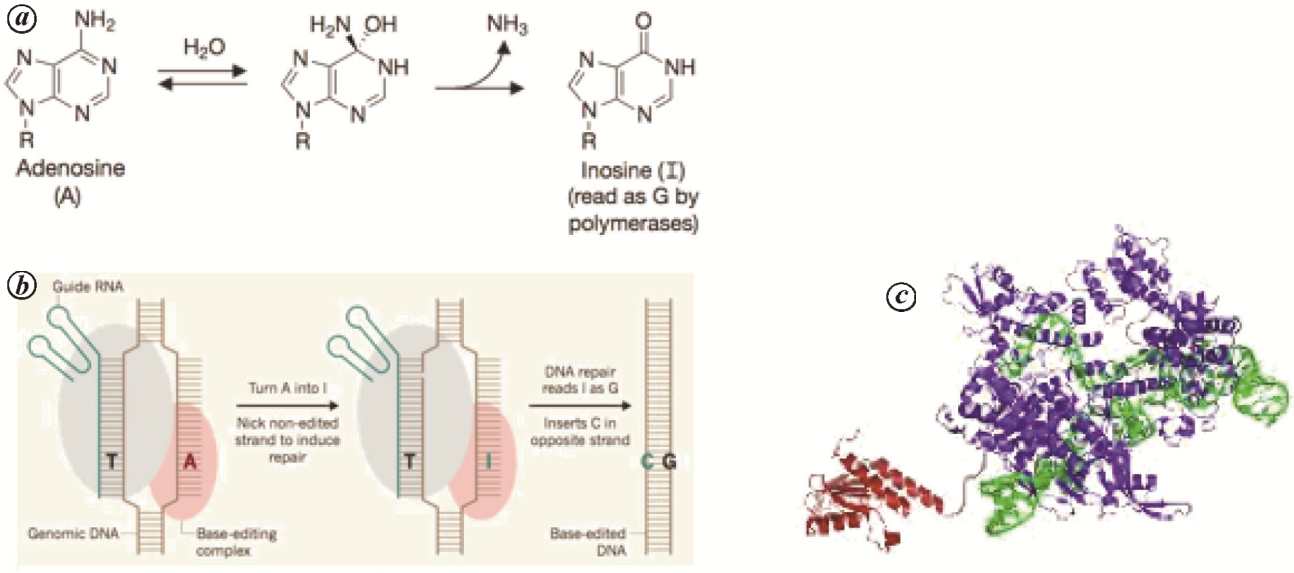

Figure 8. Adenine base editotor (ABE): ***a***, Conversion of adenine into inosine using synthetic adenine deaminase; ***b***, Conversion of T : A base pair into C : G base pair using ABE; ***c***, ABE showing different components including an adenine deaminase (red), gRNA (green), and Cas9 nickase (violet)[1].

synthesized in the laboratory, ABEs could be developed for conversion of adenine into inosine (I); the latter mimics guanine (G) during DNA replication, so that T : A base pair could be edited into C : G base pair (Figure 8 *a* and *b*). The detailed structure of an ABE is depicted in Figure 8 *c*.

Several generations of ABEs were developed, so that ultimately the seventh generation improved ABEs included the following four ABEs: ABE 7.3, ABE 7.8, ABE 7.9 and ABE 7.10 (ref. 11). Of these, ABE7.10 was the most active editor, with high level of editing efficiency (53%). The editing window, however, was still 4–7 in

ABE 7.10 and 4–9 in the other three seventh generation ABEs. These ABEs did not display any significant A to non-G conversion at target loci, because the removal of inosine through BSE from DNA is not as common as that of uracil (U) due to the presence of UNG. ABEs also performed better than many BEs in terms of off-target editing and frequency of indels produced during editing. In an actual study, ABE 7.10 modified only 4/12 off-targets with a frequency of 1.2% indels, compared to 9/12 known off-targets with a frequency of 14% indels in other BEs.

Base editing in RNA and 'REPAIR' technology

In October 2017, along with the report of development of ABEs by the group led by David Liu[11], another research paper was published, which reported the development of BEs for editing RNA transcripts. These RNA base editors were developed by the group led by Feng Zhang (Broad Institute), who has also been visible in recent years due to CRISPR patent battle[26]. In most studies on base editing, target site occurs in genomic DNA, but this technology has been limited by the requirement of canonical PAM-NGG site at the target locus, although BEs for sites with non-canonical PAM were also developed (discussed earlier). Keeping this in view, RNA base editors were developed using the programmable type VI CRISPR-associated RNA-guided RNase Cas13b; the technology has been described as REPAIR [RNA editing for programmable A to I (G) replacement][12]. The naturally occurring ADAR (adenosine deaminase acting on RNA) was used with disabled Cas13 (dCas13) and guide RNA (gRNA) for endogenous A → I editing at specific transcript targets in mammalian cells. The specificity of dCas13b was further improved through mutagenesis to generate REPAIRv1 and then REPAIRv2, which had very high level of specificity. Hopefully, REPAIRv2 or other improved forms of REPAIR will be utilized in future for a variety of purposes, particularly for transient alteration in the transcripts, where a permanent alteration of the genome is not required/desired (e.g., temporary relief from a disease).

Applications of base editing

A variety of BEs and ABEs (described as molecular machines) have the potential of being used for a wide variety of applications including human healthcare and crop improvement. Some of the initial studies conducted during 2016–2018 demonstrated this potential. Using model systems and all major crops, it has already been shown that specific genes can be modified by single base alterations using BEs. However another possibility both in healthcare and crop improvement is promoter engineering involving alteration in a single base, because promoter engineering can be used to change a constitutive promoter into a promoter that may cause temporal and/or spatial regulation of gene expression in a desirable manner. Some details of the studies already conducted are summarized in Tables 3 and 4. Other possible uses are described in the following sections.

Application of base editing in human health

The base editors including BEs and ABEs can correct each of the following four 'transition' mutations; C → T, T → C, A → G, or G → A, which together account for almost two-thirds of all disease causing point mutations in human genome. There must be another proportion of disease causing mutations, which are due to transversions, which include the following mutations C → A, C → G, T → A, T → G. The available BEs do not cover all these mutations; but more BEs will be developed in future to cover all diseases caused due to single base alterations. Thus a large number of diseases (e.g. sickle cell anemia, genetic blindness, cystic fibrosis and several neurological and metabolic disorders) are associated with single nucleotide polymorphisms (SNPs) and no treatment is currently available for most of these diseases, suggesting that editing at single base level may ameliorate patients from a number of diseases. The base editors can also help in the future development of gene-therapy approaches[11]. Additional research is, however, needed to enable BEs and ABEs to target as much of the genome as possible.

The ability of BEs for correction of single base genetic defects has been demonstrated by using several model systems including mouse, rat, zebrafish and *Xenopus*. Efficiency and effectiveness of these BEs has also been demonstrated using human cell lines and zygotes/embryos (Table 3). For instance, in a seminal recent study, using mouse/human cells, it was shown that the mutations (C → T or G → A) associated with Alzheimer's (APOE4, Cys158Arg mutation: potent Alzheimer's risk factor), and some cancers (p53 Tyr163Cys mutation, associated with cancer) can be treated through base editing.

Several reports are available, where corrections of individual genes have been demonstrated even at the organismal levels in model systems like mouse, rat, zebrafish and *Xenopus*, where altered embryos were transplanted in pseudopregnant surrogate mothers and mutant offspring obtained[24]. The gene *tyr* encoding tyrosinase enzyme causing albinism has been successfully used in several of these studies.

Although the pioneering work on BEs was undertaken in USA, scientists in China worked more aggressively for clinical application of the base editing technology. In September 2017, a single base mutation in *HBB* gene causing β-Thalassemia, was corrected in human embryos using a base editor[27]. Similarly, in August 2018,

Table 3. Gene editing studies (C to U) in living cells, zygotes, embryos and organisms relevant to human health; vector was delivered generally through injection (rarely through electroporation); sometimes lipid-mediated delivery of RNPs was also tried[19]

Animal system	Gene edited	Vector used	Reference
Mouse: HEK293T	HEK293 site 3 and RNF2	BE1m BE2, BE3	18
Mouse: HEK293T	Number of loci	Five new BE3 editors	39
Mouse embryo and offsprings	*Dmd* or *tyr*	BE3 (rAPOBEC1-nCas9-UGI)	40
Mouse zygote, embryos and pups: HEK293 site 3 and RNF2 293T cells	*tyr*	HF2-BE2 (rAPOBEC1-XTEN-dCas9-HF2-UGI)	24
Xenopus laevis embryos	*trya, tryb* (tyrosinase) *tp53*	BE3, rAPOBEC1-nCas9-UGI	40
Humans Myeloid leukemia cells	BCR-ABL	dCas9-AIDX-UGI	41
HEK298 cells; mouse inner ear; zebrafish	EMX1, FANCF, HEK3	HF-BE3 (RNPs)	25
Humans: FT Cells, HeLa Cells		UGI + BE3	17
Human HAP1 cells; HAP1 UNG cells	EMX1, FANCF, HEK2, HEK3, HEK4, or RNF2 locus	BE4, SaBE4, BE4-Gam and SaBE4-Gam	11
Human cells HEK293T	6–17 genomic sites	ABE1s to ABE7s	12
Mammalian cells; HEK293FT	Cluc	Cas13-ADAR; REPAIR	12
HEK293FT cells	EMX1, FANCF, HEK2, HEK4, or RNF2 locus	eBE + UGI; eBE-S1, eBE-S3	18

tyr, tyrosinase (for melanin pigment for body colour (black versus white; mutantion causes albinism); *tp53*, a tumour suppressor gene, causing apoptosis; HAP1, a haploid human cell line; HEK, Human embryonic kidney cells; Cluc, Cypridina luciferase (a reporter gene).

Table 4. A summary of base editing studies in crop plants reported in 2017

Crop	Gene	Reference
Wheat	*TaLOX2*	38
Rice	*OsCDC48*, *OsNRT1.1B* *OsSPL14*	38
	NRT1.1B (nitrate transporter) *SLR1* (DELLA protein for plant height)	42
	OsPDS (phytoene desaturase; one site) *OsSBEIIb* (starch branching enzyme IIb; 2 sites)	43
	ALS (acetolactate synthase for herbicide resistance) *FTIP1e*	31
Maize	*ZmCENH3*	38
Tomato	*DELLA* (Solyc11g011260) *ETR1* (Solyc12g011330) Both genes regulate plant hormone signalling	31

NRT1.1B, N transporterl NUE; *SLR1*, DELLA = reduced plant height.

the results were published, where the defect responsible for causing Marfan syndrome (an autosomal dominant disorder involving a single base change, from a 'healthy' T to a 'pathogenic' C in the gene *FBN1*) was successfully corrected in viable human embryos[28]. Two research teams from East China Normal University and Sun Yat-Sen University in China also developed and improved the base editing system using mouse and rat models that mimic Dunchenne Muscular Dystrophy (DMD) and glycogen storage disease (GSD). Promising results were obtained in these efforts also[29].

In USA also, Beam Therapeutics, Inc. (a new startup) was recently started by David R. Liu, Feng Zhang, and J. Keith Joung, who developed the base editors for the first time. The company has been granted a worldwide license by Harvard University to utilize base editing technologies and commercialize the same for treatment of human diseases. This will allow rapid mobilization of resources to fully develop and exploit the technology in this field.

Application of base editing in crop improvement

Examples of successful base editing using protoplasts followed by regeneration of plants including rice, wheat, maize and tomato are also available in plants (Table 4). In most cases, a BE3 variant with nickase nCas9 fused with a cytidine deaminase and a UGI were used for base editing. Since delivery of template DNA can sometimes be a problem in plants, a target-AID was used as cytidine deaminase[30], and the fusion product was codon optimized for plants (cereals); these base editors were, therefore, described as plant base editor (PBE). In 2017, base editing was successfully attempted in one study from Japan and three studies from China. Altogether nine genes in rice, one gene in wheat, one gene in maize and two genes in tomato were targeted. Protoplasts or calli were used for Agrobacterium-mediated delivery for gene editing and the edited protoplasts or calli were regenerated to produce mutant plants, which exhibited the mutant phenotype. In most cases, the editing efficiency was much higher than earlier reported for CRISPR/Cas9 mediated editing, suggesting that base editors can be profitably utilized for crop improvement with much higher precision (see Table 4 for details). In some cases, the mutant plants produced through base editing have been described as transgenic plants[30,31], but we will avoid calling them transgenic, since no transgenes have been inserted in the genome, and only genes have been altered, making them equivalent to mutant plants, rather than transgenic plants or GMO crops. In view of this, these mutant plants should

not be subjected to the regulation that is used in case of transgenic (GM) crops, thus making it easier for these improved crops to reach the farmers.

During 2017 and 2018, a number of base editors for rice were also developed (in China) through codon optimization of genes encoding deaminases (APOBEC1) and glycosylase inhibitors (UGI). A series of pairs of these rice base editors were developed, which included the following: (i) rBE3 and rBE4: rat APOBEC1 gene encoding cytosine deaminase and phage UGI gene encoding UGI were codon optimized for rice; also the promoter sequences used included maize Ubi and CaMV35S[32]; (ii) rBE5 and rBE9: human hAID (cytosine deaminase) was optimized with priority for target sites with C in the following context: GC, AC, CC, TC rather than TC alone preferred by APOBEC1; hAID variant, hAID*Δ carrying several mutations with improved catalytic activity was also used; UGI fused to 3' end of eBE5 gave rBE9[33]; (iii) rBE14 and rBE15, carried ABE TadA and the optimized variant TadA*7 (ref. 34); (iv) rBE17, rBE18: A142N and P152R mutations were imported into Tad*7.10 to create TadA*7.8, which was used for developing rBE17 and rBE18 (ref. 34). A large number of rice genes (including *OsCERK1*, *OsCERK2*, *ipa1*, *Pi-ta* and *BRI1*) were successfully edited using these rice base editors (rBEs).

Base editors as cell recorders

A novel use of BEs as cell recorders was demonstrated in a recent report[35]. Two approaches for recording cell's history were described. In one approach, BEs can be used as recorder of a response of cells to different stimuli including biotic and abiotic stresses. The technique was described by the acronym CAMERA (**C**RISPR-mediated **A**nalog **M**ulti-event **R**ecording **A**pparatus). The stimuli and changes in the cell including those due to antibiotics, nutrients, viruses, light and changes in signaling events could be recorded with mere 10–100 cells, instead of large number of cells needed earlier for similar study.

Ethical and safety issues

Ethical and biosafety issues involving genetically modified organisms (GMO) have been the subject of discussion for the last almost three decades. Consequently GMOs have faced stringent regulations globally, thus restricting commercialization of many products that were developed. However, recently it has been argued that the products of gene editing following CRISPR technology should not be subjected to the same regulatory restrictions, which are used for GMOs. This argument has been accepted in Americas, so that the products of gene editing in Americas no longer require to undergo the regulations that are required for GMOs. Consequently, the issue has been examined globally, particularly in Americas (USA, Canada, Argentina) and Europe. Consequently, in USA these products no longer require the regulatory clearance like GMOs[36], but in Europe, the highest court recently decided that the gene-edited crops should be subject to the same stringent regulations that are used for GMOs[37]. It is not surprising, since Europe has always been conservative in dealing with the subject of the release of GMOs. One would expect that the products of base-editing (discussed in this article) would be treated like any other mutant product of conventional breeding, since only a single base is altered (as in base substitution mutants).

Summary and perspectives

The BEs were developed for the first time by David Liu and his coworkers starting in late 2016 followed by the development of a number of other base editors in 2017 and 2018. These base editors involved deamination of Cytosine (C) into Uracil (which mimics Thymine) and deamination of Adenine (A) into Inosine (I). Although naturally occurring cytosine deaminases could be initially used for converting C : G into T : A, but DNA adenine deaminases had to be developed in the laboratory through mutations in an *Escherichia coli* RNA adenine deaminase TadA (since no DNA adenine deaminases occurred in nature).

The above technique was found to be capable of editing individual bases in specific genes in a programmable manner with precision. The progress in this field has been revolutionary and phenomenal, where the utility of these BEs was demonstrated for human healthcare and crop improvement. Base editing technology has also been developed for editing RNA transcripts, so that a disease of a patient can be addressed for temporary relief without causing any permanent alteration in the genome. Some highlights of the development of major BEs are summarized in Table 5.

Base editing technology for editing DNA and RNA, both represent a modification of CRISPR/Cas9 system, which revolutionized genome editing during the last five years, after it was first proposed in 2012–2013. A comparison of CRISPR/Cas genome editing, DNA base editing and RNA base editing is depicted in Figure 9. In the DNA/RNA base editing technology, some of the limitations of CRISPR/Cas9 technology have been addressed, so that alteration in a gene is now possible at a single base level in a predictable manner both at the DNA level and RNA level. The technology has already been successfully used for base alterations in several animal and plant systems.

Although initially David Liu and his coworkers at Harvard (USA), developed the technique by mainly keeping in mind human healthcare, it was picked up quickly in China and Japan and used for demonstrating their utility for crop improvement. Successful base-editing was

Table 5. Highlights of the development of major base editor variants

Organism used	Base editors with main features	Reference
Human cells (HEK293)	BE1, BE2, BE3: Progressing improvement in editing efficiency; BE3 had highest efficiency, but also higher indel frequency relative to BE2	10
Yeast BY4741	Hybrid Target-AID. Edits 3–5 base window surrounding – 18 position upstream of the PAM	8
Mouce cell lines	SaB → 3, BE3 PAM variants, BE3 editing window variants Greatly expands the number of target loci for base editing	39
HEK298 cells	HF-BE3: HF-BE3 and ribonucleoprotein delivery decrease BE3 off-target activity	25
Human HAP1 cells; HAP1 UNG cells	BE4 and BE4-Gam; AID, CDA1 and APOBEC3G BE3 variants: A second copy of UGI improves product purity. Gam decreases indel frequency	17
Human cell lines	DNA Adenine base editors (ABE). A → I (A → G) editing with high product purity and low off-target editing	11
Mammalian cells	RNA editors (ABE: RNA guided dCas13, A → I (A → G); ADAR2; RNA transcripts in mammalian cells (REPAIR)	12

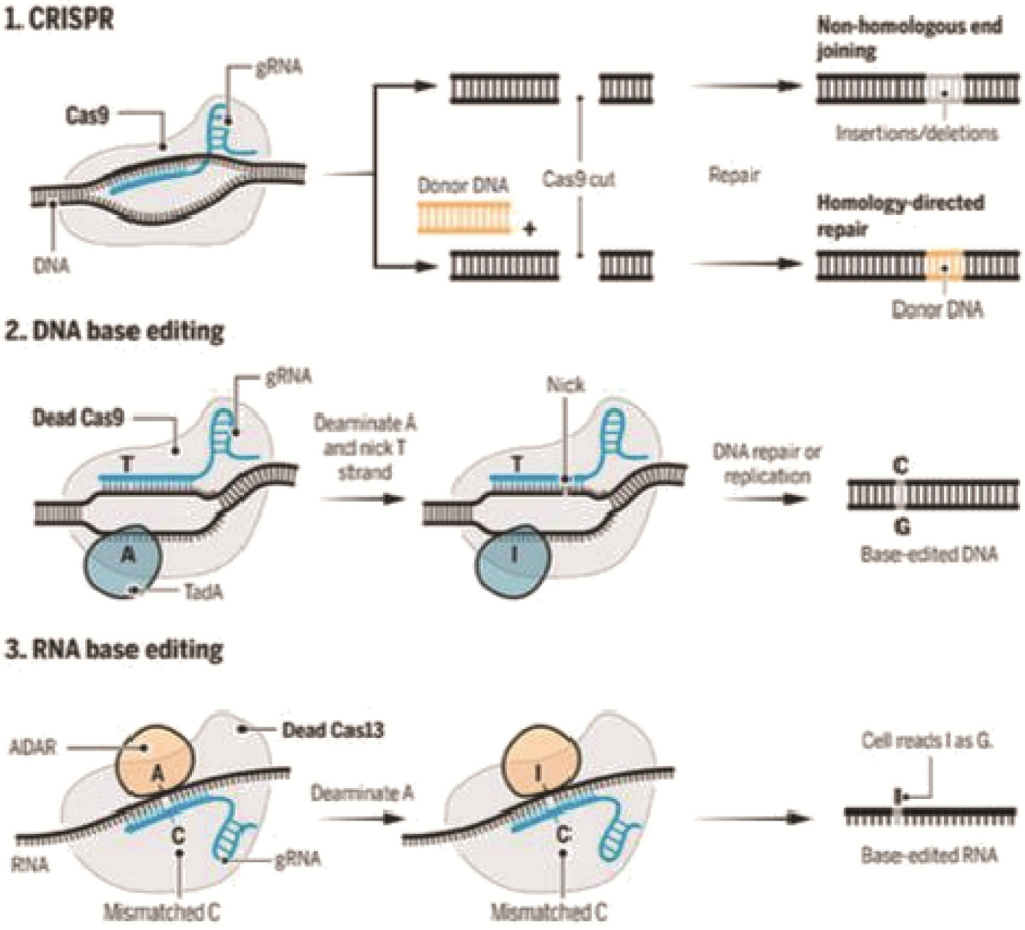

Figure 9. A comparison of three approaches for correcting mutations: CRISPR/Cas9 gene editing, DNA base editors and RNA base editing.

demonstrated in several crops including wheat, maize, rice, tomato, cotton and tobacco. In late 2017 and early 2018, a number of specific rice base editors (rBE3, rBE4, rBE5, etc) were also developed for specific base editing in rice for a large number of genes. In all these base editors, one of the limitations was the requirement of the presence of PAM sequence in the vicinity of the target site, which occurs with a frequency of only 10% of the possible target sites thus limiting their utility. In order to overcome this limitation, an xCas9 variant was developed in 2018 through a technique called **P**hage-**A**ssisted **C**ontinuous **E**volution (PACE), so that xCas9 variants can now be developed and used without any restriction of the availability of PAM in the proximity of the target site. The technology can be immediately used in plant systems for crop improvement, although in case of human

healthcare, it may take some time before delivery systems are developed, side effects are examined and other regulatory/ethical issues are addressed.

1. Gaj, T., Gersbach, C. A. and Barbas, C. F., ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends Biotech*., 2013, **31**, 397–405.
2. Zetsche, B. *et al*., Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. *Cell*, 2015, **163**, 759–771.
3. Zaidi, S., Mahfouz, M. M. and Mansoor, M., CRISPR-Cpf1: A new tool for plant genome editing. *Trends Plant Sci*., 2017, **22**, 550–553.
4. Waltz, E., CRISPR-edited crops free to enter market, skip regulation. *Nat. Biotechnol*., 2016, **34**, 582.
5. Waltz, E., Gene-edited CRISPR mushroom escapes US regulation. *Nature*, 2016, **532**, 293–293.
6. Hall, S. S., Editing the mushroom. *Sci. Am*., 2016, **314**, 57–63.
7. Shi, J. *et al*., ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotech. J*., 2017, **15**, 207–216.
8. Nishida, K. *et al*., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, 2016, **353**, aaf8729.
9. Satomura, A. *et al*., Precise genome-wide base editing by the CRISPR nickase system in yeast. *Sci. Rep*., 2017, **7**, 2095; doi:10.1038/s41598-017-02013-7.
10. Komor, A. C. *et al*., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 2016, **533**, 420–424.
11. Gaudelli, N. M. *et al*., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, 2017, **551**, 464–471.
12. Cox, D. B. T. *et al*., RNA editing with CRISPR. *Science*, 2017, **358**, 1019–1027.
13. Saey, T. H., New CRISPR gene editors can fix RNA and DNA one typo at a time. *Sci. News*, 25 October 2017.
14. Ledford, H., David Lui: Gene corrector in '10 people who mattered this year'. Nature's 10. *Nature*, 2017, **552**, 315–324.
15. Kleinstiver, B. P. *et al*., Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 2015, **523**, 481–485.
16. Kim, D. *et al*., Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods*, 2015, **12**, 237–342.
17. Komor, A. C. *et al.*, 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*., 2017, **3**, eaao4774.
18. Wang, L. *et al.*, Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res*., 2017, **27**, 1289–1292.
19. Ran, F. A. *et al*., *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature*, 2015, **520**, 186–191.
20. Nishimasu, H. *et al*., Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, 2018; 10.1126/science.aas9129 aas9129.
21. Hu, J. H. *et al*., Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, 2018, **556**, 57–63.
22. Kleinstiver, B. P. *et al*., Broadening *Staphylococcus aureus* Cas9 targeting range by modifying PAM recognition. *Nat. Biotechnol*., 2015, **33**, 1293–1298.
23. Kleinstiver, B. P. *et al*., High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 2016, **529**, 490–495.
24. Liang, P. *et al*., Effective gene editing by high-fidelity base editor 2 in mouse zygotes. *Protein Cell*, 2017, **8**, 601–611.
25. Rees, H. A. *et al*., Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun*., 2017, **8**, 15790, doi:10.1038/ncomms15790.
26. Hall, S. S., The embarrassing destructive fight over biotech's big breakthrough. *Sci. Am*., 4 February 2016.
27. Liang, P. *et al*., Correction of *β*-thalassemia mutant by base editor in human embryos. *Protein Cell*, 2017, **8**(11), 811–822.
28. Zeng *et al*., Correction of the marfan syndrome pathogenic *FBN1* mutation by base editing in human cells and heterozygous embryos. *Mol. Ther.*, 2018, **11**; https://doi.org/10.1016/j.ymthe.2018.08.007.
29. Liang, P. *et al.*, Effective and precise adenine base editing in mouse zygotes, *Protein Cell*, 2018, **9**(9); doi:10.1007/s13238-018-0566-z.
30. Cohen, J., Novel CRISPR-derived 'base editors' surgically alter DNA or RNA, offering new ways to fix mutations. *Sci. News*, 25 October 2017.
31. Shimatani, Z. *et al*., Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol*., 2017, **35**, 441–443.
32. Ren, B. *et al*., A CRISPR/Cas9 toolkit for efficient targeted base editing to induce genetic variations in rice. *Sci. China. Life Sci*., 2017, **60**, 516–519.
33. Ren, B. *et al*., Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. *Mol. Plant*, 2018, **11**, 623–626.
34. Yan, F. *et al*., Highly efficient A•T to G•C base editing by Cas9n-guided tRNA adenosine deaminase in rice. *Mol. Plant*, 2018, **11**, 631–634.
35. Tang, W. and Liu, D. R., Rewritable multi-event analog recording in bacterial and mammalian cells. *Science*, 2018, **360**, eaap8992; doi:10.1126/science.aap8992
36. Kwon, D., USDA will not regulate CRISPR-edited crops. *The Scientist*, 2 April 2018
37. Callaway, E., EU law deals blow to CRISPR crops, *Nature*, 2018, **560**, 16.
38. Zong, Y. *et al*., Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol*., 2017, **35**, 438–440.
39. Kim, Y. B. *et al*., Increasing the genome- targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol*., 2017, **35**, 371–376.
40. Park, D. S. *et al*., Targeted base editing via RNA-guided cytidine deaminases in *Xenopus laevis* embryos. *Mol. Cells*, 2017, **40**, 823–827.
41. Ma, Y. *et al*., Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods*, 2016, **13**, 1029–1035.
42. Lu, Y. and Zhu, J. K., Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant*, 2017, **10**, 523–525.
43. Li, J. *et al*., Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol. Plant*, 2017, **10**, 526–529.

ACKNOWLEDGEMENTS. This article was written, when the author was holding the position of INSA Honorary Scientist at the Department of Genetics and Plant Breeding, CCS University, Meerut. Special thanks are due to Dr Neeraj Kumar Vasistha for extending help in formatting the final manuscript.

Received 24 February 2018; revised accepted 2 November 2018

doi: 10.18520/cs/v116/i3/386-397