Chapter 2

Application of CRISPR/Cas to Understand Cis- and Trans-Regulatory Elements in Plants

Felix Wolter and Holger Puchta

Abstract

The recent emergence of the CRISPR/Cas system as a genome editing tool enables simple, fast, and efficient induction of DNA double-strand breaks at precise positions in the genome. This has proven extremely useful for analysis and modification of protein-coding sequences. Regulatory sequences have received much less attention, but can now be quickly and easily disrupted as well. Editing of cis-regulatory elements (CRE) offers considerable potential for crop improvement via fine-tuning of gene expression that cannot be achieved by simple KO mutations, but its widespread application is still hampered by a lack of precise knowledge about functional motifs in CRE. As demonstrated for mammalian cells, CRISPR/Cas is also extremely useful for the identification and analysis of CRE in their native environment on a large scale using tiling screens. Transcriptional complexes are another promising target for crop genome editing, as demonstrated for pathogen resistance and regulation of flowering. The development of more diverse and sophisticated CRISPR/Cas tools for genome editing will allow even more efficient and powerful approaches for editing of regulatory sequences in the future.

Key words Cis-regulatory element, Promoter, Enhancer, Transcription factor, CRISPR/Cas, Genome editing

1 Introduction to CRISPR/Cas

The recent emergence of the CRISPR/Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated) system as a biotechnological tool has revolutionized molecular biology. Its virtue lies in its ability to precisely and efficiently induce DNA double-strand breaks (DSB) at any site in the genome. Although targeted DSB induction was possible before with engineered proteins like zinc-finger nucleases [1] and transcription-activator like effector nucleases (TALENs) [2], the emergence of the CRISPR/Cas system as a genome editing tool was a breakthrough: its outstanding advantage is RNA-guided nuclease specificity, enabling a very simple, fast, and cheap design process. After DSB induction, the cell's own DSB repair mechanisms, categorized into nonhomologous end joining (NHEJ) and homology-directed repair

Nobutoshi Yamaguchi (ed.), *Plant Transcription Factors: Methods and Protocols*, Methods in Molecular Biology, vol. 1830, https://doi.org/10.1007/978-1-4939-8657-6_2,

(HDR), can be harnessed for genome editing purposes, which was demonstrated already 20 years ago [3, 4]. NHEJ is non-template directed, and involves direct re-ligation of the exposed DNA ends. This is error prone and frequently causes small insertions or deletions (InDels), which can lead to gene knockout (KO) by frameshift mutations. HDR-mediated repair of the DSB is error-free but requires the availability of a template with homology to the DSB. Naturally, this is the sister chromatid or the homologous chromosome. However, in a technique known as gene targeting, a donor DNA molecule containing a desired change as well as sequences homologous to the insertion site serves as artificial repair template for the cell, leading to precise modifications at the desired locus. This process can be exploited for both insertion of additional sequences or precise alterations of the existing sequence [5, 6].

The CRISPR/Cas system is based on an adaptive immune system found in prokaryotes. CRISPR arrays were already identified in the *Escherichia coli* genome in the 1980s [7]. They are found in 40% of sequenced bacterial and 90% of sequenced archaeal genomes [8]. However, their biological function remained unknown for a long time. Only in 2005 the suggestion of a role in adaptive immunity was made due to homology between the spacers in the CRISPR array and viral sequences [9]. CRISPR/Cas systems are highly diverse, current classification differentiates two classes, six types and 19 subtypes of CRISPR [10], with the typeII system of *Streptococcus pyogenes* most commonly used for biotechnological purposes. In the natural system, a complex of the nuclease Cas9 and two short RNAs, tracrRNA and crRNA, is able to cleave foreign plasmid or viral DNA entering the cell, with the latter determining specificity.

The transition of the CRISPR/Cas system from a biological phenomenon to a genome engineering tool was enabled by the work of Jinek et al. [11]. They successfully reprogrammed the specificity of the crRNA by changing the 20 nucleotides responsible for DNA targeting. Furthermore, they combined the properties of the crRNA (specificity) and the tracrRNA (structural stability) in a single chimeric guide RNA (sgRNA) by fusing the 3′end of the crRNA to the 5′end of the tracrRNA with a GAAA tetraloop (*see* Fig. 1). At the same time, all regions of the crRNA and tracrRNA not required for guiding Cas9-mediated DNA cleavage were truncated. This further reduced the complexity of the system from three to two components, at increased efficiency. Cleavage is dependent on the presence of a short sequence motif next to the target sequence termed protospacer adjacent motif (PAM), which is NGG for *S.p.* Cas9, although NAG can also be recognized at low efficiency. Cas9 exhibits 2 nuclease domains, RuvC and HNH, each cleaving one strand 3 bp upstream of the PAM [11].

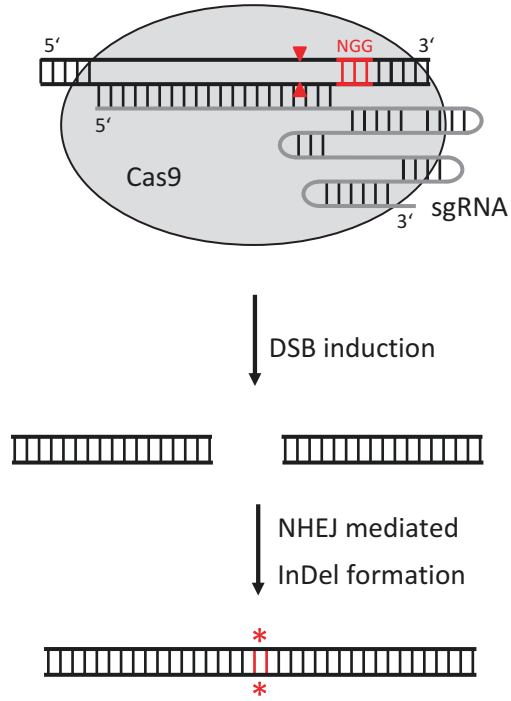

Fig. 1 Using CRISPR/Cas for targeted mutagenesis. The sgRNA is programmable via alteration of its 5′ 20 nt. The custom sgRNA is expressed in addition to Cas9 which forms a complex with the sgRNA that specifically binds the target sequence. Two different nuclease domains of Cas9 then separately cleave the two DNA strands, leading to a blunt DSB. Most breaks are repaired via NHEJ which usually leads to small insertions or deletions (InDels)

After the first successful demonstration of CRISPR/Cas functionality in human cells [12], the system was successfully applied to plant cells [13–15]. The demonstration of stable inheritance of induced mutations in plants followed the year after [16, 17]. Up to now, CRISPR/Cas was primarily used for targeted mutagenesis to disrupt gene function, which is now a routine procedure. This is mainly used to analyze gene function, but it can also be used to improve traits in crops (see Scheben et al. [18], for an overview). However, the CRISPR/Cas system is a much more versatile tool [19]. Cas9 can be transformed to a DNA-binding protein by mutating its two nuclease domains, thereby deactivating its DNA cleavage activity without affecting its DNA-binding potential. By fusing this nuclease "dead" Cas9 (dCas9) to an effector domain, it can then be used to guide diverse enzymatic functions to any specific site in the genome [20]. The effector domain can also be fused to an RNA-binding protein that is able to bind an aptamer sequence integrated into the sgRNA [21]. Either way, it is now possible to use CRISPR/Cas for manipulation of expression of any genes [22, 23], for site-specific manipulation of epigenetic modifications [24, 25], or for

site-specific imaging of genomic loci in live cells [26]. Recently, a cytidine deaminase fusion was successfully targeted to specific genomic sites using CRISPR/Cas, which enables site-specific "base-editing" without DSB induction by conversion of cytidine to uridine, ultimately leading to substitution of cytosine with thymine or guanine with adenine [27, 28].

For a long time, regulatory sequences have taken a back seat to protein-coding sequences. But, in recent years, evidence that they are crucial to a wide array of phenotypic traits, including human disease, piled up, together with a requirement for improved methods to characterize their function [29]. The CRISPR/Cas system now opens the possibility to rapidly and easily disrupt regulatory sequences on a large scale. Accordingly, regulatory sequences can be studied in their native environment, which will help to unlock the still neglected potential of editing the regulatory part of the genome.

2 Using CRISPR/Cas to Modify Cis-Elements

Already in the 1960s, it was hypothesized that evolutionary change is driven to a large extent by mutations that influence gene expression [30]. Now, it is clear that changes in gene expression are sufficient to cause dramatic phenotypic variation. Cis-regulatory-elements (CRE) are noncoding DNA containing binding sites for transcription factors (TFs) or other regulatory molecules that affect transcription. The most common and best studied examples are promoters and enhancers. While promoters are usually bound by a standard set of highly conserved transcriptional regulators, enhancers tend to be much more variable. Thus, enhancers are assumed to account for most of the cis-regulatory divergence in evolution. Enhancers can be located far away and not only upstream but also downstream and even in the introns of the gene they regulate [31]. In addition to recruiting activating transcription factors, enhancers are able to alter chromatin state and physically interact with target genes [32].

Studies have shown that cis-regulatory mutations conveying altered expression levels are widespread in nature and contributed markedly to the evolutionary process of crop domestication via reshaping of the transcriptome (e.g., [33, 34]). Meyer et al. report that almost half (26 of 60 mutations analyzed) of the mutations associated with crop domestication reside in CRE [35]. By now, even more CRE mutations causative for the process or crop domestication have been described (summarized in [36]). However, for most of the CRE that were mutated during domestication, the interacting TFs or micro RNAs remain unknown. An example of a well-studied case is the *GRAIN WIDTH 7* (*GW7*) gene of rice, which is controlling cell division in the spikelet hulls.

GW7 expression is controlled by the repressing TF GRAIN WIDTH8 (GW8), which binds a CRE in the *GW7* promoter. A mutation in this CRE leads to reduced affinity for GW8 during panicle development, leading to increased *GW7* expression (*see* Fig. 2) and ultimately to the production of a higher amount of more slender grains. Hence, this CRE mutation is associated with enhanced yield and also grain quality, as long slender grains are usually preferred [37, 38].

Complete loss- or gain-of-function mutations frequently display deleterious pleiotropic effects [35]. In contrast, CRE mutations offer the promising possibility to fine-tune gene expression without pleiotropic effects. E.g., the increased fruit size associated with domestication of tomato depended to a large extent on mutations supporting altered expression of multiple components in the classical CLAVATA-WUSCHEL pathway [39]. Consequently, CREs constitute an interesting target for CRISPR/Cas-mediated genome editing. In addition, it should be noted that CREs are likely to be easily targetable with CRISPR/Cas: Using a method called ChIP-seq (chromatin immunoprecipitation followed by sequencing) which determines the genomic sites bound by a known protein, it was found that Cas9 preferentially binds to open chromatin [40]. As CREs are characterized by open chromatin (which can be determined by DNaseI-seq as described below), CREs should be easily accessible for Cas9.

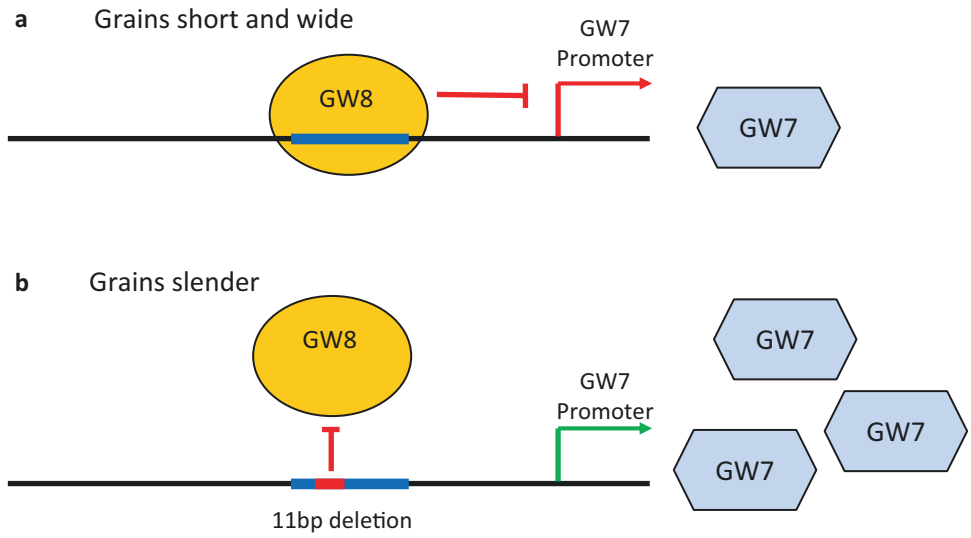

Fig. 2 Example for the potential of CRE variability for crop improvement. The *GW8* transcription factor represses *GW7* expression by binding to a specific motif (blue bar in the figure, core sequence GTACGTAC) in the *GW7* promoter. Reduced *GW7* expression leads to a short and wide grain form (**a**). A 11 bp deletion in the *GW8*-binding site (marked red, 2 bp upstream of the GTACGTAC core) prevents *GW8* binding and consequently its repressive activity. Increased *GW7* expression leads to a slender grain form (**b**) [36, 37]

Up to now there are only a limited number of studies for successful editing of CRE for crop improvement available. Li et al. showed a detailed example [41]. The Transcription Activator-Like Effector (TALE) AvrXa7 from *Xanthomonas oryzae* binds a CRE in the *OsSWEET14/Os-11N3* promoter, activating transcription. Mutagenesis in this CRE using TALE nucleases was able to mediate rice resistance against *Xanthomonas* strains carrying the AvrXa7 effector. Interestingly, complete KO mutants of *OsSWEET14* also mediate resistance, but are accompanied by several defects like delayed growth and small seeds [42], which is a good example to demonstrate that CRE mutagenesis can be superior to complete KO.

Soyk et al. provide a convincing demonstration of the usefulness of CRE mutations [43]. Inflorescence architecture in tomato can be improved by combining two mutations conveying weakened expression of closely related MADS-box TF genes. The first mutation is an insertion of a transposable element in the tomato homolog of the Arabidopsis floral organ identity gene *SEPALLATA4*. The second mutation is a 564 bp insertion in the homolog of Arabidopsis *FRUITFULL*. The improved inflorescence architecture resulted not only in considerably enhanced fruit number and yield (around 30 and 55%, respectively), but also in increased fruit weight (around 20%) while sugar content remained unaltered. Importantly, improved inflorescence architecture in the form of moderately increased branching was dependent on alleles that supported reduced expression, one of them being in a heterozygous state. In contrast, homozygously combining CRISPR/Cas-mediated complete KO alleles resulted in excessively branched inflorescences that produced infertile flowers. However, newly identified weak transcriptional alleles supporting a range of different expression strengths could be generated by targeting CREs of above-mentioned genes. The authors also identify a further promising target for CRE editing, *LONG INFLORESCENCE* (*LIN*), which is another tomato *SEPALLATA4* homolog. Alleles conveying reduced LIN expression might enable subtle increases in the number of flowers [43]. The fact that rice carries a homolog of *LIN* that controls panicle architecture and grain production [44] suggests that the approach might be extended to other crop species.

Unfortunately, a lack of precise information about CRE-TF interactions is hampering the widespread application of CRISPR/Cas for fine-tuning of gene expression via CRE editing. The most important methods to obtain the required knowledge include the following: If a TF involved in a process to be edited is known, chromatin immunoprecipitation followed by sequencing (ChIP-seq) is the method of choice to elucidate the DNA regions bound and the genes regulated by the TF [45]. If knowledge about TFs involved in the process to be edited is lacking, this information can

be inferred from the correlated or anticorrelated expression profile of candidate regulators, an approach that becomes increasingly accessible with the increasing availability of RNA-Seq data [46]. The increasing amount of sequenced plant genomes allows prediction of CREs from sequence conservation [47]. Furthermore, CREs regions can be predicted on a genome-wide scale via DNaseI-seq. This method is mapping chromatin accessibility using the nonspecific endonuclease DNaseI, and TF occupied regions are characterized by a hypersensitivity to DNaseI due to the associated open chromatin structure [48]. However, when candidate enhancers are identified, it is difficult to ascertain their connected target genes since enhancers are often located remote from the genes they regulate. Chromosome conformation capture (3C) is a method that can predict distal enhancers that are brought into spatial proximity to their regulated promoter via DNA looping [49]. Finally, active enhancers feature transcription of their own loci, resulting in so-called enhancer-associated RNAs (eRNAs) [50]. This can be exploited to identify active enhancers, as eRNA expression can be detected by methods that measure nascent RNAs.

3 Using CRISPR/Cas to Identify and Analyze Cis-Elements

Recently, CRISPR/Cas-based methods to find and analyze CREs were developed. Most importantly, these CRISPR-based approaches enable the analysis of CREs in their native chromatin environment. The basic strategy is to tile putative CREs with many sgRNAs to disrupt it on its entire sequence to map its functional domains [51]. This is a valuable expansion to traditional tests of enhancer function that mainly rely on ectopic heterologous reporter assays (e.g., [52, 53]). In the first proof of concept [54], Canver et al. used pooled sgRNA libraries for saturating mutagenesis of the human and mouse *BCL11A* composite enhancer to identify essential motives. They demonstrated that functionally important sequences within the enhancers are highly sensitive: single nucleotide alterations in these sequences can markedly modify gene expression. Furthermore, they investigated the effect of targeted deletions and inversions of individual constituents of this composite enhancer, achieved by inducing two DSBs simultaneously. In a similar approach [55], Vierstra et al. analyzed the same enhancer and were able to derive a consensus sequence for a TF recognition site. Seruggia et al. successfully deleted a genomic insulator upstream of the mouse tyrosinase gene by targeting both flanking sequences, leading to a dramatic decrease in tyrosinase gene expression in genome-edited mice [56]. Duplications and inversions were also demonstrated for mammalian cells [57], in this case for a CRE of the Pcdhα cluster.

After these first proof of concept studies, three more recent studies further developed the tiling screen approach used by Canver et al. [54] for large scale application.

Korkmaz et al. used a high-throughput CRISPR–Cas9 tiling screen (1116gRNAs) to dissect ChIP-Seq predicted distal enhancer regions of the p53 target gene CDKN1A19, an essential tumor suppressor [58]. They constructed a lentiviral sgRNA library for transduction of human BJ cells and performed NGS to detect gRNA enrichment or depletion via their effect on cell proliferation after 4 weeks of culturing. Using this approach, they were able to detect six enhancer elements and precisely map functional domains within enhancer elements.

Whereas Korkmaz et al. focused only on enhancer elements, Rajagopal et al. [59] tiled sgRNAs across a broad range of cis-regulatory regions surrounding target genes (3908 sgRNAs per target gene). They designed a screen named multiplexed editing regulatory assay (MERA) which employs a unique strategy to ensure that a single sgRNA is incorporated into each cell for tiling of the target loci. Here, a dummy sgRNA integrated cell line is constructed, followed by its replacement with library sgRNAs through homologous recombination. They use MERA to characterize CREs and their regulatory effect on the expression of four different mouse genes. In their approach, the investigated genes are tagged by GFP and the effect of CRISPR-induced genomic variation on target gene expression is quantified by GFP cell sorting and NGS of sgRNAs from cell populations with no or diminished GFP expression. Interestingly, they found that neighboring gene promoters can contribute substantially to gene expression and they identified unmarked regulatory elements that control gene expression but do not have typical enhancer epigenetic or chromatin features. These unmarked regulatory elements were often over 1 kb in length and produced a loss of GFP comparable to that induced by distant enhancers.

What is common to all of the above-mentioned high-throughput screenings is that by far the largest share of sgRNAs were not enriched or depleted, and the sgRNAs that did show a strong signal colocalized to discrete genomic regions. This is a strong indication that enhancer elements consist of many redundant and only a few critical sequences [51]. But these critical sequences can be highly sensitive to small mutations generated by single DSBs, which enabled these high-throughput functional screens.

Instead of screening CRISPR-induced sequence variability, CRISPR interference can also be used to characterize the regulatory functions of CREs in their native contexts. Fulco et al. developed a system based on a dCas9-KRAB fusion [32]. The KRAB domain mediates transcriptional repression by causing epigenetic modifications. They tiled 98.000 sgRNAS across >1 megabases in

the vicinity of two essential genes, the transcription factors GATA1 and MYC, and quantified their repressive capacity via a cellular proliferation screen. They identified 9 distal enhancers, and demonstrated that a single enhancer can have regulatory activity on multiple genes. In addition, they found competition between neighboring promoters.

4 Using CRISPR/Cas to Modify Transcription Factors and Related Complexes

Trans-elements or trans-regulatory elements are DNA sequences that regulate target gene expression indirectly by coding for TFs. This is opposed to CREs, which also regulate expression of a target gene but instead, serve as TF-binding sites. TFs are a very interesting target for editing as their capacity to regulate whole metabolic pathways can have dramatic phenotypic outcomes.

While constitutive overexpression of specific TFs using the classical transgenic approach can lead to improved varieties, especially concerning tolerance to abiotic stress, this often leads to negative side effects like dwarfing, late flowering, and lower yields [60]. The usage of stress-inducible promoters might alleviate this limitation [60]. In addition to overexpression, the CRISPR/Cas technology now also enables simple targeted knock out of TFs to engineer crops with desired traits (Table 1).

A convincing demonstration is pathogen resistance. Rice blast caused by the fungus *Magnaporthe oryzae* is one of the most devastating rice diseases. CRISPR/Cas-mediated KO of the ethylene response factor OsERF922 led to strongly enhanced plant resistance: the number of blast lesions formed after infection was significantly decreased in all mutant lines when compared to wild-type. Other agronomically relevant traits such as yield were not affected in the mutant lines [67]. Abiotic stress resistance is another trait that could successfully be improved. Shi et al. used Cas9-induced DSBs for site-specific insertion of a native promoter supporting low constitutive expression into the 5'UTR of maize *ARGOS8*, a negative regulator of ethylene responses [61]. This leads to modest *ARGOS8* overexpression which improved yield under drought stress conditions in field trials.

The trait most widely improved by TF editing is flowering time, regulated by the photoperiodic pathway (or florigen pathway). This pathway begins in the leaves, where photoreceptors sense changes in day length. Dependent on the species, either long day or short day conditions lead to accumulation of CONSTANS protein (in *Arabidopsis*) in the phloem cells of leaves. CONSTANS in turn activates expression of a phloem mobile signal protein (previously called florigen), which is FLOWERING LOCUS T (FT) in *Arabidopsis* or HEADING DATE 3A (HD3A) in rice. Upon translocation to the shoot apical meristem, FT forms a complex with

Table 1
Crop improvements achieved via modification of transcription factors

Target TF	Organism	Modification	Effect	Study
ETHYLENE RESPONSE FACTOR 922	Rice	CRISPR KO	Enhanced pathogen resistance	Wang et al. [60]
ARGOS8	Maize	CRISPR mediated promoter insertion	Improved drought resistance	Shi et al. [61]
SINGLE FLOWER TRUSS,SELF PRUNING, *Solyc02g083520*	Tomato	Natural and EMS mutations	Enhanced yield	Park et al. [62]
FLOWERING LOCUS T	Arabidopsis	CRISPR KO	Delayed flowering, POC	Hyun et al. [63]
FLOWERING LOCUS T 2a	Soybean	CRISPR KO	Delayed flowering	Cai et al. [64]
HEADING DATE 2 (HD2), HD3, HD5	Rice	CRISPR KO	Earlier flowering	Li et al. [65]
SELF-PRUNING 5G	Tomato	CRISPR KO	Reduced photoperiod response, earlier yield	Soyk et al. [66]

FLOWERING LOCUS D (FD), a bZIP TF expressed in the meristem. The FT/FD heterodimer activates expression of downstream target genes that promote floral identity [68].

Flowering time was already a promising target without CRISPR: Park et al. identified a way to enhance the yield of tomatoes by altering the balance between a factor involved in the flowering pathway and its antagonist [62]. They identified a mutant allele of the florigen gene SINGLE FLOWER TRUSS (SFT) showing reduced expression and two mutations in a bZIP (leucine zipper) TF. By combining heterozygous mutations, they achieved a productive balance of flowering signals that ultimately lead to enhanced yields.

The first application of CRISPR/Cas9 to edit flowering pathway components followed soon after: Targeted KO of *FLOWERING LOCUS T* (*FT*) led to late flowering phenotypes in *Arabidopsis* plants homozygously carrying novel null alleles [63]. Homozygous mutants of the soybean *FT* ortholog Gm*FT2a* showed a similar effect, exhibiting a late flowering phenotype under natural, short day, and long day conditions [64]. Since the geographical range of soybean cultivation is limited by its high sensitivity to photoperiod, such new varieties can expand the range of soybean cultivation. A similar problem impedes the northward

expansion of rice cultivation in China. The day length extension accompanied with northward cultivation is preventing the tropical short-day plant rice from properly completing flowering and seed setting. For this purpose, early-maturing rice cultivars with low photoperiod response are required. In a recent study [65], KO of three TF involved in photoperiodic flowering pathway and negatively regulating the heading date of rice (*HEADING DATE 2* (*HD2*), *HD4*, and *HD5*) led to significantly earlier flowering. This finding will accelerate not only the northward expansion of rice cultivation, but also local breeding programs in China's norther provinces. This region has only 70 years of rice cultivation history with little available germplasm, but now the introduction of elite rice germplasm resources from southern regions to China's northern provinces becomes possible by editing of rice *HEADING DATE* genes [65].

In tomato, domestication was associated with loss of day-length sensitive flowering. Fine-mapping of a QTL region linked to photoperiod response identified the flowering repressor *SELF-PRUNING 5G* (*SP5G*) to cause delayed flowering during long days in tomato [66]. Interestingly, variation in a CRE leading to reduced induction of *SP5G* under long days is responsible for the loss of day-length sensitive flowering in cultivated tomato. CRISPR/Cas-mediated KO of *SP5G* accelerated flowering under long day conditions and reduced time to harvest by 2 weeks. The approach could be extended to another tomato cultivar, where *SP5G* KO generated early-yielding plants in one generation. Thus targeting *SP5G* homologs could serve as a first step toward domestication of wild tomato relatives with agricultural potential, enabling them to grow in more northerly latitudes. The approach can even be extended to other crops, since flower-repressing florigen paralogs play similar roles in diverse crops such as potato [69], sugar beet [70], and sunflower [71]. Accordingly, *SP5G* is a promising candidate for a generally applicable, simple, and fast CRISPR/Cas-mediated one-step breeding approach to avert day-length sensitivity and engineer early-maturing varieties. Targeting other flowering regulating genes might allow quick customization of day-length sensitivity in elite-germplasm of many crops to extend the cultivation range [66].

5 Future Directions and Perspectives

It was demonstrated that CRISPR/Cas is an efficient tool for editing CREs. However, when targeting small CREs, the target site restriction by the PAM required by *S.p.* Cas9 is hampering efficient CRE mutagenesis [36]. This is especially the case when saturating mutagenesis is required for dissection of enhancer function at nucleotide resolution [54]. However, an increasing amount of

Cas9 orthologs with different PAM requirements is becoming available, e.g., Cas9 from *Streptococcus thermophilus* requires NNRGAA [72] and *Staphylococcus aureus* Cas9 requires NNGRR(T) [73], both of which were already used successfully in plants [74]. Alternatively, *S.p.* Cas9 variants with altered PAM specificities created by directed evolution are available [75]. Finally, another nuclease effector from a typeII CRISPR system, named Cpf1 (CRISPR from Prevotella and Francisella1), was recently characterized [76] and successfully employed in plants [23]. Opposed to the G-rich PAM required by Cas9 orthologues, Cpf1 orthologues recognizes a T-rich PAM (TTTN for Cpf1 from *Lachnospiraceae bacterium* ND2006 and from *Acidaminococcus sp. BV3L6*). Additionally, for Cpf1 the PAM is located upstream instead of downstream of the target sequence as is the case for Cas9. Together, these characteristics greatly increase the number of possible targets, enabling edits to be made precisely at the intended spot.

The same limitation constrains tiling screens for comprehensive analysis of CRE function: since only targets with PAMs are targetable, the coverage density in certain areas might be too scarce for a saturation screen [29]. Thus, making use of different orthologues of Cas9 and Cpf1 could address this issue. Another limitation of tiling screens is that *Streptococcus pyogenes* Cas9 creates only very small indels, mostly 1 bp insertions [17]. In the case of regulatory sequences, such small indels might not be sufficient for causing phenotypic change. As Zhou et al. point out [29], this problem could be resolved by modification of the library design in a way that paired sgRNAs instead of single sgRNAs are expressed. This way, mutations could be generated at two loci at the same time or larger fragments could be precisely deleted by inducing two DSB at close positions simultaneously. In addition, the repair outcome from different nucleases is not identical, e.g., *Staphylococcus aureus* Cas9 tends to induce a larger share of longer deletions than the standard *S.p.* Cas9, although this characteristic of *S.a.* Cas9 might be PAM or target sequence dependent [74]. The same applies for Cpf1 compared to *S.p.* Cas9. Larger deletions might be more useful for a phenotypic outcome when regulatory sequences are dissected in tiling screens.

The paired nickase approach [77–79] might also be useful for editing of regulatory sequences because it generates a much more diverse mutation profile in plants (*see* Fig. 3) that can be influenced by the distance of the nicks. In this approach, a nickase version of Cas9 is used where one of the two nuclease domains (RuvC) is inactivated. Consequently, a single strand break (SSB) is generated instead of a DSB. The generation of two SSBs in close proximity results in a mutagenic DSB. Originally, this approach was developed for increasing the specificity of genome editing as SSBs are usually repaired error-free. However, its

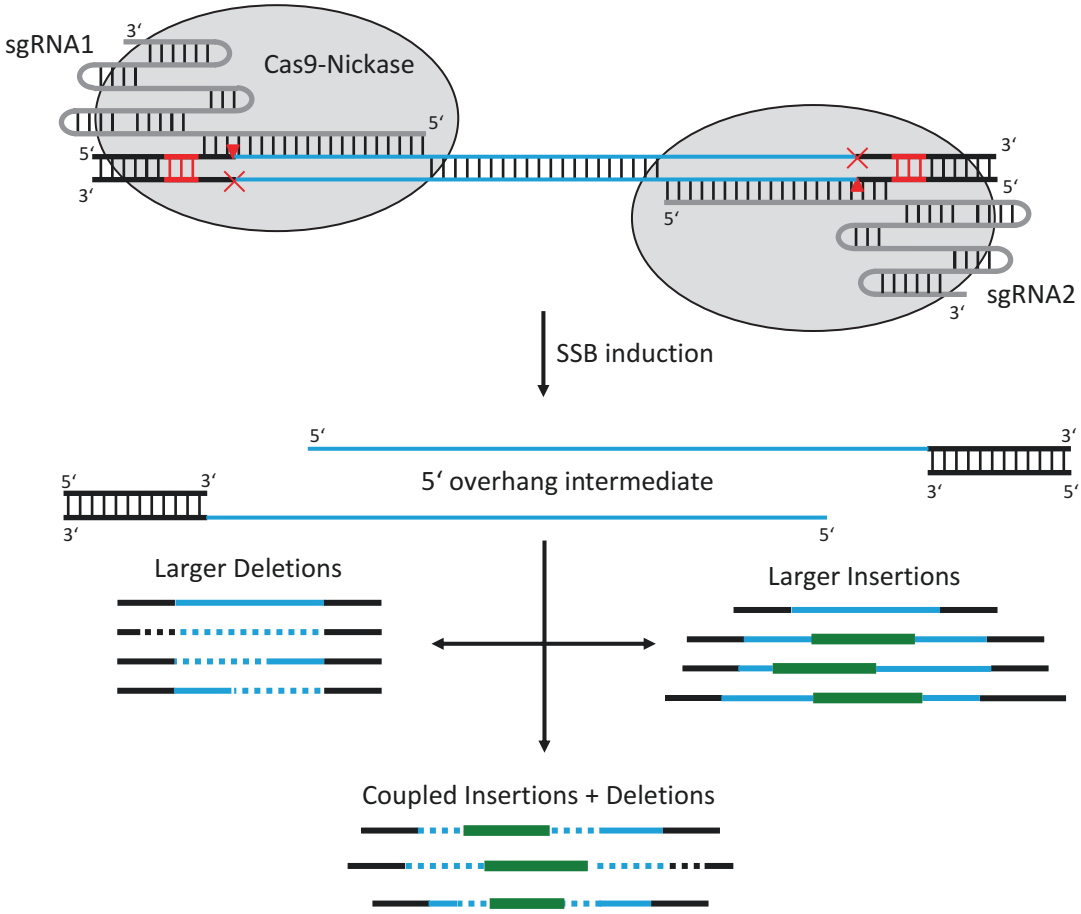

Fig. 3 The paired nickase approach and its diverse mutation profile. Two close nicks on opposite strands are induced in such a way that a mutagenic DSB with long 5′ overhangs is generated. Repair of this staggered DSB generates primarily large deletions, but also large insertions and combinations of both. The insertions are usually tandem duplications, which might be used for duplicating TF-binding sites

unusual mutation spectrum makes it attractive for CRE editing as well. Paired nickases induce primarily large deletions but also insertions that arise mainly from tandem duplications [80]. The generation of tandem duplications might be used for duplication of TF-binding sites.

Besides indel formation, epigenetic modifications can be used for tiling screens, as shown for dCas9-KRAB fusion [32]. Fusion of the catalytic core of acetyltransferase p300 to dCas9 led to robust H3K27 acetylation [24]. Opposed to previous CRISPR-based transcriptional activation approaches, H3K27 acetylation was shown to enable robust transcriptional activation not only from promoters, but also from proximal and distal enhancers regions with a single sgRNA. Thus, H3K27 acetylation might be another option for tiling screens beside indel formation to further expand the methodological toolbox for enhancer dissection.

As noted by Soyk et al. [43], new genome engineering tools like CRISPR/Cas enable engineering a range of alleles with different types and expression strengths. Such weak or strong alleles could improve a wide range of agronomic traits in crops by allowing customized gene dosage effects. Since expression strength of TFs affects many downstream genes up to whole metabolic pathways, dosage effects could be especially useful for editing TFs, as demonstrated for the tomato homologs of the MADS-box-TFs *SEPALLATA4* and *FRUITFULL* [43] and the flowering repressor *SP5G* [66]. Thus, combining editing of cis- and trans-regulatory elements by editing CREs of TFs seems especially promising.

It was demonstrated that CRISPR/Cas for the first time enables in-depth functional analysis of CREs in their native context. In addition, CRISPR/Cas-mediated editing of cis- as well as trans-regulatory elements has a considerable potential for crop improvement. However, up to the present this potential did not receive much attention and was left mainly untouched. It can be concluded that the regulatory part of the genome is a valuable extension of genome editing targets for future breeding programs.

References

1. Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci 93(3):1156–1160. https://doi.org/10.1073/pnas.93.3.1156
2. Boch J, Scholze H, Schornack S et al (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326(5959):1509–1512. https://doi.org/10.1126/science.1178811
3. 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 U S A 93(10):5055–5060
4. Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J 17(20):6086–6095. https://doi.org/10.1093/emboj/17.20.6086
5. Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56(409):1–14. https://doi.org/10.1093/jxb/eri025
6. Pacher M, Puchta H (2016) From classical mutagenesis to nuclease-based breeding - directing natural DNA repair for a natural end-product. Plant J 90(4):819–833. https://doi.org/10.1111/tpj.13469
7. Ishino Y, Shinagawa H, Makino K et al (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol 169(12):5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987
8. Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 8:172. https://doi.org/10.1186/1471-2105-8-172
9. Bolotin A, Quinquis B, Sorokin A et al (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151(Pt 8):2551–2561. https://doi.org/10.1099/mic.0.28048-0
10. Shmakov S, Smargon A, Scott D et al (2017) Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol 15(3):169–182. https://doi.org/10.1038/nrmicro.2016.184
11. Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816–821. https://doi.org/10.1126/science.1225829

12. Le C, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823. https://doi.org/10.1126/science.1231143
13. Li J-F, Norville JE, Aach J et al (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol 31(8):688–691. https://doi.org/10.1038/nbt.2654
14. Nekrasov V, Staskawicz B, Weigel D et al (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol 31(8):691–693. https://doi.org/10.1038/nbt.2655
15. Shan Q, Wang Y, Li J et al (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31(8):686–688. https://doi.org/10.1038/nbt.2650
16. Feng Z, Mao Y, Xu N et al (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proc Natl Acad Sci U S A 111(12):4632–4637. https://doi.org/10.1073/pnas.1400822111
17. 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(2):348–359. https://doi.org/10.1111/tpj.12554
18. Scheben A, Wolter F, Batley J, Puchta H, Edwards D (2017) Towards CRISPR/Cas crops – bringing together genomics and genome editing. New Phytol 216:682–698. https://doi.org/10.1111/nph.14702
19. Puchta H (2016) Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. Curr Opin Plant Biol 36:1–8. https://doi.org/10.1016/j.pbi.2016.11.011
20. Gilbert LA, Larson MH, Morsut L et al (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154(2):442–451. https://doi.org/10.1016/j.cell.2013.06.044
21. Konermann S, Brigham MD, Trevino AE et al (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517(7536):583–588. https://doi.org/10.1038/nature14136
22. Piatek A, Ali Z, Baazim H et al (2015) RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. Plant Biotechnol J 13(4):578–589. https://doi.org/10.1111/pbi.12284
23. Tang X, Lowder LG, Zhang T et al (2017) A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. Nat Plants 3:17018. https://doi.org/10.1038/nplants.2017.18
24. Hilton IB, D'Ippolito AM, Vockley CM et al (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 33(5):510–517. https://doi.org/10.1038/nbt.3199
25. Thakore PI, D'Ippolito AM, Song L et al (2015) Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat Methods 12(12):1143–1149. https://doi.org/10.1038/nmeth.3630
26. Dreissig S, Schiml S, Schindele P et al (2017) Live cell CRISPR-imaging in plants reveals dynamic telomere movements. Plant J 91(4):565–573. https://doi.org/10.1111/tpj.13601
27. Komor AC, Kim YB, Packer MS et al (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533(7603):420–424. https://doi.org/10.1038/nature17946
28. Zong Y, Wang Y, Li C et al (2017) Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat Biotechnol 35(5):438–440. https://doi.org/10.1038/nbt.3811
29. Zhou Y, Wei W (2016) Mapping regulatory elements. Nat Biotechnol 34(2):151–152. https://doi.org/10.1038/nbt.3477
30. Britten RJ, Davidson EH (1969) Gene regulation for higher cells: a theory. Science 165(3891):349–357. https://doi.org/10.1126/science.165.3891.349
31. Wittkopp PJ, Kalay G (2011) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet 13(1):59–69. https://doi.org/10.1038/nrg3095
32. Fulco CP, Munschauer M, Anyoha R et al (2016) Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science 354(6313):769–773. https://doi.org/10.1126/science.aag2445
33. Koenig D, Jimenez-Gomez JM, Kimura S et al (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci 110(28):E2655–E2662. https://doi.org/10.1073/pnas.1309606110
34. Hufford MB, Xu X, van Heerwaarden J et al (2012) Comparative population genomics

of maize domestication and improvement. Nat Genet 44(7):808–811. https://doi.org/10.1038/ng.2309
35. Meyer RS, Purugganan MD (2013) Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet 14(12):840–852. https://doi.org/10.1038/nrg3605
36. Swinnen G, Goossens A, Pauwels L (2016) Lessons from domestication: targeting cis-regulatory elements for crop improvement. Trends Plant Sci 21(6):506–515. https://doi.org/10.1016/j.tplants.2016.01.014
37. Wang S, Li S, Liu Q et al (2015) The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. Nat Genet 47(8):949–954. https://doi.org/10.1038/ng.3352
38. Sakamoto T, Matsuoka M (2008) Identifying and exploiting grain yield genes in rice. Curr Opin Plant Biol 11(2):209–214. https://doi.org/10.1016/j.pbi.2008.01.009
39. Xu C, Liberatore KL, MacAlister CA et al (2015) A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nat Genet 47(7):784–792. https://doi.org/10.1038/ng.3309
40. Wu X, Scott DA, Kriz AJ et al (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol 32(7):670–676. https://doi.org/10.1038/nbt.2889
41. Li T, Liu B, Spalding MH et al (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol 30(5):390–392. https://doi.org/10.1038/nbt.2199
42. Antony G, Zhou J, Huang S et al (2010) Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. Plant Cell 22(11):3864–3876. https://doi.org/10.1105/tpc.110.078964
43. Soyk S, Lemmon ZH, Oved M et al (2017) Bypassing negative epistasis on yield in tomato imposed by a domestication gene. Cell 169(6):1142–1155.e12. https://doi.org/10.1016/j.cell.2017.04.032
44. Liu C, Teo ZWN, Bi Y et al (2013) A conserved genetic pathway determines inflorescence architecture in Arabidopsis and rice. Dev Cell 24(6):612–622. https://doi.org/10.1016/j.devcel.2013.02.013
45. Kaufmann K, Muiño JM, Østerås M et al (2010) Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat Protoc 5(3):457–472. https://doi.org/10.1038/nprot.2009.244
46. Goossens A (2015) It is easy to get huge candidate gene lists for plant metabolism now, but how to get beyond? Mol Plant 8(1):2–5. https://doi.org/10.1016/j.molp.2014.08.001
47. de Witte D, van de Velde J, Decap D et al (2015) BLSSpeller: exhaustive comparative discovery of conserved cis-regulatory elements. Bioinformatics 31(23):3758–3766. https://doi.org/10.1093/bioinformatics/btv466
48. John S, Sabo PJ, Canfield TK et al (2013) Genome-scale mapping of DNase I hypersensitivity. Curr Protoc Mol Biol Chapter 27: Unit 21.27. doi: https://doi.org/10.1002/0471142727.mb2127s103
49. Dekker J, Rippe K, Dekker M et al (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. https://doi.org/10.1126/science.1067799
50. Kim T-K, Hemberg M, Gray JM et al (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465(7295):182–187. https://doi.org/10.1038/nature09033
51. Lopes R, Korkmaz G, Agami R (2016) Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. Nat Rev Mol Cell Biol 17(9):597–604. https://doi.org/10.1038/nrm.2016.79
52. Melnikov A, Murugan A, Zhang X et al (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol 30(3):271–277. https://doi.org/10.1038/nbt.2137
53. Patwardhan RP, Hiatt JB, Witten DM et al (2012) Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol 30(3):265–270. https://doi.org/10.1038/nbt.2136
54. Canver MC, Smith EC, Sher F et al (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature 527(7577):192–197. https://doi.org/10.1038/nature15521
55. Vierstra J, Reik A, Chang K-H et al (2015) Functional footprinting of regulatory DNA. Nat Methods 12(10):927–930. https://doi.org/10.1038/nmeth.3554
56. Seruggia D, Fernandez A, Cantero M et al (2015) Functional validation of mouse tyrosinase non-coding regulatory DNA elements by CRISPR-Cas9-mediated mutagenesis. Nucleic Acids Res 43(10):4855–4867. https://doi.org/10.1093/nar/gkv375

57. Li J, Shou J, Guo Y et al (2015) Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. J Mol Cell Biol 7(4):284–298. https://doi.org/10.1093/jmcb/mjv016
58. Korkmaz G, Lopes R, Ugalde AP et al (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat Biotechnol 34(2):192–198. https://doi.org/10.1038/nbt.3450
59. Rajagopal N, Srinivasan S, Kooshesh K et al (2016) High-throughput mapping of regulatory DNA. Nat Biotechnol 34(2):167–174. https://doi.org/10.1038/nbt.3468
60. Wang H, Wang H, Shao H et al (2016) Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology. Front Plant Sci 7:67. https://doi.org/10.3389/fpls.2016.00067
61. Shi J, Gao H, Wang H et al (2017) ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotechnol J 15(2):207–216. https://doi.org/10.1111/pbi.12603
62. Park SJ, Jiang K, Tal L et al (2014) Optimization of crop productivity in tomato using induced mutations in the florigen pathway. Nat Genet 46(12):1337–1342. https://doi.org/10.1038/ng.3131
63. Hyun Y, Kim J, Cho SW et al (2015) Site-directed mutagenesis in Arabidopsis thaliana using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles. Planta 241(1):271–284. https://doi.org/10.1007/s00425-014-2180-5
64. Cai Y, Chen L, Liu X et al (2017) CRISPR/Cas9-mediated targeted mutagenesis of GmFT2a delays flowering time in soybean. Plant Biotechnol J 16(1):176–185. https://doi.org/10.1111/pbi.12758
65. Li X, Zhou W, Ren Y et al (2017) High-efficiency breeding of early-maturing rice cultivars via CRISPR/Cas9-mediated genome editing. J Genet Genomics 44(3):175–178. https://doi.org/10.1016/j.jgg.2017.02.001
66. Soyk S, Muller NA, Park SJ et al (2017) Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. Nat Genet 49(1):162–168. https://doi.org/10.1038/ng.3733
67. Wang F, Wang C, Liu P et al (2016) Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene OsERF922. PLoS One 11(4):e0154027. https://doi.org/10.1371/journal.pone.0154027
68. Turck F, Fornara F, Coupland G (2008) Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annu Rev Plant Biol 59:573–594. https://doi.org/10.1146/annurev.arplant.59.032607.092755
69. Abelenda JA, Cruz-Oró E, Franco-Zorrilla JM et al (2016) Potato StCONSTANS-like1 suppresses storage organ formation by directly activating the FT-like StSP5G repressor. Curr Biol 26(7):872–881. https://doi.org/10.1016/j.cub.2016.01.066
70. Pin PA, Benlloch R, Bonnet D et al (2010) An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. Science 330(6009):1397–1400. https://doi.org/10.1126/science.1197004
71. Blackman BK, Strasburg JL, Raduski AR et al (2010) The role of recently derived FT paralogs in sunflower domestication. Curr Biol 20(7):629–635. https://doi.org/10.1016/j.cub.2010.01.059
72. Esvelt KM, Mali P, Braff JL et al (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods 10(11):1116–1121. https://doi.org/10.1038/nmeth.2681
73. Ran FA, Le C, Yan WX et al (2015) In vivo genome editing using Staphylococcus aureus Cas9. Nature 520(7546):186–191. https://doi.org/10.1038/nature14299
74. Steinert J, Schiml S, Fauser F et al (2015) Highly efficient heritable plant genome engineering using Cas9 orthologues from Streptococcus thermophilus and Staphylococcus aureus. Plant J 84(6):1295–1305. https://doi.org/10.1111/tpj.13078
75. Kleinstiver BP, Prew MS, Tsai SQ et al (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523(7561):481–485. https://doi.org/10.1038/nature14592
76. Zetsche B, Gootenberg JS, Abudayyeh OO et al (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163(3):759–771. https://doi.org/10.1016/j.cell.2015.09.038
77. 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(6):1139–1150. https://doi.org/10.1111/tpj.12704
78. Mikami M, Toki S, Endo M (2016) Precision targeted mutagenesis via Cas9 paired nickases

in rice. Plant Cell Physiol 57(5):1058–1068. https://doi.org/10.1093/pcp/pcw049

79. Wolter F, Edelmann S, Kadri A et al (2017) Characterization of paired Cas9 nickases induced mutations in maize mesophyll protoplasts. Maydica 62:2_15

80. Schiml S, Fauser F, Puchta H (2016) Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes. Proc Natl Acad Sci U S A 113(26):7266–7271. https://doi.org/10.1073/pnas.1603823113