An overview of designing and selection of sgRNAs for precise genome editing by the CRISPR-Cas9 system in plants

Ajay Prakash Uniyal¹ · Komal Mansotra¹ · Sudesh Kumar Yadav² · Vinay Kumar¹

Received: 19 September 2018 / Accepted: 13 May 2019 / Published online: 21 May 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

A large number of computational tools have been documented in recent years for identification of target-specific valid single-guide (sg) RNAs (18–20 nucleotide long sequence) that is an important component for the efficient utilization of the CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats—CRISPR-associated Protein) system. Despite optimization of Cas9, other major concerns are on-target efficiency and off-target activity that depend upon the sequence(s) of target-specific sgRNA(s). However, a very little attention has been paid for identification of the best-hit sgRNA for precise targeting as well as minimizing the off-target effects. The aim of this present work is to offer comparative insight into existing CRISPR software tools with their unique features (including targeted genome) and utilities. These available web tools were found to be designed based upon only a few limited mathematical models. Among all these available web tools, three (Benchling, Desktop and CRISPR-P) have been curated as exclusively available for plant genome-editing purpose. These three software tools have been comprehensively described and analyzed with single same target enquiry from two randomly selected genes (*IDM2* and *IDM3* from *Arabidopsis thaliana*). Interestingly, all these selected tools generated different results (sgRNAs) even for the same query. In fact, the sequence of sgRNA is considered an important parameter to determine the efficiency and specificity of sgRNAs for precise genome editing. Thus, there is an urgent requirement to pay attention for a validated sgRNA-designing tool for precise DNA editing in plants. In conclusion, this work will encourage building up a consensus for developing a universal valid sgRNA designing for different organisms including plants.

Keywords Genome editing · sgRNA · Web tools · CRISPR-Cas9 · Plants

Abbreviations

Cas	CRISPR-associated protein/gene
CRISPR	Clustered regularly interspaced short palindromic repeats
sgRNA	Single-guide RNA
crRNA	CRISPR RNA
PAM	Protospacer adjacent motif
CFD	Cutting frequency determination
ZFN	Zinc finger nuclease
CLD	CRISPR library designer
TALEN	Transcription activator-like effector nuclease
tracrRNA	<i>Trans</i> -activating CRISPR RNA

DESKGEN	Desktop genetics
NGS	Next-generation sequencing

Introduction

Genome editing is a group of techniques that changes the DNA of organisms including plants, bacteria, and animals. Either inserting of novel DNA or replacing with other potential DNA can achieve the editing of DNA of any organism of interest. This method of genome editing requires the comprehensive knowledge of DNA of both recipient and donor organisms. Another promising method of genome editing is deleting DNA fragment of interest from genome of experimental organisms (Zhang et al. 2014). A number of genome-editing systems including the zinc finger nucleases (ZFNs), the TALENs (transcription activator-like effector nucleases) and the clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated 9 (Cas9) systems are being used for precisely or efficiently editing of

✉ Vinay Kumar
vinayihbt@gmail.com; vinay@cup.edu.in

¹ Department of Plant Sciences, School for Basic and Applied Sciences, Central University of Punjab, Bathinda 161001, India

² Center of Innovative and Applied Biotechnology, Mohali 140306, India

DNA of any targeted organism. Among these, the CRISPR had quickly exiled its predecessors because of its unprecedented efficiency, simplicity and ability to target DNA as well as RNA (Ran et al. 2013). The CRISPR naturally occur in bacterial system and provides protection against infection mediated by plasmids and viral attacks by degrading the foreign DNA. The six microbial CRISPR-Cas systems (I–VI) have been reported till today (Wright et al. 2016). However, all the reported CRISPR-Cas9 systems have common set of core features with two different functional modules including adaptive and effector modules. The adaptive module incorporates a small degraded genomic sequence(s) into CRISPR arrays. These incorporated sequences named as repeated array are then transcribed into crRNA [CRISPR RNA; 18–20 nucleotide (nt) spacer region with repeated region of 12 nt]. The effector module guides direct cleavage of invaded DNA using different variants of Cas nuclease as reported in different organisms (Kumar and Jain 2014). These six types are also divided into Class 1 and 2, depending upon the participation of a single or multiple Cas proteins in effector module. The Class 1 possesses multisubunit effector complexes, while Class 2 possesses single-protein effector as documented by Shmakov et al. (2015). Only two Class 2 effectors, named Cas9 and Cpf1, are known as so far and Cas9 has been explored for precise DNA targeting.

Among different microbial CRISPR-Cas9 types (I–VI), type II is extensively investigated system and has advantage over others due to minimal basic requirement for DNA targeting. The minimum assembly is a three component system that includes crRNA, tracrRNA (trans-activating CRISPR RNA; 64 nt) and Cas9 nuclease. A portion of tracrRNA molecule base pairs with the direct repeats (12 base pairs) of pre-crRNA and helps in maturation of crRNA with duplex formation by the activities of conserved endogenous RNaseIII and Cas9 protein. The matured crRNA (20-nt-long sequence) was found to be responsible for searching the complementary target sequences for site of introduction of cleavage within the targeted DNA of organism of interest (Ran et al. 2013). The third most important component is the Cas9 protein, a multifunctional nuclease protein which consists of two functional distinct lobes REC (recognition lobe; RuvC) and NUC (nuclease lobe; HNH) (Chen et al. 2014). A special part of REC (recognition lobe), α -helical region recognizes specific 10–12 nucleotide 'seed region' of sgRNA and plays critical role during guiding of Cas9 for the introduction of cleavage at specific site. The HNH nuclease domain induces the cleavage at precise site on DNA strand that is complementary to crRNA while other RuvC domain induces the cleavage in the opposite complementary DNA strand. In conclusion, both nuclease domains cleave at two different strands of target DNA. By catalytically inactivation of the domain(s) of Cas9, either HNH (H840) or RuvC (D10A), the fully functional Cas9 protein can be converted

into half functional, termed as nickase and can be used for inducing the cleavage at specific defined strand of target DNA (Gopalappa et al. 2018).

The guide RNA in the CRISPR-Cas9 system is composed of tracrRNA and crRNA. These two RNAs (a tracrRNA and crRNA) are streamlined into a single-RNA chimera known as sgRNA (single-guide RNA), with dual-tracrRNA and crRNA secondary structure and used favorably in genome editing. In the target gene knockout by the CRISPR-Cas9 system in rice, the sgRNA construct generates more efficiently than the dual-crRNA: tracrRNA construct (Miao et al. 2013). In fact, the specific single-nucleotide sequence of chimeric sgRNA is a duplex of crRNA and tracrRNA (tetra loop) in which the 3' end of crRNA is fused with 5' end of tracrRNA, forms a functional complex with the optimized variant Cas9 and guides the Cas9 to match a 18–20 nt complementary sequence on target DNA site. The cleavage has been reported to the upstream of a required 5'-NGG, PAM (Protospacer Adjacent Motif) sequence which is a recognition site of sgRNA binding. The PAM is 2–6-nt-long sequence reported just next to the target sequence which influences the cleavage efficiency (Mojica et al. 2009). Interestingly, it has also been observed that the Cas9 could not bind to those target sequence, which are not followed by PAM sequence. In addition, the PAM sequence also helps to distinguish the self from non-self-DNA sequence. The PAM sequences have also been manipulated to alter the specificity for targeting to reduce off-target effects. A brief development history of research on sgRNA designing is shown in Fig. 1.

Any mismatching in sgRNA sequence can impair target binding and is responsible for numerous off-target mismatches (Jinek et al. 2014). Thus, in addition to occurrence of a unique PAM sequence, the precise designing of sgRNA also influences the cleavage efficiency of the optimized Cas9 nuclease. The sgRNA is composed of four different components named as target sequence (crRNA sequence), crRNA repeat sequence, tetra loop sequence and tracrRNA sequence and has six secondary structural modules (spacer, lower stem, bulge, upper stem, nexus, and hairpins) as confirmed by mutational analysis of 77 sgRNA variants (Briner et al. 2014). Among them, bulge and nexus are most sensitive to disruption as well as important for DNA cleavage. The upper stem exhibits the DNA cleavage activity even after large deletion mutation. Extension to stem loop increases the sgRNA stability and improves its assembly with SpCas9. The sequence composition of all parts of sgRNA and dual RNA has been shown in Fig. 2. Eventually, potential role of sgRNA designing in cleavage efficiency encourages the researchers for in vitro designing of sgRNA and this step seems to be an essential step in optimization of the CRISPR-Cas9 system. The length (number of nucleotides) and composition (nature of present nucleotides) of sgRNA majorly contribute to the precise editing with least off-target

Fig. 1 The diagrammatic representation of research progress of sgRNA designing. **a** Charpentier and Doudna reported that the crRNA and the tracrRNA could be fused together to create sgRNA (2012). **b** In the same year, Jinek et al. documented that the 20 nucleotides sequence of sgRNA determine the nucleotide sequence and by changing its sequence, Cas9 can be guided to any location. **c** The interaction between sgRNA and Cas9 is found to be due to REC domain of Cas9 (Jinek et al. 2014). **d** The sgRNAs tolerate certain mismatches and can produce off-target mutations as documented by two independent reports (Cong et al. 2013; Zhang et al. 2013). **e** Fu et al. (2014) reported that truncating sgRNA by 2 or 3 nucleotide can increase the specificity of Cas9. **f** Ran et al. (2013) also documented that combined D10A mutated Cas9 with truncated sgRNA was found to be responsible for reduction in off-targets. **g** Doench et al. (2014, 2016) built predictive sgRNA model and developed tool for sgRNA designing

Fig. 2 A representation of sgRNA structure. **a** Sequence composition of sgRNA (crRNA and tracrRNA components) in CRISPR-Cas9 type II system. The crRNA consists of spacer sequence (blue; 18–20 nt in length) and crRNA repeat sequence (brown 22 nt). Complementary spacer sequence pairs with the target sequence and complementary repeat sequence pairs with tracrRNA (purple). **b** Chimeric sgRNA sequence composition. The sgRNA is formed by the fusion of crRNA and tracrRNA with engineered tetra loop structure (green). The sgRNA sequence consists of four different sequences, spacer (blue), crRNA repeat (brown), tetra loop (green) and tracrRNA (purple). It also has six structural secondary structure, spacer, lower stem, bulge, upper stem, nexus and hairpins. **c** The repeat sequence of crRNA dictates recognition lobe of spCas9 localization when there is complementary base pairing between target DNA and guide RNA. Nexus region lies in center of sgRNA: Cas9 interaction and helps in DNA cleavage. The hairpin secondary structures support in formation of this stable complex

effects (Jiang and Doudna 2017). The truncating of sgRNA sequence at the 5' end up to 17 or 18 nucleotide complementary to target sequence has also been documented with least off-target effects (Fu et al. 2014). In conclusion, the designing of target-specific sgRNA molecule is the only major essential requirement for precise DNA editing other than optimized Cas9 variant using the CRISPR-Cas system.

To minimize off-target effects mediated by Cas9 protein, several improvements including fusion of inactivated the Cas9 with FokI nuclease have been carried out in optimization of the Cas9 protein (Cong et al. 2013). A number of Cas9 variants (SpCas9, SaCas9, NmCas9, StCas9, StCas9) have been used with human codon optimization (Nowak et al. 2016). These have been explored for catalytically inactive and nickase version. The structure-guided protein engineering approach has been used to generate "enhanced specificity" SpCas9 (eSpCas9) version with reduce off-target effects and precise on target cleavage (Slaymaker et al. 2016). Cpf1 nuclease instead of the Cas9 protein is also reported in type II. The Cpf1, a new Cas9 classmate from Class 2, has small size and has only a single-nuclease domain, which does not need tracrRNA. Unlike other Cas9 variants, Cpf1 cuts the target region downstream of a thymidine-rich PAM site (TTTN), produces 5'-protruding ends and enables multiple round of DNA cleavage (Zetsche et al. 2017). AsCpf1 isolated from *Acidaminococcus* sp. recognizes a 5' T-rich (TTTN) PAM and documented for robust activity in human cells. All these advantages over Cas9 protein make it a preferable variant for precise genome editing (Fagerlund et al. 2015; Sontheimer and Wolfe 2015). One step ahead, the fusion of the dCas9 (dead) with effector molecules (*VPC4* and *KRAB*) also made this system as the either transcriptional activation or repression efficient (Plummer et al. 2018). The efficiency of the Cas9 nuclease activity has also been found to be altered with presence of guanine next to 3' end of PAM (Doench et al. 2014). In conclusion, the optimized Cas9 variant is guided by specific nucleotide sequence named as sgRNA for precise targeting.

The CRISPR-Cas9 system not only provides a molecular means for inquiring into biological problems in-depth but also enables the development of creative and realistic diverse applications in organisms. In 2017, this technology has been successfully used for gene editing in human cell lines, embryo and animals including mosquitoes, monkey, and rabbit for many beneficial applications. Recent reported lists of potential gene-edited animals have been listed in Table 1. Due to diverse applications, this technology has been also exploited in plants (Miao et al. 2013; Zhang et al. 2013). However, there is the requirement of improvement in its specificity and applicability for genome editing in plants. A few model plant species including *Arabidopsis thaliana*, *Oryza sativa*, *Nicotiana tabacum* and Sorghum have been tested for the feasibility of the CRISPR-Cas9 technology

in plant systems (Feng et al. 2013; Miao et al. 2013). The CRISPR-Cas9 system is also so popular technique due to its advantage over GM crops that it does not leave a trace of foreign DNA element in the genome of testing plants without the trouble of costly deregulation processes (Feng et al. 2013; Yin et al. 2017). The improvement of the CRISPR-Cas9 system is being in progress to make it more flexible and powerful platform for genome engineering in diverse living systems including plants (Feng et al. 2013; Miao et al. 2013; Yin et al. 2017).

The major concern of the CRISPR-Cas9 system is the occurrence of relative high amount of off-target effects in tested organisms. The off-target activity mainly depends on both sgRNA sequence as well as experimental conditions. The unique sequence of nucleotides (8–12) at 3' end of sgRNA is essential requirement for binding and cleavage of target sites (Sander and Joung 2014). However, more than one target site in genome can have possibility of occurrence of similar complementary nucleotides resulted in undesirable off-target effects. Thus, apart from the length (number of nucleotides) and composition (type of nucleotides) of sgRNAs, the nature (uniqueness) of sgRNA also affects the on-target efficiency as well as off-target effects. Therefore, consideration of uniqueness of sgRNAs as another essential characteristic multiple sgRNAs is encouraged for precise genome editing.

Considerable efforts have been made to apply the CRISPR-Cas9 genome-editing tool to various applications including preclinical studies, biotechnology, and therapeutic usages. However, the numerous approaches for effective designing of Cas9 and their choice for perfection are still underway. Numerous computational tools and methods have been emerged in the past few years to help the designing and assembly mechanism of these CRISPR nucleases and make a calculated decision on their off-target activity and efficiency in model organisms especially animals (Doudna and Charpentier 2014; Fine et al. 2014; Periwai 2016; Mendoza and Trinh 2017). Basically, all major explored computational tools for sgRNA designing have been based upon the sequence of model organisms especially animals and only a few has been reported which are based upon plant genome. Actually, the precise tool for the evaluation of the efficiency of sgRNA editing is still lacking in plants. The comparative screening of sgRNA libraries in animal system is less time consuming than plant system because it takes long time for development of transgenic products.

This report provides a comprehensive overview of the CRISPR databases and their mode of action execution. The second part of this manuscript is focused on plant-based sgRNA-designing databases (Benchling, Deskgen and CRISPR-P). In this part, the comparison on the basis of features and commands has been carried out. In addition, two randomly selected genes (*IDM2* and *IDM3*) were exploited

Table 1 Summary of the CRISPR-Cas9-mediated effects on target organisms (human, animals and plants)

No	Human/animals	Targets	Results	References	
1.	Human	Cell lines	Human embryonic stem cell line WA01(H1)	Knockout of <i>Smo</i> gene leads to shift in open-reading frame and generation of SMO-mutant cell line (Wae001-A-16)	Wu et al. (2017)
2.			<i>Androgen receptor (AR)</i> in androgen-positive prostate cancer cells	Disruption of the <i>AR</i> gene for therapeutic uses against prostate cancer	Wei et al. (2018)
3.		Embryo	<i>OCT4</i> gene	Genetically ablate the <i>OCT4</i> gene in human preimplantation embryos and documented key differences from its function	Fogarty et al. (2017)
		Zygote	Human dual pronuclear, <i>2PN</i> zygotes	Efficient homologous recombination-mediated correction of point mutation in <i>HBB</i> and <i>G6PD</i> , also reveal limitations of this correction procedure	Tang et al. (2017)
4.	Pig		Foot and mouth disease (FMD) virus and porcine reproductive and respiratory syndrome (PRRSV)	Multi-resistance against viral diseases	Oh et al. (2017)
5.			Inactivate all endogenous retroviruses (ERVs)	Improve the feasibility of clinical trials of xenotransplantation	Ross and Coates (2017)
6.	Mosquitoes		Site-specific mutagenesis of the white gene	Generation of biallelic stable mutants for reverse genetic analysis	Li et al. (2017a, b)
7.	Sheep		<i>PDX1</i>	Creation of pancreatogenesis disabled sheep for interspecies organ generation	Vilarino et al. (2017)
8.			Fibroblast growth factor 5, <i>FGF5</i>	Loss of EGFs function leads to increase in wool length and yield	Li et al. (2017a, b)
9.	Rabbit		<i>FOXN1</i> , <i>RAG2</i> , <i>IL2RG</i> or <i>PRKDC</i>	Productive of immunodeficient rabbit/double gene mutant founders	Song et al. (2017)
10.	Monkey		<i>Arntl</i> and <i>Prot2</i>	To establish monkey model systems for genome-editing purposes	Zou et al. (2017)
No	Plants	Targets	Results	References	
1	Rice	<i>PYL</i> genes (<i>PYL 1–6</i> , <i>PYL 7–11</i> , <i>PYL 12</i> and <i>PYL 13</i>)	Increase in growth and grain improvement	Miao et al. (2018)	
		<i>C287</i>	Herbicide-resistant rice	Shimatani et al. (2017)	
		<i>OsAnn3</i>	Resistant rice	Shen et al. (2017)	
2	Wheat	<i>DsRed</i> , <i>TaLox2</i> and <i>TaUbiL1</i>	Acceleration in potency of gene discovery and crop breeding	Bhowmik et al. (2018)	
		Conserved region adjacent to coding region of <i>alpha-gliadin</i> genes	Low-gluten, non-transgenic wheat	Sánchez-León et al. (2018)	
		<i>TaGW2</i> , <i>TaLpx-1</i> and <i>TaMLO</i>	Increase in seed size and grain weight	Wang et al. (2018)	
		<i>TaDREB2</i> and <i>TaERF3</i>	Potentiality for crop improvement	Kim et al. (2018)	
3	Maize	<i>MS8</i>	Transgene-free maize male sterile lines	Chen et al. (2018)	
		<i>Zmzb7</i>	Albino plants	Feng et al. (2016)	
		<i>ZmTMS5</i>	Thermosensitive male sterile maize	Li et al. (2017a, b)	
		<i>ARGOS8</i>	Improvement in maize grain yield	Shi et al. (2017)	
4	Soybean	<i>GmFT2a</i> and <i>GmFT25</i>	Gene functional analysis	Cai et al. (2018)	
5	Arabidopsis	<i>Coding region of CMVR1</i> , <i>CMVR2</i> , <i>CMVR3</i> , and <i>CMV 3'UT</i> , and <i>TMV coding region</i>	Reduce viral accumulation	Zhang et al. (2018a, b)	
		<i>elF(iso)4E</i>	Development of <i>Turnip mosaic virus</i> (TMV)-resistant plants	Pyott et al. (2016)	
6	Cotton	<i>Gh14-3-3d</i>	Resistance to fungal pathogen (<i>Verticillium dahliae</i>)	Zhang et al. (2018a, b)	
		<i>Conserved intergenic region</i>	Cotton leaf curl viruses	Uniyal et al. (2019)	
7	Tomato	<i>SIORRM4</i>	Delayed fruit ripening	Yang et al. (2017)	
		<i>SIIAA9</i>	Parthenocarpic tomato	Ueta et al. (2017)	

for designing of the relevant total sgRNAs using these three selected web tools and results were analyzed on the basis of respective output against same query. This work encourages the CRISPR research community for development of universal web tools for effective sgRNA designing in plants.

Concept of sgRNA designing

The CRISPR-Cas9 system can be defined to cleave any target site preceding a specific PAM sequence. Within the genome, the number of specific PAM-targeting loci can be variable so it does not always succeed with precise editing. Thus, a number of off-target sites that share sequence similarity with targeted sites in the genome are a major concern for precise editing. In fact, the occurrence of non-canonical PAM sequence that lies within the tolerance limit of the nucleotide mismatch in-between sgRNA (only crRNA sequence component of sgRNA) and target site is primary reason for off-target effects. Based upon the extent of nucleotide mismatching of sgRNA against a target sequence, sgRNA populations can be divided into two groups, i.e., efficient and non-efficient sgRNA populations based upon their specificity and efficiency. Thus, designing of sgRNA (only crRNA sequence component) against any target effect is an important step and relevant information is shown as flow chart in Fig. 3. The unique genomic sequence of experimental organism(s) is found to be the major reason for all possible DNA target sites against a single sgRNA and efficiency of specific sgRNAs with specific target site (Doench et al. 2014; Gang et al. 2016).

The efficiency of specific sgRNA against target site majorly depends on the unique nucleotide sequence as well as secondary structure of sgRNA. In animals, the nucleotide preference of sgRNAs has also been documented, however, similar observations were not found in plants. The GC content of sgRNA is also an important criterion affecting the efficiency of sgRNA. Liang et al. (2016) reported that 97% of total investigated sgRNA have 30–80% GC content. This range of GC content was found to contribute the efficiency of sgRNA with target site.

To understand the possible link between secondary structure and editing efficiency of sgRNA, assessment of all sgRNA secondary structures for validated sgRNA have been carried out. It was documented that all validated sgRNAs have unique secondary (three crucial stem loops named as intact RAR (repeat and anti-repeat) 2 and stem loop 3) structure (Fig. 2). In fact, sgRNA sequence has been documented to be composed of crRNA- and tracrRNA-sequences joined by an artificial tetra loop (Liu et al. 2016; Nishimasu et al. 2014). The repeat and anti-repeat region (RAR) of sgRNA initiates processing of crRNA with the help of RNaseIII and later induces the cleavage at target site with the help of Cas9 (Nishimasu et al. 2014).

Fig. 3 A systematic representation of different steps required for designing and validation of sgRNAs. **a** Fix the target (gene/genomic region) of interest of identified organism. **b** Get a list of possible sgRNA(s) using parameters “on-target” and “off-target” activities. **c** Utilize all potential sgRNA(s) for further cloning using individually or combinatorial approach. **d** Direct delivery of individual constructs (Cas9 and sgRNA vectors) or combined single construct into plant cells. **e** Validation of sgRNA(s) using sequencing and RFLP

The base pairing of guide sequence (complementary to target site) with sgRNA in secondary structure is not found more than 12 total base pairing (TBP) and even not more than seven conservative base pairs in most of documented cases (Liang et al. 2016). Apart from this interaction, internal base pairs (IBP) within the guide sequence of sgRNA have also been documented to interfere with the recognition of target.

In conclusion, The GC content (30–80%), secondary structure (RAR and stem loop 2 and 3) and base pairing scores (12, TBP; 7, CBP; 6, IBP) are major limits that need to be considered for choice of efficient sgRNAs for precise targeting.

Established algorithms for designing of sgRNAs

In 2013, Hsu et al. characterized the targeting specificity of SpCas9 in human cells for the identification of target sites as well to avoid off-target effects. More than 700 sgRNA variants having indel mutation level with more than 100 score at predicted genomic off-target loci have been evaluated. The mismatching between target site and SpCas9 was found to be flexible with distribution, position and number of mismatches. Hsu et al. (2013) also documented that modulating the amount of the Cas9 and sgRNA minimized the off-targets effects.

In addition, Hsu et al. (2013) also proposed web tools for selection of target sequences, validation of selected sgRNAs as well as off-target analyses of targeted DNA. Four different constraints have been considered for identification of the ‘off-target’ genomic sequences to maximize specificity of Cas9 for precise editing. The first is that off-target should not be followed by PAM (5′ NGG or 5′ NAG) sequences. Second, the sequence similarity of the target with all potential off-targets should be minimized. The mismatches (more than three) in-between guide sequence and off-target loci should also be avoided. As third point, two mismatches within the PAM-proximal region of the off-target site should be permissible. Fourth, the presence of maximal number of mismatches should be either consecutive or occurred less than four bases apart. On the basis of above-proposed guidelines, a frame-scoring algorithm was framed for computing the off-target loci and selection of valid sgRNA in the form of a computational tool (accessible from www.genome-engineering.org). In this web tool, the enquiry-based operation generates output as specificity scores (‘0’ for highly non-specific to ‘100’ for highly specific). The specificity score above ‘50’ for a query reflects reasonable good designing of sgRNA(s) and assumed that it has no complementary site elsewhere in the global genomic sequence (Hsu et al. 2013).

In 2014, Doench et al. selected endogenous human and mouse gene(s) for generation of sgRNAs against their target sites. The ability to produce null alleles for complementary genes was then assessed. A number of sequence features of sgRNAs have been considered for improving the activity of the SpCas9 for precise editing. Furthermore, a predictive model of sgRNA activity for improving sgRNA designing was also constructed based upon results collecting from selected 1841 sgRNAs (on basis of ability of targeting the CDS). These selected 1841 sgRNAs were also examined on the basis of parameters including position-dependent nucleotide preference and flanking target sequence. The most significant difference among nucleotide preferences in selected sgRNAs was found at position 20, very adjacent to the PAM, and guanine is

very strongly preferred nucleotide at this position. The additional differences were found at position 3 (cytosine disfavored), 16 (preference for cytosine) and exact middle of sgRNAs (preference for adenine). In addition, a preference in PAM (variable nucleotide) was also documented. In total, the full set of features includes nucleotide features (120), dinucleotide features (464) and the G/C count (2) as reported by Doench et al. (2014). The predictive model for sgRNA activity was built up using a logistic regression classifier for differentiating most efficient sgRNAs (highest activity quintile of sgRNAs) using a set of sequence features. This model used support vector machine (SVM) with logistic regression for differentiating higher scored sgRNAs. In all these studied genes, sgRNAs were ranked according to parameter “percentage rank”, in which, worst sgRNA is ranked as ‘0’ value and best sgRNA will have percentage score approaching ‘1’ value. All these developed rules for sgRNA designing were proposed collectively as “Rule Set 1”.

Doench et al. (2016) published another report in “Nature Biotechnology” journal for optimization of sgRNA design to increase on-target activity and reduce off-target effects. In this report, along with “Rule Set 1”, specificity lying within protein-coding regions and the target site location within the gene were also used as criterion for selection of sgRNAs from selected endogenous genes. The authors used important features collectively known as feature set for predicting activity other than modeling approach. The single and dinucleotide position-specific nucleotides, and GC count of sgRNAs are considered same as previously in published report (Doench et al. 2014). However, few additional features such as position-independent nucleotide counts and location of the sgRNA target site within the gene, and localized thermodynamics properties were also considered for improved prediction along with regression model. The authors proposed ‘Rule Set 2’ with off-target avoidance metric for improving on-target activity prediction and used for designing of fully optimized libraries named as Brunello and Brie (human and mouse genomes). After incorporation of “Rule set 2”, the prediction performance extends to CRISPRa and CRISPRi screens. Doench et al. (2016) also tested library using well-established library-screening system and ranked sgRNAs by their log₂ fold change on their abundance in the plasmid DNA pool and on the basis of average ranks from the two replicates.

All these studies reveal some of important rules related to sequence composition for sgRNA activity. However, it has been limited due to a small numbers of genomic loci and a single Cas9 system. Therefore, an in vivo library-on-library methodology has been developed to simultaneously assess single-guide RNA activity across 1400 genomic loci (Chari et al. 2015). Later, another model based on elastic net algorithm has been documented for predicting the efficiency of

sgRNA (Kuan et al. 2017). Harnessing the knowledge from oligonucleotide design and nucleosome occupancy models, it has been postulated that position-dependent dinucleotide features are an important feature for improving the designing of sgRNA. In addition, the frequency of T and TT was also documented to determine the efficiency of sgRNA activity. In 2015, CCTop, a predictive tool was also documented for identification of effective sgRNA according to their off-target quality (Stemmer et al. 2017).

Online available sgRNA-designing web tools

The numbers of computational tools are being employed for designing of precise sgRNA for effective genome editing. The selection criteria for the curation of most efficient sgRNA are majorly determined by predicted on-target and off-target activity score curated by these web tools. Comprehensive analysis of all possible available sgRNA tools are reviewed here for in-depth understanding of sgRNA-designing strategies (Table 2).

ZiFit

The ZiFit is an acronym for Zinc Finger Targeter Identifies. Initially, it was developed by Zinc Finger Consortium in 2010 for selection of potential ZFP- and ZFN-binding sites. The recent updated version of ZiFit (version 4.2) also allows features for designing of sgRNA plasmids and identification of potential CRISPR target sites in input sequence (Sander et al. 2010). Query sequences are input in form of FASTA and raw text. Uppercase letter in input sequence denotes exon and lowercase letter denotes introns. The nucleotides where users want to induce breaks are indicated with brackets. ZiFit also allows features for identifying the potential off-target sites, their position and number of time this potential off-target site can occur. This web tool supports target genome of *Homo sapiens*, *Rattus*, *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *C. elegans*, *Culicidae* and *E. coli*. Link: <http://zifit.partners.org/ZiFit/ChoiceMenu.aspx>.

CRISPR design

The CRISPR design is a web tool generated by Zhang Lab (2013) for simplifying the selection of sgRNAs in the input sequence by identifying possible off-target sites and annotating the best sgRNA (Hsu et al. 2013). This web tool is user friendly as users have to just enter DNA sequence and choose a target genome. It also allows input of 23–100 bp of interested target sequence for designing of sgRNA. Output results present list of all sgRNAs against target sequence with on-target activity and off-target scores. Scores are

computed on basis of cleavage specificity and total number of mismatches (also their respective position, distribution and the distance between mismatches to clarify for steric inference of mismatches in interaction of sgRNA with the sequence). The sgRNA(s) with the aggregate score above 50% (green colored) are considered suitable for use. The rest of sgRNA (less than 50%; red and yellow colored) are generally avoided. It supports target genome of diverse types of organisms including *Homo sapiens*, *Arabidopsis thaliana* and *C. elegans*. Link: <http://crispr.mit.edu:8079/>.

CRISPR direct

This web tool is developed with purpose of designing guide RNA with reduced off-target sites (Naito et al. 2015). The perfect target sequences with low off-targets are counted by jellyfish algorithm. Jellyfish is an evaluation algorithm that use command line program (runs on 64 bit Linux) for reading DNA sequences in FASTA file format. The output data for all potential sgRNAs are displayed in three columns—20mer + PAM, 12mer + PAM and 8mer + PAM. 20mer + PAM column display the perfect matching of guides with whole 20 nt target sequences while 12mer + PAM and 8mer + PAM columns display guides with perfect matching with seed sequence only. The output column also displays GC content, melting temperature and the presence of TTTT sequence. The sequences showing the presence of TTTT are annotated as gray and considered as avoidable. It hosts genomic sequences of various organisms including *Homo sapiens*, *Rattus*, *Saccharomyces*, *Gallus*, *Rana*, *Drosophila*, and *Bombyx*. Link: <http://crispr.dbcls.jp/>.

Cas designer

This web tool is exploited for designing guide RNAs for various orthologues of type II Cas9 nucleases along with their potential off-targets (Park et al. 2015). The unique feature of this web tool is that it can also search for off-target sites in DNA and RNA bulges. Cas designer also assists researcher in selecting the appropriate target sites with variable PAM sequences such as NGG, NRG, NNAGAAW and NNGRRT. It assigns “microhomology associated out of frame score” that helps in choosing suitable sites for gene knockout (<http://www.rgenome.net/mich-calculator/>). In addition, “Frame score” also provides the ranked list of potential targets using features as percentage of GT, SNPs and uniqueness of guide. These features add up flexibility for selection of appropriate target sites. Link: <http://www.rgenome.net/cas-designer/>.

Table 2 All available well-established web tools for designing of the crRNA component of sgRNA of CRISPR-Cas9 system in plants

S.no.	Potential sgRNA design tools	Option for target search from whole genome is available or not	Input (sequence format)	Output (prediction for sgRNA off-target and other details)	Web source
1	ZiFit	Yes (only eight genomes including human genome)	Nucleotide sequence (raw text or FASTA format)	Yes	http://bindr.gdcb.iastate.edu/ZiFit/
2	Optimized CRISPR design	Currently not available			http://crispr.mit.edu:8079/
3	CRISPR direct	Yes (more than 200 genomes including many plant species)	Accession number or genome location or nucleotide sequence in FASTA format	Yes (highlight target highly specific and have lower off-target hits)	http://crispr.dbcls.jp/
4	Cas9 online designer	Yes (more than 300 genomes including many plant species)	Nucleotide sequence (raw text or FASTA format)	Yes	http://www.rgenome.net/cas-designer/
5	CHOPCHOP	Yes (more than 300 genomes including many plant species)	RefSeq/ENSEMBL, Gene ID or genomic co-ordinates	Yes	https://chopchop.rc.fas.harvard.edu/
6	E-CRISP	Yes (more than 50 genomes including a few plant species)	Gene symbol or nucleotide sequence (FASTA format)	No	http://www.e-crisp.org/E-CRISP/
7	sgRNAcas9	No details available	Any genomic sequence with FASTA header	Yes	www.biotechtools.com
8	sgRNA designer/CRISPRko	Yes (only human and mouse)	Transcript IDs, Gene IDs/Symbol or raw DNA sequence	Yes, but only for CRISPRko applications	http://portals.broadinstitute.org/gpp/public/analysis-tools/sgma-design
9	CRISPRseek	No details available	Nucleotide sequence (FASTA format)	Yes	http://www.bioconductor.org/
10	CRISPR MultiTargeter	No details available	Nucleotide sequence (FASTA format)	Yes	http://multicrispr.net/
11	CRISPR ERA	Yes (only nine genomes)	Gene name/location or gene sequence in FASTA format	Yes	http://crispr-era.stanford.edu/
12	Cas OT	Yes (any given gene/user provided sequence)	Raw data, Ensembl and genome sequence	Yes	http://casot.cbi.pku.edu.cn/
13	GuideRNA sequence design platform	Yes (only 32 genomes)	Nucleotide sequence (FASTA format)	Yes	http://cas9.cbi.pku.edu.cn/
14	Fly CRISPR	Yes	Any genomic sequence that includes ATGC and without any headers or special characters	No	http://flycrispr.molbio.wisc.edu/
15	Cas-OFFINDER	Yes (more than 300 genomes including many plant species)	Query sequence in FASTA format without any header	Yes	http://www.rgenome.net/cas-offinder/
16	Cas Finder	Yes	FASTA DNA sequence files	Yes	http://crispr.i2bc.paris-saclay.fr/Server/

Table 2 (continued)

S.no.	Potential sgRNA design tools	Option for target search from whole genome is available or not	Input (sequence format)	Output (prediction for sgRNA off-target and other details)	Web source
17	CRISPR design tool	Yes (only two including human)	FASTA DNA sequence files	Yes	https://apps.thermofisher.com/apps/crispr/index.html
18	CRISPR-P	Yes (especially for plant species; 50 genomes)	Locus tag, position or sequence in FASTA format	Yes	http://cbi.hzau.edu.cn/crispr/
19	Benchling	Yes	Target location	Yes	https://benchling.com
20	Deskgen	Yes	Gene transcript	Yes	https://www.Deskgen.com/

CHOPCHOP

The CHOPCHOP is web tool used for both CRISPR and TALEN genome editing. This tool accepts wide range of inputs as genomic regions, DNA sequence, gene name or co-ordinates. This web tool offers different model (optimized and advanced) for basic and advanced level users. The advanced features are for targeting sub region of gene as 5' UTR, 3' UTR, and splice site (Montague et al. 2014). CHOPCHOP searches for the off-targets by "Bowtie". The program "Bowtie" aligns short DNA sequences (35 bp reads) to large genomes and uses "Burrow Wheeler Transform", an algorithm for compressing data and for aligning the sequences. CHOPCHOP also give scores for target sites and as per their scores, and targeted sites are highlighted. In addition, it also has tool for primer designing. This web tool include model organisms as *Homo sapiens*, *C. elegans*, *Drosophila* and *Xenopus*. Link: <https://chopchop.rc.fas.harvard.edu/>.

E-CRISP

The E-CRISP is a web tool developed for designing sgRNA sequences. This web application identifies the complementary target sequences ending with 3' NG PAM (Heigwer et al. 2014). This tool also uses "Bowtie2" alignment for searching off-targets and homology in target sites and provides the list with detailed information about the targeted sequences. E-CRISP is also used for re-evaluating the CRISPR constructs for off-targets. This tool also assesses genomic contexts such as exons and CpG islands of putative designs. E-CRISP is available for diverse organisms such as *Arabidopsis*, *Brachypodium*, *C. elegans*, *Chlamydomonas*, *Drosophila*, *Leishmania*, *Neurospora*, *Macaca*, *Mus*, *Homo sapiens* and *Gallus*. Link: <http://www.e-crisp.org/E-CRISP/>.

sgRNAs9

The sgRNAs9 is a pearl script-based tool for designing of sgRNA with minimal off-target sites (Xie et al. 2014). It searches for target site with user-defined parameters and predict Cas9 potential off-target sites in targeted genome. The potential off-target sequences are classified as type I (1–5 mismatches in non-seed region), type II (1–5 mismatches in both seed and non-seed region) and type III with 1–3 mismatches. Link: <http://www.biotoools.com>.

sgRNA designer (CRISPRko)

The sgRNA designer (CRISPRko) is a tool constructed for identification of the sgRNA sequence for a particular target for only CRISPRko (CRISPR-knockout) application

(Doench et al. 2014, 2016). This tool attempts to maximize the on-target activity. The “Rule set 2” method has been used for on-target scoring (Doench et al. 2016). “Rule set 2” method employs “regression model” that ranks all on target activity. Top-scored sgRNAs will have highest on-target activity; while off-target sites are evaluated simultaneously by CFD (cutting frequency determination). The CRISPRko program allows input sequence only from human and mouse. The Cas9 used in CRISPRko application is only from *S. pyogenes* and *S. aureus* CRISPR-Cas9 families. Link: <http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>.

CRISPRSeek

The CRISPRSeek is a bioconductor package source that helps in designing of the sgRNA and also examines the potential off targets (Zhu et al. 2014). The different predicted sgRNAs are ranked on the basis of “Cleavage score” (range 0–100) determined by various published metrics. It also allows variations in PAM and users can input their own mismatch penalty matrix. It uses the biostrings for searching mismatches. It is a best tool, when sgRNA has to discriminate between two related sequences. As per the command, the features are also set in the target sequences, as addition of guanine for in vivo synthesis of sgRNA using U6 Promoter. Filtering features such as oriented and spaced pairs are also provided for further improvement of cleavage specificity. Link: <http://www.bioconductor.org>.

CRISPR MultiTargeter

The CRISPR MultiTargeter is generally used for finding the common (highly similar) and unique sgRNAs in the input sequences corresponding to queries (Prykhodzhiy et al. 2015). The multiple sequence alignment is used for common target search while string comparison algorithm is used for identification of unique sequence(s) among all input sequence. The key point of this alignment approach is that it searches for the sequence that does not show pairing with other sequences. This tool also provides information about melting temperature (T_m) for DNA: RNA hybrid. However, this tool does not predict off-targets and it has additional tool named as Cas-OFFinder for this purpose. Input sequence can be included from *Gallus*, *Xenopus*, *Rattus*, *Drosophila*, *C. elegans*, *H. sapiens* and *Arabidopsis*, which helps in designing of unique targets in alternative exons in transcripts of isoforms. Link: <http://multicrispr.net/>.

CRISPR ERA

In CRISPR ERA, the acronym “ERA” refers to the editing, repression and activation (Liu et al. 2015). This web tool is based on sgRNA design and user(s) can predict the most efficient and specific sgRNA based upon a specific sequence of interest. In addition, user can also generate sgRNA for modulating gene transcription (its activation and repression with editing components). The CRISPR ERA includes genomic resources from model organisms (*E. coli*, *B. subtilis*, *S. cerevisiae*, *C. elegans*, *Zebra fish*, Mouse and Fruit fly, *Rattus*, *Homo sapiens*) for identification of valid sgRNAs. Link: <http://crispr-era.stanford.edu/>.

CasOT

The CasOT is a tool designed for searching of potential target site(s) in the user-provided sequence according to type of PAMs and mismatches number in seed and non-seed regions (Xiao et al. 2014). This tool offers three modes for sgRNA searching such as single sgRNA, paired sgRNA of the target, and off-target. Using this tool, any genetic sequence can be targeted using two paired sgRNAs than single sgRNA. Link: <http://casot.cbi.pku.edu.cn/>.

GuideRNA sequence

GuideRNA sequence is a design platform that is used for the designing of sgRNAs to find potential targets for user-provided sequence(s). This tool ranks pool of sgRNAs as per their distinctiveness, AT content and single-nucleotide polymorphism parameters. This also offers option for sharing of experimental results related to most sgRNA sequences (Ma et al. 2013). Link: <http://cas9.cbi.pku.edu.cn/>.

Fly CRISPR

Fly CRISPR is exclusively available web tool for utilization of the CRISPR-Cas9 system in *Drosophila* based upon homology-directed repair approach. It facilitates the incorporation of large DNA sequences with aim of genome engineering (Bassett and Liu 2014). Link: <http://flycrispr.molbiol.wisc.edu/>.

CRISPR Cas-OFFinder

CRISPR Cas-OFFinder is a flexible sgRNA design RNA tool (algorithm) that searches for potential off-target sites in the given genome or user-defined sequence. It supports user-defined PAM sequence and sgRNA length which offers opportunity to improve targeting range and specificity. The algorithm used for on-target and off-target scoring is also integrated for identification of highly specific sgRNAs (Bae

et al. 2014). The output results are presented with explanation of chromosome positions and potential off-targets associated with specific sgRNAs. Link: <http://www.genome.net/cas-offfinder/>.

CRISPRFinder

CRISPRFinder is specialized web service tool for detection of the CRISPR (even shortest length of 1–2 motifs). It provides additional features for obtaining flanking sequences and blasts them against Genbank databases (Grissa et al. 2007). This web tool also checks that whether direct repeats (DR) is found elsewhere in prokaryotic genome or not. The Vmatch software is used to find the maximal repeats (<http://www.vmatch.de/>) and subsequently CLUSTALW program compares the repeats with the spacer, and eliminates the tandem repeats from DR (direct repeats). This web tool can be accessed from <http://crispr.u-psud.fr/Server/CRISPRfinder.php>.

CRISPR design tool

The CRISPR design tool is used for selecting the CRISPR guide by determining off-targets and highlighting sgRNAs having high-target specificity. The target sites are also ordered according to predicted off-target sites and specificity scores. Link: <https://apps.thermofisher.com/apps/crispr/index.html>.

SSFinder

SSFinder is python-scripted database used for searching specific CRISPR Cas target site in small as well as large genomic data. The tool is also compatible with windows interface. This web tool first analyze the DNA sequence for 23-nt sequences adjacent to PAM and then screened for distinct 12 bp seed sequence (Upadhyay and Sharma 2014). The screened sequences are then classified on the basis of nature of start and end nucleotides. The output result is displayed in seven columns that include specific target sequence, potential off-target sites, position of initial and last nucleotide in the target sequence and classification on the basis of it. Link: <https://code.google.com/p/ssfinder>.

Off-spotter

Off-spotter assists in designing of the potential sgRNAs by detecting potential off-targets for each candidate sgRNA (Pliatsika and Rigoutsos 2015). This tool offers features including multiple display options (e.g. column display) and multi-sorting for designing and effective selection of sgRNAs. It operates currently on human, mouse and yeast. The outputs are also annotated that includes pool of unspliced

pre-mRNA, unspliced and spliced RNA and mRNA. The output result also displays the position in which sgRNA is displayed reverse complement of sgRNA strand on which it is found, and histogram showing mismatches for each sgRNA. Link: <https://cm.jefferson.edu.offspotter>.

CASPER

CRISPR-associated software for pathway engineering and research (CASPER) implements an algorithm for prediction of both on- and off-target activities irrespective of experimental conditions of sgRNA. For on-target activities, the formulated scores were given by dividing CRISPR scan score with penalty score (Mendoza and Trinh 2017). This scoring method depends only upon target sequence and hence do not rely on experimental conditions. This tool has expanded functions for multi-targeting analysis and multi-population analysis. Link: <https://github.com/TrinhLab/CASPER>.

Selected plant-based databases

A database (OMICS TOOLS) was identified that shows all available tools developed till today for the CRISPR genome editing and it has been found that only three databases (Deskgen, Benchling and CRISPR-P) were claimed for screening of sgRNAs from plants and discussed in our best knowledge as follows:

CRISPR-P

The CRISPR-P is the first web tool documented for sgRNA designing for plant species. It searches for specific target sites and provides the visually displayed off-target sites (Yang et al. 2014). The off-targets have been identified with the integration of “Off-target rules” and sequence alignment has been carried out using “Batmis program”. The “Batmis” (Basic Alignment Tool for mismatches) is C/C++ written algorithm for finding the mismatches. This tool also provides the restriction enzyme cutting sites for each sgRNA and scores the guide sequences. A guide sequence with high “guide score” and low off-target score seems to be fit for further editing experiment. The CRISPR-P offers the sgRNA designing in more than 20 plant species (*Arabidopsis*, *Brassica*, *Chlamydomonas*, *Oryza* and *Sorghum*). The CRISPR-P allows users to search for highly specific Cas9 target sites in user-defined DNA sequences. This tool also provides off-target loci prediction for further analysis and marks restriction enzyme cutting sites for ease of experiment. The CRISPR-P has many benefits over other existing sgRNA design tools including CRISPR DESIGN, Cas-OFFinder, E-CRISP and Cas designer (Bae et al. 2014; Heigwer et al. 2014; Park

et al. 2015). The CRISPR-P also provides off-target sites prediction for every possible sgRNA. Mismatch sites and detailed information for off-target loci are also presented as output results. CRISPR-P is available at <http://cbi.hzau.edu.cn/crispr>.

Benchling

The Benchling is another potential web tool that offers the designing of sgRNAs in the plants (Pellegrini 2016). It allows sharing the experiments and tests with other researchers, and especially constructed for pharma and biotech companies. The Benchling designs the CRISPR sgRNAs by analyzing parameters including target location, specificity, and efficiency feature. This web tool is integrated with the CRISPR sgRNA design tool to specify a gene or genome co-ordinates and automatically annotate given genome with the exonic information from 40 + reference genomes. The scoring of sgRNAs covering on-target and off-target effects could be retrieved quickly (within seconds) and need to browse different sequences based upon scores. A good score can help for identification of sgRNA for experiment. The different potential sgRNAs as oligos could be organized using tags and folders, and stored on Benchling. This useful information can be subsequently used for cloning the sgRNAs in the CRISPR plasmids. Alternatively, sgRNAs, along with scores and off-target sites could be exported, for easy use in one's own spreadsheets. It can be accessed from <https://benchling.com>.

Deskgen

The Deskgen is user-friendly web tool for academic users for designing of sgRNAs. The “knockout” and “knockin”, new features offered by the Deskgen, is also supported for some plant genomes and very useful for functional validation of genes (Hough et al. 2016). Currently, the genomic resources of maize, rice, wheat and barley are supported by Deskgen. This web tool is organized into two modules named as “EXPERIMENTS” and “REPOSITORIES”.

The desktop genetics (Deskgen) was founded in 2012 by three graduates (University of Cambridge; UK) with the goal of marrying modern genomics with rapid advancements in data science. This is a continuously curated platform that inculcates the recent updates in sgRNA design algorithms and serves as a proof-of-concept testing ground for an ever-changing CRISPR dogma. The “Knockout” and “Knockin” tools facilitate a range of genomes including eukaryotes and prokaryotes, as well as alternative CRISPR nucleases such as Cas9 orthologs and Cpf1.

The Deskgen platform features multiple experiment-focused design settings (while most online single-guide RNA design tools do not). It incorporates advanced strategies to improve precision editing rates via homology-directed repair. To more precisely edit the genome, the Deskgen develops CRISPR libraries with weighted off-target scoring. CRISPR libraries can be designed to target coding and noncoding (regulatory) regions of the genome. Validating CRISPR experiments with NGS deep sequencing can ensure causality and identify or confirm off-target effects. The more data that flow through the Deskgen, the better the CRISPR cognitive tool gets due to machine learning (Hough et al. 2016). Link: <https://www.Deskgen.com/>.

Assessment of plant-based web tools for precise sgRNA designing

These three plant-based web tools can be analyzed by their features and comparing the results against a single common user-defined query. Initially, comparative analysis of features offered by these three web tools has been discussed to later understand their respective operation by compiling a set of functions. In addition, the output comparisons also help their basis designing and could also offer opportunity for construction of new efficient web tools for plant research community. For feature-based comparisons, eight major categories were selected to describe the functionality as well as user interface. These categories are further described briefly to sensitize users about their function and usage for analysis purposes. The Benchling, Deskgen and CRISPR-P were comprehensively analyzed in terms of different outputs of sgRNA designing.

Basic functions

These functions embody the primary goals of the software and should be the first place to look to determine the suitability or applicability of any web tool. A number of basic functions have been predicted to make easy operation of relevant web tool. The basic function complied as single-target design (design of one sgRNA to one target), multi-target design (design of one sgRNA to many targets), paired nickase design (able to design a pair of sgRNA for a second nickase approach), off-target aware (considers maximum possible off-targets), on-target aware (considers binding or cleaving efficiency of on-targets), adjustable sgRNA length (able to design sgRNAs other than the canonical length of 20 nt), small mismatch limit (finds off-targets up to M mismatches distant, where M is

less than 5 and typically three due to use of fast read mappers such as bowtie), high mismatch limit (finds off-targets up to M distant, where M is equal to 5 or more), exact mapper (guaranteed to take into account all off-targets matching parameters), approximate design (applies some fair rules to determine the value of one sgRNA design over another), and empirical design (makes use of empirically derived or prior knowledge to inform the qualities of what defines sgRNA design and both). In addition to these functions, the Doench root square (considers the Doench root on-target aware score), pairwise score, hybrid off-target aware score, single-PAM design (only designs sgRNA for one PAM motif, usually NGG), multi-PAM design (designs sgRNA for several PAM motif), universal PAM model (designs sgRNA for any motif) and degenerate off-target aware (degenerate PAM motifs are considered when calculating off-target scores) have also been documented. All these functions reflect the feasibility and suitability of enlisted web tools and must be treated as common features for successful execution of operation.

Advanced functions

These functions are not entirely necessary but should be considered extremely useful, depending on designing for goals. These are predicted and enlisted as, feature aware (reports context of results regarding genomic annotations, a score of 0.5 indicates limited functionality), sgRNA nucleotide locking (allows specific nucleotide position in the sgRNA to be specified either exactly as ACGT or as mixed nucleotide IUPAC code), exogenous off-target (able to check for possible off-targets in exogenous DNA), SNP aware (an organism's SNP are either taken into account when designing, or contextualize in the results), restriction enzyme aware (provides context of restriction sites), secondary structure-aware (sgRNA secondary structure is either taken into account during design, or reported) and microhomology aware (microhomology is taken into account during design). All these functions are advanced and provide more in-depth information about the different aspects of query.

Utility functions

These functions include batch and multiplex design (multiple single- or multi-target designs without restarting the workflow or application), flanking FASTA (allows one to get the sequence surrounding the target site). The primer design (suggests optimal primer design), plasmids design (helps design a custom plasmids containing the sgRNA for CRISPR) and design classes (provides sensible parameter

sets based on intended use of CRISPR) are also useful to harness for different usages. However, the most common utility function is batch design feature followed by primer designing.

User interaction

The user interactive features such as “offline”, “online”, “offline-device”, “offline: Linux”, “offline: MAC OS X”, “offline: Windows”, “commercial”, “command line interface” and “graphical user interface” are enlisted as major features offers for making any web tool as user interactive.

Input flexibility

The input flexibility feature offers the easy selection of user-defined sequence(s) or genome for further analysis. Sometimes the genome of interest is absent and that software is no furthermore usable for in-depth analysis. Generally, in software, this feature offers via two different approaches, either the software is relying on a database, or some parameters are dependent on the selected organisms. In continuation, visual track (input sequence or region by click-dragging on genome track), sequence (options: copy, paste, upload, open, FASTA formatted file and text), identifier (use unique gene identifier), region (use genomic co-ordinates), attribute (use the meta-date found in either GFF or GTF annotation file to select a gene), full genome FASTA file (indexed or not) are also be specified, usually a hallmark of offline tools and load previous session (can load a previously saved file).

Output diversity

This includes HTML: HTML formatted report, FASTA: FASTA formatted report, either GFF or GTF (provides targetable sites in genomic annotation format) visual track, text, plot (a graphical plot of the statistical aspects of the results), and tabular (provides tabular data easily exported for use in spreadsheet and statistical analysis software). A desirable feature in the “the output diversity” section is the ability to analyze the results further or provide any form of enhancing exploration of the data and all command line interfaces related to this feature should be automatically given. The other features including exportable (results can be exported into various formats), save current session (the results may be saved to a file that can be between users and loaded using the same software) should also provided by different web tools.

Community exclusivity

Organism agnostic (the tool can be used for any organism in exact the same way), special focus (the tool is clearly designed for use in one particular organism or a group of organisms), model organisms (the tool has been implemented for maximum model organisms), short sequence (tools is designated to detect target sites within a small sequence, typically a coding region), small genome (the tool is capable of analyzing small genomes, i.e., <0.5 GB), large genome (the tool is capable of analyzing large genomes, i.e., <2 GB), enormous genome (the tool is capable of analyzing vast genomes, i.e., >2 Gbps), user extended for any genome (the user can extend the software to organisms that was not initially released with, or the software applies to any genome and only constrained to by computational resources) and organism-biased scoring. All these exclusive options are integrated by different web tools according to different applications.

Comparisons of tools in terms of output with a single enquiry

Two genes {*IDM2* (*AT1G54840*) and *IDM3* (*AT1G20870*)} were randomly selected for testing and understanding of operation of these three selected computational tools. These selected genes are located on chromosome number one of *Arabidopsis* genome. *IDM2* (*AT1G54840*): increase in DNA methylation 2 is a protein-coding gene (1657 bp), and exons

(1–3) and promoter regions of *IDM2* are used for sgRNA designing (Li et al. 2015). Another selected target is *IDM3* (*AT1G20870*: increase in DNA methylation 3) that is also a protein-coding gene with 1674 bp (Li et al. 2015; Duan et al. 2017). The exon (1 and 2) and promoter regions of *IDM3* are used for sgRNA designing.

Output analysis from the Benchling

Using *IDM2* sequence as an enquiry with the Benchling, a total numbers of 111 possible sgRNAs were generated and further only eight were selected as best hits on the basis of criteria for potential sgRNAs (on-target score, off-target score, sequence properties and GC%). Among eight curated sgRNAs, only four sgRNAs showed 100% score. Similarly, sequence of *IDM3*, was also used as user-defined enquiry and out of total generated sgRNA populations (145), only seven sgRNAs were identified as most potential on the basis of same selection criteria as discussed above in the same subsection (Table 3).

Output analysis from the desktop genetics

Using desktop genetics as web tool, a population of 85 sgRNAs were generated using gene sequence of *IDM2* as input sequence. Only three sgRNAs (which have maximum score) were founded as best hits for usage based upon selection criteria as adopted earlier. With genetic sequence of *IDM3* as an enquiry, among total generated 145 sgRNAs, only 3

Table 3 Potential sgRNAs (only crRNA component) selected from Benchling using *IDM2* and *IDM3* genes as input queries

No	Sequences	PAM type	Predicted score	GC content (%)	Mismatches
<i>IDM2</i> as an input query					
1.	TCTCTCATACAAGAAGTATC	TGG	100	65	0
2.	CTCTTCCGTGTATCGCTTCC	TGG	100	55	0
3.	AATGCGGGTACTTGCAC TAG	CGG	100	50	0
4.	GTTTAAGCGGCTTACGGGAT	TGG	100	50	0
5.	TCTGTCTCGTTCCGCCTCCC	CGG	100	65	0
6.	ACGCGGAGTCTCTGCCCTCC	CGG	100	70	0
7.	CAGGCGGGAAGCGAGTTAAA	CGG	100	55	0
8.	GCTGTGAAGTAGAAGACAAT	GGG	100	55	0
<i>IDM3</i> as an input query					
1.	CCGTAATCTTTGCGGACACC	GGG	100	55	0
2.	GCGGATCAACTGGTCCCGGG	AGG	100	70	0
3.	GTTAGCACTTAGACCCGATG	TGG	100	50	0
4.	CCGATCAGATGGTATCTTCG	AGG	100	50	0
5.	ATCCGGAAGCTGTGTCCGCC	TGG	100	65	0
6.	TACTTCCATACACGCCACAT	CGG	100	45	0
7.	TCCTTCATCTCCGAAGAAAA	AGG	100	40	0

Table 4 Potential sgRNAs (only crRNA component) selected from Deskgen using *IDM2* and *IDM3* genes as input queries

No	Sequences	Type of PAM sequence	Presence of exon	Activity	Score of off-target	GC content (%)
<i>IDM2</i> as an input query						
1.	GTGAAATCTGAAGCTCATGG	TGG	2	72	100	45
2.	GAAACGCAATGATTCTGAACA	GGG	2	68	100	40
3.	CATTGGGTATAATCGCCACG	TGG	2	74	100	45
<i>IDM3</i> as an input query						
1.	TAATGGAGTTGTCTACTAACG	AGG	1	75	100	40
2.	GTTCCAGTAGTTACAACCGA	CGG	1	70	100	45
3.	GCGGATCAACTGGTCCCGG ^a	AGG	2	71	100	70

^aThis sequence is exceptionally 19-nt-long sequence as potential sgRNA, while all others are 20-nt-long sequence

Table 5 Potential sgRNAs (only crRNA component) selected from CRISPR-P using *IDM2* and *IDM3* genes as input queries

No	Sequences	PAM type	Predicted Score	Score of off-target
<i>IDM2</i> as an input query				
1	GTTTAAGCGGCTTACGGGAT	TGG	100	0
2	ACGCGGAGTCTCTGCCCTCC	CGG	100	0
<i>IDM3</i> as an input query				
1	ATACCATCTGATCGGAAGTT	AGG	100	0
2	CCGATCAGATGGTATCTTCG	AGG	100	0
3	TACTTCCATACACGCCACAT	CGG	100	0
4	CACAGCTTGAATGGTCCAGG	CGG	100	0
5	CCCGGTGTCCGCAAAGATTA	CGG	100	0
6	GCCGCGGATCAACTGGTCCC	GGG	100	0

sgRNAs (with maximum score) were identified as potential sgRNAs (Table 4).

Output analysis from the CRISPR-P

Using *IDM2* sequence as an input enquiry, the CRISPR-P database generated 111 sgRNAs. Subsequently, only two sgRNAs were screened as best hits. The sgRNAs with unique scores (off-target sites score, 0; on-target score, 100) were identified as best hits. While using *IDM3*, a total number of 145 sgRNAs were generated; only 6 sgRNAs were identified as potential sgRNAs (off-target sites score, 0; on target score, 100) as represented in Table 5.

Comparison analysis of output from web tools

All three databases (Benchling, Deskgen and CRISPR-P) have different features for generation and analysis of sgRNAs against user-defined query. In present investigation, the sequences of two genes (*IDM2* and *IDM3*) were

selected for comparative analysis of generated corresponding sgRNAs (common and unique). Using these three selected tools (Benchling, Desktop and CRISPR-P), it was also found that the Benchling display unique eight different sgRNAs (selected on the basis of maximum score) and unique seven different sgRNAs for user-defined query as *IDM2* and *IDM3*, respectively. In contrast, the Deskgen web tool gives only unique three different sgRNAs for each *IDM2* and *IDM3* as query. In addition, only two different sgRNAs were generated with maximum score against *IDM2* gene as query and six different gRNAs against *IDM3* as query using third selected web tool, the CRISPR-P. Interestingly, all considered generated sgRNAs for *IDM2* gene using two different web tools (Benchling and Deskgen) are totally different from each other in terms of nucleotide sequence. With same query (*IDM2*), the sequence comparison among all considered selected sgRNAs using web tools (Deskgen and CRISPR-P) was also found to be totally dissimilar in nature. Only nucleotide sequences of two generated sgRNAs by both Benchling and the CRISPR-P were found to common with same query (*IDM2*).

Considering *IDM3* sequence as query for generation of sgRNAs using all three web tools, only sequence of one sgRNA, generated via Deskgen has been found to be exactly similar in sequence to the sgRNA generated from the Benchling. After sequence comparison of all considered potential sgRNAs using web tools (the CRISPR-P and Deskgen), only one sgRNA via CRISPR-P has been found to be common to the sgRNA generated by the Deskgen. Similarly, the two sgRNAs generated via CRISPR-P have been found to be common in their sequence with the sgRNAs generated via the Benchling.

The nucleotide sequence of sgRNA is very important to determine the efficiency and specificity of sgRNA for precise genome editing and surprisingly, these three selected tools generated different results (sgRNAs) even for common gene query. Thus, it is urgent requirement to pay the attention to construct potential sgRNA population using a well-validated tool for precise editing. In conclusion, two web tools (Benchling and CRISPR-P) selected from the pool of computational tools offered somehow common results on the basis of a number of generated sgRNAs.

Future perspectives

A detail study of research on CRISPR-Cas-mediated modifications which occur in plants still remain unclear. More investigations are still needed as far as the specificity of the CRISPR-Cas system is concerned. This can be fulfilled by testing more sgRNA-mediated genome editing, in models as well as in other plant species. If off-target effects are found in other potential sgRNA-mediated genome editing and in other plants, so there is an urgent requirement of attention for drafting novel methodologies to reduce the off-target effects. However, the generation of sgRNA library for their testing in terms of efficiency and specificity in plant is somehow difficult in comparison of animal system and very time consuming. Using truncated sgRNAs and the Cas9-nickase, this system with off-target mutagenesis can be reduced (Fu et al. 2014). It is also feasible that CRISPR-Cas can be used to place markers such as a *GFP* gene in a desired location, providing important tools for basic research (Schwinn et al. 2017). The sgRNA libraries need to be constructed for plants as well, to enhance plant-breeding programs at least in model *Arabidopsis*. Among a long list of the sgRNA-designing tools, only few qualify to design guide for plant system. But it is difficult to validate each and every web tools because of non-uniformity in their outputs. Therefore, the construction of precise sgRNA web tool for usage in plant system is highly desirable at this moment.

Acknowledgements The authors express deepest gratitude to Vice Chancellor of Central University of Punjab, India for providing

financial support during course of this work. “UGC-BSR start up grant” sanctioned to Vinay Kumar, who sponsors this research. The authors thank anonymous reviewers and editors for critical reading the manuscript and suggesting substantial improvements.

Author contributions VK conceived and designed the present research. VK and KM conducted the experiments. KM and VK analyzed the data. APU, SKY and VK wrote the manuscript. All the authors read and approved the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Bae S, Park J, Kim JS (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30:1473–1475
- Bassett AR, Liu JL (2014) CRISPR-Cas9 and genome editing in *Drosophila*. *J Genet Genom* 41:7–19
- Bhowmik P, Ellison E, Polley B et al (2018) Targeted mutagenesis in wheat microspores using CRISPR-Cas9. *Sci Rep* 8:6502
- Briner AE, Donohoue PD, Goma AA et al (2014) Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* 56:333–339
- Cai Y, Chen L, Shi S et al (2018) CRISPR-Cas9-mediated deletion of large genomic fragments in soybean. *Int J Mol Sci* 19:3835
- Chari R, Mali P, Moosburner M, Church GM (2015) Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* 12:823
- Chen H, Choi J, Bailey S (2014) Cut site selection by the two nuclease domains of the Cas9 RNA guided endonuclease. *J Biol Chem* 289:13284–13294
- Chen R, Xu Q, Liu Y et al (2018) Generation of transgene-free maize male sterile lines using the CRISPR-Cas9 System. *Front Plant Sci* 9:1180
- Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR-Cas systems. *Science* 339:819–823
- Doench JG, Hartenian E, Graham DB et al (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 32:1262–1267
- Doench JG, Fus IN, Sullender M et al (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 34:184–191
- Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096
- Duan CG, Wang X, Xie S et al (2017) A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation. *Cell Res* 27:226–240
- Fagerlund RD, Staals RHJ, Fineran PC (2015) The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biol* 16:251
- Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P, Zhu JK (2013) Efficient genome editing in plants using a CRISPR-Cas system. *Cell Res* 23:1229
- Feng C, Yuan J, Wang R et al (2016) Efficient targeted genome modification in maize using CRISPR-Cas9 system. *J Genet Genom* 43:37–43
- Fine EJ, Cradick TJ, Zhao CL, Lin Y, Bao G (2014) An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Res* 42:e42

- Fogarty NM, McCarthy A, Snijders KE et al (2017) Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* 550:67
- Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32:279–284
- Gang L, Zhang H, Lou D, Yu D (2016) Selection of highly efficient sgRNAs for CRISPR-Cas9-based plant genome editing. *Sci Rep* 6:21451
- Gopalappa R, Suresh B, Ramakrishnan S, Kim HH (2018) paired D10A Cas9 nickases are sometimes more efficient than individual nucleases for gene disruption. *Nucleic Acids Res* 46:e71
- Grisa I, Vergnaud G, Pourcel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57
- Heigwer F, Kerr G, Boutros M (2014) E-CRISP: fast CRISPR target site identification. *Nat Methods* 11:122–123
- Hough SH, Ajetunmbi A, Brody L, Humphries-Kirilov N, Perello E (2016) Desktop genetics. *Per Med* 13:517–521
- Hsu PD, Scott DA, Weinstein JA, RanFA KonermannS, Agarwala V, Cradick TJ (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31:827
- Jiang F, Doudna JA (2017) CRISPR-Cas9 structures and mechanisms. *Annu Rev Biophys* 46:505–529
- Jinek M, Jiang FG, Taylor DW et al (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997
- Kim D, Alptekin B, Budak H (2018) CRISPR-Cas9 genome editing in wheat. *Funct Integr Genom* 18:31–41
- Kuan PF, Powers S, He S, Li K, Zhao X, Huang B (2017) A systematic evaluation of nucleotide properties for CRISPR sgRNA design. *BMC Bioinform* 18:21
- Kumar V, Jain M (2014) The CRISPR-Cas system for plant genome editing: advances and opportunities. *J Exp Bot* 66:47–57
- Li Q, Wang X, Sun H, Zeng J, Cao Z, Li Y, Qian W (2015) Regulation of active DNA demethylation by a methyl-CpG-binding domain protein in *Arabidopsis thaliana*. *PLoS Genet* 11:e1005210
- Li J, Zhang H, Si X et al (2017a) Generation of thermosensitive male-sterile maize by targeted knockout of the ZmTMS5 gene. *J Genet Genom* 44:465–468
- Li WR, Liu CX, Zhang XM, Chen L et al (2017b) CRISPR-Cas9-mediated loss of FGF5 function increases wool staple length in sheep. *FEBS* 284:2764–2773
- Liang G, Zhang H, Lou D, Yu D (2016) Selection of highly efficient sgRNAs for CRISPR-Cas9-based plant genome editing. *Sci Rep* 6:21451
- Liu H, Wei Z, Dominguez A, Li Y, Wang X, Qi LS (2015) CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression, and activation. *Bioinformatics* 31:3676–3678 (btv423)
- Liu X, Homma A, Sayadi J, Yang S, Ohashi J, Takumi T (2016) Sequence feature associated with cleavage efficiency of CRISPR-Cas 9 system. *Sci Rep* 6:19675
- Ma M, Ye AY, Zheng W, Kong L (2013) A guide RNA sequence design platform for the CRISPR-Cas9 system for model organism genomes. *Biomed Res Int* 2013:1–4
- Mendoza BJ, Trinh CT (2017) Enhanced guide-RNA design and targeting analysis for precise CRISPR genome editing of single and consortia of industrially relevant and non-model organisms. *Bioinformatics* 34:16–23
- Miao J, Guo D, Zhang J, Huang Q, Qin G, Zhang X, Wan J, Gu H, Qu L (2013) Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res* 23:1233
- Miao C, Xiao L, Hua K et al (2018) Mutations in subfamily of abscisic acid receptor genes promote rice growth and productivity. *PNAS* 115:6058–6063
- Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155:733–740
- Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E (2014) CHOPCHOP: a CRISPR-Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* 42:W401–W407
- Naito Y, Hino K, Bono H, Ui-Tei K (2015) CRISPRdirect: software for designing CRISPR-Cas guide RNA with reduced off-target sites. *Bioinformatics* 31:1120–1123
- Nishimasu H, RanFA Hsu PD, Konermann S, Shehata SI, Dohmae N et al (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156:935–949
- Nowak CM, Lawson S, Zerez M, Bleris L (2016) Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Res* 44:9555–9564
- Oh JN, Choi KH, Lee CK (2017) Multi-resistance strategy for viral diseases and in vitro shRNA verification method in pigs. *Asian Australas J Anim Sci*. <https://doi.org/10.5713/ajas.17.0749>
- Park J, Bae S, Kim JS (2015) Cas-Designer: a web-based tool for the choice of CRISPR-Cas9 target sites. *Bioinformatics* 31:4014–4016
- Pellegrini R (2016) Edit single bases with Benchling! <https://blog.benchling.com/2016/07/18/base-editor>. Accessed 18 July 2016
- Periwal V (2016) A comprehensive overview of computational resources to aid in precision genome editing with engineered nucleases. *Brief Bioinform* 18:698–711
- Pliatsika V, Rigoutsos I (2015) Off-Spotter: very fast and exhaustive enumeration of genomic look alikes for designing CRISPR-Cas guide RNAs. *Biol Direct* 10:4
- Plummer RJ, Guo Y, Peng Y (2018) A CRISPR reimagining: new twists and turns of CRISPR beyond the genome-engineering revolution. *J Cell Biochem* 119:1299–1308
- Prykhodzhiy SV, Rajan V, Gaston D, Berman JN (2015) CRISPR multitargeter: a web tool to find common and unique CRISPR single RNA targets in a set of similar sequences. *PLoS One* 10:e0138634
- Pyott D, Sheehan E, Molnar A (2016) Engineering of CRISPR-Cas9-mediated potyvirus resistance in transgene-free *Arabidopsis* plants. *Mol Plant Pathol* 17:1276–1288
- Ran FA, Hsu PD, Lin CY et al (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154:1380–1389
- Ross MJ, Coates PT (2017) Using CRISPR to inactivate endogenous retroviruses in pigs: an important step toward safe xenotransplantation. *Kidney Int* 93:4–6
- Sánchez-León S, Gil-Humanes J, Ozuna CV et al (2018) Low-gluten, nontransgenic wheat engineered with CRISPR-Cas9. *Plant Biotechnol J* 16:902–910
- Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 32:347–355
- Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res* 38:W462–W468
- Schwinn MK, Machleidt T, Zimmerman K, Eggers CT, Dixon AS, Hurst R et al (2017) CRISPR-mediated tagging of endogenous proteins with a luminescent peptide. *ACS Chem Biol* 13:467–474
- Shen C, Que Z, Xia Y et al (2017) Rapid generation of genetic diversity by multiplex CRISPR-Cas9 genome editing in rice. *Sci China Life Sci* 60:89–93
- Shi J, Gao H, Wang H et al (2017) ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol J* 15:207–216
- Shimatani Z, Kashojiya S, Takayama M et al (2017) Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat Biotechnol* 35:441–443
- Shmakov S, Abudayyeh OO, Makarova KS et al (2015) Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol Cell* 60:385–397

- Slaymaker IM, Gao L, Zetsche B et al (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* 351:84–88
- Song J, Yang D, Ruan J, Zhang J (2017) Production of immunodeficient rabbits by multiplex embryo transfer and multiplex gene targeting. *Sci Rep* 7:12202
- Sontheimer EJ, Wolfe SA (2015) Cas9 gets a classmate. *Nat Biotechnol* 33:1240–1241
- Stemmer M, Thumgerger T, del Sol Keyer M, Wittbrodt J, Mateo JL (2017) CCTop: an intuitive, flexible and reliable CRISPR-Cas9 target prediction tool. *PLoS One* 12:e0176619
- Tang L, Zeng Y, Du H, Gong M (2017) CRISPR-Cas9-mediated gene editing in human zygotes using Cas9 protein. *Mol Genet Genom* 292:525–533
- Ueta R, Abe C, Watanabe T et al (2017) Rapid breeding of parthenocarpic tomato plants using CRISPR-Cas9. *Sci Rep* 7:507
- Uniyal AP, Yadav SK, Kumar V (2019) The CRISPR–Cas9, genome editing approach: a promising tool for drafting defense strategy against begomoviruses including cotton leaf curl viruses. *J Plant Biochem Biotechnol*. <https://doi.org/10.1007/s13562-019-00491-6>
- Upadhyay SK, Sharma S (2014) SSFinder: high throughput CRISPR-Cas target sites prediction tool. *Biomed Res Int* 14:742482
- Vilarino M, Rashid ST, Suchy FP, McNabb BR et al (2017) CRISPR-Cas9 microinjection in oocytes disables pancreas development in sheep. *Sci Rep* 12:17472
- Wang W, Pan Q, Akhunova A et al (2018) Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. *CRISPR J* 1:65–74
- Wei C, Wang F, Liu W, Zhao W, Yang Y, Li K, Xiao L, Shen J (2018) CRISPR-Cas9 targeting of the androgen receptor suppresses the growth of LNCaP human prostate cancer cells. *Mol Med Rep* 17:2901–2906
- Wright AV, Nunez JK, Doudna JA (2016) Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 164:29–44
- Wu F, Ge Gao, Pan T, Yang Z et al (2017) Generation of a *SMO* homozygous knockout human embryonic stem cell line WAe001-A-16 by CRISPR-Cas9 editing. *Stem Cell Res* 24:89–93
- Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, Gao G, Zhang B (2014) CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* 30:1180–1182 (**btt764**)
- Xie S, Shen B, Zhang C, Huang X, Zhang Y (2014) sgRNAs9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS One* 9:e100448
- Yang L, Li L, Hai-Yang L, Sen L, Feng X, Ling-Ling C (2014) CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant* 7:1494–1496
- Yang Y, Zhu G, Li R (2017) The RNA editing factor SIOORM4 is required for normal fruit ripening in tomato. *Plant Physiol* 175:1690–1702
- Yin K, Gao C, Qiu JL (2017) Progress and prospects in plant genome editing. *Nat Plants* 3:17107
- Zetsche B, Heidenreich M, Mohanraju P et al (2017) Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nature Biotechnol* 35:178
- Zhang Y, Heidrich N, Ampattu BJ et al (2013) Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol Cell* 50:488–503
- Zhang F, Wen Y, Guo X (2014) CRISPR-Cas9 for genome editing: progress, implications and challenges. *Hum Mol Genet* 23:R40–R46
- Zhang Z, Ge X, Luo X et al (2018a) Simultaneous editing of two copies of Gh14-3-3d confers enhanced transgene-clean plant defense against *Verticillium dahliae* in allotetraploid upland cotton. *Front Plant Sci* 9:842
- Zhang T, Zheng Q, Yi X et al (2018b) Establishing RNA virus resistance in plants by harnessing CRISPR immune system. *Plant Biotechnol J* 16:1415–1423
- Zhu LJ, Holmes BR, Aronin N, Brodsky MH (2014) CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One* 9:e108424
- Zou E, Cai YJ, Li K, Wei Y (2017) One-step generation of complete gene knockout mice and monkeys by CRISPR-Cas9-mediated gene editing with multiple sgRNAs. *Cell Res* 27:933–945