Invited Expert Review

The CRISPR/Cas revolution continues: From efficient gene editing for crop breeding to plant synthetic biology

Running title: CRISPR/Cas: From breeding to synthetic biology

Jochen Kumlehn¹, Janine Pietralla², Goetz Hensel¹, Michael Pacher² and Holger Puchta²*

¹. Plant Reproductive Biology, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) 06466 Seeland OT Gatersleben, Germany

². Botanical Institute, Molecular Biology and Biochemistry, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

*Correspondence: Holger Puchta (holger.puchta@kit.edu)

Edited by Uwe Sonnewald, Friedrich-Alexander University, Germany
Abstract Since the discovery that nucleases of the bacterial CRISPR/Cas system can be used as easily programmable tools for genome engineering, their application massively transformed different areas of plant biology. In this review, we assess the current state of their use for crop breeding to incorporate attractive new agronomical traits into specific cultivars of various crop plants. This can be achieved by the use of Cas9/12 nucleases for double strand break induction, resulting in mutations by non-homologous recombination. Strategies for performing such experiments – from the design of guide RNA to the use of different transformation technologies – are evaluated. Furthermore, we sum up recent developments regarding the use of nuclease-deficient Cas9/12 proteins, as DNA-binding moieties for targeting different kinds of enzyme activities to specific sites within the genome. Progress in base deamination, transcriptional induction and transcriptional repression, as well as in imaging in plants, is also discussed. As different Cas9/12 enzymes are at hand, the simultaneous application of various enzyme activities, to multiple genomic sites, is now in reach to redirect plant metabolism in a multifunctional manner and pave the way for a new level of plant synthetic biology.
INTRODUCTION

A major challenge for crop improvement is the limited genomic variability within elite breeding material. Since the 1950s, the application of genotoxic agents, such as chemicals or irradiation, were used to induce new alleles by random mutagenesis. This technology was time-consuming and ineffective; multiple mutations were introduced in the same genome, simultaneously, and many resulted in adverse effects (Pacher and Puchta 2017). This situation changed only when it became possible to introduce precisely located breaks into DNA, in vivo, by the use of site-specific nucleases. Targeted rearrangements require transient opening of the genetic information, with a precise cut, followed by changes catalysed by the cellular DNA repair machinery. Base editing, directed transcriptional regulation and epigenetic modifications are further developments arising from the ability to target specific genomic sites.

This portfolio of genome modifications became available mainly within the last decade with the discovery of molecular effectors that: (1) can be targeted easily to specific sequences in a genome or transcriptome, (2) possess intrinsic nuclease activity, or (3) are modified to derivatives that deliver defined biochemical reactions or properties to genomic loci or transcripts. This molecular toolbox can be applied both in plant biotechnology, to explore and expand the genetic potential of crops as a major source of nutrition and renewable energy, as well as facilitating and accelerating breeding processes to increase the economic value of products thereof.

In this review, we provide a state-of-the-art overview of the available molecular tools, including their derivatives, their properties and mode of action, their handling in wet-lab activities and advances in plant biotechnological applications, based on these revolutionary techniques. We conclude with an outlook of possible future applications of these tools and considerations as to how the existing toolset might evolve, providing unprecedented possibilities to the growing field of plant synthetic biology.

SEQUENCE-SPECIFIC INDUCTION OF DNA DOUBLE-STRAND BREAKS

The ability to introduce unique genomic site-specific DNA double-strand breaks (DSBs), in vivo, dates back to the discovery and characterization of the homing endonuclease I-SceI in Saccharomyces cerevisiae (Jacquier and Dujon 1985; Perrin et al. 1993). Expression of this molecular scissor, in plants, harbouring its 18mer target sequence, led not only to fundamental progress in the understanding of DSB repair processes, in somatic plant cells (Puchta 2005), but also to the demonstration that genes can be knocked out (Salomon and Puchta 1998), or genomic loci be changed in a controlled manner (Puchta et al. 1996).

The vast majority of DSBs are repaired via non-homologous end joining (NHEJ), a highly mutagenic mechanism of partial end-resection and re-ligation of the free ends. Nevertheless, the intrinsic DSB repair machinery has the potential to remove breaks also in an error-free manner, namely homologous recombination (HR), which utilizes allelic or ectopic sequence homologies as a repair template. The
equilibrium between HR and NHEJ is strongly biased to the latter in somatic plant cells, as a result of evolutionary requirements to ensure genomic integrity in highly complex and partially repetitive genomes (Cavalier-Smith 2005; Puchta 2005). Thus, the knowledge of repair pathway utilization, key factors involved therein and the superimposed regulatory mechanisms is crucial to any rational design aiming for efficient generation of modified plant genomes (Puchta and Fauser, 2015). So far, both DSB repair pathways have been experimentally addressed to obtain various genetic alterations.

The prevalent DSB repair pathway NHEJ has been largely exploited for obtaining targeted mutations, but furthermore, its potential to specifically invert or delete genome fragments, break genetic linkage groups, or reshuffle chromosome orientations has been recognized (Pacher et al. 2007; Lee et al. 2012; Le Cong et al. 2013; Li et al. 2013; Mali et al. 2013; Wang et al. 2013; Weinthal et al. 2013; Filler Hayut et al. 2017). In addition, HR-mediated DSB repair has been applied for targeted sequence alterations (Steinert et al. 2016). This has been achieved by making use of recombination substrates, such as extrachromosomal templates provided, for example, via an incoming T-DNA from Agrobacterium tumefaciens, transfected plasmids, biolistic delivery of DNA sequences, or delivery by viral replicons (Baltes et al. 2014; Čermák et al. 2015; Butler et al. 2016; Čermák et al. 2017; Gil-Humanes et al. 2017). Both the controlled introduction of very minor sequence modifications and the insertion of new genes in transgenic and natural loci has also been achieved with the in planta Gene Targeting (GT) method (Fauser et al. 2012; Fauser et al. 2014; Wolter et al. 2018), in combination with the use of viral replicons (Dahan-Meir et al. 2018), or in combination with developmentally-controlled promoters (Miki et al. 2018; Wolter et al. 2018).

While initial efforts for targeted genome rearrangements, applying rare cutters like I-SceI, were limited by the need of artificially inserting the target site as part of a transgene, the development and availability of engineered zinc-finger nucleases (ZFN) (Kim et al. 1996; Smith et al. 2000), transcription activator-like effector nucleases (TALENs) (Boch et al. 2009; Moscou and Bogdanove 2009; Cermak et al. 2011) and, most recently, the CRISPR (clustered regularly interspaced palindromic repeat)-associated (Cas) system, enabled the generation of new genome engineering tools for addressing nearly all genomic loci with unprecedented accuracy and efficiency (Voytas 2013; Mahfouz et al. 2014; Belhaj et al. 2015; Schaeffer and Nakata 2015; Lee et al. 2016; Weeks et al. 2016; Pacher and Puchta 2017; Puchta 2017; Schindele et al. 2018).

The availability of such new tools not only allowed the development of a better understanding of DNA repair pathways (e.g. Schiml et al. 2016), but also provided means for generation of more than one defined break, simultaneously, within a genome to induce more complex changes, such as deletions, inversions and translocations. The following section provides a state-of-the-art overview on the tool set, its intrinsic properties and artificially designed derivatives thereof, for enhancing the possibilities of plant synthetic biology.
CLASSES OF CAS-ENDONUCLEASES

Clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated (Cas) proteins form a prokaryotic defense mechanism against bacteriophages and mobile genetic elements (Wiedenheft et al. 2012). This adaptive immune system is able to integrate fragments of the invading genomes, as so-called spacers, in the CRISPR array and provides a sequence-based memory of previous infections (Hille and Charpentier 2016; Barrangou and Gersbach 2017; Murugan et al. 2017). Two proteins, Cas1 and Cas2, are essential for most CRISPR/Cas systems to deliver the foreign DNA fragments to the CRISPR array (Koonin et al. 2017). During the second infection, the CRISPR locus, encoding the effector Cas proteins and spacers, is transcribed into the so-called precursor crRNA (pre-crRNA). Various mechanisms exist to process the pre-crRNA, in the different CRISPR/Cas systems, although they are not yet completely understood.

Independent of the manner of processing, the crRNA contains a spacer, which is complementary to viral DNA and provides the target specificity of the Cas system. The crRNA interacts with a Cas effector and guides the effector protein accurately to the DNA or RNA target motif, where the complex performs its cleavage function (Jinek et al. 2012). As a consequence of reciprocal selective pressure, a wide variety of defense strategies against viral invaders has evolved and, therefore, there exists a high diversity of structural varieties between the individual CRISPR/Cas systems (Wiedenheft et al. 2012; Makarova et al. 2015).

CRISPR/Cas systems essentially comprise two classes: the most abundant CRISPR/Cas system in bacteria and archaea is referred to as class 1, which contains multi-subunit effector complexes and includes type I, type III and type IV systems. Class 2 consists of three types of single-protein effector modules: type II, type V and type VI. Each type can also be subdivided into several subtypes (Makarova et al. 2015; Koonin et al. 2017). The most extensively studied Cas protein is the single effector endonuclease, Cas9, of the type II-A system (Jinek et al. 2012). Fonfara et al. 2014). A second RNA-guided effector nuclease was identified with bioinformatics studies (Schunder et al. 2013) and was classified as Cpf1 (now Cas12a), a class 2 type V CRISPR/Cas system (Makarova et al. 2015). Recently, three novel class 2 effectors, C2c1, C2c2 and C2c3, were identified by a computational pipeline (Shmakov et al. 2015). Two of these proteins, C2c1 and C2c3, showed similarity to the already known type V effector, Cas12a, and were recently renamed as Cas12b and Cas12c (Murugan et al. 2017). The third effector, C2c2 (now Cas13a), shows a higher divergence and was assigned to the new class 2 type VI CRISPR/Cas system (Shmakov et al. 2015).

The RNA-guided endonuclease Cas9 of the class 2 type II CRISPR/Cas system from *Streptococcus pyogenes* has become the best characterized single effector protein and an important biotechnological tool for molecular biology (Jinek et al. 2012). Like most class 2 effectors, Cas9 consists of two lobes. A recognition domain (REC) is connected over an arginine-rich bridge helix with the nuclear domain (NUC) (Jinek et al. 2014). The NUC lobe contains two catalytically active subdomains,
RuvC and HNH, which are required for the induction of the double-strand break (DSB). The HNH domain is responsible for the cleavage of the DNA strand that is complementary to the spacer region of the guide RNA (gRNA), whereas the RuvC-like domain cleaves the non-complementary DNA strand (Jinek et al. 2012).

In order to find foreign DNA, the Cas9 nuclease requires this gRNA, which is formed by the crRNA and a tracrRNA and a short G-rich sequence motif, termed protospacer adjacent motif (PAM) positioned next to the gRNA-specific part of the target sequence (Figure 1A). Upon formation of the Cas9/gRNA complex, it scans the DNA for the PAM sequence. Due to the nature of the CRISPR/Cas system, the effector protein is able to target multiple loci, simultaneously. Upon PAM identification, the complex separates the DNA double strand and checks whether the spacer on the crRNA is complementary to the sequence next to the PAM site. After formation of a hetero-duplex, the nuclease domains of Cas9 are activated and induce a blunt-end DSB three base pairs (bp) upstream of the PAM (Garneau et al. 2010; Jinek et al. 2012; Chaudhary et al. 2018).

An additional well-characterized single RNA-guided endonuclease is Cas12a (formerly Cpf1), which is assigned to the class 2 type V CRISPR/Cas system. Despite the same function and nearly identical size, there are some important differences compared to the Cas9 protein. In contrast to Cas9, Cas12a requires a shorter crRNA and there is no evidence that a tracrRNA is required (Zetsche et al. 2015). Furthermore, Cas12a is able to process the pre-crRNA itself (Fonfara et al. 2016). As opposed to type II CRISPR-Cas systems, where the crRNA scaffold is located at the 3´-end, the crRNA scaffold of the type 5 CRISPR/Cas systems is located at the 5´-end (Figure 1B) (Zetsche et al. 2015). Additionally, the protein creates a double strand break by DNA cleavage distal from the PAM sequence and the seed region, where the crRNA is more tolerant of mismatches (Gao et al. 2016). The cleavage occurs 18 nucleotides downstream of the PAM, at the non-target strand, and 23 nucleotides downstream, at the target strand, which leads to a five nucleotide 5´-overhang (Zetsche et al. 2015). Presumably the overhang is due to the sole cleavage activity of the RuvC-domain (Swarts et al. 2017). Similar to Cas9, Cas12a also appears to have two nuclease domains, RuvC and a Nuc domain (Yamano et al. 2016), but recent structural analyses suggest that the Nuc domain supports DNA binding and lacks catalytic activity (Swarts et al. 2017). However, in contrast to Cas9, Cas12a requires T-rich PAMs, which increases the total number of potential target sites. Recently, a single-stranded DNA (ssDNA) cleavage activity of Cas12a was discovered, whereby Cas12a completely degrades non-specific ssDNA after being activated upon target binding (Chen et al. 2018).

What was previously only possible in the DNA world with Cas9 and Cas12a can now be extended into the RNA world with the RNase Cas13. This type of Cas protein was identified by bioinformatics analyses of bacterial and archaeal genomes (Shmakov et al. 2015). The programmable RNA-guided RNA-targeting CRISPR effector, Cas13a (previously C2c2), cleaves single-stranded RNA molecules in a sequence-specific manner (Figure 1C) (Abudayyeh et al. 2016). Regardless of the similar origin, the
protein shows no strong structural correlation to the previously characterized Cas proteins. It also
displays a bi-lobed structure, but contains the NUC lobe and two conserved higher eukaryote and
prokaryote nucleotide-binding (HEPN) domains.

Cas13a also differs in the location of the catalytic sites; in an activated state, the HEPN domain is
located at the external surface, whereas the nuclease domains of Cas9 and Cas12a are located inside the
protein (Liu et al. 2017b; Wolter and Puchta 2018). Like Cas12a, Cas13a can also process the
pre-crRNA by itself and possesses no tracrRNA (East-Seletsky et al. 2016). Furthermore, Cas13 shows
general non-specific degradation of RNA, after activation, via target binding. This function is not
affected by deactivation of the HEPN domains, indicating that there are two different catalytic sites
(East-Seletsky et al. 2016). However, the ‘collateral’ cleavage activity seems not to be present in
eukaryotic cells (Abudayyeh et al. 2016), making Cas13a a suitable tool for targeting and degrading
RNA in a highly specific manner and, thus, expanding the toolset to address the transcriptome as well.

Despite the necessity of PAM sequences for Cas9 and Cas12a, such a requirement could not be
observed for all Cas13 orthologs, but some Cas13 proteins seem to require a PAM-like motif, called
protospacer flanking site (PFS). This has been established so far for the three Cas13 orthologs: Cas13a
from Leptotrichia wadei (LwaCas13a), Cas13b from Prevotella sp. (PspCas13b) and Cas13d from
Eubacterium siraeum (EsCas13d) (Abudayyeh et al. 2017; Cox et al. 2017; Konermann et al. 2018).

Taken together, the molecular tool-set of sequence-specific programmable DNA and RNA nucleases
provides unprecedented possibilities for both the targeted generation of DNA breaks, or triggering the
degradation of selected RNA molecules, thereby increasing the potential to modify genes, genome
structure, or selectively-controlling the transcriptome (Schindele et al. 2018). In the following section,
we will discuss both the applications of these targeted nucleases, as well as derivatives thereof, which
can be generated artificially to change PAM requirements, or enhance the specificity of protospacer
recognition to reduce potential off-target cleavage. Additionally, we consider the possibilities to target
one Cas effector to more than one locus, simultaneously, or operating different Cas orthologs with
variable modifications to induce single-strand breaks (SSBs), or targeting fused protein effectors to loci
of interest in multiplex approaches.

SELECTION OF TARGET MOTIFS

Selection criteria

Based upon the interaction of the gRNA and Cas endonuclease and the binding of their complex to target
DNA, there are general and sequence-specific requirements, and preferences, to be considered when
target motifs are selected. The most decisive limitation for the choice of target motifs derives from the
requirement of a protospacer adjacent motif (PAM), which is the nucleotide triplet specifically bound
by the Cas endonuclease. In the case of canonical Cas9, the PAM comprises one flexible nucleotide, followed by two guanosine residues. Utilizing plant species that enable reasonably efficient genetic transformation, simple knockout approaches are fairly straightforward. For site-directed mutagenesis, it may be sufficient to simply select two to three target motifs within the coding sequence of the target gene, with each target comprising any 21 nucleotides upstream of a double G. In more challenging approaches, such as multi-target, double Cas or repair template-directed modifications, it is advisable, however, to select target motifs for which high cleavage activity of the gRNA/Cas complex is predicted.

Secondary structure formation of the gRNA plays a crucial role in Cas cleavage activity, because not only must the gRNA 5’-terminal part be capable of base-pairing with the target DNA, but also the binding between Cas endonuclease and gRNA is essentially dependent on two-dimensional structures formed within the gRNA 3’-terminal scaffold (Ma et al. 2015b). Online tools, such as RNAfold (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) and MFOLD (Zuker 2003), are available to predict the most likely occurring secondary structure variants of gRNAs \textit{in silico}. Taking advantage of this opportunity, substantial interactions between the target-compatible 5’-terminal part of the gRNA and its 3’-scaffold can largely be avoided.

In Figure 2, examples of predicted 2D-structures of gRNAs with high (A) and compromised functionality (B) are shown. The extent of intramolecular pairing of nucleobases within gRNAs is also considered in the validation of target motifs by online platforms specifically devoted to the use of customized endonucleases, such as CRISPR-GE (http://skl.scau.edu.cn/) (Xie et al. 2017). Using rice as the experimental model, Liang et al. (2016) have shown that, in efficient gRNAs, no more than 12 nucleotides of the target-specific gRNA 5’-terminal part are involved in base-pairing with the invariable gRNA scaffold. Moreover, a maximum of seven consecutive nucleotides of the variable part are tolerated to pair with complementary bases of the gRNA scaffold. Importantly, three stem-loops formed within the invariable gRNA scaffold have been proven to be essential for proper interaction of gRNA and Cas endonuclease. By contrast, any further secondary structures, more or less frequently occurring within the gRNA scaffold, were not associated with high cleavage activity of the gRNA/Cas9 complex and, hence, can be neglected when it comes to the selection of target motifs (Liang et al. 2016). Wong et al. (2015) showed that a CTT triplet, residing at the position −1 to −3 immediately upstream of the PAM, severely disturbs the functionality of gRNAs as it is complementary and, thus, readily base-pairs with an AAG triplet present on the gRNA scaffold where one of its functional stem-loops needs to be formed.

Preferences of the various nucleotide positions along the target motifs for particular nucleobases have been elucidated by functionally testing large gRNA libraries, followed by the differentiation of fractions of functional versus non-functional gRNAs, or subpopulations of most efficient gRNAs (Doench et al. 2014; Moreno-Mateos et al. 2015). Conspicuous and independently reoccurring patterns have mostly been detected in the immediate sequence context of the PAM that, itself, constitutes the
most striking feature of useful gRNA/Cas target motifs. In the case of the best studied Cas9, guanine at the position −1 upstream and cytosine at the position +1 downstream of the PAM were shown to be associated with high cleavage activity. On the other hand, the following nucleobases were determined to occur particularly rarely in the given positions of efficiently cleaved target motifs: cytosine and thymine at −1 upstream of the PAM, thymine at the N-position of the PAM and guanine at +1 downstream of the PAM (Xu et al. 2015).

Inconsistent results have been published regarding a number of features of useful target motifs, which is likely to be due to various basic criteria for gRNA selection, as well as to different experimental models, with the latter having specific preferences in terms of DNA repair mechanisms (Xu et al. 2015). For example, an attempt to confirm nucleotide preferences of target motifs established in various vertebrate cell systems failed in rice, where no statistically significant associations of nucleotide composition of the target specific 5'-part of the gRNA with cleavage efficiency were observed, albeit different average occurrences of the four available nucleobases were seen in some positions within the target sequences of efficient vs. inefficient gRNAs (Liang et al. 2016).

Likewise, there are inconsistent results as to the role of GC content of the target-specific 5'-part of gRNAs. Whereas Wang et al. (2014), using human cells, showed that the fractions of gRNAs featuring lower than 45 and higher than 75% GC content of their target-specific 5'-part were associated with reduced cleavage efficacy, Gagnon et al. (2014) reported that the fraction of significantly more efficient gRNAs in zebrafish have a GC content higher than 50%. Using Drosophila as experimental system, Ren et al. (2014) showed that a GC content of over 50% is beneficial only within the six nucleobases positioned immediately upstream of the PAM. Then again, Labuhn et al. (2018) reported a significant effect of the GC content in the domain comprising positions −8 through to 17 upstream of the PAM in human cells.

Whereas the vast majority of approaches to establish selection rules for target motifs have been generated using association principles, rather than causal relationships, some of these rules have been validated at least using hypothesis-driven approaches; e.g., on the topological interaction of gRNA and target DNA (Xu et al. 2017).

The repair machinery preferentially uses microhomologies between the free ends produced by DNA cleavage. It is, therefore, possible to take advantage of this phenomenon to predict sequence modifications, or even to use such patterns strategically, be it to enhance cleavage efficiency, to preferentially obtain in- or out-of-frame mutations, or to precisely create particular sequence modifications (Bae et al. 2014).

Online platforms for the selection of efficient and specific target motifs
The selection of target motifs for Cas endonucleases is greatly facilitated by a variety of online platforms. While WU-CRISPR (Wong et al. 2015), SSC (Xu et al. 2015), CRISPR MultiTargeter (Prykhozhij et al. 2015), sgRNA Scorer (Chari et al. 2015) and Desktop Genetics (Doench et al. 2016) are particularly reliable in recommending gRNAs and predicting their cleavage efficacy, other platforms have been tailored for the specific requirements of certain organisms. For example, CRISPR-P 2.0 (Liu et al. 2017a) and CRISPR-GE (Xie et al. 2017) were established for plants. Notwithstanding their utility, the predictive power of these online tools is still limited, since their algorithms have been developed based on the association of target features with cleavage activity, rather than on experimentally established causal relationships. Yet another limitation is that results and findings which are highly dependent on experimental background conditions and cellular particularities cannot readily be extrapolated from one experimental system to another.

The establishment of selection criteria for particularly efficiently cleaved target motifs and the development of computational algorithms balancing such criteria, has resulted in the above-mentioned tools for the prediction of useful target sites. A further improved predictability of cleavage activity may be achieved by the elucidation of causative relations between structure and function. To this end, systematic comparisons of two or more targets that differ in just one feature would be required. However, owing to the limited presence of all theoretically possible target motifs, in a given organism, and the influence of chromatin structures and their position effects, this will be technically very challenging. Nonetheless, it will be worthwhile to develop experimental concepts in this particular field, since an increased prediction accuracy of target efficacy is key to improved and more sophisticated methods of site-directed genome modification.

**Simultaneous modification of multiple genomic targets**

Various solutions exist to express multiple gRNAs within the same cell. This principle is of particular value in allopolyploid species where the genetic similarity, across all partial genomes, is often not high enough to allow for the identification of identical and useful target motifs present in all homeoalleles of a given target gene. Moreover, it is sometimes required to simultaneously cut more than one target motif in the same gene (e.g. for double nickase setups), or in two or more target genes, to achieve multiple genetic modifications. Approaches to the modification of entire gene families or to pathway engineering are the most challenging in this context.

The first approaches for the simultaneous use of multiple gRNAs were to deliver each gRNA driven by its own promoter (Schiml et al. 2014; Xing et al. 2014; Ma et al. 2015b). The limitation herein is that, with a polymerase (Pol) III promoter for each gRNA, the required constructs become comparatively large. Three different concepts to avoid this challenge are based on the principle of a polycistronic gene. The expression of multiple functional gRNAs as a polycistronic gene, under the control of Pol II promoters in human cells, was tested by utilizing the RNA-binding and RNA-endonuclease capabilities
of Csy4, whereby the gRNAs could be released from their transcript (Nissim et al. 2014). In this context, it is worthy to consider the requirement of a second endonuclease to separate the gRNAs, which also significantly contributes to the length of DNA constructs. Alternatively, a polycistronic construct was developed utilizing the endogenous tRNA processing system in rice. The tRNA sequences are located in between the gRNA sequences and are necessary for the cleavage of two host RNases (Xie et al. 2015).

With the tandem-like array of self-cleaving ribozymes and gRNAs, the separation of the gRNAs after transcription is also enabled. The advantage here is that no further components are required (Tang et al. 2016). Flanking gRNAs with ribozymes can also be of interest for individual gRNAs, as the gRNA does not necessarily have to be expressed by a Pol III promoter. It will then be possible to use Pol II promoters, which are better studied and can also permit tissue specific or inducible expression (Gao and Zhao 2014).

Using a polycistronic tRNA-gRNA gene expression strategy in rice, Xie et al. (2015) demonstrated the simultaneous modification of as many as eight target sites. Later, this group took advantage of the same principle to study the roles of four stress-related MAP kinases of rice, via Cas9-mediated knockout (Minkenberg et al. 2017). Lowder et al. (2018) took a multiplex gene activation approach in Arabidopsis and rice using a chimeric dCas9-VP64-EDLL, aiming to activate gene-specific expression. However, the authors came to the conclusion that the level of transcriptional activation achieved is remarkably dependent on the individual genes manipulated.

Čermák et al. (2017) compared Cas9 and Cas12 with regards to the simultaneous genome modification at multiple targets in plants. It was shown that Cas12a exhibits endoribonuclease activity and is able to process its pre-crRNA by itself (Fonfara et al. 2016; Swarts et al. 2017). Due to this property, Cas12a is naturally qualified for multiplex gene editing, compared to Cas9. Using an artificial CRISPR array, with alternating direct repeats and crRNA, this functionality was proven in mammalian cells (Zetsche et al. 2017) and plants (Wang et al. 2017b). A further advantage is the smaller size of the Cas12a-crRNA, which means either a smaller size of the delivered vector or an enhancement of the amount of crRNAs. In general, Cas12a can compete with Cas9 concerning DNA cleavage activity, mutation frequencies and DNA-binding (Zetsche et al. 2017). Recently, multiplex gene editing was shown in rice with FnCas12a, LbCas12a and SpyCas9. Nucleases and gRNAs were co-expressed by a Pol II promoter with no additional processing components (Wang et al. 2018).

**Delivery systems for Cas endonucleases**

For site-directed genome modification of plants, genomic integration of gRNA- and Cas endonuclease-encoding transgene expression units is the most widely used approach. Therefore, conventional plant genetic transformation methods, based upon Agrobacterium, can be applied. In experimental model systems, such as isolated protoplasts, these transgenes can also be introduced via poration of the plasma
membrane using voltage pulses or high concentrations of polyethylene glycol. In protoplast systems, transgene expression is only required for a limited period. Therefore, it is sufficient that much of the transgene expression derives from non-integrated DNA.

Alternatively, non-integrating virus replicons carrying units of gRNA, Cas9 and synthetic template (for homology-directed DNA break repair) have been used (Baltes et al. 2014; Ali et al. 2015). The major advantage of this method is that the self-replicating vectors facilitate high transgene expression, which increases the efficiency of site-directed genome modification. However, the challenge remains to find useful viruses that are infectious in particular plant species of choice and to eventually generate modified plants that are entirely virus-free.

Particularly high transgene expression levels can also be achieved during a limited episode upon biolistic gene transfer. Taking this approach, Zhang et al. (2016) produced wheat plants carrying targeted mutations in the yield-related DEP1 and GASR7 genes. Of note, no selective conditions were required for mutant regeneration, at reasonable efficiency, and quite a proportion of the primary mutant plants proved to be non-transgenic for the gRNA and Cas9 genes.

In conventional methods, Cas endonucleases and gRNAs act as product of their coding DNA sequences previously transferred into host cells. A viable alternative, however, is to in vitro transcribe gRNA and Cas endonuclease genes, or to pre-produce Cas protein using microbial expression platforms and, in contrast to transgenesis, to use these non-heritable molecules for transfection. It should be noted here that such reagents collectively constitute nothing but target-specific mutagens. In this context, gRNA and Cas endonuclease can also be assembled prior to transfection to form ribonucleoprotein (RNP) complexes.

The use of RNPs has some advantages over the principle of gRNA and Cas transgene expression. First, there is no need to create species- or cell-specifically tailored expression vectors for gRNA and Cas endonuclease genes. Second, the cellular dosage of gRNA and Cas endonuclease is independent of the specificity and strength of available promoters. Third, as the cellular implementation of transgenes into accumulating gene products is rendered unnecessary, gRNA and Cas endonuclease exert their full capacity to generate double-strand breaks immediately after transfection. Fourth, the efficacy of RNP complexes can be strong and is yet confined to a comparatively short temporal window after administration of the reagents, because these are non-replicating molecules that are prone to both dilution, as a result of cell proliferation, and cellular degradation pathways. Consequently, the occurrence of superfluous on-target mutations beyond the germline, as well as unwanted off-target mutations in germline cells is much reduced during later stages of plant development. Fifth, once mutants are generated, there is no need to screen the progeny for segregants that have lost the coding sequences for gRNA and Cas, i.e., all mutant individuals obtained are transgene-free and hence available for any further analyses, investigations and applications.
A first proof of this concept was demonstrated by Woo et al. (2015), who introduced gRNA/Cas9 RNPs to protoplasts from a variety of plant species and, thereby, showed indeed that targeted mutations had been induced. In an attempt to transfect in vitro transcribed RNAs for both gRNA and Cas9, one percent of the biolistically-transformed cells of bread wheat proved to carry mutations in the GW2 target gene (Zhang et al. 2016). As a further outcome of this study, about one third of the obtained plants carried mutations in all six GW2 alleles that are present in hexaploid wheat, which is particularly remarkable, albeit this approach was ca. 60% less efficient than the compared one that involved transient and stable expression of gRNA and Cas9 transgenes.

In maize, Svitashev et al. (2015) achieved another milestone, in that they mutated four different genes, simultaneously, via biolistic transfer of RNPs into re-generable cells of immature embryos, which was conducted in comparison with gRNA and Cas9 transgenesis. In this study, both principles (RNPs vs. transgenes) were on a par in terms of mutagenesis efficiency (Svitashev et al. 2016). A particularly valuable application of using gRNA/Cas9 RNP complexes was demonstrated by Andersson et al. (2018), who generated an amylopectin potato via knock-out of GBSS. In addition, the utilization of Cas12a has been exemplified along with gRNA in the format of RNPs that were transfected into protoplasts to induce targeted mutations in the two FAD2-1 gene variants of the palaeopolyploid soybean (Kim et al. 2017).

The cleavage efficiency of the enzyme variant derived from Lachnospiraceae bacterium (LbCas12a) proved to be about 10 times more efficient as compared to the one of Acidaminococcus sp. BV3L6 (AcCas12a). In the same study, tobacco protoplasts were used as well to mutate the AOC gene. In this case, however, LbCas12a and AcCas12a caused a comparable mutation rate of about 1%. Cas/gRNA RNPs were also shown to be effective in approaches to homology-driven genome editing using artificial DNA templates that feature the modified sequences to be introduced to the host genome. In this context, protein and DNA have to be co-precipitated to particles used for biolistics or co-transfected via PEG-mediated poration of the plasma membrane, which renders this procedure quite challenging. Moreover, owing to the co-transfer of template DNA, a particular advantage of using RNPs is actually lost, namely that otherwise no transgenes are introduced to the host genome.

The use of RNPs to implement the DNA sequence of a repair template was demonstrated in maize, where herbicide tolerance was established by modifying the ALS2 gene. In this study, Svitashev et al. (2016) used a single-stranded repair template that was 127 DNA polynucleotides in length. In another novel approach, pre-assembled RNP complexes were also used to identify individuals carrying targeted modifications. To this end, target-specific PCR amplicons were in vitro digested by corresponding RNPs, which allowed for gel-electrophoretic distinction between wild-type and mutated plants with unprecedented reliability. This method was demonstrated via site-directed mutagenesis of the GW2 and Cer9 genes of wheat and the PDS1 gene of rice (Liang et al. 2018b).
This approach may be of particular advantage in species carrying large and complex genomes. In addition, it provides much increased freedom as to the choice of target sequences, in comparison to the use of targets that carry recognition sites of conventional restriction enzymes. A further incentive for the use of RNPs for approaches to targeted genome modification is that the commercial provision of customized gRNA and a variety of Cas proteins is nowadays well established. On the other hand, a broader application of RNPs in the field of genome engineering is still hampered, because suitable transfection methods for regenerable cells have been established only in a limited range of plant species.

APPLICATIONS OF TARGETED GENOME MODIFICATION BY NHEJ

Target sequence-specific genome modification employing RNA-guided Cas endonucleases for DSB induction has been established as routine in many laboratories in which plant research is performed. A comprehensive survey of applications of the technology would thus go beyond the scope of this review, in which only representative, or particularly impressive examples are given.

Fundamental research

Ever since the RNA-guided Cas9 endonuclease emerged as a viable tool for research and development in plants, this technology has been used for the functional validation of gene functions. Recent examples of studies devoted to plant development, using Arabidopsis as experimental model, were given for example by Saito et al. (2018), who elucidated the role of BZR1 in vascular differentiation, via Cas9-mediated knockout, by Osakabe et al. (2016), who manipulated stomata closure by knockout of OST2, by Liang et al. (2018a), who knocked out GGAT1 so as to alter the photorespiration pathway, and by Yu et al. (2018), who created triple mutants with a single gRNA, which could not be created by conventional crossbreeding.

Comprehensive genomic data, resources and tools are now available in the majority of important cultivated plants, and so the dominance of Arabidopsis research has been somewhat declining in recent years. Instead, as one of the major crop species, and representative of agriculturally most important monocots, rice has been playing an increasing role in fundamental research. For example, Zeng et al. (2018) established the function of OsBBS1, in the context of leaf senescence, and the response to salt stress, by knocking out this gene. Taking a base-editing approach using nCas9-PBE, Zong et al. (2017) elucidated the role of OsCDC43 in senescence and programmed cell death.

Further work was devoted to the functional validation via site-directed mutagenesis of the nuclear genes OsSLA4 and OsFLN1 involved in chloroplast development (He et al. 2018; Wang et al. 2018c). Site-directed mutagenesis triggered by Cas9 endonucleases was also used in further species to validate gene function. In cotton, for example, the function of CLA1, another gene involved in chloroplast formation, was studied via Cas9-mediated knockout (Wang et al. 2018a). In barley, knockout lines of
were used to study the role of this gene in dormancy. In tomato, a novel mechanism of interspecific incompatibility was recovered by knockout of the pollen-expressed \emph{FPS2} gene (Qin et al. 2018). Other examples are soybean and \emph{Medicago truncatula}, in which the processing of small RNAs in the context of post-transcriptional gene silencing was studied using Cas9-mediated knockout of \emph{GmDrb2a}, \emph{GmDrb2b} and \emph{MtHen1} (Curtin et al. 2018). Zhou et al. (2017) produced rice plants mutated in a variety of micro-RNA genes, thereby providing new insights into miRNA function and complex gene regulation pathways.

**Herbicide resistance**

The establishment of herbicide resistant crops was among the first applied approaches taking advantage of site-directed genome modification by means of RNA-guided Cas9. On the other hand, the readily selectable herbicide resistance also proved to be particularly useful in establishing Cas endonuclease technology. Whereas resistance to the herbicide Bentazon was simply achieved by knockout of \emph{OsBEL} (Xu et al. 2014), in the case of the herbicide targets ALS and EPSPS, precise modifications were required to prevent their inactivation by the respective herbicides while retaining their essential function for the plant. And since the development of herbicide resistance is one of the most simple genetic modifications that can be associated with a selective advantage, \emph{in vitro}, this principle was the first to be used to establish methods of homology-directed genome editing; e.g., in maize (Svitashev et al. 2015), rice (Li et al. 2016a; Sun et al. 2016), and flax (Sauer et al. 2016). Basically the same experimental approach was then also used to establish homology-directed genome modification, via delivery of gRNA and Cas9 in the format of preassembled ribonucleoprotein (Svitashev et al. 2016).

**Yield-related traits**

Customized RNA-guided Cas9 has also been broadly used to modify traits that are determinants of crop yield. Modern cultivars in numerous crops represent semi-dwarf variants that are less prone to lodging, have an increased capability of implementing nitrogen fertilizer and have an improved harvest index; i.e., an increase in the portion of harvested seeds or grains from the total above-ground biomass. The generation of more genetic diversity in plant height is still of high interest to further fine-tune this trait for different environmental and regional demands. Dwarf and semi-dwarf plants have been generated via knock out of \emph{Bol.C.GA4a} in rapeseed (Lawrenson et al. 2015), \emph{ER1}, \emph{ER2} and \emph{SEC3a} in rice (Ma et al. 2018; Zhang et al. 2018) and \emph{Dep1} in wheat (Zhang et al. 2016). Semi-dwarf rapeseed was also created by selecting mutants carrying deletions in the \emph{BaA6.RGA} gene, whereby the translational reading frame is retained (Yang et al. 2017). This principle provides the opportunity to achieve functional modifications rather than an entire loss-of-function of the target gene.
The temporal regulation of the developmental switch from the vegetative to the reproductive phase of both flowering time and fruit maturation, are further important determinants of yield, which have to be considered when tailoring cultivars for different agricultural systems and regional conditions. Accelerated maturity in rice and tomato was achieved via knockout of OsHd2, Hd4, Hd5 (Li et al. 2017b) and of SlSP5G (Soyk et al. 2017), respectively. By contrast, knockout of GmFT2 entailed delayed flowering time in soybean (Cai et al. 2018). Li et al. (2018a) demonstrated that mutagenesis of the long non-coding RNA1459 results in delayed tomato fruit ripening.

Grain or seed number and weight represent even more direct determinants of yield. Aiming to improve both of these traits in rice, Li et al. (2016b) produced mutants for the genes OsGn1a, DEP1, GS3 and IPA1. Alternatively, grain weight was enhanced via knock out of GW2, GW5 and TGW6 in rice (Xu et al. 2016), and of GASR7 in wheat (Zhang et al. 2016). Increased locule and seed number per silique was obtained in rapeseed via knockout of both homeoalleles of the CLV3 gene (Yang et al. 2018). Taking an even more sophisticated approach in tomato, Rodríguez-Leal et al. (2017) demonstrated that a whole range of variability in locule number and fruit size can be obtained by producing an array of CLV3 promoter variants with different fragment deletions. Cas9-mediated knock out of AGL6 caused facultative parthenocarpy in tomato, which results in seedless fruits (Klap et al. 2017). Reduced seed shattering was achieved by knocking out the ALCATRAZ gene of rapeseed (Braatz et al. 2017).

**Tolerance to abiotic stress**

Using Cas9-induced knockout lines, Huang et al. (2018) have validated the role of rice NCED3 for salt tolerance. In tomato, Wang et al. (2017a) revealed the implication of MAPK3 in drought tolerance also by generating and analysing respective knockout lines. In a particularly sophisticated approach, enhanced drought-tolerance was achieved by homology-directed repair of Cas9-induced DNA breaks leading to enhanced expression of the ARGOS8 gene of maize (Shi et al. 2017).

**Resistance to pathogens**

Resistance of *Nicotiana benthamiana* to pathogenic geminiviruses was achieved by Baltes et al. (2015) and Ji et al. (2015) via mutagenesis of target motifs residing within the viral genomes. Upon mutational modification of the previously known susceptibility factor eIF4F, Chandrasekaran et al. (2016) generated cucumber with resistance to various ipomo- and potyviruses.

Site-directed mutagenesis using RNA-guided Cas endonucleases was also used to establish resistance to bacterial plant diseases. In rice, for example, resistance to bacterial blight and rice blast, respectively, was achieved by knockout of the sugar transporter SWEET13 (Zhou et al. 2015) and of ERF922 (Wang et al. 2016). Jia et al. (2017) established resistance to citrus canker by knocking out the LOB1 gene. And more recently, Wang et al. (2018b) rendered grapevine resistant to *Botrytis cinerea*.
through the knockout of the transcription factor WRKY52. The establishment of powdery mildew resistance in wheat has been among the agriculturally most relevant achievements using RNA-guided Cas endonucleases (Zhang et al. 2017b). In this study, all three EDR1 homeoalleles of hexaploid wheat were disrupted. Also in tomato, powdery mildew resistance was achieved by knockout of the susceptibility gene MLO1 (Nekrasov et al. 2017).

**Improvement of nutritional value**

Site-directed genome modification has also been shown to be instrumental for the improvement of the nutritional quality of feed and food. For example, Cas9-mediated knockout of the maize IPK gene resulted in a reduction of phytic acid biosynthesis, which is a phosphorous storage compound indigestible by mammals (Liang et al. 2014). Targeted knockout of the SBEI and SBEIIb genes of rice resulted in a substantial reduction of the amyllopectin content in favour of the amylose fraction in grains, which may contribute to reduced occurrence of diabetes II, since the resultant, so-called “resistant” starch is preferentially digested in the colon rather than in the small intestine (Sun et al. 2017).

A comparatively challenging approach was taken by Sánchez-León et al. (2018), who simultaneously knocked out as many as 35 of the α-gliadin genes of bread wheat, leading to low-gluten content of grains with much reduced immunoreactivity. The fatty acid composition has a significant impact on human health. In the emerging oil seed plant Camelina sativa, the FAD2 gene was knocked out thereby increasing the fraction of monounsaturated oleic acid to over 50% at the cost of less desirable polyunsaturated fatty acids (Jiang et al. 2017; Morineau et al. 2017). This approach is another example where the power of Cas9 technology was essential, because three homeoalleles of the hexaploid plant species had to be mutated. In yet another study in Camelina, the content of very long fatty acids was reduced by knocking out the FAE1 gene (Ozseyhan et al. 2018). An enrichment of the nutritionally highly valuable compound lycopene was achieved in tomato fruits via multiplex site-directed mutagenesis of several genes involved in the carotenoid biosynthesis pathway (Li et al. 2018b).

**Industrial plant products**

The formation of amyllopectin-enriched starch was demonstrated by knockout of the potato GBSS gene (Andersson et al. 2017, 2018). Due to the resultant reduction in the amylose fraction, a particular starch quality was obtained that is highly useful for the production of paper and other materials. The same principle was also used to produce a new maize variety, in which the GBSS gene was knocked out in a current elite genetic background to make waxy corn with higher agricultural performance, as compared to conventional breeds with the same trait. This crop is likely to be amongst the very first ones produced by Cas endonuclease technology to enter the market (Waltz 2016). An example of generating plants
with improved utility for molecular farming approaches was provided by Kapusi et al. (2017), who altered the N-glycosylation of barley-produced proteins by Cas9-mediated knockout of \textit{ENGase}.

**Plant breeding technology**

Targeted genome modification has also been used to improve plant breeding technology. In the context of hybrid breeding, the opportunity of switching between male fertility and sterility is an essential prerequisite. In numerous studies, different genes of maize (\textit{MS26, MS45}) and rice (\textit{CSA, TMS5, PKS2}) involved in pollen functionality have been knocked out to form male sterility (Svitashev et al. 2015; Li et al. 2016c; Svitashev et al. 2016; Zhou et al. 2016; Zou et al. 2018). In this context, a novel gene dosage-dependent mechanism for indica-japonica rice hybrid male sterility was revealed by knockout of one or two of the three tandem-repeated copies of the \textit{Sc-i} gene of indica rice, which may facilitate the utilization of heterosis in hybrid rice (Shen et al. 2017). Haploid technology also belongs to the most powerful means to support plant breeding programs. A recent study showed that haploidy inducer lines of rice can be established via knockout of the rice ortholog of the maize \textit{MATL} gene. When the \textit{matl} mutant lines were used as pollinator, as much as 6% of the progeny proved to be haploids (Yao et al. 2018).

**MODIFICATIONS AND EXTENSIONS OF CAS ENDONUCLEASES**

Due to the presence of two individual DNA cleavage domains, both Cas9 and Cas12a enzymes can be converted into nickases (nCas9/nCas12a) by the induction of point mutations in the catalytic sites of either the HNH or RuvC domain of Cas9, or in the Nuc domain of Cas12a, respectively. Thus, these nickases are able to induce site-specific DNA single-strand breaks (SSBs) (Jinek et al. 2012; Yamano et al. 2016). While publications for the utilization of nCas12a are not yet available, nCas9 was tested in different approaches. To induce a DSB with single-strand overhangs, the nickase can be combined with two gRNAs, each recognizing one target on one of both DNA double strands with few base pairs distance from each other. Thus, both nickases generate an adjacent SSB, resulting in a DSB with single-strand overhangs (Figure 3).

Attempts to boost \textit{in planta} Gene Targeting (GT), by nCas-mediated induction of SSBs instead of DSBs, did not result in comparable targeting efficiencies (Wolter et al. 2018). Consequently, it is tempting to speculate whether the utilization of Cas12a, with its intrinsic property to induce short 5’-overhang DSBs, can improve the efficiency of \textit{in planta} (GT), over the frequencies obtained with Cas9 (Wolter et al. 2018). Due to its cleavage activity distal to its PAM, multiple rounds of DSB induction are possible after InDel formation, via NHEJ repair, as the crRNA is more tolerant to
mismatches at this side of the target sequence. Therefore, the crRNA might bind again and a stimulation of the targeted site for HR-mediated GT is still possible (Schindele et al. 2018).

A further advantage of engineering a DSB with paired nickases is the increase in specificity to the target by duplicating the recognition site requirements from 23 to 46 nt in total (Mali et al. 2013; Ran et al. 2013). Moreover, potentially induced SSBs, at off-target sites, can be accurately repaired by the base excision repair pathway, thus any unintended sequence alterations in non-target regions can be abolished (Fauser et al. 2014). For optimal mutagenic results, the two SSBs should be generated at a distance of 50-100 bp to each other, with most mutagenic outcomes being deletions, compared to mostly small insertions with the Cas9 nuclease. The induction of DSBs by paired SSBs leads to single-strand-overhangs that are degraded, leading to longer deletions. Besides deletions, the formation of sequence duplications was discovered in plants directly at the DNA breakage, presumably caused in accord the patch-mediated double-strand-break creation model followed by NHEJ (Vaughn and Bennetzen 2014; Schiml et al. 2016).

Early studies with Cas9 revealed a relatively high off-target DNA cleavage frequency in human cells (Fu et al. 2013). To reduce such off-target activity, it is advisable to use bioinformatics tools to predict potential off-target sites within the genome. Interestingly, hardly any off-target events were identified so far in plants by sequencing respective sites (Feng et al. 2014). Likewise, large-scale whole-genome sequencing analysis showed barely any off-target events for SpyCas9 and LbCas12a (Tang et al. 2018). Nevertheless, off-target activity remains a concern to be considered, especially for targeted nuclease applications in crops with large and partly polyploid genomes, such as maize, rice and soybean (Scheben et al. 2017), where highly repetitive sequences represent special challenges for site-specific mutagenesis or GT. Therefore, other efforts to improve the on-target/off-target ratio even further focus on the gRNA-target binding.

In particular, the binding strength between Cas9 and the target DNA appears to influence the specificity of the nuclease. Therefore, several attempts have been undertaken to influence the strength of this bond by truncating the sequence of the gRNA at its 5’-end (Fu et al. 2014), or to destabilize the interaction between the helicase domain and the non-complementary DNA strand by mutations between the HNH and RuvC domains in SpyCas9 (Slaymaker et al. 2016). Truncation of the guide sequences revealed that a length of 17 or 18 nt exhibit efficient on-target recognition and a reduction in off-target cleavage (Fu et al. 2014). Structural analyses of SpyCas9 showed that the nuclease forms contacts, including hydrogen bonds, to the phosphate backbone of the target DNA (Anders et al. 2014; Nishimasu et al. 2014), and it was tested whether the disruption of these contacts is able to influence cleavage of off-target sites (Kleinstiver et al. 2016).

To weaken the binding between protein and target DNA, single amino acids in the SpyCas9-DNA contact sites were exchanged. This disruption might affect the binding-energy of the complex, resulting in the induction of DSBs with perfect matches with the target DNA, while the bond is not sufficient to
perform a break when mismatches are present. This so-called SpCas9-HF (HF: high fidelity) variant shows comparable on-target DNA cleavage to wild-type SpyCas9, whereas almost no off-target events where observed (Kleinstiver et al. 2016). Previous studies from the same group revealed that mutations in the PAM interacting domain lead to more specificity in the PAM recognition.

These findings suggest less off-target events compared to wild-type SpyCas9. In addition, these mutations cause an alteration in the PAM site to be recognized (Kleinstiver et al. 2015). Further SpyCas9 variants (xCas9) were evolved using phage-assisted continuous evolution (PACE). The enzymes developed in this way recognize a range of different PAM motifs including NG, GAA and GAT, and nevertheless show significantly higher DNA specificity than SpyCas9. Moreover, the off-target activity is considerably reduced compared to the commonly used SpyCas9 (Hu et al. 2018).

In addition to the utilization of orthologous Cas9 proteins, altered PAM sites extend the selection of potential target sites. The commonly used Cas9 nuclease from *Streptococcus pyogenes* is naturally recognizing ‘NGG’ as PAM sequence; however, many different Cas9 orthologs with various PAM requirements have also been characterized. The Cas9 orthologs from *Streptococcus thermophilus* and *Staphylococcus aureus* seemed most promising for genome editing (Esvelt et al. 2013; Kleinstiver et al. 2015; Ran et al. 2015) due to the variability of their PAM sequences, by which a broader range of accessible targets is provided (Ran et al. 2015). The Cas9 orthologs of *Staphylococcus aureus* (SauCas9) and *Streptococcus thermophilus* (Sth1Cas9) require longer PAM sequences than SpyCas9 (SauCas9: NNGGGT; Sth1Cas9: NNRGAA), which should reduce the rate of potential off-target events (Steinert et al. 2015). Another advantage for both of these orthologs is the smaller size with 3.2 kb for SauCas9 and 3.4 kb for Sth1Cas9, compared to the 4.1 kb from SpyCas9, which is especially favorable when the delivery is performed via viral-based systems (Baltes et al. 2014; Ran et al. 2015).

Two different CRISPR open reading frames (ORFs) were identified for SthCas9 orthologous to SpyCas9, called CRISPR1 and CRISPR3. CRISPR1 is smaller and shows a higher activity compared to CRISPR3 (Horvath et al. 2008) and has been successfully used in bacteria, human cells and plants (Jinek et al. 2012; Esvelt et al. 2013; Kleinstiver et al. 2015; Ma et al. 2015a; Ran et al. 2015; Steinert et al. 2015). The application of Sth1- and SauCas9 in plants revealed that both orthologs achieved targeted mutagenesis, via NHEJ, with similar high efficiency as with SpyCas9, and for SauCas9 an even higher mutagenesis rate was established in *Arabidopsis* (Steinert et al. 2015). SpyCas9, SthCas9 and SauCas9 belong to the class 2 type II CRISPR/Cas system and are also divided into the same subtype A, but are assigned to different clusters (Fonfara et al. 2014; Louwen et al. 2014). The mentioned orthologs differ in the sequences of their associated RNAs (Ran et al. 2015) and form special structures, like hairpins or additional stem loop structures, which are required for species-specific detection (Briner et al. 2014).

In addition to their DNA cleavage properties, Cas proteins are very versatile tools. As mentioned previously, nuclease domains can be deactivated by point mutations within the active sites. For Cas9, inducing point mutations in both the RuvC and HNH domains is sufficient to obtain a site-specific DNA
binding protein, termed “deadCas9” (dCas9) (Jinek et al. 2012; Qi et al. 2013). This is also applicable for Cas12a, although only a point mutation within the RuvC domain is required (dCas12a) (Zetsche et al. 2015). As a result, the Cas proteins can be targeted to almost any desired DNA locus, by exchanging 20–23 nucleotides of the gRNA, for binding without cleavage.

Functional units can be fused to the dCas proteins, allocating a desired effect at a specific locus of the DNA by targeted binding. Thereby, these enzymes can specifically repress gene expression by sterically blocking the transcription machinery, as was shown in *Escherichia coli* and in human cells (Qi et al. 2013; Zhang et al. 2017a), or be used as a platform to guide any enzymatic activities to the desired DNA loci (Figure 4). This property can be used for catalytic activities, such as the visualization of specific genomic loci of living cells by the fusion of a fluorescent protein to a dCas protein (Chen et al. 2013). The application of multiple Cas9 orthologs fused with fluorescent proteins of different colors enabled the multicolored detection of various genomic loci within mammalian cells (Anton et al. 2014; Ma et al. 2015a; Chen et al. 2016), and also recently in plant cells (Dreissig et al. 2017).

With the intention of obtaining a detectable fluorescent signal, a novel repeating peptide array, called SunTag, was developed to serve as a platform for multiple fluorescent protein-antibody fusions (Tanenbaum et al. 2014). The linkage of a catalytic effector to Cas9 can either be achieved by direct fusion or indirect fusion, which includes a modified gRNA with sequence-specific aptamer motifs enabling the binding of the catalytic effectors to the gRNA. Protein domains such as MS2, originating from an RNA phage coat protein, recognize and specifically bind this RNA sequence motif (Peabody 1993). By fusing the respective aptamer-binding domain to a catalytic effector, the complex is also directed to the DNA in a sequence-specific manner by gRNA-Cas9 interaction.

Aptamer-binding domains bind as dimers on the respective aptamer sequence, whereby twice as many effectors can bind, compared to a direct Cas fusion. Another advantage is that various aptamers and their corresponding binding domains can be used, enabling the utilization of several effector proteins, simultaneously (Konermann et al. 2015). Thus, an approach similar to the SunTag system, to increase the fluorescent signal by the integration of up to 16 MS2 binding motifs to the gRNA, is especially suitable for imaging non-repetitive targets (Qin et al. 2017).

In order to improve the gene repression properties of dCas9 in human cells, a repressive KRAB domain was fused to Spy-dCas9 and showed reduced expression of the targeted genes. To test the so-called CRISPRi platform, as a versatile tool for specific transcriptional regulation, the same group fused the transcription activator domain VP64 to dCas9 and was able to activate reporter gene expression (Gilbert et al. 2013). Targeting multiple loci of one gene of interest by applying various gRNAs can enhance the desired effect (Perez-Pinera et al. 2013). Cas9-based specific transcriptional regulators fused with the activator domains VP64, EDLL and TAL effectors or the SRDX repressor domain were
also successfully applied in plants, to either activate or suppress endogenous genes in *Nicotiana benthamiana* leaves and *Arabidopsis* plants (Lowder et al. 2015; Piatek et al. 2015).

It was shown that the highest repression was obtained by using three different gRNAs, targeting the promotor region and first exon of the endogenous gene in *N. benthamiana* leaves (Piatek et al. 2015). The VP64 transcriptional activator domain fused to dCas9 leads to a strong activated expression of a reporter gene in tobacco leaves, as well as in protein-coding and non-protein-coding genes in *Arabidopsis* plants (Lowder et al. 2015). Through recruitment of multiple transcriptional activators, the dCas-based transcriptional activator system can be improved further (Konermann et al. 2015; Zalatan et al. 2015), as shown using the SunTag system (Tanenbaum et al. 2014). Furthermore, indirect fusions of transcriptional activators, like VP64 to modified gRNAs, have allowed robust transcriptional multiplex gene activation, *in planta* (Lowder et al. 2018).

In parallel, Cas12a is also suitable as a DNA-binding protein and for delivering catalytic effectors to targeted genomic locations. Lb-dCas12a was fused to either the strong synthetic VPR activator derived from the VP16 activator, Rta- or p65 activator domains, for robust transcriptional activation in human cells. In the same study, it was demonstrated that transcriptional activation is inducible by drug-treatment, using the split Lb-dCas12a-DmrA complex and a DmrC-VPR or -p65 complex. Taking advantage that Cas12a can process its own crRNA in a multiplex single transcript, the same group confirmed the synergistic activation when targeting either the same or different promoters (Tak et al. 2017). In *Arabidopsis*, dCas12a was already successfully applied as a transcriptional repressor fused with the SRDX repressor domain. The results indicate that Lb-dCas12a, and particularly As-dCas12a, can efficiently bind to the DNA and regulate gene expression (Tang et al. 2017).

Regulation of gene expression can also be achieved by the fusion of epigenetic regulators to Cas effectors causing chromatin manipulation. The fusion of histone acetyltransferase p300 to the C-terminus of dCas9 enabled transcriptional activation of targeted genes, in human cells, by catalyzing acetylation of histone H3 lysine 27 (H3K27). Binding of the dCas effector protein, at the selected promoter or enhancer region, activates the targeted gene by specific chromatin modification (Hilton et al. 2015). Fusion of the p300 core domain to Lb-dCas12a showed robust and specific gene activation in human cells and was able to activate three genes, simultaneously, using three different gRNAs (Zhang et al. 2017a). Vice versa, targeted epigenetic repression in mammalian cells can be achieved by demethylation of H3K4/K9 through the histone demethylase LSD1 (Kearns et al. 2015). The potential of developing new epialleles for generating traits of interest has been successfully demonstrated in *Arabidopsis* by targeted demethylation of a specific cytosine residue in the promoter of the *FWA* gene, resulting in a heritable late-flowering phenotype (Gallego-Bartolomé et al. 2018).

In plants, genome editing, based on DSB-induced HR, to incorporate a template with a desired sequence modification can be a tedious and inefficient process; thus, the newly developed Cas-dependent ‘base editing’ could be an adjuvant solution for some traits to be developed. This method is
based on the properties of cytidine deaminases or adenosine deaminases and allows for the directed conversion of one targeted DNA base. The fusion of a cytidine deaminase enzyme to a dCas9 effector, or a Cas9 nickase, mediates the targeted conversion of cytidine to uridine, resulting in a C-to-T substitution. Therefore, converting base pairs C/G to T/A does not require any DSBs nor HR-mediated repair processes. Cytidine deaminases require RNA or single-strand DNA as substrate, therefore the Cas effector is not only the key for targeted DNA binding, but also for unravelling of the DNA during R-loop formation.

Experimentally, the cytidine deaminase rAPOBEC1 was fused to the N-terminus of Spy-dCas9 (BE1) and Spy-nCas9. Fusion of the effector to a nickase causes the incorporation of the new base in the non-edited strand, while the SSB-induced mismatch repair mechanism (MMR) synthesizes along the edited strand. To counteract the intrinsic cellular repair mechanism and optimize the editing result, a uracil DNA glycosylase inhibitor (UGI) was fused to the C-terminus of the targeting tool to suppress base-excision repair (BE2/ BE3) (BER) (Komor et al. 2016) by preventing a possible removal of uracil, as the most deaminated bases are restored by BER (Liu [Man] 2009).

Currently, the commonly used tool for base editing is APOBEC-XTEN-nCas9-UGI (BE3) (Komor et al. 2016). Whereas the BE3 approach catalyzes cytidine deamination within a window of 5 bp around the -15 position upstream of the PAM sequence (Komor et al. 2016), similar results were obtained with a C-terminal fusion of an activation-induced cytidine deaminase (AID), which catalyzes deamination 3–5 bp around -18 position upstream of the PAM sequence (Nishida et al. 2016). Based on the BE3 approach, with APOBEC or the one with AID, many different base editing attempts have been undertaken; e.g., by targeted induction of STOP codons for gene disruption in human cells, silkworm or E. coli (Billon et al. 2017; Kuscu et al. 2017; Li et al. 2018d; Zheng et al. 2018). Furthermore, base editing attempts were also achieved with the BE3 system in such crop plants as rice, wheat and maize (Li et al. 2017a; Lu and Zhu 2017; Zong et al. 2017; Ren et al. 2018; Yan et al. 2018). Thus, herbicide-resistant rice as well as marker-free edited tomato plants were obtained through the BE-AID approach (Shimatani et al. 2017).

To expand base editing applications, further variants (BE4) have been developed using a Cas9 ortholog of Staphylococcus aureus (SaBE4), which increases the efficiency of the conversion from C/G to T/A. The unintended formation of InDels during editing could be successfully reduced by fusing the DSB binding bacteriophage Mu protein Gam, to BE3 and BE4 systems (Komor et al. 2017). The challenge of off-target editing events was addressed by combining the increased target sequence specificity requirements of Cas9-HF with the BE3 system in a study delivering the editor as RNPs (Rees et al. 2017). Although many different cytidine deaminases exist, no naturally adenine deaminases are known (Gaudelli et al. 2017). The development of a tRNA adenosine deaminase to mediate the conversion of A/T to G/C in genomic DNA expands the base editing tool portfolio by combining it with the BE3 and BE4 base editing designs, respectively (Gaudelli et al. 2017). The application of LbCas12a
and AsCas12a as “dead-” and nickase-variants in the aforementioned BE-variants also provides efficient options to alter the DNA sequence specifically without the induction of DSBs (Li et al. 2018c).

Next to the targeted editing of individual bases to modify genes, the Cas effector tools developed so far also allow for sequence specific regulation of transcription, utilizing the recently discovered Cas13 effector, which operates exclusively on RNA and mediates its cleavage. Three Cas13 orthologs studied so far showed efficient targeted RNA knockdown comparable to RNAi, whereas these proteins exhibit a higher specificity. LwaCas13a was applied for targeted RNA knockdown in rice protoplasts with up to 50% knockdown efficiency, 48 h post-transformation.

Post-transcriptional gene knockdown, instead of transcriptional regulation, can be interesting when only certain splicing variants are to be regulated, as all isoforms are repressed during transcriptional regulation (Mahas et al. 2018). Additionally, gene activity suppression by post-transcriptional regulation provides a fast, effective and highly specific means of regulation: existing cytoplasmic mRNAs get cleaved, whereas transcriptional regulation suppresses the production of further mRNA generations, while already produced mRNAs remain active (Wolter and Puchta 2018; Schindele et al. 2018). An applied example of such post-transcriptional repression was recently shown to successfully combat an RNA virus in plants, using the Cas13a ribonuclease from \textit{Leptotrichia shahii} (LshCas13a). \textit{Nicotiana benthamiana} leaves were infiltrated with a \textit{Turnip mosaic virus} (TuMV), LshCas13a and crRNAs targeting the virus. Seven days-post-inoculation, the viral GFP reporter signal was reduced by 50%, whereas plant vitality was not affected. However, there was a strong variance in the effectiveness of tested crRNAs, indicating a significant influence of the secondary structure of the mRNA on the cleavage activity (Aman et al. 2018). Compared to LwaCas13a and PspCas13b, the recently discovered RNA-binding effector EsCas13d has so far achieved the strongest gene knockdown in human cells (Konermann et al. 2018). Whether Cas13d is as efficient in plants has yet to be tested.

Similar to Cas9 and Cas12, the catalytic residues in the HEPN domains of Cas13 can be deactivated by induction of point mutations resulting in a dCas13, a site-specific RNA binding protein. These RNA binding proteins can also be fused to functional units; e.g., fluorescent proteins to allow for specific imaging of RNAs in live cells (Abudayyeh et al. 2017). Furthermore, dCas13 enables precise RNA base editing by the fusion of an adenosine deaminase, as previously shown with Cas9 (Cox et al. 2017). In this way, sequence alterations can be incorporated, at the transcript level, without a permanent modification of the encoding genome sequence.

\textbf{PERSPECTIVE}

\textit{Multiple Cas applications in parallel (3D principle)}
The intrinsic specificity of Cas orthologs for their corresponding gRNA, combined with the availability of modified proteins (nickase, dCas) and the feasibility of directly and indirectly fusing effector proteins, or functional domains, allows for not only simultaneous targeting of various loci in a genome, but also delivering locally different enzymatic activities to elevate, or suppress transcription of genes, edit individual bases and, thus, modify the ORF of a given gene to improve the encoded protein functions, or enable epigenetic modifications. Together with the site-specific induction of SSBs or DSBs and the possibility to transiently modify the required DNA repair machinery, we now have efficient and specific multifunctional tools available that can be applied to modulate complete pathways and, thus, change the metabolome to synthetically improve crop traits. To improve the efficiency of such highly sophisticated genome reshuffling efforts, 3D portfolio properties of Cas-orthologs could be exploited (Puchta 2016). Such an approach could enable boosting specific DSB repair pathways, by transient upregulation of desired repair factors, or downregulating competing pathways, which might shift the general DSB repair equilibrium in favor of the desired outcome.

Approaches of this nature for orthogonal regulation of gene expression became even more flexible with the availability of fusion gRNAs (fgRNAs) that can recruit different Cas orthologs, simultaneously, and thus, facilitate multiplexing with no detectable increase in off-target activity (Kweon et al. 2017). Additionally, the successful application of an orthogonal tri-functional CRISPR system in \textit{S. cerevisiae} for metabolic engineering, unambiguously demonstrates the versatility of combining different targeted activities, simultaneously, in one cell (Lian et al. 2017). Another intriguing possibility has been described as the proxy-CRISPR approach, which combines dCas9 orthologs to change chromatin structure, in a genomic region of interest, to restore the nuclease activity of another targeted effector, therein.

This targeted activation bears strong potential to selectively edit a single gene in a genome harboring multiple identical copies, by adapting the adjacent chromatin microenvironment (Chen et al. 2017). Furthermore, modulation of gene expression can be achieved by targeting engineered bivalent dCas9 complexes to distal cis-regulatory elements in order to support the formation of large DNA loops and, thereby, improving enhancer-promoter contact, as demonstrated in \textit{E. coli} reporter strains (Hao et al. 2017). Additionally, specific recombination factors might be targeted to the loci needed, to locally enhance the availability of required repair processes, as recently demonstrated in mammalian cells for the key HR factor CtIP fused to Cas9 (Charpentier et al. 2018). Furthermore, we can envisage combining the aforementioned factors and even physically enhancing the probability of desired recombination events by directly, or indirectly fusing Cas effectors to tether chromosomal fragments of choice.

**Applications for highly sophisticated genome restructuring and breeding acceleration**
The ability to induce multiple simultaneous DSBs in one genome, by targeting Cas9 nucleases to more than one locus, allows for the precise induction of segmental chromosome inversions, or reciprocal translocations. Inversions of genomic sequences would not change the overall gene pool of a given species, however, it might pose a means to render certain regions silent for meiotic recombination, due to the rearranged sequence context. Such artificially created recombination blocks might be suitable to stabilize advantageous alleles in breeding processes and, thus, might reduce developmental costs. Next to that, targeted induction of reciprocal translocations between non-homologous chromosomes can be applied to create new tight linkage groups between advantageous traits to stabilize these as one unit for breeding purposes. Vice versa, undesired traits occurring in natural close linkages can be broken to allow simple separation by segregation. The technology can be applied in a way that the obtained recombinants do not contain any transgenic sequences. Thus, the molecular tools described above should allow for reshaping genomes in order to better exploit the natural gene pool and, additionally, to accelerate crop breeding processes.

ACKNOWLEDGMENTS

We thank Amy Whitbread for critical comments on the manuscript. Funding of our cooperative research by the German Federal Ministry of Education and Research (FKZ 031B0192) is gratefully acknowledged.
REFERENCES


Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J* 79:348–359


Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci USA 93:5055–5060


Schiml S, Fauser F, Puchta H (2014) The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. *Plant J* 80:1139–1150


Xie K, Minkenberg B, Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc Natl Acad Sci USA 112:3570–3575


**Figure legends:**

**Figure 1. Schematic representation of the three CRISPR Class 2 single-protein effectors, Cas9, Cas12a and Cas13a**

(A) After hybridization of the small RNAs, tracrRNA and crRNA, a complex with Cas9 is formed. The crRNA contains a 20 nucleotide guide sequence (Spacer) and mediates sequence-specific DNA binding at the recognition site upstream of the PAM motif. Subsequent to Cas9 recruitment to the target DNA, a sequence-specific DSB is induced. (B) In contrast to Cas9, Cas12a requires only a single crRNA for DNA targeting and the 23-25 nucleotide recognition site is located downstream of the PAM sequence. The Cas12a nuclease-induced DSB possesses 5 nt single-strand DNA overhangs, 18 to 23 nucleotides distal and downstream to the PAM in the spacer region. (C) The Cas13a-mediated RNA-specific cleavage activity requires one single crRNA. Binding of the Cas13a/crRNA complex to its target RNA is mediated by the guide sequence of the crRNA. RNA cleavage occurs at the catalytically active sites of Cas13a located on the outer protein surface. (A–C) All three single-protein effectors can be specifically directed to almost any target DNA/RNA by exchanging the specific recognition sequences of the crRNA.

**Figure 2. Secondary structure of single guide RNAs**

(A) Example of guide RNA with high functionality due to limited base pairing of the target-specific part and to the presence of all three stem loop structures required to correctly interact with Cas9 endonuclease. Probability of the individual bases to pair is indicated by a color code. (B) For comparison, a guide RNA with severely compromised functionality, owing to a high number of paired nucleobases of the target-specific part and to the absence of an essential stem loop structure (namely the 3rd one according to A). The structure models were generated using the RNAfold online platform (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi).

**Figure 3. Cas9 paired nickases approach**

A single point mutation renders one nuclease domain of Cas9 inactive, resulting in the generation of a SSB-inducing enzyme (nCas9, nickase). Providing two different gRNAs targeting adjacent positions on opposite DNA strands, two SSBs can be obtained in close proximity, resulting in a double strand break with single-stranded overhangs.
Figure 4. Applications for catalytically inactive Cas9 DNA-binding proteins

The nuclease domains of Cas9 can be deactivated by point mutations to convert the protein into a sequence-specific DNA-binding protein. The resultant inactive Cas9 (dCas9) can serve as a platform for targeting other effectors to the DNA by adapting the single guide RNA (sgRNA). These effectors can be fused to dCas9, via a short peptide linker, and perform various functions on the DNA. By fusing fluorescent proteins to dCas9 (A), specific DNA sequences can be labeled. The fusion of transcriptional repressors (B), or activators (C), allows for targeted modulation of gene expression. Gene activity can also be regulated, epigenetically, by fusing histone modifiers to dCas9 (D). Single bases can be edited by the targeted activity of deaminases (E). Alternatively, factors can be indirectly bound to extended crRNAs (aptamers), via compatible peptide motifs (RBP) (F). Instead of direct protein-linker-mediated fusion of effectors to dCas9, factors can be indirectly bound to extended crRNAs (aptamers), via compatible peptide motifs.
Figure 1

A

B

C

Figure 1
Figure 4

A. Imaging
B. Transcriptional repression
C. Transcriptional activation
D. Epigenetic regulation
E. Base editing
F. Aptamer fusion