Current Capabilities, Precision, and Risks of Genome Editing Techniques in Agricultural Systems

Morgan Carter[1]

[1]School of Plant Sciences, University of Arizona, Tucson, AZ; (morgancarterphd@gmail.com)

Abstract

We are in a new chapter of crop and livestock improvement with the emergence of genome editing. This latest generation of molecular tools can be used to make targeted changes in a genome including insertions, deletions, and mutations. With new advances comes new risks for unintended changes and impacts, thus the need for appropriate risk assessment for product development and to inform regulatory measures. Though CRISPR/Cas has arisen as the predominant technology, there are multiple types of genome editing tools each with pros and cons depending on the organism and desired outcome. Furthermore, each editing tool differs in specificity as they may edit non-intended sites, referred to as off-target edits. The consensus of the agricultural editing community is to avoid off-target editing through design and detection, instead of determining whether off-target editing in each case is detrimental. The design of a targeting component, the tool chosen, and the identification of the edit(s) made are the critical factors in avoiding off-target edits and confirming intended edits in final products that are released commercially. The limited amount of head-to-head comparisons of genome editing tools in diverse crops and livestock make it difficult to develop broad conclusions and best practices, which is further compounded by the diversity of techniques, targets, and processes. Developers and breeders should consult the literature and test as needed to determine which editing technology will be the most effective for their purposes, especially as more tools with altered efficiency and specificity become available. Yet, **the lack of off-target edits in studies that employed careful design of targeting components followed by wide testing for on- and off-target edits bodes well for the use of genome editing with proper precautions of target selection and screening.**

Keywords: genome editing; agriculture; crispr; talen; specificity; off-target

Introduction

An explosion of interest in genome editing from the nonprofit, industry, and academic sectors has resulted in many different methods for genome editing that vary in ease and efficacy, yielding a complex and quickly moving landscape of possibilities, applications, and risks. Presented in this report is an overview of genome editing and the most popular tools, with a focus on conditions and attributes that contribute to editing specificity, as well as the detection of edits. This introduction section establishes many of the terms and concepts related to specificity and efficiency of genome editing that will be used throughout this report when exploring technologies and methods in further depth.

Genome Editing

Genome editing is characterized by the ability to direct changes to a specific region of an organism's DNA (**genome**), contrasting with the random nature of previous genetic engineering technologies. Desired changes include deleting, inserting, or mutating one or more individual units of DNA (**bases/nucleotides**) ranging from small edits that convert one nucleotide to another, to large edits causing sizeable segments of genetic material to move between parts of the genome. The potential uses range from moving a copy of a gene from a wild tomato to a cultivated variety to improve disease resistance, to making a slight change to gene regulation to deal with drought or heat tolerance, to removing a gene entirely to create hypoallergenic cow milk. To edit an organism, genome editing components must be introduced into plant or animal cells where they can create the edit and then be removed by the developer or by the cell itself. Possible methods for introduction include microinjection into individual animal cells and bombardment of plant cells with a "gene gun". Such a broad spectrum of possible edits through a variety of tools that vary in origin, process, and efficacy complicates conversations around where genome editing fits into breeding techniques, food safety, and environmental impact (CAST, 2018).

Genome editing tools are often referred to colloquially as "molecular scissors" and require both targeting and cutting components. Depending on the technology, a protein or RNA component will bind to the DNA sequence of interest, acting as the targeting component directing the cutting component to the correct site in the genome. The cutting component is a **nuclease** that cleaves DNA to create a **double stranded break (DSB)** if it cuts both strands, or a **nick** if only one strand is cut. The editing process then relies on the edited cell's machinery to repair the damage. The basic repair mechanism used by cells to fix DSBs is **non-homology end joining (NHEJ)**, which simply sticks the two ends back together in a way that can introduce mutations including the **insertion or deletion (indel)** of one or more DNA bases, possibly disrupting or altering any role of that sequence (*Figure 1*). NHEJ is an effective way to render a gene or other genetic feature nonfunctional. An alternative repair mechanism, **homology directed repair (HDR)**, relies on a piece of DNA that is similar (homologous) enough to the sequence around the break that it can bind and be integrated where the cut occurred, repairing the break (*Figure 1*). HDR is useful for inserting new DNA sequences into a genome, including genes from other organisms or variations of a gene, called alleles, that may be taken from the same species or close relatives.
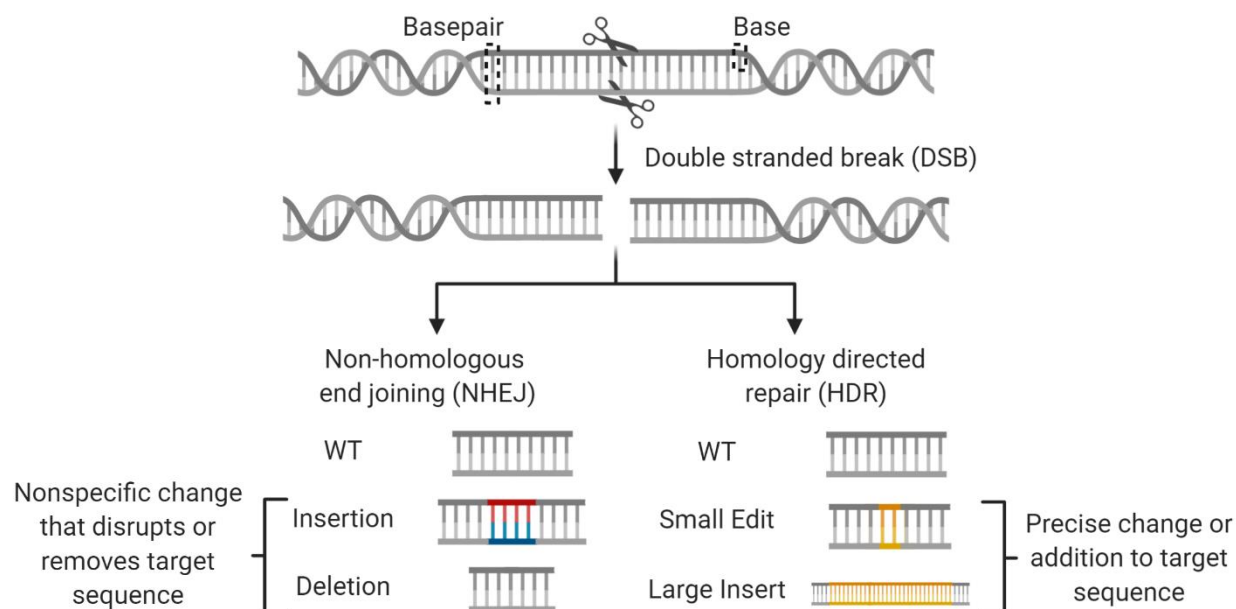
Figure 1. Possible outcomes from repair of a double stranded break (DSB) in the genome and how these are typically used in genome editing. Grey DNA can be repaired back to wildtype (WT) or repaired in ways that introduce specific (yellow) or nonspecific (blue/red) insertions or deletions.

Specificity and off-target edits

        Specificity refers to the ability of a genome editing tool to only make changes that are intended at a target sequence and not cause changes at other locations, referred to as **off-target edits**. Off-target edits generate concern because of their unknown impact depending on where they occur and in what way, possibly turning off or on other genes in negative ways. When considering the specificity of genome editing, there are general attributes that vary between approaches: length of target sequence, selection of target sequence, target-adjacent sequence requirements, type of targeting component, and method for introducing genome editing tools. Improving genome editing technologies to reduce off-target edits and increase efficiency is largely driven by human therapeutic applications where edits must be precise to be deployed in a clinical setting. Yet, improved efficiency and reduction of off-target editing are valuable in agriculture as well, though they are less of a concern because of the ability to select for desired outcomes through both screening of many individual events and breeding after the initial editing.

        First, the **length of the target sequence** impacts specificity, as longer sequences will have a reduced frequency of occurring randomly in a genome. Second, **selection of target sequence** should be done carefully to reduce potential off-target edits as predicted computationally by avoiding sequences that occur many times throughout the genome. Third, the **necessity of certain sequences or nucleotides adjacent to the target sequence** for cutting to happen can also impact the number of possible off-target edits (see PAM for CRISPR/Cas and thymine requirement for TALENs). Fourth, **whether the targeting component is RNA or protein impacts specificity** as protein-DNA interactions (TALENs, ZFNs, Meganucleases) appear to be less specific than nucleotide pairing through RNA-DNA binding (CRISPR/Cas systems). Finally, **the way in which genome editing tools are introduced to a cell or organism** and how long they stay active in cells or resulting organisms can dramatically alter the

efficiency and risk of off-target edits. Taken together, both the genome editing tool and the process of editing a plant or animal contribute to the risk of off-target edits. Ideally, unintended changes will be avoided entirely by design and screening. However, creating edited organisms for agricultural purposes does allow for removing unwanted changes that do occur through additional breeding. Thus, the general outlook on specificity in genome editing for agricultural applications is to minimize or eliminate off-target editing, followed by accurate detection of edit occurrence, instead of differentiating between good or bad off-target edits.

Editing by DNA-RNA interaction: CRISPR/Cas systems and derivatives
CRISPR/Cas basics

The increasingly popular **CRISPR** (clustered regularly interspaced short palindromic repeats)**/Cas** (CRISPR associated) system for gene editing is an application of a natural anti-viral system present in many bacteria. A Cas protein serves as the nuclease cutting component, with Cas proteins from different bacteria varying in their specificity and sequence requirements for functionality. To know where to cut, the Cas protein relies on one or more pieces of RNA that bind to a targeted region of DNA, directing Cas cleavage (*Figure 2*). The most commonly used CRISPR-Cas system uses **SpCas9** *(Figure 2A)* to create a DSB in DNA; SpCas9 is a nuclease from a CRISPR/Cas system from *Streptococcus pyogenes*.

In its natural configuration, SpCas9 requires two RNAs, one constant that interacts with SpCas9 and one changeable that binds the targeted sequence for cleavage. When CRISPR/Cas9 was developed into a gene editing tool, these two RNAs were combined into a single guide RNA (**sgRNA**) which retains functionality and simplifies application (Jinek et al., 2012). **To target SpCas9 to a gene or sequence of interest, a sequence of 20 bases can be identified within the sequence of interest, synthesized in a lab, and incorporated into the sgRNA.**

To cut at a site that matches the sgRNA, all Cas proteins require a neighboring sequence (**PAM**; protospacer adjacent motif) in the DNA. In the case of SpCas9, the PAM needs to be after the target sequence (*Figure 2A*) and is composed of the sequence "NGG" where N is any of the four DNA bases followed by two guanosine bases. SpCas9 typically cuts three basepairs before the PAM site, within the target sequence. The necessity of a PAM undesirably limits sites that can be targeted, but also reduces off-target risk. **If there is no PAM, even when the sgRNA directs a Cas protein to a sequence, it will not actually cut the sequence.**

Cas variations

> **Key Points: CRISPR/Cas**
> - CRISPR refers to genome editing tools that use a designable RNA guide (sgRNA) to identify target regions in DNA and a Cas nuclease to create a DSB at that site.
> - Each Cas protein needs a specific sequence (PAM) to be adjacent to the target sequence for it to cut, limiting which sequences can be targeted, but also reducing edits at off-target binding sites.

Altering SpCas9 to increase specificity, change PAMs, or improve efficiency has been of high interest in the genome editing world (Naeem et al., 2020). For example, a version of SpCas9, xCas9, was developed to recognize a wider array of PAMs increasing the sites that could be targeted with the protein. Surprisingly, xCas9 was demonstrated to be more specific in mammalian cells, reducing off-targets to zero or substantially less than SpCas9 depending on target site (Hu et al., 2018). Diversity of PAMs is one of the reasons additional Cas9 proteins from other bacteria have been investigated, such as SaCas9 cloned from *Staphylococcus aureus* which has also been modified to recognize more PAMs while maintaining specificity
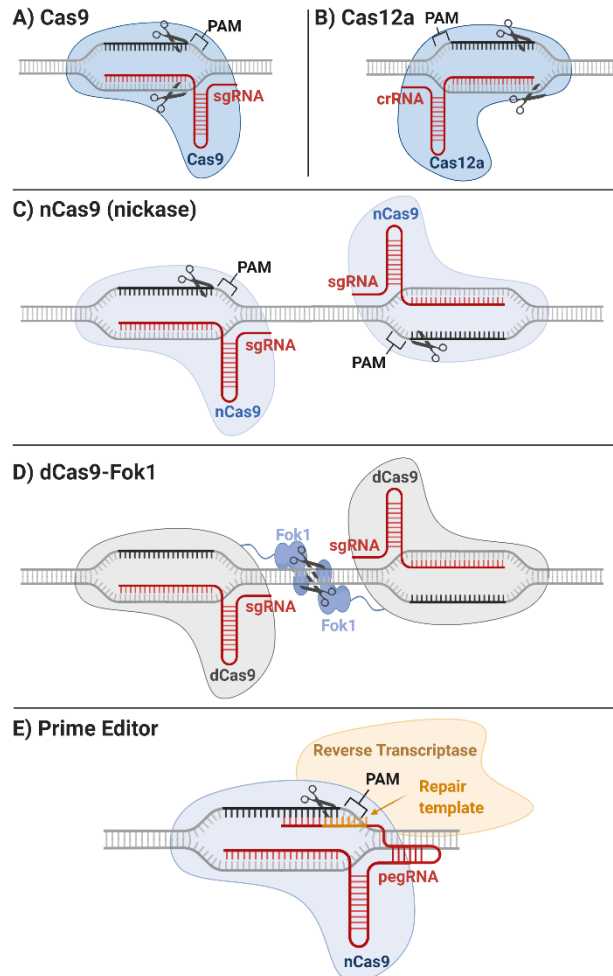
Figure 2. Examples of CRISPR/Cas genome editing tools and how they interact with and cut DNA. Scissors indicate sites where grey DNA strands would be cut near or in black target sequences. Cutting components (proteins) are indicated in blue and targeting components (guide RNAs) in red. PAM indicates the DNA sequence required by Cas for it to cut the RNA-bound DNA. Yellow components relate to repairing the DSB using a provided template and another protein component called reverse transcriptase.

Key Points: Cas variations

- There are a variety of Cas proteins, both natural and artificial, that can create DSBs or nicks.
- The most popular Cas protein, SpCas9, creates an even DSB, while Cas12a creates an uneven DSB, which can be more useful if trying to insert new DNA.

(Kleinstiver et al., 2015). Anzalone et al. (2020) provides a comprehensive scientific review of the development and diversity of CRISPR/Cas systems beyond what is presented here, including other synthetic Cas9 derivatives with alternative PAM specificities. However, **SpCas9 remains the most popular and widely used tool for genome editing in crops and livestock**, given its history as the first CRISPR/Cas tool, ease of sgRNA design, and continued use for proof-of-concept of genome editing in new organisms.

One of the most popular Cas proteins used besides Cas9 is **Cas12a** (also **Cpf1;** CRISPR from *Prevoltella* and *Francisella* 1) (Zetsche et al., 2015). The natural guide RNA (crRNA) for Cas12a is shorter than the synthetic one needed for SpCas9. The PAM also varies from Cas9 and is before the target sequence (*Figure 2B*) and typically thymine-rich, making it useful for targeting areas of the genome that may not have the guanosine-rich site needed by SpCas9. Certain mutations of Cas12a broaden the possible PAMs, similar to SpCas9 (Alok et al., 2020). The most appealing attribute of Cas12a is that it creates an uneven DSB where one of the DNA strands has more bases than the other (note the staggered scissors in *Figure 2B*). **The DNA "overhang" created by Cas12a increases the efficiency of HDR instead of NHEJ and is useful if the intention is to add in new genetic material, not just to disrupt the target sequence**. Cas12a variants from different bacteria have been successfully deployed in many plant species as has been recently reviewed in Alok et al. (2020), including LbCas12a from *Lachnospiraceae bacterium*. While less examples exist for Cas12a usage in livestock, a modified CRISPR/Cas12a approach was successfully used to edit rabbits and pigs with no detected off-target edits (Wu et al., 2018).

Off-target considerations

The inherent specificity and efficiency of CRISPR/Cas systems depends on the (1) **target organism and (2) target sequence.** The first is difficult to predict but can be determined experimentally, while the second is easy to predict computationally. CRISPR/Cas editing is regarded as fairly specific in plants assuming stringent target selection (Hahn & Nekrasov, 2019), but efficiency may be increased for certain Cas proteins, depending on the organism to be edited (*Table 1*). In rice, off-target edits by both Cas9 and Cas12a appear to rarely occur, but Cas12a may be more efficient for on-target edits (Yin et al., 2017, Tang et al., 2018, Banakar et al., 2020). However, in maize, SpCas9 was more efficient for on-target edits, with no detectable off-target mutations identified in plants from the generation after editing (Lee et al., 2019). As more Cas9 and Cas12a proteins are discovered or engineered, their specificity across crop plants and livestock will have to be assessed, informing the decisions of which Cas protein(s) to use within a certain organism.

Mutations to the Cas protein or guide RNA that reduce off-target edits often result in lower efficiency of on-target cleavage, and will likely not work well in combinations (Anzalone et al., 2020). For example, engineered sgRNAs that are truncated (Fu et al., 2014) or contain added structures (Kocak et al., 2019) increase cleavage specificity but likely by reducing binding stability. **One strategy to reduce off-target editing by CRISPR/Cas9 is to lengthen the target sequence by using two proteins, instead of one, to create a DSB**. SpCas9 has two domains with nuclease activity, which each cut one of the DNA strands. By mutating one of these, a nickase Cas9 (nCas9; *Figure 2C*) is created, requiring two nCas9s to create a DSB. A specific sequence of 40 basepairs will occur much less often in the genome that a sequence of 20 basepairs; thus, paired nCas9s with sgRNAs targeting adjacent sequence would result in a DSB with less potential for off-target edits given that two separate sgRNAs must bind adjacent regions. Like Cas12a cleavage, DSBs caused by nCas9s will create an uneven cut and can promote HDR. In mammalian cells, nCas9s reduced off-targets by 50- to 1,500-fold (Ran et al., 2013). In rice, off-target edits did not occur when using nCas9, but at the expense of lower on-target edits compared to SpCas9 (Mikami et al., 2016). A similar strategy employs deactivated Cas9 fused to a **Fok1** nuclease domain from the bacterium *Flavobacterium okeanokoites* (dCas9-Fok1; *Figure 2D*). Two dCas9-Fok1 proteins targeting adjacent sequences must join together (dimerize) for the Fok1s to create a DSB (Guilinger et al., 2014). Both nCas9 and dCas9-Fok1 require more and/or larger components limiting their potential usefulness in cell delivery methods with restrictions on the size of editing reagents, such as delivery by viruses.

Guide RNAs are designed with consideration of off-target sites through computational prediction (see *Target selection*), including whether or not a PAM is present at potential off-target binding sites. **In pigs, only one out of four tested sgRNAs resulted in off-target edits, but at two computationally predicted locations that each had a PAM, highlighting the role that off-target prediction and sgRNA selection can have in reducing off-target edits** (Carey et al., 2019). Not all off-target sites are equivalently likely to be edited as it depends on the number and locations of mismatches with the sgRNA. The more mismatches and the closer they are to the PAM, the more disruptive. Given that Cas9 PAMs are after the targeted sequence and

> **Key Points: Off-target considerations**
> - CRISPR/Cas tools vary by specificity and efficiency depending on the target organism.
> - Using nickases like nCas9 or dCas9-Fok1 will require targeting a larger sequence, increasing specificity.
> - Similarity to off-target sequences needs to be considered when picking a target sequence.

Cas12a PAMs are before, the tolerance for mismatches with the sgRNAs will be opposite for these two enzymes; i.e. mismatches in the beginning of the sgRNA will inhibit cleavage by Cas12a, but similar ones might be tolerated by Cas9 resulting in off-target edits.

Table 1. Comparison of key features of Cas9 and Cas12a, including outcomes of studies testing both.

Features	Cas9	Cas12a
Cut	Even	Uneven
PAM Location	After target	Before target
PAM Composition	G-rich	T-rich
Variants with Increased PAMs	Yes	Yes
Additional Notes	Original Cas protein used for genome editing	Promotes HDR more than Cas9
Direct comparisons		
Rice	High specificity (Tang et al., 2018, Yin et al., 2017)	High specificity (Tang et al., 2018) **Higher** efficiency (Yin et al., 2017)
Maize	High specificity and **higher** efficiency (Lee et al., 2019)	High specificity(Lee et al., 2019)
Tomato	Similar efficiency, but specificity not determined (Bernabé-Orts et al., 2019)	Similar efficiency*, but specificity not determined (Bernabé-Orts et al., 2019)
Pigs	Similar efficiency and high specificity (Yang et al., 2016)	Similar efficiency and high specificity (Wu et al., 2018)

*LbCas12a – another Cas12a was also tested and was less efficient

Base editing

An even more targeted approach to making changes is to use base editing tools developed from deactivated Cas proteins that change single basepairs without causing a DSB. A deaminase enzyme is attached to dCas9 (no cutting activity) and is directed to the target base by a sgRNA. Two common types of base editors are cytosine base editors (CBE) that convert C●G base pairs to T●A and adenosine base editors (ABE) that convert A●T to G●C. While typically considered to be very specific, base editing can result in various non-intended edits including indels and editing of other bases near the target. Each of these non-intended edits can also occur at off-target locations, necessitating careful selection from the wide variety of base editors available depending on the application (Anzalone et al., 2020). Base editing has been applied to a variety of genes and traits in plants with little-to-no off-target edits as reviewed in Mishra et al. (2020) and has been used successfully used in sheep (Zhou et al., 2019) and pigs (Xie et al., 2019), with no off-target edits detected in sheep and one detected in pigs (but with no apparent deleterious effect in live piglets).

Prime editing

For edits where a very specific outcome is desired, such as directed mutations for protein engineering, HDR has been used to incorporate a provided DNA template. However, HDR can be very inefficient, and methods that do not involve any DNA components are of interest given the potential for undesired random insertion of provided DNA. Prime editing is a new Cas-derived tool that has been shown to make targeted changes in mammalian models (Anzalone et al., 2019) and rice and wheat (Lin et al., 2020) with a nickase Cas9 (only cuts one strand) domain fused to a reverse transcriptase domain (*Figure 2E*). The prime editing guide RNA (pegRNA) includes the desired sequence to be incorporated into the repair, but as an RNA copy. The reverse transcriptase incorporates a DNA copy of the repair template once the target DNA is nicked. The initial study found that off-targets were reduced with prime editing, as compared to Cas9, in human cell lines (Anzalone et al., 2019). While very new, prime editors are already being improved by altering the pegRNA as well as the Cas and reverse transcriptase domains to improve efficiency of editing (Anzalone et al., 2020). Prime editing has been replicated by other groups in mammalian models with no or reduced off-targets (Anzalone et al., 2020), and in rice but with no investigation of off-targets (Hua et al., 2020, Butt et al., 2020, Xu et al., 2020).

Editing by DNA-protein interaction: TALENS, ZFN, and Meganucleases

TALENs

TALENs (transcription activator-like effector nuclease) are synthetic proteins that combine an engineerable DNA-recognition domain as a targeting component and a nonspecific nuclease domain that can cut DNA once bound, both derived from bacterial proteins (Christian et al., 2010). In nature, TALE proteins (without the nuclease domain) are made by plant pathogenic bacteria to upregulate genes in their plant host. Their DNA-recognition domain is made of highly conserved repeats, each with two variable amino acids (repeat variable diresidue; RVD). Following a deciphered code, the RVD dictates the specificity of that repeat, i.e. which DNA base(s) that repeat can bind (Moscou & Bogdanove, 2009). The breakthrough in using TALEs for genome editing came from fusing this engineerable DNA-recognition domain to a FokI nuclease domain, creating a nuclease with designer specificity (Christian et al., 2010). TALENs were developed before CRISPR/Cas systems and the intellectual property around them was decided much more quickly. Thus, they have been used in livestock and plants as early as 2012 (Carlson et al., 2012b, Tan et al., 2012, Zhang et al., 2013).

Key Points: TALENs and their specificity
- TALENs are a combination protein made from Fok1 nuclease and a DNA-binding protein called a TALE that can be modified to target different sequences.
- Two TALENS are required for a DSB, necessitating an optimal target sequence of 30-40 bases, which has a low likelihood of occurring more than once in a genome.
- Using modified Fok1 nucleases that cannot interact with themselves further prevents off-target editing.

For each sequence to be edited, two TALENs targeting adjacent sequences must be designed and built to create a DSB, like dCas9-Fok1 (*Figure 3*). TALENs are built by cloning repeats that correspond to the target DNA sequence in a 1-to-1 manner into a backbone that contains the FokI nuclease domain (Cermak et al., 2011). Kits containing the different TALE repeats and TALEN backbones are available, and companies like ThermoFisher Scientific offer synthesis options complete with design tools. Like the PAM requirement in CRISPR/Cas, typically TALEs require a thymine nucleotide in the position before the target sequence to facilitate

strong binding, thus limiting both target and off-targets sites. However, mutational studies have found variants of TALENs that do not have this requirement (Lamb et al., 2013). Notably, unlike CRISPR/Cas systems that are present across a variety of bacterial species with many derivatives to investigate for biotechnology applications, TALE proteins are largely restricted to two genera of plant pathogenic bacteria: *Xanthomonas* spp. and *Ralstonia solanacearum*. Typically, *Xanthomonas* TALE repeats are used for genome editing as they were the first used to generate TALENs and have a wider variety of resources available (de Lange et al., 2014).

Off-target considerations and specificity

Because a protein domain binds the DNA to direct cleavage, the specificity is less rigid compared to DNA-DNA or DNA-RNA binding. Single RVDs preferentially bind certain nucleotides but often can bind others. For example, the RVD "NN" can bind adenosine or guanine and is commonly used when designing TALENs. This has driven expansion beyond the natural TAL variation to identify RVDs that have the highest binding affinity and/or desired specificities. All 400 RVD combinations (20 amino acids for each of the 2 positions) have been tested, as well as effects of positionality (Miller et al., 2015, Yang et al., 2014, Rogers et al., 2015).

The need for two TALENs for each edit largely decreases the chances of off-targets by increasing the overall size of the target site. **TALENs are generally considered to have a low probability of off-target editing because of the length of the sequence being targeted and the relative ease of computationally designing them to minimize this potential.** The length of each TALE DNA-recognition domain also contributes to the specificity, but in an interesting manner. Increasing the target sequence is usually thought to increase specificity, yet TALEs increase in ability to bind non-target sequences as they get larger; their general affinity for DNA overcomes the specificity requirements as they grow. This specificity tradeoff results in TALEs that target 15-19 nucleotides being the optimum size for specific targeting (Rinaldi et al., 2017). Another interesting possibility is if two of the same TALENs bind DNA near each other allowing for interaction with each other, referred to as homodimerization, an off-target edit can occur. To prevent homodimerization, FokI nuclease domains have been created so that only distinct TALENs, not two identical ones, can interact to make the cut as is needed for the intended edit (Cade et al., 2012).
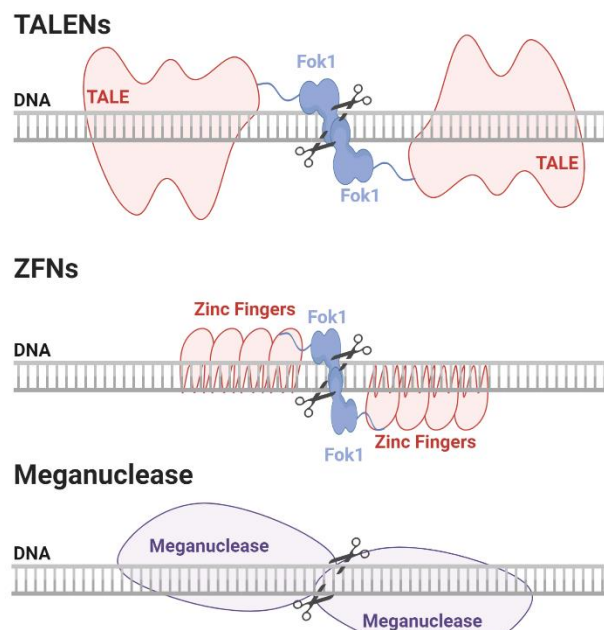

Figure 3. Genome editing tools that rely on DNA-protein binding for targeting. Cutting components are indicated in blue and targeting components in red. Purple indicates that a single component is responsible for targeting and cleavage. Scissors indicate sites where grey DNA strands would be cut.

Zinc Finger Nucleases

Zinc finger nucleases (ZFNs; *Figure 3*) are similar to TALENs, in that a zinc finger DNA binding domain is added to a FokI nuclease and that two ZFNs are required to create a DSB (Kim et al., 1996). Each "finger" binds to 3 nucleotides, allowing for targeting of

sequences of 9-18 bases depending on the number of zinc fingers incorporated, and overall target sequence of 18-36 bases across the two ZFNs. Interactions between fingers contribute to the specificity and make design more challenging and costly as modular assembly is inefficient (Ramirez et al., 2008). A major issue with ZFNs is their tendency to cause cell death during edited, likely due to off-target cleavage (Porteus & Carroll, 2005). However, ZFN editing to knockout the *BLG* gene in cattle to make hypoallergenic milk found no off-target edits (Sun et al., 2018). **ZFNs became less popular once TALENs were introduced as a more efficient, easier alternative to a protein-DNA genome editing approach, but they are still being optimized, especially for human therapeutic applications** (Mussolino et al., 2011, Carlson et al., 2012a, Paschon et al., 2019). A recent report of an optimized ZFN system with new linker options saw no or low off-target cleavage of ZFNs in human T-cells, with an on-target cleavage rate of ~80% (Paschon et al., 2019).

Meganucleases

      Meganucleases (also homing endonucleases; *Figure 3*) are proteins that naturally bind and cleave DNA sequences of 12-40 bases. They are the smallest genome editing tool, as they only involve one component, a single protein, but this means that alterations to change specificity can impact cleavage efficiency. Mutagenesis to create artificial meganucleases targeting diverse sequences has been undertaken with some success (Smith et al., 2006). Meganucleases are considered to be fairly highly specific for their target sequence and there is some evidence that mixtures of domains from different meganucleases could diversify specificities available (Stoddard, 2011). **While their size makes them useful for size-limited delivery to cells, the difficulty in engineering specificity while retaining cleavage ability has limited the general interest and resource development for these proteins.**

> **Key Points: ZFNs and Meganucleases**
> - ZFNs are a combination protein made from Fok1 nuclease and a DNA-binding protein made of "zinc fingers" that can be assembled to target different sequences.
> - Like TALENS, two ZFNs are required to create a DSB. However, TALENS surpassed ZFNs in popularity because they are easier to design and are less toxic to cells.
> - Meganucleases are single proteins that bind and cut DNA but are difficult to engineer for use and improvement.

Generating edited plants and animals

      The following section explains how genome editing occurs once a tool has been selected to alter an organism. The process used can dramatically change specificity and efficiency of genome editing depending on whether genome editing tools are introduced and how many edits are done at one time. For example, if editing tools make edits late in an organism's development, it can lead to **mosaicism** where different parts of a single organism have different genotypes, from being edited and repaired in different ways. In general, the method for editing can easily be considered as critical as the editing tool chosen for facilitating intended edits without introducing unintended ones.

DNA-free versus transgenic editing approaches

      The traditional way to modify or edit a genome is to move DNA instructions (genes) that encode desired components into the organism of interest when it is only a single cell. Depending on the method used, this DNA may or may not integrate into the host genome, making the organism transgenic, and need to be removed at a later point if that is not desired in the final

organism. In the case of editing, the cell uses the DNA instructions to make editing targeting and cutting components which will begin to edit the genome until they are deactivated by the cell and the genes are removed. To avoid the added time of removing any transgenes, reduce the potential for off-target edits from prolonged presence of the editing components, and eliminate the possibility of unintentional insertion of provided DNA within the genome, DNA-free methods to edit plants and animals have been developed. **DNA-free methods move the editing components themselves, as RNA or protein, directly into the cell with no risk of creating a transgenic organism unless a DNA repair template is provided.**

There are various protocols for introducing genome edited tools into animals to create edited animals, with novel methods being actively researched (McFarlane et al., 2019). Conventionally, plasmid DNA encoding proteins and RNAs necessary to make changes is injected into a **zygote**, the initial developmental stage of an organism where only one cell is present. An alternative method is called electroporation where an electrical current is used to shock the cell into taking up the DNA. The provided DNA can integrate randomly at low rates into the genome, or at higher rates by using DNA viruses to promote integration (West & Gill, 2016). Such approaches and others that use viral vectors are employed because of their relative ease of use including in more difficult systems like avian livestock such as chickens and quail (McFarlane et al., 2019, Lee et al., 2020). In contrast, **direct injection or electroporation of RNA and/or protein components instead of DNA reduces off-target editing and the possibility that DNA will be incorporated into the genome in an unintended way.** DNA-free methods are increasingly considered to be a better approach to editing animals, especially because of the difficulties arising from mosaicism as discussed further in the next section.

The typical approach to genetic engineering or genome editing in plants is to use *Agrobacterium*-mediated transformation, a method that uses the natural ability of a plant pathogenic bacterium to insert DNA into a host plant's genome. Another option is to bombard DNA constructs into immature plant embryos or **callus** (undifferentiated plant cells) using a gene gun. **Both *Agrobacterium*-mediated transformation and bombardment create transgenic plants that will have to be bred to remove the DNA encoding editing components**. Retaining the editing component DNA within the genome does allow for editing to occur over time or in the next generation. For example, editing a plant variety that is not as amenable to laboratory manipulation can be done by breeding it with a transgenic plant encoding editing components. A study of maize where CRISPR/Cas9 was introduced by *Agrobacterium*-mediated transfer found that new on-target edits occurred each generation that the transgenes were still present (Char et al., 2017). However, this leads to concerns about the increased risk of off-target edits from active editing components. Yet, a whole-genome sequencing analysis of Cas9 and Cas12a-edited rice plants found that plants from the generation after editing did not have more off-target edits despite the fact that the Cas9 or Cas12a transgene remained (Tang et al., 2018). No edits were found in studies that looked at predicted off-target sites in grapevine (Ren et al., 2016) and wheat (Zhang et al., 2018b), after *Agrobacterium*-mediated transfer of CRISPR components. In cotton, four off-target edits were detected across 14 plants edited with six sgRNAs each (Li et al., 2019). Thus, specificity of transgenic approaches may depend on the target sequence, tool, and organisms used.

In plants, it is possible to deliver purified Cas9 complexed with sgRNAs directly into zygotes or into wall-free cells called **protoplasts**; this is often referred to as RNP delivery. When done in lettuce protoplasts, there were no off-target edits detected while on-target editing was detected up to 46% efficiency (Woo et al., 2015). Using DNA-free methods such as this to edit

whole plants relies on the researchers being able to regenerate whole plants from protoplast or being able to isolate zygotes, both of which can be technically challenging even where possible. RNPs can also be bombarded into callus. In a recent review of DNA-free genome editing in plants, all studies that surveyed for off-target edits found none, across a variety of plant species (Metje-Sprink et al., 2019). **Generally, DNA-free methods are seen as the best direction for optimization to reduce off-target edits.** A direct comparison of off-target edits in maize resulting from methods that were DNA-free or not showed that compared to editing processes involving DNA, off-target edits were reduced to zero in plants generated by RNP delivery, but that the on-target editing rate was also reduced (Young et al., 2019).

DNA viruses are another option for introducing genome editing tools into a plant, as well as repair templates to insert new sequence during HDR (Zaidi & Mansoor, 2017). Use of a modified wheat dwarf virus was able to dramatically increase the editing efficiency and insertion of template during repair in wheat, without the viral DNA integrating into the genome (Gil-Humanes et al., 2016). New research shows the ability to use RNA viruses as a vector for genome editing reagents in the tobacco relative *Nicotiana benthamiana*, though viral delivery methods tend to be size limited and have host ranges that may require tailoring viral vectors to plant species of interest (Ma et al., 2020). A comparison of ways to genome edit plants can be found in *Table 2*.

Table 2. Summary of categories for genome editing of plants.

Category	What is Delivered to the Cell	Methods	Specificity and Efficiency Trends	Pros	Cons
Transgenic	DNA encoding the editing components	*Agrobacterium*-mediated transfer; bombardment	Efficiency is increased Specificity is reduced	Can continue to edit in the next generation	Need to remove unwanted DNA after generating edits
DNA-free	RNA or protein gene editing components	Bombardment; protoplasts	Efficiency is reduced Specificity is increased	No need to remove DNA and no risk of unintended integration	Decreased efficiency can mean having to produce and screen more plants to find desired edits
Viral	Viral genomes encoding the editing components	DNA or RNA plant viruses	Efficiency is increased Specificity is not well assessed	Potential for use in intact plants instead of callus or zygotes; can carry repair template as well	May result in unintended DNA integration; typically need to use smaller editing components (size restriction)

The challenge of mosaicism

A major challenge in reliably editing a multicellular organism comes from mosaicism - when a single organism has more than one genotype, due to individual cells across the organism varying from mutations or other events during development, including the presence of gene editing machinery. When the individual arising from a gene edited embryo (F0) is tested for genetic changes, the results may be incorrect if only sampled from one tissue or may be difficult to interpret if mixtures of unedited and edited cells are present. While possible in plants, mosaicism is typically more of an issue for generating genome edited animals.

Mosaicism is a larger risk when plasmid DNA or nuclease RNA are used in the editing process. Plasmid DNA must be transcribed and the resulting RNA for that transcription or provided nuclease RNA must be translated before editing occurs. If this is delayed or prolonged, editing may happen at a multicellular stage, thus only affecting the tissues that those cells differentiate into. **Once at the multicellular stage, DNA cuts will cause repair events that may introduce different mutations or no mutation in each individual cell as the breaks are repaired, resulting in mosaicism.** It is recommended to screen efficacy of editing in the F1 animals that result from breeding the F0 individual to a non-edited individual. Each F1 individual will come from a single cell from the F0 parent and should therefore not have mosaicism from the editing process that the F0 individuals may have.

Reducing the occurrence of mosaicism will be different for different species and depend on the gene editing technology used. The cell types that are being manipulated can impact specificity and efficiency. Direct germline editing is done in a variety of ways including modification of sperm-producing cells (spermatogonial stem cells) or injection of immature egg cells (oocytes), depending on the animal of interest (Mehravar et al., 2019). In one study, injecting oocytes directly resulted in high levels of mosaicism in pigs (Sato et al., 2015). An alternative is indirect editing of germ cells, done through **somatic cell nuclear transfer** (SCNT) which has been historically used for cloning animals (Bartlett, 2014). **Genetic material from previously edited cells is moved into enucleated, or genetically empty, germ cells that can be implanted for development, removing the possibility of further editing during development**. SCNT is considered to have a low development rate in that many altered cells will not result in a live animal, but high editing rate in that the animals that are produced are likely to have the intended edits (Menchaca et al., 2020). Indirect editing by CRISPR/Cas9 via SCNT has been efficient at achieving desired edits while eliminating mosaicism in pigs (Sheets et al., 2016) and goats (Ni et al., 2014).

Reducing the length that an organism is exposed to editing nucleases seems to be key, with less mosaicism observed when Cas9 protein is injected into a cell, compared to *cas9* RNA or DNA encoding Cas9 (Mehravar et al., 2019). Using *in vitro* fertilization where Cas9 protein and sgRNAs are electroporated into the cell before it divides was very successful to eliminate mosaicism in mice (Hashimoto et al., 2016). Other approaches include using Cas9 that is flagged for quick breakdown by the animal cell. For a more thorough

> **Key Points: Animal Editing and Mosaicism**
> - Mosaicism can occur when editing components make edits in different cells over the course of development, resulting in one organism having cells with different genetic makeup.
> - Reducing the amount of time that editing components are active in a cell/organism can reduce the possibility of mosaicism occurring.
>   - DNA-free methods and indirect germ cell editing reduce mosaicism.

review of mosaicism in genome edited animals including a collation of data across studies and model species, please refer to Mehravar et al. (2019).

Generating multiple edits simultaneously

Given the diversity of traits that are of interest in livestock and crops, as well as the interplay of genes and pathways, it is not surprising that genome editing applications have quickly expanded from editing one target to editing multiple (I.e. multiplex genome editing). Any genome editing technology could be used to edit single traits one at a time, the products of which could then be stacked together through breeding. **Of more interest is the ability to edit multiple genes simultaneously by including multiple targeting components during the editing process**. Because of the size of TALENs and other protein-DNA targeting technologies, they are not as readily adaptable to editing multiple targets at a time. CRISPR/Cas systems rely only on the small RNA component to identify target genes and use a universal large protein component. Thus, adding in multiple guide RNAs each corresponding to a different target allows for the common Cas protein to edit multiple target sites. This can be done by providing multiple RNAs instead of a single one in the DNA-free approach or including an array of guide RNAs to be transcribed in an approach using DNA transformation.

Multiplex genome editing can be used to edit sites close together in the same gene or far apart in completely different genes or chromosomes, at the same time (*Table 3*)**.** The simplest example is the use of two sgRNAs targeting the same gene with the goal of removing the small piece of the DNA between the cut sites, rendering the gene nonfunctional. Multiplex genome editing was able to edit three distinct genes in sheep simultaneously with two guide RNAs targeting each gene (Wang et al., 2016). Multiplex genome editing has been widely used in plants from potato to grapevine, as is reviewed by Armario Najera et al. (2019). Additionally, some guide RNAs are intentionally designed to target a gene family or gene with many copies in the genome, such as in polyploid plants. In this case, ZFNs or TALENs can be quite useful because of their tolerance of different basepairs based on a protein-DNA interaction instead of an RNA-DNA interaction. The "promiscuity" of the recognition sites of these proteins can be exploited to tolerate mismatches between multiple homeologs (copies from genome duplication) in a genome that may vary slightly. Wheat is notoriously difficult for genetic experimentation because it has six of each chromosome, two copies of three genomes that can each differ in sequence for single genes; in one case, a single TALEN pair edited all three copies of a single gene by tolerating a mismatch across the three homeologs (Wang et al., 2014) .

For each new targeting component added to multiplex genome editing, the possibility of off-target edits directed by that component is also introduced. **Increasing the number of targeting components used in a single editing process implies an increase in the probability of off-target edits. Yet each targeting component is still individually designed and chosen to be as specific as possible**. In a genome-wide assessment of multiplex genome editing of 14 targets in the model plant Arabidopsis, no off-target edits were detected (Peterson et al., 2016). In contrast, multiplex genome editing of three genes in Arabidopsis found high off-target editing from one sgRNA despite being predicted to be quite specific; it is possible that the off-target editing was promoted by on-target editing of nearby genes targeted by the other sgRNAs (Zhang et al., 2018a). Clearly, adequate identification of edits becomes more essential as multiplex genome editing increases (see *Confirming target and non-target edits*).

Table 3. Example targets for multiplex genome editing.

Target	Preferred Tool	Method	Example Study
Single gene	CRISPR/Cas	Design two sgRNAs to target different sites in a single gene resulting in the DNA between them being cut out	A single rice gene was modified by targeting two sites within it with Cas12a (Li et al., 2018).
Many unrelated genes	CRISPR/Cas	Design one or more sgRNAs for each gene resulting in indels or removal of DNA at each gene	Three unrelated genes were disrupted in sheep with Cas9 (Wang et al., 2016) .
Multiple copies of a single gene (homeologs)	TALENs	Design a pair of TALENS to accommodate mismatches across the gene copies resulting in edits at each copy	Three homeologs were edited in wheat with a pair of TALENs (Wang et al., 2014).

Identifying efficiency and specificity of edits

Given all the possible tools and methods for genome editing in plants and animals, each situation is likely to be unique. But many of these decisions largely inform efficiency – instead, the selection of a target sequence is possibly the most important way to encourage specificity. Furthermore, screening for intended and unintended edits will ultimately reveal whether genome editing was successful and the outcome can be used in scientific research or further developed for market. In this section, target selection and design considerations as well as the ways in which on- and off-target edits can be identified and assessed will be discussed.

Target selection

Target sequence selection and guide component design are regarded as the most critical ways to reduce off-target editing. Editors design the targeting components of editing tools using computational methods to predict efficacy and off-targets. There are many tools that have been developed by industry and academic groups to computationally predict where in a genome a targeting component, such as a sgRNA or TALEN, will bind to (Bradford & Perrin, 2019, Bogdanove & Booher, 2016). The variety of target prediction tools ranging from scoring-based to alignment-based models that are current developed for CRISPR/Cas9 was recently summarized by Naeem et al. (2020). Use of these computational tools is both crucial and common practice when designing components to reduce the possibility of off-target edits and to identify genomic locations that will need to be checked for edits if whole genome sequencing is not feasible. For example, CRISPR editing in plants is largely considered to have a low risk of off-target editing if the guide RNA is designed so every potential off-target has at least 2 mismatches, especially in the bases closest to the PAM, as demonstrated in many papers including by LeBlanc et al. (2018). Thus, selecting a target where the corresponding sgRNA does not correspond to off-target sites with two or fewer mismatches chiefly reduces the likelihood of off-target edits occurring.

Furthermore, the on and off-target *efficiency* of a single guide component can vary greatly. In one study of two guide RNAs targeting the same gene in sweet basil (*Ocimum basilicum*), one resulted in over 90% of plants with edits while the other resulted in negligible edits (Navet & Tian, 2020). The reasons behind this are still being researched, but DNA

packaging within the cell is likely to play a role. **Thus, it can be difficult to computationally predict which sites will be most amenable to edits and which will not, so multiple guide components may be designed and tested that target neighboring sequences to achieve the same effect**. Temperature also appears to affect efficiency of editing: on-target editing with SpCas9 is more efficient at a higher temperature of 37°C in Arabidopsis and citrus (LeBlanc et al., 2018), while editing of pig and cattle fibroblasts with TALENs works better at a lower temperature of 30°C (Tan et al., 2013).

Preliminary testing for edits and off-targets

When trying to edit a plant gene, a typical first step is to edit the gene in protoplast cell culture and then check for editing efficiency and off-targets. **Commonly, multiple targeting components are tested for specificity in small-scale, cell-culture laboratory experiment before generating whole organisms**; similarly, protoplasts are often used to test new technologies in a system as well, such as trying out a new Cas protein for the first time in a species. Protoplasts of Arabidopsis, tobacco, lettuce, and rice were all used as a first test for delivering sgRNA-Cas9 complexes into plants in a DNA-free way (Woo et al., 2015). In some cases, whole plants can even be regenerated from protoplasts, though an important caveat is that not every plant is amenable to protoplast generation or regeneration from protoplasts. Similar to protoplasts, embryo or tissue sequencing can be done with livestock before generating embryos that will be transferred for gestation (Menchaca et al., 2020). In one such case, Wang et al. (2016) screened sheep fibroblasts for off-target edits, finding no edits at the top predicted off-target sites for a total of six sgRNAs targeting three genes, before moving to production of gene-edited sheep.

Furthermore, there is a myriad of other techniques that can be used to test for genome-wide DSBs caused by a genome editing tool (Naeem et al., 2020). One possibility for testing efficiency of genome editing tools in any organism's genome is to use *in vitro* methods to test out potential off-targets in a cell-free way (Cameron et al., 2017, Tsai et al., 2017). However, a genome editing tool in an actual cell will likely make less off-target edits than in an *in vitro* assay, given the cellular environment and packaging of DNA obscuring some potential sites. A popular cell-based assay called GUIDE-seq puts a specific sequence where all DSBs occur, acting as a flag to identify all locations in the genome that were cut (Tsai et al., 2015). **The applicability of these testing methods is in assessing genome-wide cleavage sites as a proof of concept, but it can shed light on sites that need to be investigated for edits in a final edited organism** as was done by Young et al. (2019) in maize.

Confirming target and non-target edits

Whether testing a new guide component or checking for edits in an organism intended for distribution, the identification of target and non-target edits is critical to ensuring that no unintended edits were introduced. In an ideal world, the genomes of all edited organisms would be sequenced and thoroughly assessed for all possible off-target and on-target effects. Whole genome sequencing (WGS) determines the sequence of an individual's genome, typically by assembling sequencing reads to a known, reference genome of the organism. The benefits of WGS include having a relatively complete picture of mutations arising from the genome editing process, though the efficacy of WGS to detect both on- and off-target changes depends on the sequencing technology used as well as downstream analysis. When considering that the editing process typically generates multiple individuals to screen, especially in plants where hundreds

16

may be created to look for a single editing event, the ability to sequence each genome quickly escalates in price and practicality. Yet, **decreases in cost and increases in throughput make WGS increasingly valid especially for crops and livestock with more genetic resources, or for later-stage assessment of top-candidate individuals intended for commercial use.**

Table 4. Detection methods used for determining where edits occurred and the nature of those edits.

Detection Method	Benefits	Barriers	Caveats
Whole genome sequencing	Captures most-to-all changes in the genome that occurred during the editing process	Time and resource intensive, low throughput; requires a reference genome to compare to	Still limited by what the developer looks for and how well the genome is understood
Directed sequencing	Fairly high throughput and focused on likely changes	Requires developer to determine which sites to sequence	Limited to the chosen areas and may miss unexpected or large changes
Other directed amplification	Most high throughput and cost effective to determine where edits occurred	Does not reveal the details of an edit, such as specific sequence changes	Limited to chosen areas; should be followed up with additional sequencing

   The difficulties of WGS includes cost and feasibility depending on the organism, in addition to needing a non-edited genome to compare to. For example, plants with multiple copies of their genomes (high ploidy) like wheat or potato can be difficult to reliably sequence, so few varieties have quality reference genomes. In the cases where technical and resource limitations prevent WGS, researchers will assess editing in a directed manner by amplifying and sequencing sites in the genome that were expected to be edited based on computational analysis during guide component design. For example, if a given guide component has only four likely off-target locations in the genome, each of those can be sequenced in a much higher throughput and more cost-efficient manner than sequencing genomes for every potentially edited organism generated. However, this is limited to only those sites screened, thus is biased based on computational tools used and a developer's threshold for sites to investigate. **Notably, directed sequencing does not detect large changes or unexpected insertions, unless additional sequencing is done, such as sequencing targeted at any DNA sequences provided during the genome editing process.** Uncaptured changes can include unintended genomic rearrangements (Hahn & Nekrasov, 2019) or integrations of DNA provided during the process, such as repair template or DNA encoding editing components (Zhang et al., 2018a).

   When considering confirmation of edits and preliminary experimentation, different standards may be more or less feasible depending on whether the edited organism is a plant or animal. Typically, many individual plants are created resulting in a large number to screen with the intention of finding the individuals with the desired on-target edit and no other edits. Versions of directed amplification that do not involve sequencing are also done and can be very useful for high-throughput screening; these tend to detect general presence of an edit, but cannot always show the type of edit and are summarized by Zischewski et al. (2017), though not all have been used in plants and/or livestock. Organisms that are screened in this way are typically

investigated further with targeted or WGS once an edit is detected. In livestock systems, less individuals are produced because of the length and labor involved in the process (gestation, care, etc.) so more time is spent in preliminary testing which may also been done by directed amplification.

Unintended on-target edits

   It is worth at least noting that on-target deleterious edits are also possible, necessitating experimental confirmation that edits at the target site are expected, without unintended edits or incorporation of other DNA. In fact, perhaps the most widely known case of an edit resulting in unintended changes comes from development of cattle edited for hornlessness. Calves were edited with TALENs, using HDR to insert a polled allele of a gene, resulting in hornless cattle (Carlson et al., 2016). However, the DNA template containing the allele was part of a larger bacterial DNA plasmid, all of which was incorporated into genome-edited calves at the target site (Norris et al., 2020). Despite sequencing the genomes of these calves, the insertion was initially missed, as the sequence was likely not checked for integration of the plasmid, a clear oversight. Whether this larger insertion had any impact on the edited cattle is unknown. Similar instances have been documented in Arabidopsis when T-DNA used in *Agrobacterium*-mediated transformation was found to be unexpectedly inserted at a target site (Zhang et al., 2018a). **These examples underscore the necessity of identifying the outcomes of all edits, given the unknown impact of unintended on- or off-target changes, and the careful consideration of the use of DNA during the process of genome editing.**

Final Suggestions and Guidelines:

Editing tools should be chosen based on the current literature as to which is the best for each organism, given that it varies and will continue to change as new tools and improvements are introduced both for process and technology.

sgRNAs that have off-targets with one or less mismatches should be avoided as evidence indicates that two or more mismatches may be enough to prevent off-target editing. For protein-DNA tools, mismatches should also be limited through off-target prediction and the use of stringent binding modules.

At minimum, top off-target sites should be screened for editing. Initial screens in plant or animal cell cultures can help avoid using targeting components that edit non-target locations when generating whole organisms and can identify the most relevant sites to screen in final organisms.

Target sequence should be checked that edits are as intended and do not contain additional material such as editing components or extraneous template.

DNA intended to be used as template for HDR should be restricted as much as possible to the sequence intended for insertion, to reduce the possibility that extraneous DNA, including antibiotic resistance, is incorporated into the edited organism's genome.

In many cases, editing by providing functional components (RNA/protein) instead of DNA will result in less mosaicism, less off-target edits, and less potential for integration of DNA at an edited site or elsewhere in the genome, though usually at the tradeoff of a lower editing efficiency.

Glossary:

- Base/nucleotide – a single unit of DNA
  - Adenosine (A), guanosine (G), thymine (T), or cytosine (C)
- Basepair – two bases that connect the double strands of DNA, one in each strand
- Callus/Calli – undifferentiated plant cells that can be grown and manipulated in a laboratory and used to produce new individual plants
- Cas12a – an increasingly popular Cas protein that creates an uneven break in the DNA that can promote HDR
- CRISPR/Cas – genome editing tool that uses an RNA guide for targeting and Cas nuclease for cutting
- Double stranded break (DSB) - a cut in both strands of DNA resulting in a free end that needs to be reconnected for repair
- Efficiency – degree that the tool edits the target site as opposed to no edit occurring
- Fok1 – a small nuclease protein that makes a nick in the DNA strand; the cutting component of dCas9-Fok1, TALENs, and ZFNs
- Genome – all of an organism's genetic material
- Homology directed repair (HDR) – a cell repair mechanism to fix a DNA break using template DNA either naturally in the cell or provided during the gene editing process
- Indel – refers to insertions or deletions of DNA bases, especially during repair by NHEJ
- Meganuclease – genome editing tool consisting of a single protein that can target and cut DNA
- Mosaicism – when different cells in a single organism have different genetic makeup due to mutations occurring during development
- Nick – a cut in only one strand of the DNA; requires repair but does not create a free end
- Non-homology end joining (NHEJ) – a cell repair mechanism to fix a DNA break without using any template
- Nuclease – the protein component of an editing tool that cuts the DNA
- Off-target edits – changes to the DNA through cutting and repair outside of the area or the genome intended to be edited
- Protoplast – a intact cell missing the cell wall that in some cases can be used to generate a new plant
- Protospacer Adject Motif (PAM) – the short sequence that must be beside a target site for a Cas nuclease to recognized and make a cut
- Repeat variable diresidue (RVD) – amino acids in each repeat of a TALEN that determine what DNA base that repeat will bind to
- Somatic cell nuclear transfer (SCNT) – moving the genetic material from a non-germ cell into an "empty" germ cell that does not have any DNA
- SpCas9 – the most popular Cas nuclease; creates an even break in DNA when directed by an sgRNA
- Specificity – degree that the tool edits at *only* the target site, the best match in the genome for the guide component, and not at off-target locations
- TALENs – genome editing tool that uses a protein guide (TALE) for targeting and a Fok1 nuclease for cutting; requires two for a DSB

- Zinc Finger Nuclease (ZFN) – genome editing tool that uses a protein guide (zinc finger) for targeting and a Fok1 nuclease for cutting; requires two for a DSB
- Zygote – a single cell that is the earliest developmental stage of a new individual

Acknowledgements and Conflict of Interest

Highlighted References – OA indicates Open Access that is publicly available

Issue paper on genome editing in agriculture

- Council for Agricultural Science and Technology (CAST). 2018. Genome Editing in Agriculture: Methods, Applications, and Governance—A paper in the series on The Need for Agricultural Innovation to Sustainably Feed the World by 2050. Issue Paper 60. CAST, Ames, Iowa. **OA**

Comprehensive reviews of CRISPR-Cas-based editing and reduction of off-targets

- Anzalone, A. V., et al. (2020). "Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors." Nature Biotechnology.
- Chen, K., et al. (2019). "CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture." Annu Rev Plant Biol 70: 667-697. **OA**
- Naeem, M., et al. (2020). "Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing." Cells 9(7). **OA**

Reviews of genome editing in animal systems with summary tables of published studies

- Lee, J., et al. (2020). "Current Approaches and Applications in Avian Genome Editing." International Journal of Molecular Sciences 21(11). **OA**
- Kalds, P., et al. (2019). "Sheep and Goat Genome Engineering: From Random Transgenesis to the CRISPR Era." Frontiers in Genetics 10: 750-750. **OA**
- Zhao, J., et al. (2019). "Genome editing in large animals: current status and future prospects." National Science Review 6(3): 402-420. **OA**

Reviews of genome editing in plant systems with summary tables of published studies

- Alok, A., et al. (2020). "The Rise of the CRISPR/Cpf1 System for Efficient Genome Editing in Plants." Front Plant Sci 11(264): 264. **OA**
- Armario Najera, V., et al. (2019). "Applications of multiplex genome editing in higher plants." Current Opinion in Biotechnology 59: 93-102.
- Metje-Sprink, J., et al. (2019). "DNA-Free Genome Editing: Past, Present and Future." Frontiers in Plant Science 9: 1957-1957. **OA**

References

Alok A, Sandhya D, Jogam P, *et al.*, 2020. The Rise of the CRISPR/Cpf1 System for Efficient Genome Editing in Plants. *Front Plant Sci* **11**, 264.

Anzalone AV, Koblan LW, Liu DR, 2020. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38**, 824-44.

Anzalone AV, Randolph PB, Davis JR, *et al.*, 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-57.

Armario Najera V, Twyman RM, Christou P, Zhu C, 2019. Applications of multiplex genome editing in higher plants. *Current Opinion in Biotechnology* **59**, 93-102.

Banakar R, Schubert M, Collingwood M, Vakulskas C, Eggenberger AL, Wang K, 2020. Comparison of CRISPR-Cas9/Cas12a Ribonucleoprotein Complexes for Genome Editing Efficiency in the Rice Phytoene Desaturase (OsPDS) Gene. *Rice* **13**, 4.

Bartlett Z, 2014. Somatic Cell Nuclear Transfer in Mammals (1938-2013). In: Turriziani Colonna F, ed. *Embryo Project Encyclopedia.* Arizona State University. School of Life Sciences. Center for Biology and Society. Embryo Project Encyclopedia.

Bernabé-Orts JM, Casas-Rodrigo I, Minguet EG, *et al.*, 2019. Assessment of Cas12a-mediated gene editing efficiency in plants. *Plant Biotechnology Journal* **17**, 1971-84.

Bogdanove AJ, Booher NJ, 2016. Online Tools for TALEN Design. *Methods Mol Biol* **1338**, 43-7.

Bradford J, Perrin D, 2019. A benchmark of computational CRISPR-Cas9 guide design methods. *PLOS Computational Biology* **15**, e1007274.

Butt H, Rao GS, Sedeek K, Aman R, Kamel R, Mahfouz M, 2020. Engineering herbicide resistance via prime editing in rice. *Plant Biotechnology Journal* **n/a**.

Cade L, Reyon D, Hwang WY, *et al.*, 2012. Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Research* **40**, 8001-10.

Cameron P, Fuller CK, Donohoue PD, *et al.*, 2017. Mapping the genomic landscape of CRISPR–Cas9 cleavage. *Nature Methods* **14**, 600-6.

Carey K, Ryu J, Uh K, *et al.*, 2019. Frequency of off-targeting in genome edited pigs produced via direct injection of the CRISPR/Cas9 system into developing embryos. *BMC Biotechnology* **19**, 25.

Carlson DF, Fahrenkrug SC, Hackett PB, 2012a. Targeting DNA With Fingers and TALENs. *Molecular therapy. Nucleic acids* **1**, e3-e.

Carlson DF, Lancto CA, Zang B, *et al.*, 2016. Production of hornless dairy cattle from genome-edited cell lines. *Nature Biotechnology* **34**, 479-81.

Carlson DF, Tan W, Lillico SG, *et al.*, 2012b. Efficient TALEN-mediated gene knockout in livestock. *Proceedings of the National Academy of Sciences* **109**, 17382-7.

Cast, 2018. Genome Editing in Agriculture: Methods, Applications, and Governance. In: Bogdanove Adam J, ed. (Cast A, Ia, ed.)

Cermak T, Doyle EL, Christian M*, et al.*, 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82.

Char SN, Neelakandan AK, Nahampun H*, et al.*, 2017. An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnology Journal* **15**, 257-68.

Christian M, Cermak T, Doyle EL*, et al.*, 2010. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **186**, 757.

De Lange O, Binder A, Lahaye T, 2014. From dead leaf, to new life: TAL effectors as tools for synthetic biology. *The Plant Journal* **78**, 753-71.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK, 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology* **32**, 279-84.

Gil-Humanes J, Wang Y, Liang Z*, et al.*, 2016. High efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *The Plant Journal*, n/a-n/a.

Guilinger JP, Thompson DB, Liu DR, 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnology* **32**, 577-82.

Hahn F, Nekrasov V, 2019. CRISPR/Cas precision: do we need to worry about off-targeting in plants? *Plant Cell Reports* **38**, 437-41.

Hashimoto M, Yamashita Y, Takemoto T, 2016. Electroporation of Cas9 protein/sgRNA into early pronuclear zygotes generates non-mosaic mutants in the mouse. *Dev Biol* **418**, 1-9.

Hu JH, Miller SM, Geurts MH*, et al.*, 2018. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63.

Hua K, Jiang Y, Tao X, Zhu J-K, 2020. Precision genome engineering in rice using prime editing system. *Plant Biotechnology Journal* **n/a**.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E, 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816.

Kim YG, Cha J, Chandrasegaran S, 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* **93**, 1156.

Kleinstiver BP, Prew MS, Tsai SQ*, et al.*, 2015. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nature Biotechnology* **33**, 1293-8.

Kocak DD, Josephs EA, Bhandarkar V, Adkar SS, Kwon JB, Gersbach CA, 2019. Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nature Biotechnology* **37**, 657-66.

Lamb BM, Mercer AC, Barbas CF, 3rd, 2013. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Research* **41**, 9779-85.

Leblanc C, Zhang F, Mendez J, *et al.*, 2018. Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *The Plant Journal* **93**, 377-86.

Lee J, Kim DH, Lee K, 2020. Current Approaches and Applications in Avian Genome Editing. *International journal of molecular sciences* **21**.

Lee K, Zhang Y, Kleinstiver BP, *et al.*, 2019. Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnol J* **17**, 362-72.

Li J, Manghwar H, Sun L, *et al.*, 2019. Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. *Plant Biotechnology Journal* **17**, 858-68.

Li S, Li J, Zhang J, *et al.*, 2018. Synthesis-dependent repair of Cpf1-induced double strand DNA breaks enables targeted gene replacement in rice. *Journal of Experimental Botany* **69**, 4715-21.

Lin Q, Zong Y, Xue C, *et al.*, 2020. Prime genome editing in rice and wheat. *Nature Biotechnology* **38**, 582-5.

Ma X, Zhang X, Liu H, Li Z, 2020. Highly efficient DNA-free plant genome editing using virally delivered CRISPR–Cas9. *Nature Plants* **6**, 773-9.

Mcfarlane GR, Salvesen HA, Sternberg A, Lillico SG, 2019. On-Farm Livestock Genome Editing Using Cutting Edge Reproductive Technologies. **3**.

Mehravar M, Shirazi A, Nazari M, Banan M, 2019. Mosaicism in CRISPR/Cas9-mediated genome editing. *Dev Biol* **445**, 156-62.

Menchaca A, Dos Santos-Neto PC, Mulet AP, Crispo M, 2020. CRISPR in livestock: From editing to printing. *Theriogenology* **150**, 247-54.

Metje-Sprink J, Menz J, Modrzejewski D, Sprink T, 2019. DNA-Free Genome Editing: Past, Present and Future. *Frontiers in Plant Science* **9**, 1957-.

Mikami M, Toki S, Endo M, 2016. Precision Targeted Mutagenesis via Cas9 Paired Nickases in Rice. *Plant Cell Physiol* **57**, 1058-68.

Miller JC, Zhang L, Xia DF, *et al.*, 2015. Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat Meth* **12**, 465-71.

Mishra R, Joshi RK, Zhao K, 2020. Base editing in crops: current advances, limitations and future implications. *Plant Biotechnol J* **18**, 20-31.

Moscou MJ, Bogdanove AJ, 2009. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **326**, 1501.

Mussolino C, Morbitzer R, Lütge F, Dannemann N, Lahaye T, Cathomen T, 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Research* **39**, 9283-93.

Naeem M, Majeed S, Hoque MZ, Ahmad I, 2020. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. *Cells* **9**.

Navet N, Tian M, 2020. Efficient targeted mutagenesis in allotetraploid sweet basil by CRISPR/Cas9. *Plant Direct* **4**, e00233.

Ni W, Qiao J, Hu S*, et al.*, 2014. Efficient gene knockout in goats using CRISPR/Cas9 system. *PLoS One* **9**, e106718.

Norris AL, Lee SS, Greenlees KJ, Tadesse DA, Miller MF, Lombardi HA, 2020. Template plasmid integration in germline genome-edited cattle. *Nature Biotechnology* **38**, 163-4.

Paschon DE, Lussier S, Wangzor T*, et al.*, 2019. Diversifying the structure of zinc finger nucleases for high-precision genome editing. *Nature Communications* **10**, 1133.

Peterson BA, Haak DC, Nishimura MT*, et al.*, 2016. Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis. *PLoS One* **11**, e0162169.

Porteus MH, Carroll D, 2005. Gene targeting using zinc finger nucleases. *Nature Biotechnology* **23**, 967-73.

Ramirez CL, Foley JE, Wright DA*, et al.*, 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods* **5**, 374-5.

Ran FA, Hsu Patrick d, Lin C-Y*, et al.*, 2013. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380-9.

Ren C, Liu X, Zhang Z*, et al.*, 2016. CRISPR/Cas9-mediated efficient targeted mutagenesis in Chardonnay (Vitis vinifera L.). *Scientific Reports* **6**, 32289.

Rinaldi FC, Doyle LA, Stoddard BL, Bogdanove AJ, 2017. The effect of increasing numbers of repeats on TAL effector DNA binding specificity. *Nucleic Acids Research* **45**, 6960-70.

Rogers JM, Barrera LA, Reyon D*, et al.*, 2015. Context influences on TALE-DNA binding revealed by quantitative profiling. *Nat Commun* **6**, 7440.

Sato M, Koriyama M, Watanabe S*, et al.*, 2015. Direct Injection of CRISPR/Cas9-Related mRNA into Cytoplasm of Parthenogenetically Activated Porcine Oocytes Causes Frequent Mosaicism for Indel Mutations. *International journal of molecular sciences* **16**, 17838-56.

Sheets TP, Park C-H, Park K-E, Powell A, Donovan DM, Telugu BP, 2016. Somatic Cell Nuclear Transfer Followed by CRIPSR/Cas9 Microinjection Results in Highly Efficient Genome Editing in Cloned Pigs. *International journal of molecular sciences* **17**, 2031.

Smith J, Grizot S, Arnould S*, et al.*, 2006. A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Research* **34**, e149-e.

Stoddard BL, 2011. Homing Endonucleases: From Microbial Genetic Invaders to Reagents for Targeted DNA Modification. *Structure* **19**, 7-15.

Sun Z, Wang M, Han S*, et al.*, 2018. Production of hypoallergenic milk from DNA-free beta-lactoglobulin (BLG) gene knockout cow using zinc-finger nucleases mRNA. *Scientific Reports* **8**, 15430.

Tan W, Carlson DF, Lancto CA*, et al.*, 2013. Efficient nonmeiotic allele introgression in livestock using custom endonucleases. *Proceedings of the National Academy of Sciences* **110**, 16526-31.

Tan WS, Carlson DF, Walton MW, Fahrenkrug SC, Hackett PB, 2012. Precision editing of large animal genomes. *Advances in genetics* **80**, 37-97.

Tang X, Liu G, Zhou J*, et al.*, 2018. A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol* **19**, 84.

Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK, 2017. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nature Methods* **14**, 607-14.

Tsai SQ, Zheng Z, Nguyen NT*, et al.*, 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology* **33**, 187-97.

Wang X, Niu Y, Zhou J*, et al.*, 2016. Multiplex gene editing via CRISPR/Cas9 exhibits desirable muscle hypertrophy without detectable off-target effects in sheep. *Scientific Reports* **6**, 32271.

Wang Y, Cheng X, Shan Q*, et al.*, 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* **32**, 947-51.

West J, Gill WW, 2016. Genome Editing in Large Animals. *Journal of equine veterinary science* **41**, 1-6.

Woo JW, Kim J, Kwon SI*, et al.*, 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology* **33**, 1162-4.

Wu H, Liu Q, Shi H*, et al.*, 2018. Engineering CRISPR/Cpf1 with tRNA promotes genome editing capability in mammalian systems. *Cellular and Molecular Life Sciences* **75**, 3593-607.

Xie J, Ge W, Li N*, et al.*, 2019. Efficient base editing for multiple genes and loci in pigs using base editors. *Nature Communications* **10**, 2852.

Xu R, Li J, Liu X, Shan T, Qin R, Wei P, 2020. Development of Plant Prime-Editing Systems for Precise Genome Editing. *Plant Communications* **1**, 100043.

Yang J, Zhang Y, Yuan P*, et al.*, 2014. Complete decoding of TAL effectors for DNA recognition. *Cell Res* **24**, 628-31.

Yang Y, Wang K, Wu H*, et al.*, 2016. Genetically humanized pigs exclusively expressing human insulin are generated through custom endonuclease-mediated seamless engineering. *Journal of Molecular Cell Biology* **8**, 174-7.

Yin X, Biswal AK, Dionora J*, et al.*, 2017. CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental gene EPFL9 in rice. *Plant Cell Reports* **36**, 745-57.

Young J, Zastrow-Hayes G, Deschamps S*, et al.*, 2019. CRISPR-Cas9 Editing in Maize: Systematic Evaluation of Off-target Activity and Its Relevance in Crop Improvement. *Scientific Reports* **9**, 6729.

Zaidi SS-E-A, Mansoor S, 2017. Viral Vectors for Plant Genome Engineering. *Frontiers in Plant Science* **8**.

Zetsche B, Gootenberg Jonathan s, Abudayyeh Omar o*, et al.*, 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759-71.

Zhang Q, Xing H-L, Wang Z-P*, et al.*, 2018a. Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention. *Plant Molecular Biology* **96**, 445-56.

Zhang S, Zhang R, Song G*, et al.*, 2018b. Targeted mutagenesis using the Agrobacterium tumefaciens-mediated CRISPR-Cas9 system in common wheat. *BMC Plant Biol* **18**, 302.

Zhang Y, Zhang F, Li X*, et al.*, 2013. Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol* **161**, 20-7.

Zhou S, Cai B, He C*, et al.*, 2019. Programmable Base Editing of the Sheep Genome Revealed No Genome-Wide Off-Target Mutations. *Frontiers in genetics* **10**.

Zischewski J, Fischer R, Bortesi L, 2017. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnology Advances* **35**, 95-104.