Review article

Patterns of CRISPR/Cas9 activity in plants, animals and microbes

Luisa Bortesi1,†, Changfu Zhu2,†, Julia Zischewski1, Lucia Perez2, Ludovic Bassié2, Riad Nadi2, Giobbe Forni2, Sarah Boyd Lade2, Erika Soto2, Xin Jin2, Vicente Medina2, Gemma Villorbina2, Pilar Muñoz2, Gemma Farre2, Rainer Fischer1,3, Richard M. Twyman2, Teresa Capell2, Paul Christou2,5 and Stefan Schillberg1,*

1Institute for Molecular Biotechnology, RWTH Aachen University, Aachen, Germany
2Department of Plant Production and Forestry Science, School of Agrifood and Forestry Science and Engineering (ETSEA), University of Lleida-Agrotecnico Center, Lleida, Spain
3Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Aachen, Germany
4TRM Ltd, York, UK
5ICREA, Catalan Institute for Research and Advanced Studies, Barcelona, Spain

Received 11 July 2016; revised 5 September 2016; accepted 7 September 2016.
*Correspondence (Tel +49 241 6085 11050; fax +49 241 6085 50050; email stefan.schillberg@ime.fraunhofer.de)
†These authors contributed equally to the work.

Keywords: genome editing, mutational signature, off-target mutations, on-target mutations, sgRNA design, site-directed mutagenesis, species-dependent effects.

Summary

The CRISPR/Cas9 system and related RNA-guided endonucleases can introduce double-strand breaks (DSBs) at specific sites in the genome, allowing the generation of targeted mutations in one or more genes as well as more complex genomic rearrangements. Modifications of the canonical CRISPR/Cas9 system from Streptococcus pyogenes and the introduction of related systems from other bacteria have increased the diversity of genomic sites that can be targeted, providing greater control over the resolution of DSBs, the targeting efficiency (frequency of on-target mutations), the targeting accuracy (likelihood of off-target mutations) and the type of mutations that are induced. Although much is now known about the principles of CRISPR/Cas9 genome editing, the likelihood of different outcomes is species-dependent and there have been few comparative studies looking at the basis of such diversity. Here we critically analyse the activity of CRISPR/Cas9 and related systems in different plant species and compare the outcomes in animals and microbes to draw broad conclusions about the design principles required for effective genome editing in different organisms. These principles will be important for the commercial development of crops, farm animals, animal disease models and novel microbial strains using CRISPR/Cas9 and other genome-editing tools.

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPRs) are repetitive sequences found in bacterial and archaeal genomes interrupted by spacers captured from previously encountered virus genomes and other invasive DNA. Their function is to provide a form of adaptive immunity via CRISPR-associated (Cas) proteins that act as RNA-directed endonucleases to degrade the same type of invasive DNA if it is encountered again (Lee et al., 2015). Three major CRISPR/Cas systems have been described (Kumar and Jain, 2015) although several additional systems have been reported more recently (Makarova et al., 2011, 2015). In type II systems (Jinek et al., 2012), fragments of invasive DNA (protospacers) approximately 20 bp in length are captured due to their proximity to a short and highly degenerate sequence known as a protospacer adjacent motif (PAM) and these fragments become the spacers in the genomic CRISPR array. Transcription of the array yields a long transcript which is processed into shorter CRISPR RNAs (crRNAs), each representing a single spacer. The crRNA forms a complex with endonuclease Cas9 and a transactivating crRNA (tracrRNA) which mediates the interaction. The Cas9 ribonucleoprotein (RNP) complex then binds to DNA containing a PAM and a protospacer matching the crRNA. Cleavage occurs three nucleotides upstream of the PAM on both strands, mediated by the Cas9 endonuclease domains RuvC and HNH, respectively, introducing a precise double-strand break (DSB) with blunt ends that causes target DNA degradation (Chen and Gao, 2014; Doudna and Charpentier, 2014; Osakabe and Osakabe, 2015).

The ability of the type II CRISPR/Cas9 system to recognize specific DNA targets has been exploited to develop an RNA-guided genome-editing platform that is more versatile than equivalent platforms involving protein-based DNA-binding modules such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs; Hsu et al., 2014). The natural system has been converted into a universal genome-editing platform by reducing it to two convenient components (Figure 1a). The first is the Cas9 endonuclease, typically from Streptococcus pyogenes (SpCas9), which for eukaryotic targets is equipped with a nuclear localization signal (Belhaj et al., 2013, 2015). The second is a synthetic guide RNA (sgRNA) combining the tracrRNA and crRNA functions of the natural system into a single molecule (Belhaj et al., 2015; Sander and Joung, 2014). The sgRNA targets a unique 20-bp sequence in the genome of the host organism, and any sequence can be chosen as long as it is adjacent to a PAM. Codon-optimized versions of the cas9 gene offer maximum activity in different host species (Bortesi and Fischer, 2015) although the wild-type Cas9 is also active in...
heterologous systems such as rice protoplasts (Jiang et al., 2013b). If Cas9 retains its normal catalytic activity, a blunt DSB is generated at the genomic target site as would be the case in the natural bacterial environment. In higher eukaryotes, the DSB is usually repaired by nonhomologous end joining (NHEJ), an error-prone pathway that tends to introduce small insertions and deletions, collectively known as indels (Belhaj et al., 2013; Qu/Tier, 2016). If no donor DNA is present, these indels are the only footprints of editing and they are often used to trigger frameshift mutations by targeting an exon near the 5′ end of the gene, but if donor DNA is present the same NHEJ events can facilitate the neat insertion of a DNA cassette. These events are shown in the upper left panel of Figure 2. If one of the endonuclease domains of Cas9 is mutated (e.g. RuvC with the D10A mutation as shown in the figure or HNH with the H840A mutation, not shown) generating a Cas9 nickase. Using two sgRNAs matching adjacent genomic regions, a staggered DSB can be generated by two paired nickases. (c) If both endonuclease domains of Cas9 are mutated, the enzyme becomes catalytically inactive and is called dead Cas9 (dCas9). The dCas9 protein can still bind at its target and if fused to a nonspecific endonuclease such as FokI can generate staggered DSBs. In both (b) and (c), two precisely disposed protospacers have to be found in the genome for cleavage to occur, greatly reducing the number of possible off-target effects.

Figure 1 The engineered CRISPR/Cas9 system for genome editing. (a) Outline of the two components required for targeted cleavage: delivery of DNA constructs for transcription of the Cas9 nuclease by RNA polymerase II and the synthetic guide RNA (sgRNA) by RNA polymerase III (usually the U6 or U3 promoter) is the most common procedure, especially in plants. Alternatively, the two components can be provided as RNA or directly as a ribonucleoprotein complex (RNP; not shown). The sgRNA contains a 20-nt-long sequence complementary to the genomic target (protospacer). When the Cas9/sgRNA complex finds a matching target in the genome followed by an NGG stretch called protospacer adjacent motif (PAM), the two endonuclease domains in Cas9 (RuvC and HNH) cleave the noncomplementary and complementary strands in the target, respectively, generating a blunt double-strand break (DSB) 3-bp upstream of the PAM. The part of sgRNA proximal to the PAM (in pink) is called the seed region, and base pairing with the protospacer in this region is strictly required for recognition and cleavage of the target. Mismatches in the PAM-distal region are tolerated to some extent. (b) One endonuclease domain can be mutated (e.g. RuvC with the D10A mutation as shown in the figure or HNH with the H840A mutation, not shown) generating a Cas9 nickase. Using two sgRNAs matching adjacent genomic regions, a staggered DSB can be generated by two paired nickases. (c) If both endonuclease domains of Cas9 are mutated, the enzyme becomes catalytically inactive and is called dead Cas9 (dCas9). The dCas9 protein can still bind at its target and if fused to a nonspecific endonuclease such as FokI can generate staggered DSBs. In both (b) and (c), two precisely disposed protospacers have to be found in the genome for cleavage to occur, greatly reducing the number of possible off-target effects.
inherent in DSB repair (Fauser et al., 2014; Mikami et al., 2016; Ran et al., 2013). The other advantage of staggered DSBs is that donor DNA with matching sticky ends can be introduced into the cell or organism, yielding targeting events that facilitate the insertion of a donor DNA cassette (Maresca et al., 2013). Alternatively, the double nicks strategy generates a staggered DSB (indicated by diagonal scissors) and these tend to produce larger indels because the single-stranded tails are often involved in the repair. The indels are shown as insertions (green) or deletions (red). If donor DNA is added to the cell and is flanked by the same target sites present in the genomic locus, then compatible ends are produced which can result in a clean cassette insertion (blue). The lower panel shows the major outcomes of the homology-dependent repair (HDR) pathway if a donor DNA template is available carrying the desired modification. Donor DNA carrying a subtle change such as a nucleotide substitution (yellow) can be provided as either a duplex molecule or a single-stranded oligodeoxynucleotide (ssODN), and both will lead to allele replacement (gene conversion). Alternatively, the homology region may be used to flank a new sequence which will lead to cassette insertion. If the double nicks approach is used, the single strand overhangs may promote microhomology-mediated end joining (MMEJ) which can lead to imperfect cassette insertions with indels at the flanks (green).

Figure 2  Genome editing with CRISPR/Cas9 can have multiple outcomes depending on the nature of the double-strand break (DSB), the prevalent repair pathway and the presence of donor DNA. The upper panel shows the major outcomes of the nonhomologous end-joining (NHEJ) pathway. In the absence of donor DNA, Cas9 endonuclease generates a blunt DSB (indicated by vertical scissors) which is repaired yielding small indels. Alternatively, the double nicks strategy generates a staggered DSB (indicated by diagonal scissors) and these tend to produce larger indels because the single-stranded tails are often involved in the repair. The indels are shown as insertions (green) or deletions (red). If donor DNA is added to the cell and is flanked by the same target sites present in the genomic locus, then compatible ends are produced which can result in a clean cassette insertion (blue). The lower panel shows the major outcomes of the homology-dependent repair (HDR) pathway if a donor DNA template is available carrying the desired modification. Donor DNA carrying a subtle change such as a nucleotide substitution (yellow) can be provided as either a duplex molecule or a single-stranded oligodeoxynucleotide (ssODN), and both will lead to allele replacement (gene conversion). Alternatively, the homology region may be used to flank a new sequence which will lead to cassette insertion. If the double nicks approach is used, the single strand overhangs may promote microhomology-mediated end joining (MMEJ) which can lead to imperfect cassette insertions with indels at the flanks (green).

The provision of donor DNA can also be used to select for homology-dependent repair (HDR) events that are common in many microbes but occur once for every 10^5–10^6 NHEJ events in higher eukaryotes. This approach can be used either for the insertion of a donor cassette or the replacement of one allele with another, allowing the knockin of entire genes or the replacement of single nucleotides (Belhaj et al., 2015). HDR is sometimes resolved cleanly, but in other cases, indels may be formed at the borders of the inserted cassette due to microhomology-mediated end joining (MMEJ). This is particularly evident when the Cas9 double nicks is used because it leaves single-stranded tails (McVey and Lee, 2008; Schiml et al., 2016). These events are shown in the lower panel of Figure 2. In diploid organisms, targeted mutations can be homozygous, heterozygous or

Patterns of CRISPR/Cas9 activity 2205

© 2016 The Authors. Plant Biotechnology Journal published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 14, 2203–2216
biallelic, the latter resulting from the creation of two different mutant alleles at the target.

**Relevant differences between host species**

Many articles have been published describing the use of CRISPR/Cas9 in different species without considering the broader implications and species-dependent effects caused by target site preference, DSB structure and the characteristics of the target genome. The size of a genome determines the overall number of potential CRISPR/Cas9 targets because the larger the genome, the greater the number of PAMs. However, the larger the number of potential targets the greater the likelihood that some of them will be repeated, so the number of unique targets does not necessarily increase in proportion to genome size. In general terms, monocotyledonous plants (monocots) tend to have larger genomes than dicotyledonous plants (dicots), and vertebrates tend to have larger genomes than invertebrates, although there are many exceptions at the level of individual species (Gregory et al., 2007; Li and Du, 2014; Michael and Jackson, 2013). Metazoans tend to have larger genomes than unicellular organisms, and eukaryotes tend to have larger genomes than bacteria and archaea (Li and Du, 2014). However, large variations in genome size within clades of eukaryotes of similar biological complexity indicate variations in the amount of repetitive DNA. Because the CRISPR/Cas9 system is typically used for the editing of genes, the exome is a more relevant comparator than the whole genome. Accordingly, mutation frequencies are generally similar in all animals and plants, suggesting that genome size does not have a significant influence on the efficiency of targeted genome editing mediated by the CRISPR/Cas9 system (Xie and Yang, 2013; Zhang et al., 2016).

The GC content of the genome is known to correlate with genome size in bacteria, but in eukaryotes, the relationship is more complex due to the presence of isochores and regions containing highly repetitive DNA (Li and Du, 2014). The GC content varies greatly among different microbes but tends to fall within a relatively narrow range in animals and plants, which is higher in monocots than dicots, and higher in vertebrates than invertebrates. There is also significant variation in GC content among the chromosomes of individual animal and plant species, which is more prevalent in animals than plants, but there is no general correlation between GC content and chromosome size (Li and Du, 2014). The importance of GC content is that it has a significant impact on sgRNA efficiency (Doench et al., 2014; Gagnon et al., 2014; Ma et al., 2015). In plants, animals and microbes, sgRNAs with a GC content greater than 50% are often reported to be more efficient (Feng et al., 2014; Jiang et al., 2013a,b; Pan et al., 2016; Wang et al., 2014a; Zhang et al., 2014). By analysing sgRNAs that have been experimentally validated in plants, the GC content of most was found to lie between 30% and 80% (Liang et al., 2016). Similarly, sgRNAs with an unusually high or low GC content tend to be less effective than those with an average GC content in animals, and sgRNAs targeting the transcribed strand are less effective than those targeting the nontranscribed strand (Wang et al., 2014a). The same authors also found that Cas9 preferentially binds to sgRNAs containing purine residues at the last four positions of the spacer sequence and that the efficiency of cleavage is influenced by the affinity between the sgRNA and Cas9. Similar conclusions were drawn during the development of a bioinformatics tool to design sgRNAs for the effective targeting of mouse and human genes (Doench et al., 2014). The analysis of sgRNA nucleotide composition in animals has revealed other nucleotide preferences (Doench et al., 2014; Wang et al., 2014a; Xu et al., 2015a), but the same preferences have not been observed in plants (Liang et al., 2016). Indeed, no statistically significant preferences in nucleotide composition were observed at any of the 20 positions of the spacer region in plants, suggesting this is a key difference in design principles for sgRNAs used in animals and plants (Liang et al., 2016).

**Selection of target sites in different species**

The role of the PAM

The major PAM recognized by SpCas9 (5'-NGG-3', typically denoted as NGG-PAM) occurs once every ~10 bp in a random DNA sequence and is found every 8–12 bp in the exons of all host species investigated thus far (Anders et al., 2016). Nevertheless, the requirement for this PAM restricts targetable genomic sites to sequences immediately adjacent to this motif. This limitation has been overcome by random mutagenesis followed by screening for mutants with changes in PAM specificity; for example, VRQR-SpCas9 and VRER-SpCas9 recognize NGA-PAM and NGCG-PAM, respectively (Kleinstiver et al., 2015). Endonucleases from other sources have also been investigated, such as KKH1-SaCas9 from *Staphylococcus aureus*, which recognizes N2RRT-PAM (Kleinstiver et al., 2015), *Brevibacillus laterosporus* Cas9 (BlatCas9), which recognizes N3CND-PAM (Karvelis et al., 2015), and the above-mentioned FnCpf1, which recognizes the unusually AT-rich TTNPAM (Fagerlund et al., 2015; Zetsche et al., 2015).

The availability of complete genome sequences and oligonucleotide synthesis techniques allows the rapid design and synthesis of sgRNA libraries that can potentially target any gene in the genome (Wang et al., 2014a). Many bioinformatics tools are available for the design of sgRNAs (Hendel et al., 2015; Mohr et al., 2016), and these can often highlight the presence of potential off-target sites in the genome (Varshney et al., 2015). The number of target sequences identified by *in silico* genome analysis is influenced by the stringency of selection; that is, the number and position of any mismatches allowed between the sgRNA spacer and potential off-target sites. However, indels in the alignment between the target and the sgRNA are not taken into account by most online tools (Lin et al., 2014). Although SpCas9 primarily recognizes NGG-PAM, it also binds to NAG-PAM with a much lower affinity, so prediction software can be used to include potential off-target sites adjacent to either of these motifs. Most tools additionally provide several sgRNA sequences for each gene because the efficiency of targeting depends on many factors, including the uptake/expression of the Cas9 protein and sgRNA, the accessibility of the target and the catalytic efficiency of the enzyme. Recently, Horlbeck et al. demonstrated that nucleosomes impede Cas9 binding and cleavage both *in vitro* and *in vivo* in human cells, and they and developed an algorithm to predict highly active sgRNAs taking into account the information on nucleosome occupancy (Horlbeck et al., 2016). Several targets may need to be tested to get the best empirical balance between efficient on-target mutagenesis and the absence of off-target activity (Bassett et al., 2015).

**Selection of targets in plants**

An extensive comparative analysis of potential SpCas9 target sites in plants was carried out by *in silico* prediction to determine the impact of parameters such as genome size and GC content in the
dicots Arabidopsis (Arabidopsis thaliana), Medicago truncatula, soybean (Glycine max) and tomato (Solanum lycopersicum) and the monocots Brachypodium distachyon, rice (Oryza sativa), sorghum (Sorghum bicolor) and maize (Zea mays; Xie and Yang, 2013; Xie et al., 2014). The genome sizes of the eight species ranged from 120 to 2065 Mb which is representative of most land plants, and the GC content ranged from 34% to 47%. One of the outputs of the study was the online database CRISPR-PLANT (http://www.genome.arizona.edu/crispr/). Potential 20-bp target sequences were extracted from the current genome sequences and sorted into five different categories according to their specificity and potential for off-target activity, based on the number and position of mismatches and the presence of NGG-PAMs and NAG-PAMs. In all eight species, 5–12 NGG-PAMs were identified in every 100 bp of the genome and the total number was positively correlated with the genome size. However, the number of specific targets ranged from 4 to 11 million and correlated positively with genome size in dicots but not in monocots. The number of specific targets did not correlate with the number of transcripts or with the number of NGG-PAMs. These results indicated that although larger genomes contain more PAMs and therefore more potential targets, the new targets were more likely to align with others and would therefore lack specificity. For seven of the eight species, it was possible to design specific sgRNAs to target 83–99% of annotated transcripts (94.3% for Arabidopsis, 83.4% for M. truncatula, 89.5% for tomato, 96.4% for soybean, 98.6% for B. distachyon, 87.3% for rice and 92.6% for sorghum) and 67.9%–96% of these transcripts contained at least 10 different targetable NGG-PAM sites. This indicated that off-target effects are unlikely to present a constraint in most plant species. The exception was maize, where only 29.5% of annotated transcripts matched a specific sgRNA.

The number of annotated transcripts predicted in the reference genome, allowing the rapid selection of sgRNAs with embedded information concerning the predicted off-target sites for each 12-bp seed region (Varshney et al., 2015). Each 20-bp target site adjacent to an NGG-PAM site differs by at least three mismatches from any other target sequence adjacent to an NGG-PAM or NAG-PAM. A genome-wide sgRNA library for Drosophila (Drosophila melanogaster) has been prepared containing 40 279 sgRNAs targeting 13 501 genes (78% of all genes) including 8989 targeted by three or more independent sgRNAs (Bassett et al., 2015) and the Saccharomyces cerevisiae genome contains 645 392 specific targets (unique seed region followed by a NGG-PAM) and 108 493 less specific targets (DiCarlo et al., 2013).

Comparison of targeting parameters

Resolution and targeting efficiency in plants

The resolution of gene editing refers to the nature of the repair pathway (NHEJ, MMEJ and/or HDR) and the broad definition of the resulting mutation (insertion, deletion, replacement, inversion translocation) as shown in Figure 2. NHEJ is the preferred DNA repair pathway in somatic plant cells (Puchta, 2005); therefore, most CRISPR/Cas9 events are resolved by this mechanism, resulting in error-prone repair and the introduction of indels. The reported efficiency of indels induced by CRISPR/Cas9 in both dicots and monocots can vary significantly even within the same species (e.g. 1.1%–90.4% in Arabidopsis), but most species offer examples of efficiency approaching 100%, including Arabidopsis (Yan et al., 2015), soybean (Cai et al., 2015), potato (Wang et al., 2015a), tomato (Pan et al., 2016), petunia (Zhang et al., 2016), tobacco (Gao et al., 2015), poplar (Zhou et al., 2015), grapefruit (Jia et al., 2016), maize (Svitashev et al., 2015) and rice (Ma et al., 2015). Notably, the CRISPR/Cas9 system can efficiently induce mutations in different tissues and cell types, including embryogenic callus (e.g. rice and maize), hairy roots (e.g. soybean), protoplasts (e.g. Nicotiana benthamiana, lettuce, maize and rice), cotyledons (e.g. tomato) and leaves (e.g. N. benthamiana, petunia and poplar). Studies reporting the efficiency of different types of targeting events in plants are summarized in Table S1. This intraspecific variability partly reflects how our knowledge of the CRISPR/Cas9 system has increased since the system was first used for genome editing, resulting in more efficient experimental designs. Although direct comparisons between experiments carried out under different conditions are not possible, two major common principles have emerged.

The first principle is that the efficiency of genome editing is strongly influenced by the expression of the components. In Arabidopsis, initial experiments based on in planta transformation and Cas9 expression controlled by the constitutive Cauliflower mosaic virus (CaMV) 35S promoter resulted in low editing efficiencies and mostly somatic mutations that were not transmitted to the progeny. But the frequency of heritable mutations increased to 90.4% when the constitutive promoter was replaced with an egg-cell-specific promoter (Wang et al., 2015b), a cell-division-specific promoter (Hyun et al., 2015; Yan et al., 2015) or a germ-line-specific promoter (Mao et al., 2016). In all other plant...
systems, constitutive promoters have achieved high mutation frequencies, with biallelic and homozygous mutants readily produced in the first generation of monocots such as maize (Svitashev et al., 2015) and rice (Lowder et al., 2015; Ma et al., 2015; Miao et al., 2013; Wang et al., 2016) and dicots such as tomato (Brooks et al., 2014; Pan et al., 2016), poplar (Fan et al., 2015; Zhou et al., 2015), potato (Wang et al., 2015a), petunia (Zhang et al., 2016) and tobacco (Gao et al., 2015). These observations also suggest that mutagenesis often occurs at an early stage during the transformation process, before the first cell division. High levels of sgRNA limit the efficiency of genome editing at least in tomato (Pan et al., 2016) and Arabidopsis (Ma et al., 2015). In comparison with protoplasts, the efficiency of targeted chromosomal fragment deletion between paired sgRNA/Cas9 sites is lower in transgenic plants. This may reflect the relatively low levels of sgRNAs and Cas9 in callus tissue and regenerated plants (Xie et al., 2015).

The second principle is that the nature of the sgRNA is also an important determinant of targeting efficiency. Although the rules are not completely understood, there is little doubt that some sgRNAs are more mutagenic than others and this is a key factor in determining the outcome of each editing experiment. As stated above, sgRNAs with a GC content greater than 50% often have a high editing efficiency, possibly because of a stronger binding to their target site (Feng et al., 2014; Jiang et al., 2013a; Pan et al., 2016; Wang et al., 2014a; Zhang et al., 2014). In addition to the GC content of the spacer, the secondary structures of sgRNAs also affect the efficiency of editing (Makarova et al., 2011). The formation of a stem-loop structure in the protospacer region can inhibit the binding of the sgRNA to the target strand, reducing the likelihood of a DSB (Ma et al., 2015). Targeting one gene with multiple sgRNAs has been shown to greatly increase the mutation frequency and the recovery of homozygous mutants in rice (Wang et al., 2016; Xie et al., 2015; Zhang et al., 2014) and T0 tomato plants (Brooks et al., 2014). An extension of the culture period increased the proportion of mutated cells in Agrobacterium tumefaciens-infected rice callus (Mikami et al., 2015a,b) and in soya bean somatic embryo cultures (Jacobs et al., 2015), probably reflecting the proliferation of existing mutant cells as well as new mutations. However, this method can also reduce the regeneration capacity of the cells and increase the risk of obtaining chimeric plants (Xu et al., 2015b).

The efficiency of genome editing by HDR is generally lower than NHEJ because homologous recombination occurs $10^{-8}$–$10^{-10}$ times less frequently than repair by end ligation in plants (Figure 2). Only a handful of reports describe successful genome editing by HDR in higher plants, but they represent an interesting variety of different approaches. Gene conversion was achieved with an efficiency of 5% using short single-stranded oligodeoxyribonucleotides (ssODNs) as the repair template in Arabidopsis (Sauer et al., 2016) and with an efficiency of 9% using dsDNA as the repair template in N. benthamiana (Li et al., 2013). When directly compared in maize, a short ssODN template (127 nt) was twice as efficient as a plasmid donor (800 bp), achieving mutation frequencies of 0.4% and 0.2%, respectively (Svitashev et al., 2015). Expression cassettes flanked by 1-kb homology arms were inserted with a frequency of 4.6% in soya bean (Li et al., 2015) and 4% in maize (Svitashev et al., 2015). Gene replacement using a combinatorial dual-sgRNA/Cas9 vector to remove 255 bp of endogenous sequences and insert a ~1.9-kb cassette with homology arms of 733 and 825 bp was achieved with a frequency of 0.8% in Arabidopsis (Zhao et al., 2016).

one of the first examples of CRISPR/Cas9 genome editing, Li et al. (2013) failed to induce HDR in Arabidopsis protoplasts and attributed this to the intrinsically low efficiency of HDR in these cells compared to N. benthamiana, but more recent data suggest that the low efficiency probably reflected the relatively small number of DSBs, given that a 1.1%–5.6% mutation frequency was achieved by NHEJ with the same sgRNAs. An outstanding efficiency of 100% HDR-mediated conversion of the ALS gene was achieved in rice by Sun et al. (2016) using two sgRNAs for cleavage, flanking the homology arms on the donor with CRISPR sites to release the repair template in vivo and increasing the amount of donor DNA by introducing both the vector donor and free donor fragments (476 bp). This is by far the highest HDR frequency observed in higher eukaryotes but has yet to be tested in other species, and it will be necessary to determine whether undesirable random integration events also occur in this system. Particle bombardment appears to be up to fivefold more effective than Agrobacterium-mediated transformation for the promotion of HDR induced by CRISPR/Cas9 in maize (Svitashev et al., 2015) and rice (Sun et al., 2016). Lower levels of Cas9 but higher levels of sgRNA and repair template can increase the likelihood of resolution by HDR in yeast (Stoviczek et al., 2015), and this has been achieved by particle bombardment in rice (Sun et al., 2016) and viral replicons in tobacco (Baltes et al., 2014). Interference with the NHEJ pathway can also promote HDR as demonstrated in rice using a lig4 mutant background to abolish end-joining ligase activity, although the impairment of NHEJ may also increase the frequency of spontaneous mutations (Endo et al., 2016).

Resolution and targeting efficiency in animals

The CRISPR/Cas9 system has been used successfully in many animals, including invertebrates and vertebrates (Table S2). Unlike plants, the CRISPR/Cas9 system can be used for genome editing in particular tissues by hydrodynamic injection or by introducing the components using Adeno-associated virus (AAV) or adenovirus vectors (Rodriguez et al., 2014; Senis et al., 2014; Swiech et al., 2015; Xue et al., 2014). Although there are differences in DNA repair pathways across species, the general preponderance of NHEJ over HDR observed in plants is also observed in animals (Figure 3). Indels generated by NHEJ have therefore been introduced with an efficiency of up to ~90% in the nematode Caenorhabditis elegans (Friedland et al., 2013), Drosophila (Bassett et al., 2015), rabbit (Lv et al., 2016), chicken (Oishi et al., 2016), mouse (Yang et al., 2014) and human cells (Liang et al., 2015), whereas the insertion of donor DNA by HDR has been reported with a frequency of ~5%–20% in C. elegans (Dickinson et al., 2013), rat (Shao et al., 2014) and mouse (Platt et al., 2014). Site-specific indels were induced in zebrafish embryos by the in vivo microinjection of a sgRNA/Cas9 complex incorporating an additional tracrRNA sequence, causing mutagenesis at two sites that were impossible to edit with TALENs (Hwang et al., 2013). In mouse, the use of ssODNs instead of a plasmid donor increased the efficiency of HDR from 10%–30% to 10%–80% (Yang et al., 2014).

In human cell lines, reported mutation frequencies are often significantly lower than 5%, especially when induced pluripotent stem cells are used as the host (Miyaoka et al., 2016; Yang et al., 2013; Zhu et al., 2015). However, much higher mutation rates of ~60–90% have also been reported in these cells, which may reflect differences in the activity of specific sgRNAs (Liang et al., 2015; Veres et al., 2014). Gene knockin by HDR in human cells is
usually successful in fewer than 10% of cells, although this can be increased to 50%–66% if NHEJ is suppressed (Chu et al., 2015). The transfection of human stem cells expressing a doxycycline-inducible Cas9 gene (iCas9) with sgRNAs enabled the drug-free selection of precise HDR-mediated modifications with ssODN donors (Zhu et al., 2015). The efficiency of HDR also increased sevenfold when short hairpin RNAs were introduced expressing the Adenovirus Ad4 protein responsible for suppressing Ligase 4, which is required for NHEJ in mice and humans (Chu et al., 2015). Biallelic double-gene mutants were generated by transfecting cells with three separate plasmids and injecting ssODNs matching two of the target genes to achieve HDR-mediated mutation (Wang et al., 2013). Mice were engineered to constitutively express Cas9, and when injected with a combination of sgRNAs and HDR donors carried by AAV vectors, NHEJ and HDR events occurred at frequencies that increased over time (Platt et al., 2014).

Resolution and targeting efficiency in microbes

The CRISPR/Cas9 system is a double-edged sword in bacteria because DSBs are inefficiently repaired by NHEJ, thus targeting the bacterial genome with CRISPR/Cas9 tends to be lethal (Selle and Barrangou, 2015). However, most bacterial strains are either recombinogenic or have established functional recombineering systems, so the provision of donor DNA allows the CRISPR/Cas9 system to be used as a powerful tool for the selection of HDR events and the simultaneous counterselection of background cells, resulting in highly efficient HDR without the need for positive and negative selectable markers. Using this strategy, the wild-type SpCas9 system has been applied in both Gram-positive and Gram-negative bacteria with an HDR efficiency of 65%–100% (Table S3). In Clostridium cellulolyticum, the Cas9 nickase variant together with a repair template achieved HDR efficiencies greater than 95% (Xu et al., 2015c).

Eukaryotic microbes show diverse behaviours in response to CRISPR/Cas9, and the system must be tailored for use in different species. In contrast to higher plants, the constitutive expression of Cas9 is toxic in Chlamydomonas reinhardtii, preventing the recovery of transformants even in the absence of sgRNA. The basis of this phenomenon is unclear, but the transient expression of Cas9 and sgRNA has proven sufficient to generate indels by NHEJ (Jiang et al., 2014) as has the delivery of Cas9 RNPs (Shin et al., 2016). In contrast, Cas9 is not toxic in the marine diatom Phaeodactylum tricornutum and mutation frequencies of up to 63% have been achieved by NHEJ following stable transformation with a cas9 transgene (Nymark et al., 2016). Similarly, different yeast species show diverse behaviours in response to genome editing. S. cerevisiae is unusually amenable to HDR, so gene editing at the single nucleotide level can be achieved using short donor templates (Bao et al., 2015; Biot-Pelletier and Martin, 2016; DiCarlo et al., 2013) and the simultaneous deletion and insertion of several genes has been reported at frequencies of up to 100% (Stovicek et al., 2015; Tsai et al., 2015a). Although targeted nicks and DSBs increase the HDR frequency by up to 4000-fold in S. cerevisiae, only a 10-fold increase was observed in Pichia pastoris, where the NHEJ pathway is more prevalent (Weninger et al., 2016). The malaria parasite Plasmodium falciparum is an interesting target because it appears to be naturally deficient in the canonical NHEJ pathway, and only HDR occurs when a donor template is provided (Ghorbal et al., 2014). Genome editing using the CRISPR/Cas9 systems therefore relies on DSB-induced HDR with an external donor template, and this has been achieved with an efficiency of 50%–100% when using ssODNs.
Mutation signatures

The mutations induced by CRISPR/Cas9 in plants are mainly short deletions of 10 bp or less and single-base insertions, typically A/T in all species (Figure 3). Single-base substitutions are rare, with the exception of soya bean protoplasts where they were the most frequent mutation (Sun et al., 2015). Less frequent longer deletions may represent the results of MMEJ, indicating that gene-specific factors can influence the outcome of DSB repair (Xu et al., 2015b). In rice, mutation signatures vary according to the target (Miao et al., 2013; Xu et al., 2015b; Zhou et al., 2014). The consistent mutations observed at the same target in several independent soya bean hairy-root cultures and somatic embryos also suggest that there may be as yet undiscovered rules governing the types of mutations that are favoured at a given target (Jacobs et al., 2015). In the same study, the seven most effective sgRNAs exclusively generated short deletions, whereas those with lower efficiency were associated with more insertions and substitutions (Jacobs et al., 2015). Interestingly, all the off-target mutations found in rice by Li et al. (2016) were 1-bp insertions, indicating that the pairing of sgRNA with the target sequence may also influence the mutation type.

The picture emerging from studies in animal systems is similar to that in plants: most on-target and off-target mutations reported in animals are short deletions of up to 40 bp (An et al., 2016; Friedland et al., 2013; Kim et al., 2015). Although less common, deletions as large as 250 bp have been reported in mice (Heckl et al., 2014) and human cells (Liang et al., 2015), but insertions are usually shorter, typically 1–15 bp (Cheng et al., 2014; Friedland et al., 2013; Kim et al., 2015; Liang et al., 2015). CRISPR/Cas9 mutation signatures are not widely reported in microbes due to the lethality of the NHEJ pathway in bacteria and some microbial eukaryotes, and the relatively small number of studies.

The Cas9 double nickase generates mutations with different signatures compared to the intact enzyme because of the nature of the staggered DSB. For example, a Cas9 double nickase was used to generate a DSB with 52-nt overhangs in Arabidopsis. The average size of the resulting insertions was considerably larger (>80 nt) than for the fully functional Cas9 nuclease, and in most cases the insertions were copies of the sequence immediately upstream or downstream from the insertion site, which is indicative of MMEJ (Schmil et al., 2014, 2016). The double nickase approach also generated longer deletions in mice (39–56 bp) as well as insertions of up to 67 bp (Cheng et al., 2014). Interestingly, the two alternative Cas9 variants from S. thermophilus and S. aureus generate different mutation signatures associated with different PAMs in Arabidopsis. For StCas9, sgRNAs with a NNGAA-PAM generated 8.6% insertions (mostly 1-bp) and 3.8% deletions, whereas those with a NNGG-PAM generated 11.6% insertions (mostly 1-bp) and 3.8% deletions. For SaCas9, sgRNAs with a NNGAA-PAM generated 52.1% insertions (mostly 1-bp), whereas those with a NNGGGT-PAM generated 46.7% deletions and 21.6% insertions, both of which were generally larger than 1 bp (Steinert et al., 2015).

Off-target mutations and methods to increase efficacy and accuracy

Off-target mutations in plants

Off-target activity is generally rare in higher plants (Table S1). Where reported, it tends to involve a minority of sgRNAs even when mutations are investigated by the thorough method of whole-genome sequencing (Feng et al., 2014). A low frequency of unwanted mutations has been reported when the sgRNA features mismatches outside the seed sequence, for example in rice (Xu et al., 2015b; Zhang et al., 2014) and wheat (Upadhyay et al., 2013), indicating that such events could be avoided by designing more specific sgRNAs. Careful sgRNA design can ensure specific targeting even when the genome contains closely related paralogous genes (Baysal et al., 2016). However, in a small number of cases, unexpected cleavage has been observed at sites with one or more mismatches within the seed region, for example in Arabidopsis (Sauer et al., 2016), barley (Lawrenson et al., 2015), soya bean (Jacobs et al., 2015) and rice (Xie and Yang, 2013). Target sequences with a GC content higher than 70% may increase the likelihood of off-target effects (Tsai et al., 2015b), which might explain the unexpected mutations observed by Li et al. (2016) (GC = 65%–80%) but not those reported by Jacobs et al. (2015) (GC = 57%) or Sauer et al. (2016) (GC = 50%). Even so, the frequency of off-target mutations is much lower than that of on-target mutations, allowing the recovery of solely on-target mutations in all experiments. Interestingly, Xu et al. (2015b) detected off-target mutations only in T1 rice plants carrying the cas9 and sgRNA transgenes, but not in those where the CRISPR components had segregated, suggesting that off-target effects might be reduced or avoided by selecting appropriate T1 progeny. The frequency of unwanted mutations depends on the abundance of the Cas9/sgRNA RNP complex so the likelihood can be reduced by transient expression of the components rather than stable transgene integration, although this could reduce on-target efficiency too (Tsai et al., 2015b). The Cas9 double nickase resulted in efficient genome engineering in Arabidopsis, without off-target effects in homologous genomic regions (Fauser et al., 2014). On the whole, off-target mutations in plants are generally less frequent than the somatic mutations that arise during tissue culture (Li et al., 2016).

Off-target mutations in animals and microbes

Off-target mutations in human cells were initially reported to be up to 50% more common than mutations at the on-target site, raising concerns about the intrinsic fidelity of the CRISPR/Cas9 system (Fu et al., 2013; Hsu et al., 2013; Kim et al., 2015; Mali et al., 2013; Pattanayak et al., 2013). However, those experiments were conducted on cancer cell lines, which are often characterized by dysfunctional DNA repair mechanisms, or used sgRNAs that are known to be promiscuous (Kim et al., 2015). When stem cells were used as the host, whole-genome sequencing revealed the absence of off-target mutations (Smith et al., 2014) or only a few off-target events (Veres et al., 2014). Similarly, minimal off-target activity has been reported in zebrafish (Hruscha et al., 2013), mice (Heckl et al., 2014), chicken (Oishi et al., 2016) and rabbit (Lv et al., 2016). These studies are summarized in Table S2. However, for both animals and plants, most studies have sought off-target activity at preselected sites rather than by unbiased whole-genome sequencing, which means that off-target activity cannot be ruled out at unpredicted sites. Off-target mutations have not been directly observed in microbes but may be inferred due to their indirect toxicity (Table S3).
Methods to increase targeting efficiency and accuracy

Several approaches have been described to increase the efficiency and accuracy of CRISPR/Cas9 in plants, animals and microbes regardless of the host species and tissue, based on sgRNA design, nuclease choice and the delivery strategy. The ideal sgRNA should maximize on-target activity while minimizing off-target activity. In addition to following the general principles of sgRNA design discussed above, fidelity can be improved using a truncated spacer (~20 nt) and by adding two guanidine residues to the 5’ end of the sequence (Cho et al., 2014; Fu et al., 2013; Kim et al., 2015). The choice of nuclease can also increase targeting accuracy. Cas9 nickases must act as dimers and therefore double the length of the target site, reducing off-target activity by 50- to 1500-fold (Fauser et al., 2014; Ran et al., 2013; Schiml et al., 2014). This is also true for the hybrid endonuclease dCas9-FokI which is generated by fusing the enzymatically inactive dCas9 to the non-specific endonuclease domain of the restriction enzyme FokI (Gulingier et al., 2014; Tsai et al., 2014). The propensity of Cas9 to tolerate mismatches between the sgRNA and target site is attributed to a high binding energy, so the rational engineering of amino acid residues involved in DNA binding has produced two high-fidelity mismatch-sensitive variants of Cas9 that achieved promising results in human cells (Kleinstever et al., 2016; Slaymaker et al., 2016). The delivery of the sgRNA, nuclease and (where appropriate) the HDR donor molecule also affects the outcome of the experiment. Typically, the components are delivered as plasmid DNA and must be expressed in the cell, but they can also be introduced as RNA or as preformed RNP complexes that can be delivered efficiently by electroporation or transfection to mammalian cells (Burger et al., 2016; Chu et al., 2016; Liang et al., 2015), plant protoplasts (Subburaj et al., 2016; Woo et al., 2015) or microalgae (Shin et al., 2016). Because the RNP complexes are cleared rapidly, they are less likely to cleave at off-target sites (Liang et al., 2015). For HDR, the use of short ssODN donor molecules seems to result in higher insertion frequencies than donor delivery by plasmid, but the size is limited to ~200 nt (Sauer et al., 2016; Zhao et al., 2016). Like Cas9, endonucleases from the Argonate family have recently been found also to use oligonucleotide guides to target invasive genomes (Gao et al., 2016). The DNA-guided nuclease NgAgo binds a 5’-phosphorylated, single-stranded, ~24-nt guide DNA (gDNA) that creates site-specific DSBs without needing a PAM. Preliminary characterization suggests a low tolerance of gDNA target mismatches and highly efficient editing of GC-rich genomic targets. However, the efficiency of the system and reproducibility of the results obtained with NgAgo are being questioned (Cyranoski, 2016), so it remains to be seen whether it can be effectively used for genome editing.

More ambitious genome editing

Multiplex targeting

One of the features of the CRISPR/Cas9 system that sets it apart from ZFNs and TALENs is that multiple sgRNAs can be introduced simultaneously into a cell with little additional effort, allowing more ambitious genome-editing strategies such as the simultaneous mutation of different genes and the creation of more extensive mutations. Multiple genes can be targeted with one sgRNA by deliberately designing a promiscuous sequence, and this has been used to mutate multiple targets in related rice genes due to the tolerance of mismatches (Endo et al., 2015). Multiple sgRNAs can be introduced either as separate constructs or in the form of a polycistronic cassette, which allows any number of targets to be edited simultaneously regardless of their relationship (Xie et al., 2015). The latter strategy has been used extensively in rice, for example to generate null alleles at the SIAGO7 locus (Brooks et al., 2014), to mutate multiple genes at the YSA locus (Lowder et al., 2015), to simultaneously mutate multiple genes in the MPK family (Xie et al., 2015), to investigate the frequency of mutations in the ER922 gene (Wang et al., 2016) and to mutate the GSTU and MRP15 genes in rice and Arabidopsis (Ma et al., 2015). Gao et al. (2015) simultaneously mutated the tobacco PDS and PDR6 genes which generate an albino phenotype. Multiple genes or multiple targets within genes have also been mutated in animals. For example, Dickinson et al. (2013) simultaneously mutated four targets in the C. elegans lin-31 gene to modify the MAP kinase phosphorylation sites at the C-terminus. Platt et al. (2014) simultaneously mutated the mouse TP53, LKB1 and KRAS genes, which are the three most prevalent oncogenes in the lung, and a similar approach was used to study functional redundancy within the TET gene family (Wang et al., 2013).

Large-scale mutations

Multiple sgRNAs not only allow simultaneous targeting at different sites but they can also be combined to induce large deletions and other rearrangements. For example, chromosomal segments of up to 245 kb have been deleted in rice plants by the introduction of two sgRNAs that create DSBs on the same chromosome (Zhou et al., 2014), and similar strategies have been used to achieve deletions ranging from 65 kb to 30 Mb in mammalian cells (Esseltzibehler et al., 2014; Zhang et al., 2015). Whereas most tandem DSBs are resolved by deleting the intervening DNA, another possible outcome is the creation of a chromosomal inversion. Li et al. (2015) reported that precise inversions of DNA fragments ranging in size from ~50 bp to hundreds of kb could be generated efficiently in mice and human cells, as well as deletions and duplications resulting from transallelic recombination between DSBs on sister chromatids. Similarly, large-scale rearrangements have been generated in human cell lines to model the chromosomal hallmarks of cancer. For example, Choi and Meyerson (2014) used pairs of sgRNAs to introduce paracentric and pericentric chromosomal inversions as well as the CD74-ROS1 chromosomal translocation event often seen in lung cancer, and Maddalo et al. (2014) used a similar strategy to induce the Eml4-Alk inversion which is a hallmark of non-small-cell lung carcinoma. Chromosomal translocations resembling those associated with acute myeloid leukaemia and Ewing’s sarcoma were induced at a high frequency using pairs of sgRNAs targeting different chromosomes by Torres et al. (2014). In contrast, translocations have not yet been reported in plants, and chromosomal inversions are rare: Liang et al. (2016) found that only one of nine expression constructs producing sgRNA pairs was able to generate an inversion and such events occurred in only two of the 23 transgenic rice plants expressing this sgRNA pair. Interestingly, Li et al. (2015) reported that most chromosomal inversions in their human and mouse cell lines were accompanied by small terminal indels, suggesting that repair was promoted by MMEJ. The difference in the frequency of inversions and other chromosomal rearrangements between
mammals and plants may therefore reflect the relative activity of this repair pathway.

**Editing targets currently in commercial development**

In addition to proof-of-concept and optimization studies in model systems, there is now great interest in the commercial applications of CRISPR/Cas9, particularly in the pharmaceutical industry where it can be used to develop accurate disease models and platforms for drug screening (Jang and Ye, 2016; Tschaharganeh et al., 2016), and in the agricultural industry where it can be used to produce new crop and farm animal varieties with enhanced traits (Sovová et al., 2016). The need to characterize the outcome of genome editing in detail as a way to facilitate commercial development is therefore clear, not only to streamline the development of innovative products but also in the light of the different regulatory pathways for genome-edited crops in the USA (based on product, considered an output of plant breeding) and the EU (based on process, considered GMO technology at the moment but still under debate; Sprink et al., 2016). In the field of plant biotechnology, research is now focusing on genome editing in a broad range of crop species including citrus fruits, maize, poplar, potato, rice, sorghum, soya bean, tomato and wheat (Khatodia et al., 2016; Song et al., 2016). The purpose of such studies is the improvement of agronomic traits, but most publications describe the ability to edit the relevant target genes rather than the traits themselves (Sovová et al., 2016). For example, CRISPR/Cas9 has been used to knock out the rice sweet genes that confer sensitivity to bacterial blight (Jiang et al., 2013b), the wheat MLO genes that confer sensitivity to powdery mildew (Shan et al., 2013; Wang et al., 2014b), and the rice mpks gene that regulates defence responses (Xie and Yang, 2013). Although all these studies were successful, some were conducted on protoplasts, and even when plants were regenerated they were not tested directly for pathogen resistance. In contrast, cucumber plants in which the elf4E gene encoding eukaryotic translation initiation factor 4E was mutated using CRISPR/Cas9 have been tested for resistance against Cucumber vein yellowing virus, Zucchini yellow mosaic virus and Papaya ring spot mosaic virus (Chandrasekaran et al., 2016). Similarly, Wang et al. (2016) knocked out the rice OsERF922 gene encoding the transcription factor ERF and showed that the homozygous T2 plants were more resistant to rice blast but were otherwise identical to wild-type plants in terms of growth and yield traits. Sun et al. (2016) used CRISPR/Cas9 to induce HDR in rice, resulting in a single nucleotide substitution in the ALS gene that conferred herbicide resistance. Li et al. (2016) individually mutated four rice genes affecting yield traits (Gn1a, DEP1, GS3 and IPA1) and achieved a variety of promising mutant phenotypes in the T2 generation, including more grains (gn1a), denser panicles and semi-dwarf culms (dep1), larger grains and long awns (gs3), and a change in tiller number, either more or less depending on the precise target site (ipa1). Finally, CRISPR/Cas9 has also been used to target genes encoding polyphenol oxidases (PPOs) in mushroom. The enzyme causes browning of the fungal tissue, and knocking out one of six genes in the PPO gene family reduced overall PPO activity by 30% thus extending the shelf life (Waltz, 2016).

In academic research, the choice of genome-editing technique depends mainly on the simplicity and cost of the approach and the availability of tools and expertise, but applied research and commercial crop development must also take into account the associated intellectual property (IP) and licensing issues. Each of today’s genome-editing tools is protected by patents or patent applications (Schinkel and Schilberg, 2016), and navigation of the IP landscape is straightforward in the case of oligonucleotide-directed mutagenesis, ZFNs and TALENs (Table S4). In contrast, the IP situation for the CRISPR/Cas9 technology is strongly contested by at least three major players: Massachusetts Institute of Technology/Broad Institute, UC Berkeley and Vilnius University (Schinkel and Schilberg, 2016). The ongoing legal dispute has delayed the commercial development of crops produced using CRISPR/Cas9 technology, although DuPont Pioneer has recently received an exclusive licence (Grushkin, 2016). DuPont Pioneer has exploited CRISPR/Cas9 technology for the development of drought-resistant maize and waxy maize with an improved starch composition. In the latter case, CRISPR/Cas9 was used to knock out the Wx1 gene resulting in maize kernels that only accumulate amylopeptin. The company recently announced that they will bring the genome-edited maize to the market within the next 5 years.

**Summary and outlook**

The CRISPR/Cas9 system has been used for genome editing in a wide range of different organisms but the outcome in terms of resolution, efficiency, accuracy and mutation structure depends on various factors including target site choice, sgRNA design, the properties of the endonuclease, the type of DSB introduced, whether or not the DSB is unique, the quantity of endonuclease and sgRNA, and the intrinsic differences in DNA repair pathways in different species, tissues and cells. Species-dependent effects include the preponderance of NHEJ compared to HDR in higher eukaryotes, contrasting with the preference for HDR in bacteria and some unicellular eukaryotes, and subtle differences in the mutation signatures generated in animals and plants (Figure 3). Whereas canonical CRISPR/Cas9 predominantly introduces small deletions (<10 bp) and single-base insertions in plants, both types of indel tend to be larger in animals (deletions <40 bp and insertions of 1–15 bp) and there is a greater frequency of larger deletions. In both animals and plants, Cas9 double nickase introduces staggered DSBs and this results in even larger indels (typically <100 bp). Another difference is the relative efficiency of larger genome rearrangements in animals compared to plants. These differences are likely to reflect species-dependent aspects of the competing NHEJ, MMEJ and HDR repair pathways, suggesting that the outcome of genome editing could be influenced by modulating the activity of particular repair enzymes, as shown by the increased prevalence of HDR in cells lacking normal levels of Ligase 4 in both animals and plants. Further investigations and detailed comparisons of genome-editing outcomes in different species will provide insight into interaction between component-specific effects (nuclease activity, sgRNA design) and host-specific effects (genome structure and content, DNA repair pathways) to enable the refinement of genome-editing strategies in a context-dependent manner.

**Acknowledgements**

Work on gene targeting at the RWTH Aachen University is funded by the European Research Council Advanced Grant ‘Future-Pharma’, Grant Number 269110. Fraunhofer IME has received funding from Dow AgroSciences for research on zinc finger nucleases. Synthetic biology and genome-editing work at the UdL is supported by grants from the Spanish Ministry of Economy and
Competitiveness (BIO2014-54426-P and BIO2014-54441-P) and the Catalan Government 2014 SGR 1296 Agricultural Biotechnology Research Group.

Conflict of Interest
Fraunhofer IME has received funding from Dow AgroSciences for research on zinc finger nucleases.

References
An, L., Hu, Y., Chang, S., Zhu, X., Ling, P., Zhang, F., Liu, J. et al. (2016) Efficient generation of FvIII gene knockout mice using CRISPR/Cas9 nuclease and truncated guide RNAs. Sci. Rep. 6, 25199.
Anders, C., Bargsten, K. and Jinek, M. (2016) Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. Mol. Cell. 61, 895–902.
Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P.A. and Voytas, D.F. (2014) DNA replications for plant genome engineering. Plant Cell. 26, 151–163.
Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, Y., Sun, N., Si, S. et al. (2015) Homology-integrated CRISPR–Cas (Hi-CRISPR) system for one-step multiplex disruption in Saccharomyces cerevisiae. ACS Synth. Biol. 4, 585–594.
Basset, A.R., Kong, L. and Liu, J. (2015) A genome-wide CRISPR library for high-throughput genetic screening in Drosophila cells. J. Genet. Genom. 42, 301–309.
Baysal, C., Bortesi, L., Zhu, C., Farr, G., Schillberg, S. and Christou, P. (2016) CRISPR/Cas9 activity in the rice OsBellel gene does not induce off-target effects in the closely related paralog OsBella. Mol. Breed. 36, 108.
Belhaj, K., Chaparro-Garcia, A., Kamoun, S. and Nekrasov, V. (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas9 system. Plant Methods. 9, 39.
Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J. and Nekrasov, V. (2015) Editing plant genomes with CRISPR/Cas9. Curr. Opin. Biotechnol. 32, 76–84.
Briot-Pelletier, D. and Martin, V.J.J. (2016) Seamless site-directed mutagenesis of the Saccharomyces cerevisiae genome using CRISPR-Cas9. J. Biol. Eng. 10, 6.
Bortesi, L. and Fischer, R. (2015) The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol. Adv. 33, 41–52.
Brooks, C., Nekrasov, V., Lippman, Z.B. and Van Eck, J. (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CAS9-associated 9 System. Plant Physiol. 166, 1292–1297.
Burger, A., Lindsay, H., Felker, A., Hess, C., Felker, A., Hess, C., Anders, C., Chiavacci, E., Zaugg, J. et al. (2014) Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. Genome Res. 24, 2059–2065.
Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiacovo, M.P., Church, G.M. and Calarco, J.A. (2013) Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat. Methods. 10, 741–743.
Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reynon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas9 nucleases in human cells. Nat. Biotechnol. 31, 822–826.
Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G. et al. (2014) Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single guide RNAs. PLoS ONE. e98816.
Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y. et al. (2015) CRISPR/Cas9-mediated targeted mutagenesis in Nicotiana tabacum. Plant Mol. Biol. 87, 99–110.
Gao, F., Shen, X.Z., Jiang, F., Wu, Y. and Han, C. (2016) DNA-guided genome editing using the Natteronabacterium gregoryi Argonaute. Nat. Biotechnol. 34, 768–773.
Gasunias, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) Cas9-cRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl Acad. Sci. USA. 109, E2579–E2586.
Ghorbal, M., Gorman, M., Macpherson, C.R., Martins, R.M., Scherf, A. and Lopez-Rubio, J.-I. (2014) Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. Nat. Biotechnol. 32, 819–821.
Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G. et al. (2007) Eukaryotic genome size databases. Nucleic Acids Res. 35, D332–D338.
Grzushkin, D. (2016) DuPont in CRISPR-Cas patent land grab. Nat. Biotechnol. 34, 13.
Gulling, J.P., Thompson, D.B. and Liu, D.R. (2014) Fusion of catalytically inactive Cas9 to FokI nucleases improves the specificity of genome modification. Nat. Biotechnol. 32, 577–582.
Haeusler, M. and Concorde, J.P. (2016) Genome editing with CRISPR-Cas9: can it get any better? J. Genet. Genom. 43, 239–250.
Hedl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R.V., McDonkey, M.E., Thielke, A. et al. (2014) Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. Nat. Biotechnol. 32, 941–946.
Hendel, A., Fine, E.J., Bao, G. and Porteus, M. (2015) Quantifying on- and off-target genome editing. Trends Biotechnol. 33, 132–140.

Horbeck, M.A., Witkowski, L.B., Guglielm, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E. et al. (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. Elife, 5, e12677.

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C. and Schmid, B. (2013) Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. Development, 140, 4982–4987.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y. et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832.

Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell, 157, 1262–1278.

Hu, X., Wang, C., Fu, Y., Liu, Q., Jiao, X. and Wang, K. (2016) Expanding the range of CRISPR/Cas9 genome editing in rice. Mol. Plant, 9, 943–945.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T. et al. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31, 227–229.

Hyun, Y., Kim, J., Cho, S.W., Choi, Y., Kim, J.S. and Coupland, G. (2015) Site-directed mutagenesis in Arabidopsis thaliana using dividing tissue-targeted RNAi of the CRISPR/Cas9 system to generate heritable null alleles. Plants, 4, 271–284.

Jacobs, T.B., Lafayette, P.R., Schmitz, R.J. and Parrott, W.A. (2015) Targeted mutagenesis in rice using a CRISPR-Cas9 system for plant genome engineering. Front. Plant Sci. 6, 538.

Karlin-Neumann, G.A. et al. (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat. Biotechnol. 31, 688–691.

Li, J., Shou, J., Guo, Y., Yang, T., Wu, Y., Jia, Z., Zhai, Y. et al. (2015) Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. J. Mol. Cell Biol. 7, 284–298.

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., Lin, Q. et al. (2016) Reassessment of the four yield-related genes Gln1a, DEP1, GS3, and IPAI in rice using a CRISPR/Cas9 system. Front. Plant Sci. 7, 377.

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Shudhavan, M., Carte, J. et al. (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J. Biotechnol. 208, 44–53.

Liang, G., Zhang, H., Lou, D. and Yu, D. (2016) Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Sci. Rep. 6, 21451.

Lin, Y., Crudick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wilde, B.M. et al. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. 42, 7473–7485.

Lowder, L.G., Zhang, D., Baltes, N.J., Paul, J.W.I. III, Tang, X., Zheng, X., Voytas, D.F. et al. (2015) A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. Plant Physiol. 169, 971–985.

Lu, Q., Yuan, L., Deng, J., Chen, M., Wang, Y., Zeng, J., Li, Z. et al. (2016) Efficient generation of myosin gene mutated rabbit by CRISPR/Cas9. Sci. Rep. 6, 25029.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B. et al. (2015) A robust CRISPR/Cas9 system for convenient high-efficiency multiplex genome editing in monocot and dicot plants. Mol. Plant, 8, 1274–1284.

Maddalo, D., Manchado, E., Conception, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P. et al. (2014) In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature, 516, 423–427.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S. et al. (2011) Evolution and classification of the CRISPR/Cas systems. Nat. Rev. Microbiol. 9, 467–477.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R. et al. (2015) An updated evolutionary classification of the CRISPR/Cas systems. Nat. Rev. Microbiol. 13, 722–736.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. et al. (2013) Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838.

Mao, Y., Zhang, Q., Feng, Z., Wei, P., Zhang, H., Botella, J.R. and Zhou, J.K. (2016) Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. Plant Biotechnol. J. 14, 519–532.

Maresca, M., Lin, V.G., Guo, N. and Yang, Y. (2013) Obligate ligation-gated recombination (OBLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. Genome Res. 23, 539–546.

McVey, M. and Lee, S.E. (2008) MMEJ repair of double-stranded breaks (director’s cut): deleted sequences and alternative endings. Trends Genet. 24, 529–538.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J. et al. (2013) Targeted mutagenesis in rice using CRISPR-Cas system. Cell Res. 23, 1233–1236.

Michael, T.P. and Jackson, S. (2013) The first 50 plant genomes. Plant Genome 6, 2.

Mikami, M., Toki, S. and Endo, M. (2015a) Comparison of CRISPR/Cas9 systems with no Cas9 paired nickases in rice. PLoS ONE 10, e0132154.

Li, J.F., Norville, J.E., Aach, J., Mc Cormack, M., Zhang, D., Bush, J., Church, G.M. et al. (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat. Biotechnol. 31, 688–691.

Li, J., Shou, J., Guo, Y., Yang, T., Wu, Y., Jia, Z., Zhai, Y. et al. (2015) Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. J. Mol. Cell Biol. 7, 284–298.

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., Lin, Q. et al. (2016) Reassessment of the four yield-related genes Gln1a, DEP1, GS3, and IPAI in rice using a CRISPR/Cas9 system. Front. Plant Sci. 7, 377.

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Shudhavan, M., Carte, J. et al. (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J. Biotechnol. 208, 44–53.

Liang, G., Zhang, H., Lou, D. and Yu, D. (2016) Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Sci. Rep. 6, 21451.
NHEJ reveals effects of locus, nuclease, and cell type on genome editing. Sci. Rep. 6, 23549.

Mohr, S.E., Hu, Y., Ewen-Campen, B., Housden, B.E., Viswanatha, R. and Perrimon, N. (2016) CRISPR guide RNA design for research applications. FEBS J. 283, 3232–3238.

Nymark, M., Sharma, A.K., Sparstad, T., Bones, A.M. and Winge, P. (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae. Sci. Rep. 6, 24951.

Oishi, I., Yoshii, K., Miyahara, D., Kagami, H. and Tagami, T. (2016) Targeted mutagenesis in chicken using CRISPR-Cas9 system. Sci. Rep. 6, 23980.

Osakabe, Y. and Osakabe, K. (2015) Genome editing with engineered nucleases in plants. Plant Cell Physiol. 56, 389–400.

Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., Chen, L. et al. (2016) CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. Sci. Rep. 6, 24765.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2016) Rationally engineered Cas9 nucleases with improved specificity. Science 351, 843–845.

Patterson, J.E., Perrimon, N. (2016) CRISPR guide RNA design for research applications. Nat. Biotechnol. 34, 3194–3203.

Patterson, J.E., Perrimon, N. (2016) CRISPR-Cas9 system adapted for gene editing in marine algae. Plant Cell Physiol. 57, 901–910.

Patterson, J.E., Perrimon, N. (2016) CRISPR-Cas9 system adapted for gene editing in marine algae. Cell 165, 1543–1556.
Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.-L. (2014b) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat. Biotechnol. 32, 947–951.

Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F. and Cui, X. (2015a) Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. Plant Cell Rep. 34, 1473–1476.

Wang, Z.P., Xing, H.L., Dong, L., Zhang, H.Y., Han, C.Y., Wang, X.C. and Chen, Q.J. (2015b) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homologous mutants for multiple target genes in Arabidopsis in a single generation. Genome Biol. 16, 144.

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., Liu, Y.G. et al. (2016) Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene OsERF922. Plant Sci. 110, e0154027.

Weninger, A., Hatzl, A.M., Schmid, C., Vogl, T. and Glieder, A. (2016) Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris. J. Biotechnol. 235, 139–149.

Woo, J.W., Kim, J., Kwon, S.I., Corvalán, C., Cho, S.W., Kim, H., Kim, S.-G. et al. (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat. Biotechnol. 33, 1162–1164.

Xie, K. and Yang, Y. (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. Mol. Plant, 6, 1975–1983.

Xie, K., Zhang, J. and Yang, Y. (2014) Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. Mol. Plant, 7, 923–926.

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

Xu, H., Xiao, T., Chen, C.H., Li, W., Meyer, C.A., Wu, Q., Wu, D. et al. (2015a) Sequence determinants of improved CRISPR sgRNA design. Genome Res. 25, 1147–1157.

Xu, R.F., Li, H., Qin, R.Y., Li, J., Qiu, C.H., Yang, Y.C., Ma, H. et al. (2015b) Generation of inheritable and “transgene clean” targeted genome-modified rice in later generations using the CRISPR/Cas9 system. Sci. Rep. 5, 11491.

Xu, T., Li, Y., Shi, Z., Hemme, C.L., Li, Y., Zhu, Y., Van Nostrand, J.D. et al. (2015c) Efficient genome editing in Clostridium cellulolyticum via CRISPR-Cas9 nickase. Appl. Environ. Microbiol. 81, 4423–4431.

Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Ca, W. et al. (2014) CRISPR-meditated directed mutagenesis of cancer genes in the mouse liver. Nature, 514, 380–384.

Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J. et al. (2013) Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 41, 9049–9061.

Yang, H., Wang, H. and Jaenisch, R. (2014) Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nat. Prot. 9, 1956–1968.

Zetsche, B., Gootenberg, J.S., Abudayeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E. et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell, 163, 759–771.

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y. et al. (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol. J. 12, 797–807.

Zhang, L., Jia, R., Palange, N.J., Satheka, A.C., Togo, J., An, Y., Humphrey, M. et al. (2015) Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. PLoS ONE, 10, e0120396.

Zhang, B., Yang, X., Yang, C., Li, M. and Guo, Y. (2016) Exploiting the CRISPR/Cas9 system for targeted genome mutagenesis in petunia. Sci. Rep. 6, 20315.

Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., Li, W.X. et al. (2016) An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. Sci. Rep. 6, 23890.

Zhou, H., Liu, B., Weeks, D.P., Spalding, M.H. and Yang, B. (2014) Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Res. 42, 10903–10914.

Zhou, X., Jacobs, T.B., Xue, L.J., Harding, S.A. and Tsai, C.I. (2015) Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and redundancy. New Phytol. 208, 298–301.

Zhu, Z., Verma, N., González, F., Shi, Z.D. and Huangfu, D. (2015) A CRISPR/Cas9-mediated selection-free knockin strategy in human embryonic stem cells. Stem Cell Rep. 4, 1103–1111.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1 The efficiency, accuracy and structure of on/off-target mutations induced by CRISPR systems in different plant species.

Table S2 The efficiency, accuracy and structure of on/off-target mutations induced by CRISPR systems in different animal species.

Table S3 The efficiency, accuracy and structure of on/off-target mutations induced by CRISPR systems in different microbial species.

Table S4 Examples for the development of commercial plant products using different genome editing technologies and the involved IP.