Journal of Plant Biochemistry and Biotechnology
https://doi.org/10.1007/s13562-019-00491-6

REVIEW ARTICLE

The CRISPR–Cas9, genome editing approach: a promising tool for drafting defense strategy against begomoviruses including cotton leaf curl viruses

Ajay Prakash Uniyal[1] · Sudesh Kumar Yadav[2] · Vinay Kumar[1]

Received: 29 September 2018 / Accepted: 4 February 2019

Abstract

The CRISPR–Cas9 is emerging genome editing tool and very easy and straightforward in operation that has been tested and explored for introduction of new traits in plant systems. Recently, a number of reports have documented utilization of this technology for providing tolerance against viral diseases mediated by begomoviruses. Begomoviruses infect dicot and are transmitted by white flies and cause devastating losses to yield of important agricultural crops including tomato, cassava and cotton. An overview of genomic structure of begomoviruses has been presented to understand the potential strategy for designing of effective sgRNAs to combat the viral replication for generating resistance against infection. This review provides the introduction, recent developments, and applications of the CRISPR–Cas9 system in plants and proposes a holistic methodology for generating cotton plant an example having resistance against begomoviruses. The genome editing using CRISPR–Cas9 system against complex of *begomoviruses* collectively termed as cotton leaf curl virus, which a major contributor to reduction of the cotton yield especially in Northern India and Pakistan is also discussed thoroughly. In conclusion, this potential strategy could be a sustainable approach for development of tolerant crops against diseases mediated by DNA viruses.

Keywords Cotton · Cotton leaf curl virus · Cotton leaf curl disease · CRISPR–Cas9 · Genome editing

Abbreviations

CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR associated protein
sgRNA	Single guide RNA
crRNA	CRISPR RNA
PAM	Protospacer adjacent motif
CLCuV	Cotton leaf curl virus
tracRNA	Trans activating RNA
NW	New world
OW	Old world
CLCuD	Cotton leaf curl disease

Introduction

The CRISPR–Cas9 editing tool is an emerging approach for introducing precise modification(s) in genetic material and tested in numerous organisms (Hwang et al. 2013; Bassett et al. 2013; Yang et al. 2014; Niu et al. 2014). In-fact, this system is an integral component of adaptive immune system in bacteria and archaea to offer protection against invading plasmids and viruses. The CRISPR–Cas system has been classified into I, II and III types on the basis of presence of different assembly set of Cas proteins and mode of action against invading pathogens (Gasiunas et al. 2012). The type II CRISPR system reported from *Streptococcus pyogenes*, consists of only two different RNA molecules (CRISPR RNA; crRNA and trans activating RNA; tracRNA) and Cas9 protein as minimum assembly for action against invading pathogens and has been extensively exploited for genome editing purpose.

The utilization of major components of prokaryotic nuclease system (SpCas9, SpRNase III, pre-crRNA and tracrRNA) has been reported for targeted cleavage in

✉ 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 151001, India

[2] Center of Innovative and Applied Bioprocessing (CIAB), Knowledge City, Sector-81, Mohali, Punjab, India

Published online: 20 February 2019

mammalian genome (Jinek et al. 2012). At the same time another report documented the fusion of tracrRNA and crRNA into a chimeric single guide RNA (sgRNA) that has been demonstrated to be functional in vitro and reduced the dependency from three components (SpCas9, tracrRNA and crRNA) to two components (SpCas9 and sgRNA) for precise genome editing (Cong et al. 2013). In another finding, the expression of single chimeric sgRNA was much more efficient than expression of individual crRNA and tracrRNA in transgenic organisms (Miao et al. 2013). Thus, the two-component system (Cas9 and sgRNA) is very ease of use and offers a versatile tool for creating double stranded breaks that could facilitate targeted genome editing as shown in Fig. 1(Hwang et al. 2013; Cong et al. 2013; Cho et al. 2013; Mali et al. 2013). According to mechanism of targeted genome editing, the Cas9 nuclease protein and sgRNA complex form riboprotein structure that binds specific double stranded DNA target depends upon the availability of a conserved sequence motif NGG/NAG named as protospacer adjacent motif (PAM) at 3′ downstream site of targeted genomic region. The targeted genomic sequence must be ideally complementary to the first 17–20 nucleotides of the specific sgRNA, binds by base paring and cleavages at defined genomic site. The targeting of any DNA sequence by customization of only a specific sgRNA (20 nt) is extreme ease in genome targeting and tested in numerous organisms (Jinek et al. 2012; Feng et al. 2013; Wang et al. 2013; Nekrasov et al. 2013). The CRISPR–Cas9 system also allows for targeting of multiple sites by simultaneous expression of target specific sgRNAs along with Cas9 protein (Jinek et al. 2012; Mali et al. 2013; Wang et al. 2013; Zhou et al. 2014). The feasibility of the CRISPR–Cas9 system for gene/genome editing with various objectives of research (deletion, targeted insertions and multiplex genome modifications) has been tested in plants including model species and crops (Wang et al. 2013; Zhou et al. 2014; Shan et al. 2013; Xie and Yang 2013; Jiang et al. 2014; Upadhyay et al. 2013; Wang et al. 2014; Liang et al. 2014). In-fact, an objective specific CRISPR–Cas9 system has been designed by harnessing either non-homologous end joining (NHEJ) or homologous recombination (HR) process of cellular repair mechanism. In addition, numerous issues including cleavage efficiency, off-target effects, and potential to create large chromosomal deletions with the optimization of the CRISPR–Cas9 system have also been addressed in plant systems (Feng et al. 2013; Shan et al. 2013; Jiang et al. 2014; Fauser et al. 2012; Mao et al. 2013; Sugano et al. 2014).

For the implementation of the CRISPR–Cas9 system in plant, the constructs expressing Cas9 and/or sgRNA(s) are used to deliver in plant cells via specially designed vector (s) using a range of transformation platforms including protoplast transfection and agro infiltration as complied in reviews (Belhaj et al. 2013; Bortesi and Fischer 2015; Johnson et al. 2015; Jia and Wang 2014). A number of attempts have also been carried out to optimize the appropriate expression of constructs for both Cas9 and sgRNA in plant cells (Johnson et al. 2015; Jia and Wang 2014; Mikami et al. 2015). Different variants of Cas9 (*AtCas9*, *AteCas9*, *hCas9 and DPCas9*) have been tested for cleavage efficiencies and maximum cleavage efficiency has been achieved using hCas9 variants (Johnson et al. 2015). In addition, the type of promoter, terminator sequences, translational enhancer, Cas9 codon usage, and number and localization of nuclear localization signals have also been exploited to get efficient expression of Cas9 for precise targeting and avoid off-target effects (Belhaj

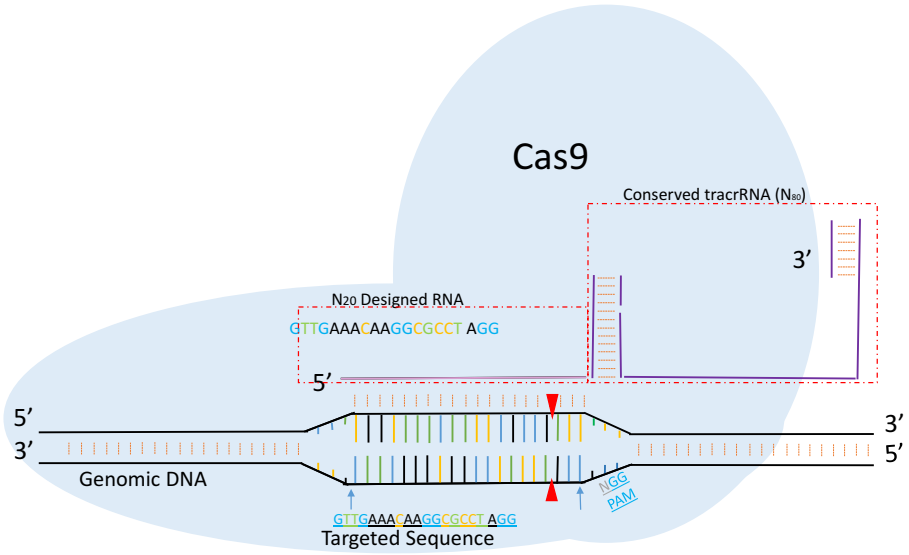


Fig. 1 The schematic representation of precise cleavage of targeted site using CRISPR–Cas9 system. The CRISPR–Cas9 system consists of two components, sgRNA and Cas9. sgRNA further consists of two component, one that is designed and shows complementary to opposite stand of targeted DNA sequence, while other is conserved and labeled as tracrRNA. The Cas9 nuclease contains two activity sites and induce cleavage at 3′ nucleotide before the PAM site on target stand

et al. 2013; Johnson et al. 2015). The outline of CRISPR Cas9 editing in plant is represented in Fig. 2.

The CRISPR–Cas9 system is most preferred genome editing system over the Zinc Finger Nucleases (ZFNs) and Transcription Activator Like Nucleases (TALENS) because it does not require any protein engineering step and is very useful for multiplexing purposes. These two later nucleases use hybrid proteins for editing purpose. Zinc finger binding protein motif is fused with FokI endonculeases to form single unit of ZFN (Carroll 2011). This single unit has binding modules around three to four in number with each module recognizing triplet nucleotides. The designing of ZFN module is little difficult because of complex interaction between amino acid residue in modules and triplet nucleotides. Similar to ZFNs, the TALE protein is fused with FokI nucleases to form single unit of TALEN (Feng et al. 2014). The construction of amino acid array and assembly of repeat sequence in TALENs makes it challenging working. In addition, user-friendly and open access policy of the CRISPR–Cas9 system has boosted the utilization of this system over ZFNs due to their heavy license fee. In addition, Johnson et al. (2015) documented the CRISPR–Cas9 system for better cleavage efficiency assays *in planta* as compared to TALENs. One more advantage of CRISPR–Cas9 system is that it can cleave methylated DNA (Hsu et al. 2013). In conclusion, all these advantages have encouraged researchers to adopt this technology for not only in-depth understanding of this system but also for their utilities for numerous applications in organisms including plants.

However, off-target effects of the CRISPR–Cas9 system are the major concern for targeting the specific DNA sequence within genome of testing organisms. One promising way to minimize off-targets is designing of specific sgRNAs that have minimum or no sequence homology to other possible sites within the genome of testing organisms. In contrast, many of the initial studies have been demonstrated that the off-target effects mediated by the CRISPR–Cas9 system occur at very low frequency in plants and require more extensive efforts to draw any conclusions about the off-target effects especially in plants (Schiffer et al. 2012). Xie and Yang (2013) proposed a general strategy for production of numerous sgRNAs with different specificity from a single polycistronic gene by exploiting endogenous t-RNA processing system and achieved efficient multiplex genome editing in transgenic rice. Thus, the CRISPR–Cas9 system will be proved as a promising tool for genetic/genomic manipulations and for diverse usages in plants.

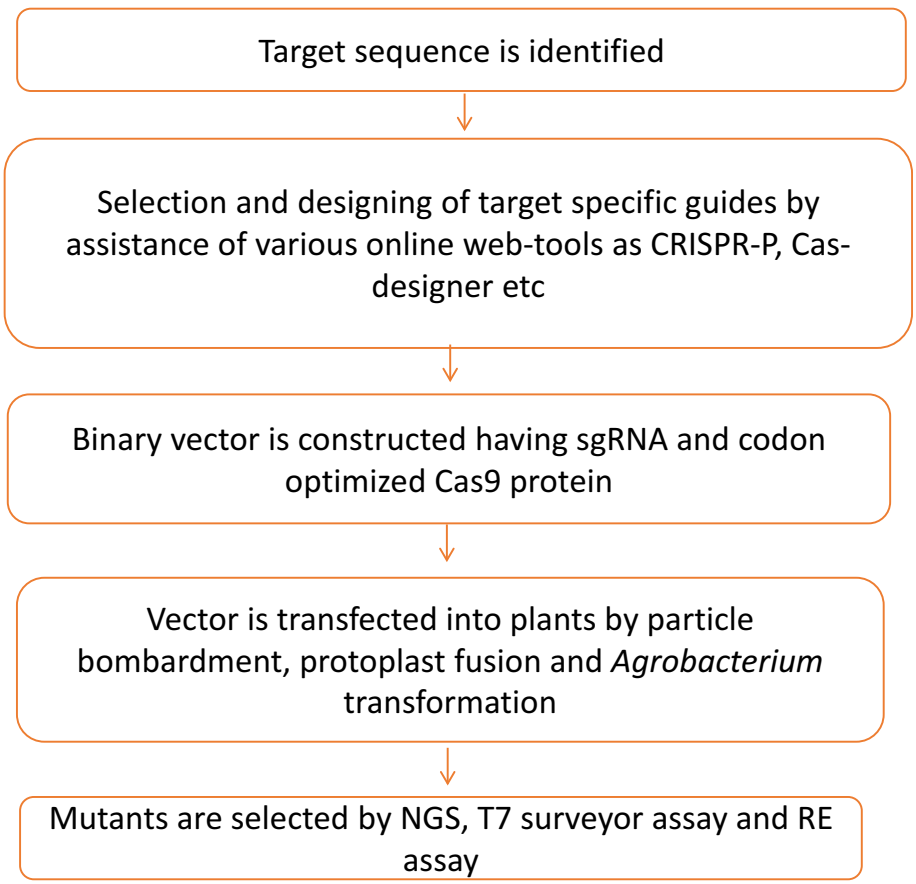


Fig. 2 The outline of CRISPR–Cas9 gene editing in plants

Besides numerous and versatile reported applications of the CRISPR–Cas9 system in functional and applied genomics of plants, this system has also been exploited for targeting viruses including human immunodeficiency virus, hepatitis B virus, epstein-barr virus and herpes simplex virus in mammals (Suenaga et al. 2014; Zhen et al. 2015; Yuen et al. 2015; Chen et al. 2014; Park et al. 2016). By using same strategy, Ali et al. (2015a, b) proposed a novel approach to boost plant immunity against geminiviruses using the CRISPR–Cas9 system. Previously, a set of artificial zinc finger proteins was used for the inhibition of geminivirus replication in transgenic plants (Sera 2015; Koshino-Kimura et al. 2009). Almost all important crops are susceptible to different viruses and damages ranges from stunted growth, reduced vigor to yield loss. In present article, an introduction of the CRISPR–Cas9 technology and its applications against development of virus resistant plants are discussed. A very basic outline for development of effective CLCuD resistant plants is also proposed. This article will broaden the scope of the CRISPR–Cas9 technology and offer a strategy to provide resistance to diseases mediated by begomoviruses including CLCuD.

Begomoviruses diseases: a major cause of yield loss in crops

Many diverse types of viruses considered as member of *Geminiviridae* family are characterized by twin icosahedral capsids with either one or two single stranded DNA genome(s) (2.5 to 3×10^3nucleotides). Zerbini et al. (2017) has classified the Geminiviruses into 9 genera on the basis of genome structure, types of insect vector and host plants. The genera include *Becurtovirus*, *Begomovirus*, *Capulavirus*, *Curtovirus*, *Eragovirus*, *Grablovirus*, *Mastrevirus*, *Topocuvirus, Turncurtovirus* and some species which are unassigned to any genus (Zebrini et al. 2017). Among these genera, Begomoviruses has large number of species that are transmitted by the whitefly *Bemisia tabaci* and responsible for heavy yield loss for many agricultural crops worldwide. Around 388 accepted virus species have been enlisted under genus begomoviruses. These species are broadly distinguished depending upon the number of genome copy into either bipartite or monopartite begomoviruses.

The begomoviruses infect dicot monocot plants. The transmission of virus occurs via insects including whiteflies. The transmission via grafting and mechanical meaning has been reported. The crumpling, curling and distortion of leaf, interveinal yellowing, yellow spot and swelled vein have been reported. In 2003, Mansoor et al. (2003) reported the influence of begomoviruses on the production of economically important crops such as tomato, cassava, legumes, peppers, cucurbits and cotton. A report published by Varma and Malathi in the same year, documented the losses of several billion dollars due to begomoviruses mediated diseases including cassava mosaic, cotton and tomato leaf curl and bean golden mosaic (Varma and Malathi 2003).

Begomoviruses mediated cotton leaf curl disease, a major threat for cotton production

Cotton (*Gossypium* spp.) plant is the richest source of fiber and an economically important crop. The cotton cultivation also contributes to food security and improves standard of living in rural regions of developing countries. Cotton is cultivated in more than 100 countries with India as world's largest country for cotton production as published in OECD/FAO (2018). However, the estimated potential of cotton yield especially in India is still far away from actual yield.

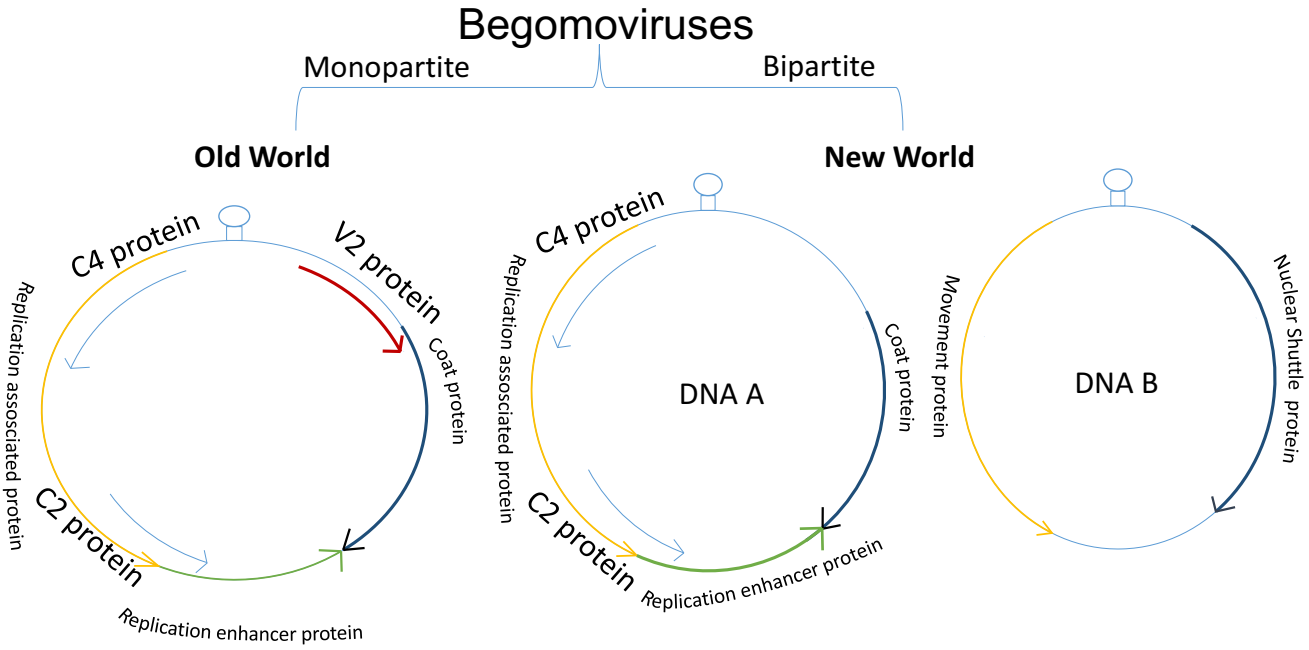
Apart from adopting the best management practices, biological interventions are majorly responsible for losses in cotton yield in major growing regions. The cotton leaf curl disease (CLCuD) has been a constant threat to cotton production in Indian sub continent during the last one and half decade and caused by a diverse species of begomoviruses including *cotton leaf curl Alabad virus* (*CLCuAV*), *cotton leaf curl Rajasthan virus* (*CLCuRV*), *cotton leaf curl Bangalore virus* (*CLCuBV*), *cotton leaf curl Burewala virus* (*CLCuBuV*), *cotton leaf curl Kokhram virus* (*CLCuKV*), *cotton leaf curl Multan virus* (*CLCuMV*) and *Papaya leaf curl virus* (*PaLCuV*). All begoviruses associated with CLCuD are collectively referred as *CLCuV* in this manuscript. Recently, white fly and *CLCuV* mediated CLCuD were also reported for potential cotton yield loss in Northern India. The CLCuD is characterized with a range of symptoms including vein swelling, and formation of leaf-like cup shaped outgrowths (enations) on the undersides of leaves with stunted cotton plant growth (Ahuja et al. 2007; Ahmad et al. 2011; Ali et al. 2013; Ji et al. 2015; Heigwer et al. 2016; Qazi et al. 2007). The seedling stage of the plant is also most severely affected by *CLCuV* with reduced flowering time, and subsequently decreased size and number of cotton bolls with negative effects on seed yield and fiber quality traits (Ahuja et al. 2007; Ahmad et al. 2011; Ali et al. 2013; Ji et al. 2015; Heigwer et al. 2016; Qazi et al. 2007). Factors such as temperature, rainfall, wind, relative humidity, plant age and vector population affect the chances of occurrence and spreading of CLCuD in cotton growing regions (Farooq et al. 2011). The CLCuD occurs in northern India, Pakistan and Africa due to poor management practices and suitable conditions for spread of transmitting vector as of

CLCuV. In last three decades, India and Pakistan have encountered major outbreaks during 1991–1996 and 2001–2010. A new variant of *CLCuV* that has emerged during recent outbreak (2009–2010), overcomes the tolerance of existing resistant that were resistant varieties against previous outbreak (Qazi et al. 2007; Farooq et al. 2011). The genetic material of diverse species of begomovirus associated with CLCuD are characterized by circular single stranded approx. 2800 nt in size, encodes six genes in bidirectional way with a non-coding intergenic region (IR). In addition, occurrence and association of DNA satellites (beta and alpha) with begomoviruses already establish their role in CLCuD. These reported DNA satellites depend upon associated begomovirus for replication and encapsidation, spreading in plant tissue as well in transmission by vector. The betasatellites contain half unit sized circular ssDNA encoding βC1 that participates in host defense suppression for enhancing the degree of virulence of pathogen resulting into systematic infection. The other DNA satellites (alpha) also contain similar sized ssDNA with a conserved genome organization with an origin of replication that encodes for a replication-associated protein. Both satellites share almost negligible sequence similarity except a specific A-rich (Adenine-rich) region. Apart from full sized betasatellites and alphasatellites, other defective vestigial satellites were also identified, however their role is not known till today. In conclusion, circular DNA of *CLCuV* supported with DNA satellites cause disparaging disease CLCuD.

Begomoviruses and their genomic structure

Begomoviruses are divided on the basis of their geographic origins into New World and Old World. Begomoviruses are also divided into two groups (bipartite and monopartite) on the basis of genome organization. All begomoviruses form New World are bipartite, while Old World begomoviruses contains either a bipartite or monopartite genome. The bipartite begomoviruses contains two genome copies designated as DNA-A and DNA-B of 2.6 kb each and both are required for efficient infection. It has been confirmed that component A encodes the entire requirement for replication and encapsidation. Monopartite begomoviruses contain single genome compartment (approx. 2.9 kb) associated with satellites DNAs and originated in the Old World. The genome structure of monopartite begomoviruses and DNA-A component of bipartite begomoviruses are almost similar and have a similar organization of six coding sequences, among them two (*CP*, and *AV2*/*V2*) on virion sense strand and other four coding sequences (*AC1*/*C1*, *AC2*/*C2*, *AC3*/*C3* and *AC4*/*C4*) on complementary sense strand (Nawaz-ul-Rehman and Fauquet 2009). The *CP* encodes coat protein for viral capsid and mediates vector transmission. In monopartite begomoviruses, CP protein also function as the nuclear shuttle protein. All monopartite begomoviruses also encode small open reading frame upstream of *CP*. The *AV2/V2* gene encodes anti defense protein to avoid the host defense mechanism (Hanley et al. Hanley-Bowdoin et al. 2013). In monopartite viruses, *V2* encoded protein also found to be involved in virus movement activity. The viral replication is initiated by replication initiator protein encoded by Rep (*AC1/C1*). The transcriptional activator protein encoded by *TrAP* (*AC2/C2*) in bipartite begomoviruses interfere with TGS and PTGS and also required for CP and NSP expression. The replication enhancer protein (REn) by *AC3/C3* is involved in viral replication in both monopartite and bipartite begomoviruses. The *AC4*/*C4* in bipartite begomoviruses is still uncharacterized. In addition, component B is also required for systemic movement and symptom production. The component B contains two coding sequences, one *BV1* on the virus sense strand and *BC1* on the complementary sense strand. The satellites DNA is either necessary for appearance of disease symptoms (beta) or no apparent or modulate symptoms (alpha). Betasatellites are well-known for pathogenicity associated with many diseases caused by monopartite begomoviruses. In continuation, this is completely dependent on genome compartment for their replication. The association of alpha-satellites with monopartite genome has also been documented in many begomoviruses mediated diseases involving cotton leaf curl disease (India and Pakistan), tomato yellow leaf curl disease (China) and ageratum yellow vein disease (South East Asia). The differences between old world and new world begomoviruses is represented in Fig. 3.

The begomoviruses genome contains divergent coding sequences, and a 5′ intergenic region containing the origin for rolling-circle replication and two RNA polymerase II promoters (Hanley et al. 1999). The opposite transcriptional units of monopartite and bipartite genomes are separated by a highly conserved intergenic region (IR) of about 200 nt. This IR includes a stem-loop structure containing an invariant nonanucleotide (replication initiation site) and conserved iterated sequences. Similarly, alphasatellites genomic organization consist a gene that encodes replication associated protein (Rep), an adenine rich region and origin of replication with conserved nonanucleotide sequence. The betasatellites encode a protein (bC1), and contain an adenine rich region (240 nt) as well as a satellite-conserved region (220 nt) that is highly conserved among betasatellites. In conclusion, rolling circle mechanism with the presence of stem-loop with a nonanucleotide origin of replication (TAATATT/AC for virus and betasatellites or TAGTATT/AC for alphasatellites) are



- Mostly, monopartite; some are bipartite
- Require satellite DNA for symptom induction (betasatellite and alphasatellite)
- Contain V2 gene
- Lack DNA B component

- All new world begomoviruses are bipartite
- No requirement of satellite DNA
- Lack V2 gene
- Contain DNA B component that encoded Nuclear Shuttle protein (NSP) & Movement Protein (MP)

Fig. 3 Diagramatic representation of difference between old world and new world begomoviruses

unique features of begomoviruses and associated satellites. In addition, begomoviruses are highly conserved in terms of genetic architecture, localization, and length of individual genes at specific locus with their respective function. The bipartite begomoviruses from NW and OW have similar genome organization except of *AV2*.

Development of begomoviruses resistant plants using genome editing

Plant DNA viruses belonging to geminiviridae family, cause destructive diseases in crop plants and a potential threat to food security and sustainable agriculture. A wide range of strategies has been employed to control geminiviruses infection including mechanical and chemical approaches to check the population of transmitting insects (*Bemisia tabaci* etc.), and development and deployment of resistant cultivars using the conventional breeding approach. The genetic engineering approach involving RNA interference, expression of viral proteins (mutated and truncated) and peptide adaptor binds to viral proteins, and genome editing approach using target specific sgRNAs have been extensively exploited for optimization and establishment of the viral interference in plants (Ahuja et al. 2007; Ahmad et al. 2011; Ali et al. 2013; Koshino-Kimura et al. 2009; Ji et al. 2015).

The replication mechanism of geminiviruses is well conserved and occurs via rolling-circle amplification mechanism or by recombination mediated replication via the host machinery for their activity. Once virus inside the plant cell, the encoded viral protein(s) bind alone or with other protein(s) to origin of replication and initiate replication cycle followed by assembly, and production of its own kind and eventually spread of infection. Among geminiviruses, *Beet severe curly top virus* (*BSCTV*), is a model DNA virus, has wide host range of dicotyledonous plants and encodes only one viral replication protein (Rep) to initiate replication process. The origin of replication site that binds with Rep protein, was targeted using genome editing approach (artificial zinc finger, AZP) for preventing *BSCTV* infection in vivo (Sera 2015). Transgenic Arabidopsis overexpressing AZP also showed improved resistance against *BSCTV* infection (Sera 2015). Similarly, an artificial zinc finger protein has been designed to target origin of replication site of the *Tomato yellow leaf curl virus* (*TYLCV*) and overexpressed in transgenic tomato plants (Koshino-Kimura et al. 2009). Recently, it was

documented that artificial zinc finger proteins designed against conserved viral sequence worked efficiently and inhibited the replication of different begomoviruses using transient expression assay (Heigwer et al. 2016).

The utilization of delivery reagent based upon *Bean yellow dwarf virus* (*BeYDV*) in tobacco has been reported with increased gene targeting frequencies of the CRISPR–Cas9 system (Baltes et al. 2014). A transient assay has also been optimized for targeting the inhibition of replication of *BSCTV* in *Nicotiana benthamiana* using the CRISPR–Cas9 system (Ji et al. 2015). The transgenic plants over expressing sgRNA-Cas9 were reported with improved resistant to disease mediated by *BSCTV* (Heigwer et al. 2016). The significant reduction of disease symptoms in infected plant against three viruses (*TYLCV*, *Beet curly top virus* (*BCTV*), and *Merremia mosaic virus* (*MeMV*), simultaneously have also been documented by using the CRISPR–cas9 system approach (Ali et al. 2015a, b). Similarly, cumulative reduction of viral copy number was also achieved with overexpression of two different designed sgRNAs in the same infected plant (Ali et al. 2015a, b). The highly conserved IR, present in the genome of all geminiviruses could be best target for designing of effective sgRNAs for viral interference using the CRISPR–Cas9 system (Zhen et al. 2015; Yuen et al. 2015; Chen et al. 2014; Park et al. 2016; Ali et al. 2015a, b). Thus, the utilization of CRISPR–Cas9 technology is very promising approach to engineer plants resistant against geminiviruses mediated diseases including CLCuD. The comparative details of all reports related to usage of genome editing against Geminivirus have been provided in Table 1.

Potential promising approach against CLCuD

The CLCuD is easily transmitted from infected to healthy plants by whitefly in a hasty manner and also spreads to new areas where conditions are more favorable using incidental transport of asymptomatic plants harboring the virus. Both the CLCuD and whitefly mainly survive on actual hosts [cultivated cotton, vegetables, tomato etc. (for CLCuD); ornamentals, hibiscus etc. (for whitefly)]. The reoccurrence of the CLCuD during the next season generally occur via transmitting vector, whitefly. The possible management of CLCuD is control of the population of transmitting vector, usage of effective approach for management of alternative hosts and through development of resistant hybrids/varieties. In addition, other promising strategies include variation in sowing timing, exploring buffer crops and proper seed treatment to prevent infection during seedling stage of cotton plants.

The native cotton, *Gossypium arboretum*, is not susceptible to the CLCuD and corresponding virus resistant genetic resources have been introgressed into *G. hirsutum* and tolerance has been improved in the F1 cotton hybrids (Ahmad et al. 2011). Similar strategy has also been documented for improved fibre strength of cotton collected form *CLCuV* infected plants (Nazeer et al. 2014). In addition, cotton germplasm lines have also been screened for tolerance against CLCuD (Ahuja et al. 2007). However, development and deployment of CLCuD resistant cultivar is difficult due to strong linkage of fruit poor quality trait with resistance locus. Besides conventional breeding approach, new emerging technologies like antisense RNA, and RNA interference (RNAi) have also been exploited for providing resistance against CLCuD (Amudha et al. 2011; Sohrab et al. 2014). The anti-sense RNA expression against

Table 1 The CRISPR–Cas9 genome editing tool and their utilization for increased immunity in plants against gemini viruses

Gemini plant viruses	Tested plants	Strategies	Targets	Results	Symptom assayed	References
Tomato yellow leaf curl virus (TYLCV), *Beet curly top virus* (BCTV) and *Merremia mosaic virus* (MeMV)	Tobacco	CRISPR–Cas9	Viral intergenic region	Replication suppressed	Yes	Ali et al. (2015a, b) and Koshino -Kimura et al. (2009)
Beet severe curly top virus (BSCTV)	Arabidopsis and Tobacco	CRISPR–Cas9	43 candidate sites [A7 (Rep), B7 (CP) and C3 (rep enhancer protein) show best result] covering 2.9 kb genome of BSCTV	Inhibit virus accumulation	Yes	Ji et al. (2015)
Bean yellow dwarf virus (BeYDV)	Tobacco	CRISPR–Cas9	RBS, hairpin, nonanucleotide sequence and three rep motifs essential for rolling circle replication (motif I, II and III)	Replication suppressed	Yes	Heigwer et al. (2016)

CLCuV DNA A specific gene has been tested and documented with improved tolerance against CLCuD (Amudha et al. 2011). With similar approach, Sohrab et al. (2014) documented transgenic cotton plants generated using antisense βC1 gene with improved tolerance against CLCuD. Artificial micro RNA technology has also been exploited to generate resistant transgenic cotton against CLCuD (Ali et al. 2013). However, no attention has been paid for the exploitation of associated DNA satellites for drafting a promising defense strategy for prevention of CLCuD. Thus, there is urgent requirement for an approach in which genomic DNA and DNA satellites could be simultaneously targeted for providing tolerance against CLCuD.

The genome editing approach has also been implemented for the inhibition of viral replication in infected cotton crop (Sera 2015; Koshino-Kimura et al. 2009; Ali et al. 2015a, b). The genome editing is a precise and efficient approach for manipulation of targeted DNA sequences from genetic material of targeted organisms and has been tested in numerous species including plants. The genome editing process occurs in two successive stages: introduction of cleavage at targeted site(s) of interest and harnessing the cellular DNA repair mechanisms for either introduction of mutations or replacement of sequence (s) with others. Two early-developed genome-editing tools, zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been exploited for inhibition of DNA viral replication in plants (Sera 2015; Koshino-Kimura et al. 2009; Mali et al. 2013; Cheng et al. 2015; Yin et al. 2015). The suppression of replication activity of selected viruses including *Beet severe curly top virus* (*BSCTV*), *Tomato Yellow leaf curl virus* (*TYLCV*) and *Tobacco curly shoot virus* (*TCSV*) has been documented using ZFNs (in *BSCTV*, *TYLCV* and *TCSV*) and TALENS (in *TCSV*) genome editing approaches (Sera 2015; Mali et al. 2013; Cheng et al. 2015; Yin et al. 2015). However, the designing of ZFNs and TALENs against a specific target is a time-consuming process. In contrast, the recently developed clustered regularly interspaced palindromic repeat (CRISPR)/CRISPR associated (Cas9) system, is easy to use and tested in many notable organisms including plants. This system requires only target specific 18–22 nucleotide long sequence with Cas9 nuclease for effective and precise targeting of sequence of interest. This technology has been demonstrated for numerous applications in plants including precise genetic/genomic modifications, activation and inactivation of single/multiple genes and identification of specific gene/genomic location. However, reports regarding the usage of this technology for development of virus resistant crops are very limited (Ali et al. 2015a, b; Ji et al. 2015). Recently, Gao et al. (2017) developed the method for fast validation of sgRNAs for CRISPR Cas9 editing in cotton experimentally in around 3 days. They used pYLCRISPR/Cas9 multiplex binary vector for targeted genome editing. The construct of vector is then transferred into *Agrobacterium* for transient and stable transformation. Genomic DNA is isolated and PCR analysis is used for amplifying genomic target and confirming mutation. Gao et al. (2017) has found that the transient transformation method allow fast sgRNA target validation in combination of emerging CRISPR viral methods in cotton. But there is no promising report of usage of the CRISPR–Cas9 with detailed methodology for drafting strategy against CLCuD.

A systematic defense strategy for utilization of the CRISPR–Cas9 technology against the CLCuD

The methodology involving a series of steps for harnessing the CRISPR–Cas9 technology against CLCuD is shown in Fig. 3. The first major step proposes a test whether the CRISPR–Cas9 system can be imported to plants to offer molecular tolerance against CLCuD or not. Subsequently, the molecular basis of *CLCuV* mediated infection, susceptibility and host defense mechanisms will be investigated for identification of possible revertants. The sequence analysis of curated revertants will be used for designing of effective sgRNAs against the potential genomic targets of *CLCuV* and further exploited for generation of transgenic cotton plants overexpressing Cas9 endonuclease and potential sgRNA(s) for providing tolerance against CuCLD as described in step three (Fig. 4).

Optimization of *CLCuV* interference in plants

The optimization of the CRISPR–Cas9 system for viral interference should be carried out using *Arabidopsis* and *Nicotiana* plant species. The shorter life span, optimized steps (agro infiltration and better regeneration) in these selected plant species could accelerate the process of development of resistant plants for better understanding of *CLCuV* interference. The selection of specific Cas9 protein for efficient nuclease activity and target specific sgRNAs is main requirement for optimization process. The plant codon optimized Cas9 protein has been exhibited higher expression than native Cas9 in *Arabidopsis* (Yin et al. 2015). In addition, selection of suitable promoter preferably ubiquitin will be helped to drive better expression and designing of construct in binary vector of interest (Ali et al. 2015a, b). The binary vector containing designed construct (s) should be transformed in *Arabidopsis* and *Nicotiana* via *Agrobacterium* mediated transformation protocols. The stable Cas9 overexpressing transgenic lines *(Nicotiana* and *Arabidopsis*) should be screened by using a specific reporter system. The confirmed transgenic lines will be

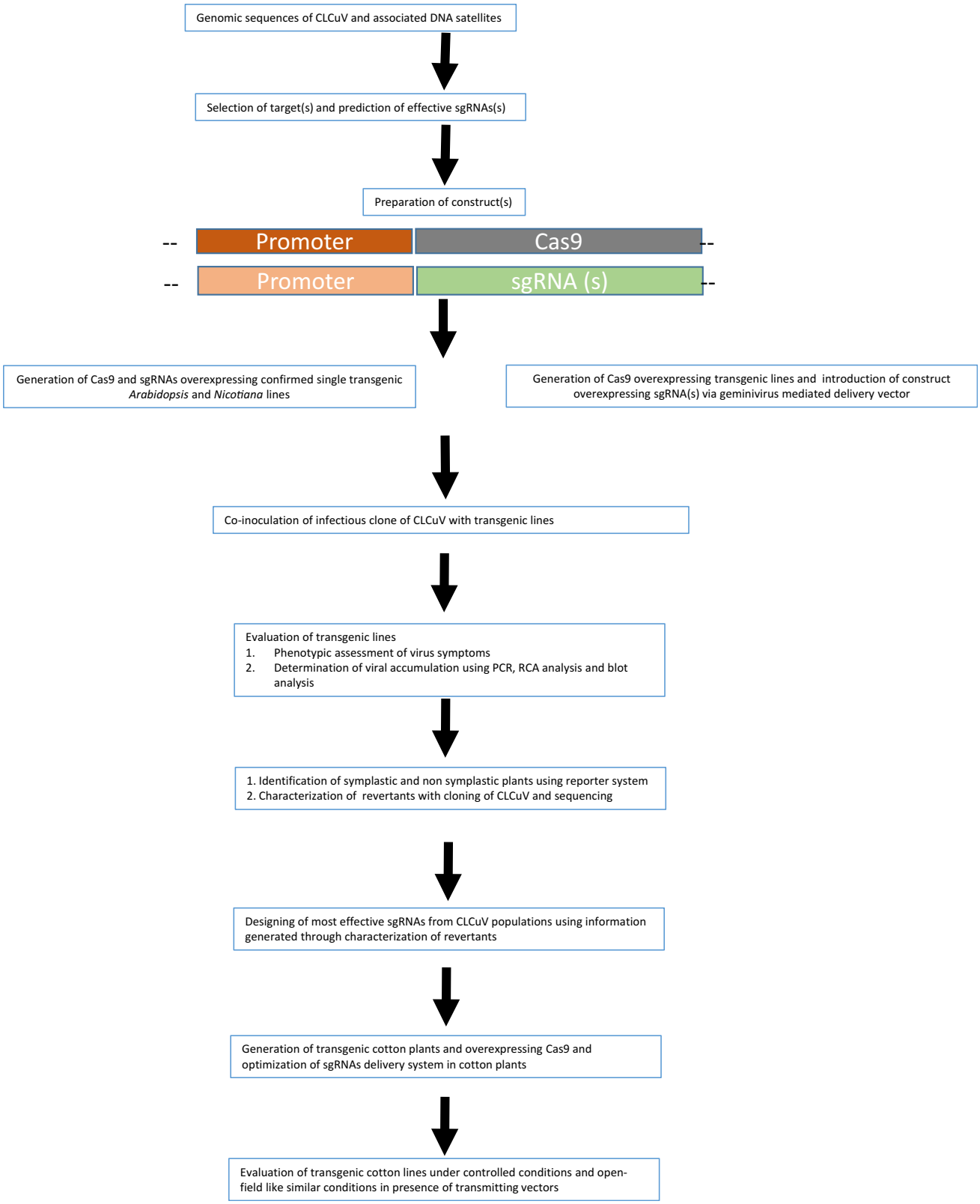


Fig. 4 The outline of defense strategy to produce viral interferene against Gemini viruses in tobacco plants

subjected for delivery of promising sgRNAs against selected genomic sites of *CLCuV* as described by Ali et al. (2015a, b).

For the prediction of best possible sgRNAs, multiple sgRNAs should be required to be generated against numerous genetic sequences including open reading frame, IR and Rep region of *CLCuV*. These targets are mainly responsible for viral replication and transcription, and could be used in propose strategy for significant reduction of CLCuD symptoms. A number of reports have been published with concept of computational modeling to derive a set of rules for prediction of more effective sgRNAs against a specific genomic region of interest (Yin et al. 2015; Cheng et al. 2015). From publically available databases, the genomic sequences of all begomoviruses and DNA satellites associated with CLCuD should be retrieved and utilized for effective sgRNAs prediction. The delivery of construct corresponding to potential targets specific sgRNAs should be carried out using effective and efficient geminivirus-based sgRNA delivery system as published earlier (Ali et al. 2015a, b; Mori et al. 2013). This strategy could reduce the tremendous time and labor involved in generating transgenic plants overexpressing sgRNAs for viral interference. For this strategy, selected sgRNAs with specific promoter U6 to derive expression should be designed and needed to insert into geminivirus-based sgRNA delivery system. Similarly another vector containing U6 promoter with scaffold that should be used as a negative control.

To access the degree of severity of CLCuD under the influence of method of introduction of specific sgRNAs, two different strategies have been adopted for employing inoculation of infectious clone. The first approach involves the systematic delivery of multiple target specific designed sgRNAs via genimivirus based delivery system and subsequently allowing their expression with second inoculation of infectious clone(s) of *CLCuV*. In second approach, co-delivering of the multiple sgRNAs and *CLCuV* infectious clone (s) should be achieved simultaneously and subjected for evaluation of the efficacy of the CRISPR–Cas9 against degree of disease severity of CLCUD via molecular and phenotypic analysis. The virus infectivity analysis could be achieved with the quantification of viral titer using semi-quantitative PCR with primer encompassing the specific genomic viral regions of interest. Subsequently, RCA (rolling circle amplification) analysis could be used effectively for testing whether targeting a specific region interference with virus accumulation or not. In addition, dot blot analysis could be also used to monitor status of interference with viral replication in transgenic plants compared with positive and negative controls. The amount of accumulation of viral genome DNA after precise targeting with relation to control could depict the efficacy of viral interferes in model plant species. The comprehensive phenotypic assessment of infected plants could also be used to correlate with molecular findings for valid conclusions. The multiplexing of potential sgRNAs should also be used to check the additive effects of multiple sgRNAs on reduction of amount of viral load and CLCuD symptoms.

Isolation and characterization of revertant forms of *CLCuV*

The *CLCuV* interference mediated by the CRISPR–Ca9 system is dependent on targeted genomic nucleotide sequences that are generally prone to mutation. Thus, the determination of interference breaking in term of mutant version(s) of *CLCuV* in transgenic lines should be investigated for determination of potential effective sgRNAs. The identification of potential relationship of different viral genomic components with degree of severity of CLCuD, could be facilitated for designing of construct with effective sgRNAs fused with potential reporter gene in a binary vector and used to generate transgenic lines overexpressing simultaneously Cas9 and sgRNAs. The stable transgenic lines should be used for agro-inoculation with infectious clone of *CLCuV* and subjected for screening fluorescence. The confirmed lines could be either symptomatic and non-symptomatic lines or both that should be exploited for cloning of isolated *CLCuV*. The sequencing of cloned viral populations should be facilitated for identification of potential revertants. In conclusion, this strategy could act as a key regulatory step for designing of an effective sgRNA (s) for targeting novel viral or relevant sequences of *CLCuV* and facilitate to study the molecular basis of virus infection, susceptibility and host defenses.

The generation of cotton crop against *CLCuV* mediated by the CRSIPR-Cas9 system

The optimization of *CLCuV* interference in model plant species and subsequently identification of potential sgRNAs targets even against revertants will be exploited to generate transgenic cotton plants by targeting simultaneously all begomoviruses associated with CLCuD. Developing CLCuD resistance cotton under open-field conditions is a major challenge due to occurrence of natural mixed viral infections compared to controlled laboratory conditions and encourages to draft a combinatorial multiplexing approach. A collection of most efficient sgRNAs targeting multiples viruses should be used to deliver into cotton plants overexpressing Cas9 protein via an efficient delivery system. This step need to be optimized in cotton plants by using numerous delivery system (tested in other plant species) as reported till today. Similarly, the generation of transgenic cotton lines via *Agrobacterium*-mediated transformation are also a challenging task. The transgenic lines will be tested for resistance against CLCuD using agro-inoculation with infectious clone and confirmed by

molecular and phenotypic analysis. The significant reduction in viral load and disease symptoms will be speculated for improved tolerance against CLCuD. The laboratory tested transgenic cotton plants with further need to test in open field viral environment mimicking the natural conditions for assessment of efficacy of the CRISPR–Cas9 system against CLCuD.

Conclusions

Cotton leaf curl disease (CLCuD), mediated by *Cotton leaf curl virus* (*CLCuV*), is a key limiting factor for cotton productivity in major cotton growing regions, worldwide. The CRISPR–Cas9 technology could be harnessed for targeted *CLCuV* interference to provide resistance against CLCuD. The expression of effective designed sgRNAs complementary to coding and non-coding genomic regions of *CLCuV* into Cas9 overexpressing cotton transgenic plants will improve tolerance and spread of CLCuD. This approach shall provide new avenues for engineering cotton and other crop against DNA viruses. The strategy could be exploited for editing of genome of other geminivirdae viruses including *Cabbage leaf curl virus* (*CLCuV*) and offers an opportunity for sustainable agriculture by providing a remedy against diseases caused by the devastating DNA viruses.

Acknowledgements Authors thank the anonymous reviewers for their critical comments and suggestions. The authors also would like to express deepest gratitude to Vice Chancellor of Central University of Punjab, India for providing unconditional supports during course of this work. This research is sponsored by "Research Seed Money" grant as sanctioned by Central University of Punjab, India.

Author's contributions VK and SKY conceived and designed present work. VK and SKY analyzed data. APU and VK wrote the manuscript. All authors read and approved the manuscript.

Compliance with ethical standards

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

References

Ahmad S, Mahmood K, Hanif M, Nazeer W, Malik W, Qayyum A et al (2011) Introgression of cotton leaf curl virus-resistant genes from Asiatic cotton (*Gossypium arboretum*) into upland cotton (*G. hirsutum*). Genet Mol Res 10:2404–2414

Ahuja SL, Monga D, Dhayal LS (2007) Genetics of resistance to cotton leaf curl disease in *Gossypiumhirsutum* L. under field conditions. J Hered 98:79–83

Ali I, Amin I, Briddon RW, Mansoor S (2013) Artificial microRNA-mediated resistance against the monopartite begomovirus Cotton leaf curl Burewala virus. Virol J 10:231

Ali Z, Abul-faraj A, Idris A, Ali S, Tashkandi M, Mahfouz M (2015a) CRISPR/Cas9-mediated viral interference in plants. Genome Biol 16:238

Ali Z, Abul-Faraj A, Li L, Ghosh N, Piatek M, Mahjoub A et al (2015b) Efficient virus-mediated genome editing in plants using the CRISPR/Cas9 system. Mol Plant. https://doi.org/10.1016/j.molp.2015.02.011

Amudha J, Balasubrami G, Malathi VG, Monga D, Kranthi KR (2011) Cotton leaf curl virus resistance transgenic with antisense coat protein gene (AV1). Curr Sci 101:300–307

Baltes NJ, Gil-Humanes J, Cermak T, Atkins PA, Voytas DF (2014) DNA replicons for plant genome engineering. Plant Cell 26:151–163

Bassett AR, Tibbit C, Ponting CP, Liu JL (2013) Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep 4:220–228

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013) Plant genome editng made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods 9:39

Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol Adv 33:41–52

Caroll D (2011) Genome engineering with zinc-finger nucleases. Genetics 188:773–782

Chen W, Qian Y, Wu X, Sun Y, Wu X, Cheng X (2014) Inhibiting replication of begomoviruses using artificial zinc finger nucleases that target viral-conserved nucleotide motif. Virus Genes 48:494–501

Cheng X, Li F, Cai J, Chen W, Zhao N, Sun Y et al (2015) Artificial TALE as a convenient protein platform for engineering board spectrum resistance to begoviruses. Viruses 7:4772–4782

Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. NatBiotechnol 31:230–232

Cong L, Ran FA, Cox D, Lin S, Barreto R, Habib N et al (2013) Multiplex genome engineering using CRISPR/Cas system. Science 339:819–823

Farooq A, Farooq J, Mahmood A, Shakeel A, Rehman A, Batool A et al (2011) An overview of cotton leaf curl virus disease (CLCuD) a serious threat to cotton productivity. Aust J Crop Sci 13:182301831

Fauser F, Roth N, Pacher M, Ilg G, Sanchez-Fernandez R, Biesgen C et al (2012) In planta gene targeting. Proc Natl Acad Sci USA 109:7535–7540

Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P et al (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res 23:1229–1232

Feng Y, Zhang S, Huang X (2014) A robust TALENS system for highly efficient mammalian genome editing. Sci Rep 4:3632

Gao W, Long L, Tian X et al (2017) Genome editing in cotton with the CRISPR/Cas9 system. Front Plant Sci 8:1364

Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci 109:E2579–E2586

Hanley-Bowdoin L, Settlage S, Orozco BM, Nagar S, Robertson D (1999) Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. Crit Rev Plant Sci 18:71–106

Hanley-Bowdoin L, Bejarano ER, RobertsonD MS (2013) Geminiviruses: masters at redirecting and reprogramming plant processes. Nat Rev Microbiol 11:777–788

Heigwer F, Zhan T, Breinig M, Winter J, Brűgermann D, Leible S, Boutros M (2016) CRISPR library designer (CLD): software for multiplexing design of single guide RNA libraries. Genome Biol 24:55

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agalwala V (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. NatBiotechnol 31:827–832

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD et al (2013) Efficient genome editing in zebrafish using a CRISPR–Cas system. Nat Biotechnol 31:227–229

Ji X, Zhang H, Zhang Y, Wang Y, Gao C (2015) Establishing a CRISPR–Cas-like immune system conferring DNA virus resistance in plants. Nat Plants 1:15144

Jia H, Wang N (2014) Targeted genome editing of sweet orange using Cas9/sgRNA. PLoS One 9:e93806

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2014) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41:e188

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821

Johnson RA, Gurevich V, Filler S, Samach A, Levy A (2015) Comparative assessments of CRISPR-Casnucleases's cleavage efficiency in planta. Plant Mol Biol 87:143–156

Koshino-Kimura Y, Takenaka K, Domoto F, Ohashi M, Miyazaki T, Aoyama Y et al (2009) Construction of plants resistant to TYLCV by using artificial zinc-finger proteins. Nucleic Acids Symp Ser (Oxf) 53:281–282

Liang Z, Zhang K, Chen K, Gao C (2014) Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. J Genet Genom 41:63–68

Mali P, Yang L, Esvelt KM, Aach J, Guell M, Dicarlo JE et al (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826

Mansoor S, Briddon RW, Zafar Y, Stanley J (2003) Geminivirus disease complexes: an emerging threat. Trends Plant Sci 8:128–134

Mao Y, Zhang H, Xu N, Zhang B, Gou F, Zhu JK (2013) Application of the CRISPR–Cas system for efficient genome engineering in plants. Mol Plant 6:2008–2011

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

Mikami M, Toki S, Endo M (2015) Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. Plant Mol Biol 88:561–572

Mori T, Takenaka K, Domoto F, Aoyama Y, Sera T (2013) Inhibition of binding of tomato yellow leaf curl virus rep to its replication origin by artificial zinc-finger protein. Mol Biotechnol 54:198–203

Nawaz-ul-Rehman MS, Fauquet CM (2009) Evolution of geminiviruses and their satellites. FEBS Lett 583:1825–1832

Nazeer W, Tipu AL, Ahmad S, Mahmood K, Mahmood A, Zhou B (2014) Evaluation of cotton leaf curl virus resistance in BC1, BC2, and BC3 progenies from an interspecific cross between *Gossypiumarboreum* and *Gossypiumhirsutum*. PLOS One. https://doi.org/10.1371/journal.pone.0111861

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. Nat Biotechnol 31:691–693

Niu Y, Shen B, Cui Y, Chen Y, Wang J, Wang L et al (2014) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell 156:836–843

OECD/FAO (2018) OECD-FAO agricultural outlook 2018–2027. OECD Publishing, Food and Agriculture Organization of the United Nations, Paris, Rome

Park J, Kin JS, Bae S (2016) Cas-Database: web-based genome-wide guide RNA library design for gene knockout screens using CRISPR–Cas9. Bioinformatics. https://doi.org/10.1093/bioinformatics/btw103

Qazi J, Amin I, Mansoor S, Iqbal MJ, Briddon RW (2007) Contribution of the satellite encoded gene βC1 to cotton leaf curl disease symptoms. Virus Res 128:135–139

Schiffer JT, Aubert M, Weber ND, Mintzer E, Stone D, Jerome KR (2012) Targeted DNA mutagenesis for the cure of chronic viral infections. J Virol 86:8920–8936

Sera T (2015) Inhibition of virus DNA replication by artificial zinc finger proteins. J Virol 79:2614–2619

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z et al (2013) Targeted genome modification of crop plants using a CRISPR–Cas system. Nat Biotechnol 31:686–688

Sohrab SS, Kamal MA, Llah A, Husen A, Bhattacharya PS, Rana D (2014) Development of cotton leaf curl virus resistant transgenic cotton using antisense βC1. Saudi J Biol Sci 23:358–362

Suenaga T, Kohyama M, Hirayasu K, Arase H (2014) Engineering large viral DNA genomes using the CRISPR–Cas9 system. Microbiol Immunol 58:513–522

Sugano SS, Shirakawa M, Takagi J, Matsuda Y, Shimada T, Hara-Nishimura I et al (2014) CRISPR/Cas9-mediated targeted mutagenesis in the liverwort *Marchantia polymorpha* L. Plant Cell Physiol 55:475–481

Upadhyay SK, Kumar J, Alok A, Tuli R (2013) RNA-guided genome editing for target gene mutations in wheat. G3 (Bethesda) 3:2233–2238

Varma A, Malathi V (2003) Emerging gemivirus problems: a series threat to crop production. Ann Appl Biol 142:145–164

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153:910–918

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C et al (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat Biotechnol 32:947–951

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

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

Yin K, Han T, Liu G, Chen T, Wang Y, Yu AYL et al (2015) Ageminivirus-based guide RNA delivery system for CRISPR/Cas9 mediated plant genme editing. Sci Rep 5:14926

Yuen KS, Chan CP, Wong NH, Ho CH, Ho TH, Lei T et al (2015) CRISPR/Cas9-mediated genome editing of *Epstein–Barr* virus in human cells. J Gen Virol 96:626–636

Zebrini FM, Briddon RW, Irdis A et al (2017) ICTV virus taxonomy profile: geminiviridae. J Gen Virol 98:131–133

Zhen S, Hua L, Liu Y-H, Gao L-C, Fu J et al (2015) Harnessing the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated Cas9 system to disrupt the hepatitis B virus. Gene Ther 22:404–412

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