AGRICULTURAL BIOLOGY,
2020, V. 56, Iss. 1, pp. 20-31
(SEL'SKOKHOZYAISTVENNAYA BIOLOGIYA)

ISSN 2412-0324 (English ed. Online)
ISSN 0131-6397 (Russian ed. Print)
ISSN 2313-4836 (Russian ed. Online)

Genomic technologies

UDC 633.853.55:575.1:577.21

doi: 10.15389/agrobiology.2021.1.20eng
doi: 10.15389/agrobiology.2021.1.20rus

SEQUENCING OF THE *U6* PROMOTERS IN CASTOR BEANS AND VECTOR CONSTRUCTION FOR CRISPR/Cas9 GENOMIC EDITING ON THEIR BASIS

O.S. ALEXANDROV ✉, G.I. KARLOV

All-Russian Research Institute of Agricultural Biotechnology, 42, ul. Timiryazevskaya, Moscow, 127550 Russia, e-mail olegsandrov@gmail.com (✉ corresponding author), karlovg@gmail.com
ORCID:
Alexandrov O.S. orcid.org/0000-0002-7146-4094 Karlov G.I. orcid.org/0000-0002-9016-103X
The authors declare no conflict of interests
Acknowledgements:
Supported by the Russian Science Foundation, Agreement No. 17-74-10233 dated July 21, 2017 on the scientific project "Adaptation of the CRISPR/Cas9 genomic editing technology elements to improve a castor bean (*Ricinus communis* L.) genome"
Received July 13, 2020

Abstract

Castor bean is an important crop in many countries. It is mainly used to obtain the castor oil, which is widely applied in various industries. The rich protein cakes and meals are remains after the oil is pressed. They are promising for use as protein additives in the fodder production. However, it is limited due to the presence of toxins. These are ricin protein and ricinine alkaloid. One of such methods can be genomic editing using the CRISPR/Cas9 system. The technology consists in cutting the target region of the intact DNA by the Cas9 enzyme with the assistance of a short guide RNA fragment. The delivery of the Cas9 and guide RNA genes into the cell of the edited plant is often carried out by plasmid vectors. For efficient synthesis of the guide RNA in such vectors, the promoters of small nuclear RNA genes are usually used. In dicotyledonous plants editing, the promoter of the *Arabidopsis* U6 gene is most often used. In this work, the amplification, sequencing and analysis of the castor bean U6 promoters were carried out at the first time. The obtained sequences were analyzed and used in the construction of CRISPR/Cas9 vectors for the ricin gene editing. Our aim was to study castor bean U6 promoters with help by bioinformatics and molecular genetic approaches. Zanzibar Green and Gibzonskaya varieties of castor bean plants were used in this work. DNA was isolated from young leaves. The preparation of sequences, alignments and homology level calculation were carried out by GenDoc program (http://www.nrbsc.org/gfx/genedoc/index.html). Primers for amplification of castor bean U6 promoters were designed by Primer3 program (http://bioinfo.ut.ee/primer3-0.4.0/). PCR was performed using a C-100 PCR machine (Bio-Rad Laboratories, Inc., USA). PCR products were separated in 1.5% agarose gel at 6 B/cm in the Sub-Cell GT electrophoresis camera (Bio-Rad Laboratories, Inc., USA). The amplicons were purified with the GeneJET PCR Purification Kit (Thermo Fisher Scientific, Inc., USA). The purified amplicons were cloned into the pAL2-T vector (Evrogen, Russia) according to protocol of manufacturer. In the CRISPR/Cas9 vector construction, the pRGE31, HindIII and SbfI restriction enzymes (Sibenzyme, Russia) and nucleotides were used. Purification of the digested products was carried out with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Inc., USA). Bioinformatic search were conducted in the GenBank base, and 12 scaffolds with the castor bean *U6* gene were found. Six promoters with intact USE and TATA-box elements of regulation were used in the primer design for amplification with cv. Zanzibar Green and cv. Gibzonskaya DNA matrices. One fragment PCR products were cloned and sequenced. The analysis of the obtained amplicon sequences reveled that promoter regions of two studied varieties were similar. The level of promoter identity from different amplicons ranged within 51-77%. In comparison of these promoters with ones from other plants, the level of homology was 42-64%. The promoter sequences with intact USE and TATA-box motifs were used in construction of the CRISPR/Cas9 vectors which can be used for efficient editing of the castor bean genes involved in the ricin and ricinine synthesis.

Keywords: castor beans, promoter, *U6* gene, sequencing, genome editing, CRISPR/Cas9, vector construction

Castor bean (*Ricinus communis* L.) is an agricultural crop in the spurge family (*Euphorbiaceae*) that is farmed in many countries of the world. It is cultivated mainly for seeds that castor oil, a valuable product used in many sectors of the economy, is produced from [1]. Castor oil virtually never dries and preserves its properties in a broad range of operating temperatures, which is why it is used as a lubricant in food machinery and in aircraft modeling engines; it is also a component of plastic lubricants [2]. Castor oil contains substances valuable for chemical synthesis: undecylenic acid and sebacic acids, components of epoxy and alkyd resins, etc. Castor oil has long been used in medicine as an effective laxative [3, 4].

Once castor beans have been pressed for oil, they retain residues rich in protein — meals and cakes, a promising source of protein in feed production. However, these residues contain toxins: ricin and ricinine (a protein and an alkaloid, respectively) [5]. Ricin is a cumulative poison that has a strong local cauterizing effect that triggers hemorrhagic gastroenteritis. Once absorbed into the blood, it causes erythrocyte agglutination and destruction, increases the risk of thrombosis, and triggers ulceration of vascular walls. Besides, ricin has a negative effect on the central nervous system, where it causes convulsions, paresis, and paralysis [6]. Ricinine is far less toxic; however, it is not a desirable feed component either. Castor bean meals and cakes are usually detoxified by destructive thermal and chemical treatment, which jeopardizes protein quality. Such treatment renders proteins less digestible and causes lysine and other scarce amino acids to hydrolyze [5]. This is why finding an alternative method to detoxify castor bean meals and cakes without conventional treatment remains a relevant research challenge.

It would be of undoubted theoretical and practical interest to obtain castor bean plants devoid of ricin and ricinine. To that end, genes involved in the synthesis of either toxin must be knocked out. CRISPR/Cas9 is an effective state-of-the-art genome editing technique that can perform such knockout. Its components are the enzyme Cas9 and a guide RNA, or gRNA [7]. Cas9 and gRNA genes are injected into a cell by plasmid vectors with corresponding promoters. *U3* and *U6* small nuclear RNA (snRNA) promoters are normally used with the gRNA gene [8]. Genome editing of monocotyledons normally uses sequences of *U3* promoters of rice; dicotyledons are edited with *U6* promoters of *Arabidopsis*. However, several papers have shown that editing is more efficient when promoters are of the edited species, and adding them to CrISPR/Cas9 vectors has become a major trend [9-13]. Research into *U6* promoters in various species shows that they contain two RNA polymerase III recognition sites: Upstream Sequence Element (USE) and TATA box [11, 13, 14]. USE has a consensus sequence RTCCCACATCG. Marshallsay *et al.* [14] proposed a system of primers for U6 promoter amplification; their system is based on the aforementioned sequence and the *Arabidopsis U6* sequence. These primers seem suitable for analyzing *U6* promoters in other species, the genomes of which are yet to be sequenced.

To date, genomes have been sequenced for several *Euphorbiaceae* species (*Jatropha curcas*, *Manihot esculentum*, *Hevea brasiliensis*, *Euphorbia esula*), including castor bean [15-19]. Ricin-coding sequences have also been sequenced and studied [20-23]. They do not contain introns, which is why a large precursor (preproricin) is translated from them; this precursor contains a single peptide and ricin A and B chains separated by a linker peptide of 12 amino acids [20]. Parts encoding the signal peptide or the beginning of the A chain are the most suitable targets for editing [23, 24].

This is the first research project to have amplified, sequenced, and analyzed castor bean *U6* promoters. The identified sequences were analyzed and used to construct CRISPR/Cas9 vectors for further editing of ricin genes.

The goal was to study castor bean *U6* promoters by bioinformation, molecular, and genetic methods.

Materials and methods. The researchers used the castor bean varieties Zanzibar Green (Gavrish, Russia) and Gibzonskaya (Gavrish, Russia).

DNA was extracted from young leaves by the method proposed by Doyle *et al.* [25], although slightly modified [26]. Sequence processing, alignment, and homology quantification were performed in GenDoc (http://www.nrbsc.org/gfx/genedoc/index.html).

Primers to amplify castor bean *U6* promoters were found by means of Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/). Polymerase chain reaction (PCR) was run on a C-100 amplifier (Bio-Rad Laboratories, Inc., USA) with the following parameters: 5 min at 95 °C; 30 s at 95 °C, 30 s at 50 °C, 1 min at 72 °C (35 cycles); 10 min at 72 °C (terminal elongation).

PCR products were separated in 1.5% agarose gel at 6 V/cm in a Sub-Cell GT horizontal electrophoresis chamber (Bio-Rad Laboratories, Inc., USA). Electrophoresis results were visualized by the gel documentation system Gel Doc™ XR+ (Bio-Rad Laboratories, Inc., USA). Amplicons were purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, Inc., USA).

The purified amplicons were cloned using the pAL2-T vector (Evrogen, Russia) according to the manufacturer's instructions.

To construct CRISPR/Cas9 vectors, the plasmid pRGE31, HindIII and SbfI restriction endonucleases (Sibenzyme, Russia), and the required oligonucleotides were used. To cut restriction products out of the gel, the GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Inc., USA) was applied.

Results. Twelve scaffolds containing the castor bean *U6* sequence were found in GenBank: AASG02000063, AASG02000163, AASG02000719, AASG02001323, AASG02002949, AASG02003223, AASG02003842, AASG02006600, AASG02019053, AASG02021200, AASG02020516, AASG02025904. However, the characteristic USE and TATA elements were found in the promoter region of the *U6* gene only in six of these scaffolds: AASG02021200, AASG02020516, AASG02006600, AASG02000719, AASG02025904, AASG02002949. Comparison to the results of the search for USE and TATA-featuring *U6* promoters in the genomes of other Euphorbiaceae species showed the number of such could vary. The genomes of *Jatropha curcas* and *Manihot esculentum* (variety W14) contain 6 such promoters, as does castor bean; however, *Hevea brasiliensis* and *Euphorbia esula* have 9 and 23, respectively.

Sequences of the scaffolds AASG02021200, AASG02020516, AASG02006600, AASG02000719, AASG02025904, AASG02002949, and that of a single USE and TATA-less scaffold AASG02000063 were used to select primers (Table 1) and further amplify the fragments containing the upstream region of the respective *U6* genes, see Figure 1. In all cases except AASG02000063, single fragments of expected length were amplified (amplicons Rc200, Rc516, Rc600, Rc719, Rc904, and Rc949). Since the case of AASG02000063 saw the amplification of two more fragments, this upstream region was subject to no further analysis. Single fragments of other variants were cloned and sequenced.

1. Primers for amplification of the *U6* gene promoter region in the analyzed castor bean (*Ricinus communis* L.) scaffolds

Primer	Nucleotide sequence	Scaffold	Amplicon	Size, bp
RcU6-200f	5′-TGGATAAGAGGAGATTCTTGAATTG-3′	AASG02021200	Rc200	392
RcU6-200r	5′-AGGGGCCATGCTAATCTTCT-3′			
RcU6-516f	5′-GTTGGCAGCCTTCAGATTTC-3′	AASG02020516	Rc516	590
RcU6-516r	5′-AGGGGCCATGCTAATCTTCT-3′			

Continued Table 1

RcU6-600f	5′-CTCCGAATTATATTTGGGGTTTT-3	AASG02006600	Rc600	280
RcU6-600r	5′-AAAAATTTGGACCATTTCTCG-3′			
RcU6-719f	5′-ACCTGTGAGGTGGCTTTCTG-3′	AASG02000719	Rc719	346
RcU6-719r	5′-CGGTGTCTGTTTGCCCTAAT-3′			
RcU6-904f	5′-AGCCCATTTTGGGGTGTTAT-3′	AASG02025904	Rc904	512
RcU6-904r	5′-AGGGGCCATGCTAATCTTCT-3′			
RcU6-949f	5′-TTCTGGGAGGTATGCATCAA-3′	AASG02002949	Rc949	401
RcU6-949r	5′-AAAAATTTGGACCATTTCTCG-3′			
RcU6-063f	5′-CAACCCGACTCCTTCATCAT-3′	AASG02000063	Rc063	584
RcU6-063r	5′-AGGGGCCCTACTTTGATACC-3′			

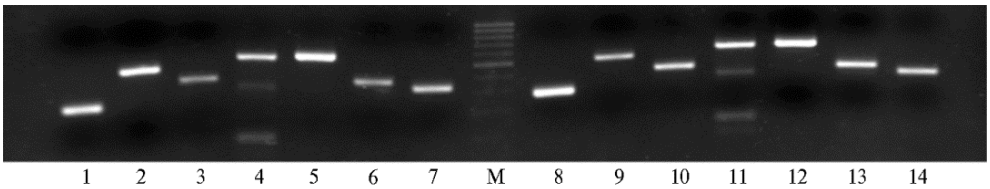

Fig. 1. PCR amplification of the *U6* gene promoter region from the template DNA of castor bean (*Ricinus communis* L.) Zanzibar Green (1-7) **and Gibzonskaya** (8-14) **cultivars using selected primers** (see Table 1)**:** 1 and 8 — Rc600f/Rc600r, 2 and 9 — Rc904f/Rc904r, 3 and 10 — Rc949f/Rc949r, 4 and 11 — Rc063f/Rc063r, 5 and 12 — Rc516f/Rc516r, 6 and 13 — Rc200f/Rc200r, 7 and 14 — Rc719f/Rc719r; M — molecular weight marker (100 bp DNA Ladder M-214S, Jena Biosience GmbH, Germany).

2. Homology (%) between the nucleotide sequences of six selected *U6* gene promoters of castor bean (*Ricinus communis* L.)

Promoter	Rc200ZG	Rc516ZG	Rc600ZG	Rc719ZG	Rc904ZG	Rc949ZG
Rc200ZG	100	56	53	77	57	50
Rc516ZG		100	63	57	67	62
Rc600ZG			100	54	63	54
Rc719ZG				100	53	51
Rc904ZG					100	59
Rc949ZG						100

Amplicon sequences obtained from Zanzibar Green (Rc200ZG, Rc516ZG, Rc600ZG, Rc719ZG, Rc904ZG, and Rc949ZG) and Gibzonskaya (Rc200Gib, Rc516Gib, Rc600Gib, Rc719Gib, Rc904Gib, and Rc949Gib) were analyzed; this analysis found these varieties to have fully matching promoter regions which also matched the corresponding sequences in the scaffolds AASG02021200, AASG02020516, AASG02006600, AASG02000719, AASG02025904, AASG02002949. However, two polymorphic sites were found outside regulatory element-containing promoter regions, the G/A substitution at a position –228 from the start site of the *U6* gene in Rc200Gib and T/A substitution at a position –351 from the start site of the *U6* gene in Rc904Gib. Apparently, the tested promoter regions are rather conservative in the varieties of the same species.

A similar pattern was observed in the case of cassava varieties AM560-2 and KU50. *U6* promoter regions from their scaffolds LTYI01019634 (nucleotides 16035-16336) and JPQF01078381 (nucleotides 1216-1517), as well as LTYI01021841 (nucleotides 27091-27392) and JPQF01070438 (nucleotides 16559-16860), were found to fully match.

Comparing sequenced castor bean *U6* promoters against each other showed medium homology varying within 51-77% (Table 2). Similar homology of 42-64% was observed when those were compared against randomly sampled *U6* promoters of other randomly picked species from different families (Tables 3 and 4).

3. Nucleotide sequences of randomly selected promoters of plant species from different families used to assess the degree of homology between them and six *U6* gene promoters of castor bean (*Ricinus communis* L.)

Code	Nucleotide sequence	GenBank acc. no.	The first nucleotide	The end nucleotide
abrus1	5′-GCCCCACATCGAACAGTATTATCAAAGCATGACACAATATATAGCAAAAGAAACACGCAGAGAGT-3′	XM_027482675	708	644
arachis1	5′-GTACCACATCGAGTAGCATCATATAACTCTGACAATATATATAGCAGAGGGTGCAAAGAAGGCTC-3′	XM_025790263	575	511
arachis2	5′-GTCCCACATCGCTTAGTATCAGACCACTCTGACAGAATATATATCAAAGGAAACACAAAAGGCTC-3′	XR_003810923	117	181
arachis4	5′-GTCTCACATCGCCCGAGTTTTGAGAAACCAATAACTTATATATCAGAGGCGAAGCAAAGGCTC-3′	XM_025845620	218	281
bras_ol1	5′-GTCCCACATCGCTCAGGTGAAGAGAAGGAGCTGCGTTTATATAGCGATGAAGTCACGAAAGTGATT-3′	LR031877	48176336	48176271
bras_rap1	5′-CTCCCACATCGCTCAGCGAAGCAAAAGAAGCTCCTGTTTATATACTTTCAGAGTCAAGAAGATGATT-3′	LR031575	1103139	1103205
bras_rap2	5′-CTCCCACATCGTTTATCAGAGAAGCAGAAGCCGAGTTTATATAGGGACGGAGTGACGAAGGAGATT-3′	LR031575	1143184	1143119
cicer1	5′-GTCCCACATCGAATACATGTATCCCATTTTCCATATTTATATAACGCAGGTTAACCATGCAGTAT-3′	CP039335	30486401	30486452
cicer2	5’-GTCCCACATCGAATACATTTATCCCTTTTTCCGTATTTATATAACGCAGGTTAACCATGGAGTTT-3′	CP039335	30494676	30494740
cicer3	5′-GTCCCACACCGAATCATCTATCATTTTTTTCGTCTTTATATAACCCATGTTAATCATTAGGTTT-3′	CP039335	30512524	30512587
cicer4	5′-GTCCCACATCGTCTAAATATTCGAATATTTAATATTTATATACAATGTTCGAGCAGTATAGTAT-3′	CP039333	18318744	18318807
cicer7	5′-GTCCCACACCGCGTACGCATAACATGTGTTCAGTGTTTATAATACCCTCGCACACATCATCAAC-3′	CP039333	31682334	31682398
cicer8	5′-GTTCCACATCGTCTACATCTATCATTATTTACGTCTTTATATTCAACCGATGAGCCATAAGGCTT-3′	CP039333	39009683	39009747
cichorium1	5′-CGTCCCATACCGACCAGTAAAGTACTTCCCGTCGCCTTATATAGCGCAGCTCGGCGACTATCATC-3′	MK455779	235	299
cichorium2	5′-GTCCCATACCGACCAGTAAAGTACTTCCCGTCGCCTTATATAGCGCAGCTCGGCGACTATCATC-3′	MK455779	236	299
cichorium3	5′-TTCCCACATCGCTCTTTGAAGCAACATCGCCATGCTTTATATAGCTTGGCTTCCAAACATATATC-3′	MK455773	235	299
cichorium4	5′-TTCCCACATCGATGATTGAAACGATTCCTCGGTGTTTTATATAGCCTGGCTTCCAATCAAATATC-3′	MK455776	238	302
cichorium5	5′-CTCCCACATCGATGATCGGAACGGTTGTTTCGTGCTTTATATAGCTCGGGTTCCAACCATTTATC-3′	MK455775	238	302
citrullus1	5′-GTCCCACATCGGTAAGTTTTGATTCTAGTTTACGCTTTATATAACTAAGACTGCAGTACAAGGCTT-3′	VOOL01000005	2250542	2250477
cynara1	5′-AATCCCACATCGCCTTTAACGATATCCAGTGCTAGCTTTATATGGCGGAGGTCGGCAGCTAAGATC-3′	XR_003069239	1	65
goss_raim1	5′-ATCCCATATCGCTAAAGAACTATAACACAGGAGCGTTTATATAAGCGAAAGAAGCAGCAAATGATT-3′	CP032562	3130166	3130101
goss_raim2	5′-ATCCCACATAGCTAAAGAATTAGGAAAATTTATTGTTTATAAAGGCAAAGGAAGAAACTTATTATT-3′	CP032562	3328533	3328468
goss_raim3	5′-ATCCCGCATCGCTAAAGAATTGAAAAAATTTTATTGTTTATATAGGAAAAACAAGCTGACTATGATT-3′	CP032562	3351518	3351452
goss_raim4	5′-ATCCCGCATCGCTAAAGAATTGAAAAAATTTTATTGTTTATATAGGAAAAACAAGCTGACTATGATT-3′	CP032562	3351518	3351452
gossipioides1	5′-ATCCCACATCGCTAAAGAACTAAAATGCCGAAGTATTTATATAAGCGAAAGAAACAGCATTAGTGT-3′	CP032252	2333488	2333423
gossipioides2	5′-ATCCCACATTGCTAAAGAATTAACAAATACTATTGTTTATATAGGCAAAAGAAACACCGTAGCAGT-3′	CP032252	2341665	2341600
gossipioides3	5′-ATCCCACATCTCTAAAGAATTAAAAAACACTATTGTTTATATGGGCAAAAGAAGCACCGTTGTATT-3′	CP032252	2343934	2343869
ipomea1	5′-CTCCCACATCGGGCGATGAAGCAGCTCTCTTCCAGTACACATACTCCGCCATTGAAGAAGAAGAAC-3′	CP025668	7620112	7620047
ipomea2	5′-CTCCCACATCGGCCAATGAGCCATCTTACTTCCAGTACATATACTCCGCCATGGAAGCTCTTATC-3′	XR_004100417	33	97
ipomea3	5′-CTCCCACATCGGCTGATGAAACAACTTGCTTCCAGTATACATACTCTATCATGGAAGCACTGAGC-3′	XM_031272161	1	65
ipomea4	5′-GTCCCACATCGGCCAATGAGCCATCTTACTTCCAGTACATATACTCCGCCATTGGAGCACTTAGC-3′	CP025668	7980996	7981060
ipomea5	5′-TTCCCACATCGGGCGATGAAGCAGCTCTCTTCCATTACACATACTCCGCCATTGAGGAAGGAGAAT-3′	CP025668	7984209	7984274
lotus1	5′-GTCCCACACCGGATAAACATACAGAAATATGAGTGTTTATAAGCAAATAGTCAGCAATAAGGTTC-3′	AP010923	70411	70347
lotus2	5′-GTTCCACATCGGCTATGTTGATTAAGATTTTATAGTTCATATATCACTACAGAACAGCAAGTATT-3′	AP010923	68417	68353
tea1	5′-GTCCCACATCGAAACTTCGACGTTATAGACATGGAGTTTATAAGAAAGAAGAAAAGAGAAGACGTT-3′	XR_003649102	150	215
tomato1	5′-CTCCCTCATCGCTTACAGAAAAAAGCTATATGCTGTTTATATTGCGAAATCTAACAGTGTAGTTT-3′	XM_004230407	25	89
tomato2	5′-CTCCCTCATCGCTTACAGAAAAAAGCTATATGCTGTTTATATTGCGAATCTAACAGTGTAGTTT-3′	X51447	198	261
vigna1	5′-GTCCCACATCGTCCAAACATGTCACAACTTCCATGTTTAAAAACGCACGCCTACTCGCTGCTGTT-3′	CP039354	53496613	53496677
vigna2	5′-GTCCCACACCGTATACTTTCACTAGAGGTTTAGTGTTTATATAGATACAGACTGCATCCAAGCTT-3′	CP039350	33742544	33742608

4. Homology (%) between nucleotide sequences of six selected *U6* gene promoters of castor bean (*Ricinus communis* L.) and randomly selected *U6* gene promoters of plant species from different families

Code	Promoter					
	Rc200ZG	Rc516ZG	Rc600ZG	Rc719ZG	Rc904ZG	Rc949ZG
abrus1	48	53	56	46	60	62
arachis1	52	53	57	47	55	49
arachis2	54	57	57	51	56	58
arachis4	54	54	63	52	49	54
bras_ol1	53	54	57	48	52	53
bras_rap1	54	55	57	56	54	52
bras_rap2	45	54	54	46	52	54
cicer1	56	54	53	53	55	52
cicer2	58	52	53	57	55	49
cicer3	59	53	50	54	51	46
cicer4	50	56	59	53	56	55
cicer7	47	64	53	50	50	55
cicer8	51	57	58	54	55	55
cichorium1	42	47	55	45	47	50
cichorium2	43	48	55	46	48	51
cichorium3	55	53	50	52	49	52
cichorium4	49	48	50	51	50	45
cichorium5	55	51	54	55	47	44
citrullus1	57	60	59	58	56	59
cynara1	52	55	63	55	56	51
goss_raim1	48	62	57	51	61	59
goss_raim2	47	58	53	50	58	56
goss_raim3	51	61	50	52	65	63
goss_raim4	50	46	50	52	65	63
gossipioides1	49	55	57	54	62	58
gossipioides2	51	60	54	51	62	61
gossipioides3	52	60	57	52	60	60
ipomea1	42	52	47	46	44	47
ipomea2	48	58	52	44	50	49
ipomea3	46	58	54	43	52	52
ipomea4	50	58	51	48	48	49
ipomea5	45	52	47	43	47	47
lotus1	46	55	58	52	63	59
lotus2	59	59	62	59	56	58
tea1	57	53	56	58	57	58
tomato1	48	47	53	48	53	51
tomato2	48	47	53	47	52	50
vigna1	52	60	57	46	53	52
vigna2	55	66	56	53	61	55

To a considerable extent, such homology was due to the presence of fairly conservative USE and TATA elements, whereas other regions had low homology (Fig. 2).

USE sequence analysis of these promoters showed castor bean to have several variants of this regulatory element: A (GAACCACATCG) of the amplicons AASG02021200ZG and AASG02021200Gib, B (A–CCCACATCG) of the amplicons AASG02020516ZG and AASG02020516Gib, and D (AT–CCACATCG) of the amplicons AASG02000719ZG and AASG02000719Gib. In the other three cases, USE motives (C, E, and F, see Fig. 2) matched the *Arabidopsis* consensus sequence RTCCCACATCG.

The identified castor bean *U6* promoters were then used to design vectors to edit the ricin gene by means of CRISPR/Cas9.

To construct CRISPR/Cas9 vectors, the research team used the plasmid pRGE31 containing the *Cas9* gene under the CaMV *35S* promoter, as well as a gRNA synthesis structure: *U3* promoter of rice (*OsU3*), a motif of two divergent sites of the Bso31I restriction endonuclease (to embed the target site once the plasmid has been hydrolyzed by the enzyme, 2xBso31I), and a sequence encoding the gRNA part interacting with the Cas9 protein (gRNAgene).

First, we searched for restriction endonuclease sites in the original plasmid pRGE31 such that one site was located before the *U3* promoter beginnin, and another site after that. Still, both sites had to be absent from the structure to ligate with pRGE31. HindIII and SbfI sites were picked as a result. Since the region between them contained the entirety of the OsU3-2xBso31I-gRNAgene structure, it was necessary to synthesize an insertion sequence that would contain 2xBso31I-gRNAgene under the corresponding castor bean promoter.

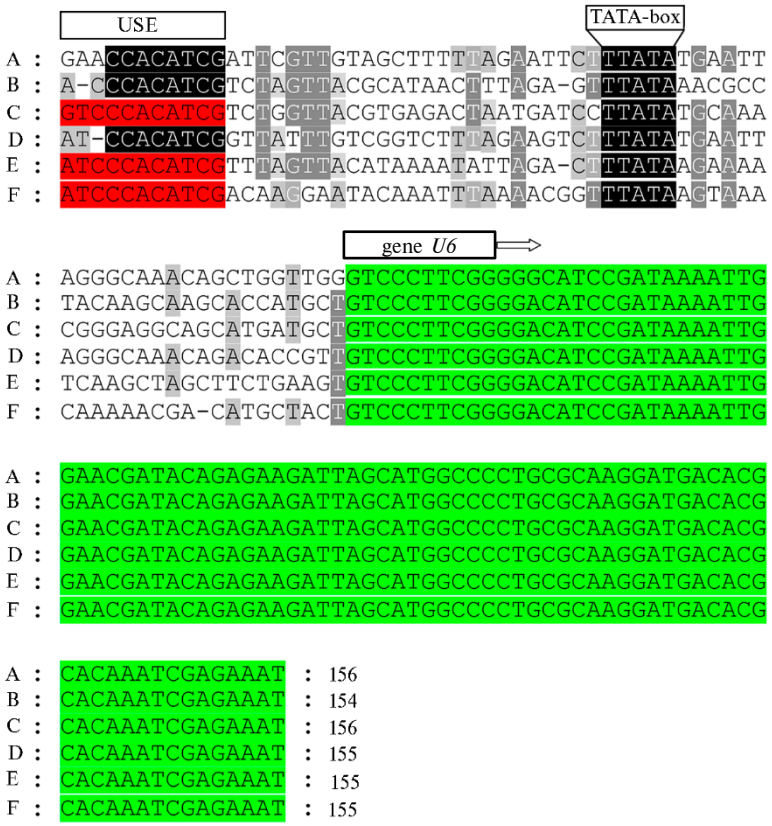

Fig. 2. Alignment of castor bean (*Ricinus communis* L.) *U6* gene promoter sequences: A — promoter from the AASG02021200ZG and AASG02021200Gib amplicons, B — AASG02020516ZG and AASG02020516Gib amplicons, C — AASG02006600ZG and AASG02006600Gib amplicons, D — AASG02000719ZG and AASG02000719Gib amplicons, E — AASG02025904ZG and AASG02025904Gib amplicons, F — AASG02002949ZG and AASG02002949Gib amplicons. The USEs with a consensus sequence RTCCCACATCG of *Arabidopsis* are highlighted in red. The part of the *U6* gene that is amplified from the Oligo3 primer [14] is highlighted in green.

To that end, six oligonucleotides (forward-n) were synthesized that contained RcU6(n)-2xBso31I-(20bp)gRNAgene, where n is the corresponding castor bean *U6* promoter variant, (20bp)gRNAgene are the first 20 nucleotides to encode the beginning of Cas9-interacting gRNA, see Table 5. A reverse gRNAgene oligonucleotide was synthesized as well. After a single PCR cycle (5 min at 95 °C, 30 s at 50 °C, 10 min at 72 °C), double-stranded RcU6(n)-2xBso31I-gRNAgene sequences were obtained for each forward oligonucleotide with a reverse oligonucleotide. This approach was chosen because it was more resource-effective compared to the complete synthesis of forward and reverse RcU6(n)-2xBso31I-gRNAgene constructs. Besides, it enabled the resource-effective synthesis of structures with any other promoter, whereby only the forward oligonucleotide would need to be replaced.

5. Oligonucleotides used for constructing and testing CRISPR/Cas9 vectors with castor bean (*Ricinus communis* L.) *U6* gene promoter s

Promer	Nucleotide sequence
	Oligos for synthesis of double strand constructs RcU6(n)-2xBso31I-gRNAgene
forward-A	5′-GAACCACATCGATTCGTTGTAGCTTTTTAGAATTCTTTATATGAATTAGGGCAAACAGCTGGTTGGGGAGACCGAGGTCTCGGTTTTAGAGCTAGAAATAGC-3′
forward-B	5′-AACCCACATCGTCTAGTTACGCATAACTTTAGAGTTTATAAACGCCTACAAGCAAGCACCATGCTGGAGACCGAGGTCTCGGTTTTAGAGCTAGAAATAGC-3′
forward-C	5′-GTCCCACATCGTCTGGTTACGTGAGACTAATGATCCTTATATGCAAACGGGAGGCAGCATGATGCTGGAGACCGAGGTCTCGGTTTTAGAGCTAGAAATAGC-3′
forward-D	5′-AATCCACATCGGTTATTTGTCGGTCTTTAGAAGTCTTTATATGAATTAGGGCAAACAGACACCGTTGGAGACCGAGGTCTCGGTTTTAGAGCTAGAAATAGC-3′
forward-E	5′-ATCCCACATCGTTTAGTTACATAAAATATTAGACTTTATAAGAAAATCAAGCTAGCTTCTGAAGTGGAGACCGAGGTCTCGGTTTTAGAGCTAGAAATAGC-3′
forward-F	5′-ATCCCACATCGACAAGGAATACAAATTTAAAACGGTTTATAAGTAAACAAAAACGACATGCTACTGGAGACCGAGGTCTCGGTTTTAGAGCTAGAAATAGC-3′
reverse	5′-AAAACAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3′
	Oligos for introducing restriction enzyme sites into RcU6(n)-2xBso31I-gRNAgene constructs
HindIII-A	5′-AAGCTTGAACCACATCGATTC-3′
HindIII-B	5′-AAGCTTAACCCACATCGTCTA-3′
HindIII-C	5′-AAGCTTGTCCCACATCGTCTG-3′
HindIII-D	5′-AAGCTTAATCCACATCGGTTA-3′
HindIII-E	5′-AAGCTTATCCCACATCGTTTA-3′
HindIII-F	5′-AAGCTTATCCCACATCGACAA-3′
SbfI-reverse	5′-CCTGCAGGAAAACAAAAAAGCAC-3′
	Oligos for testing vectors
35Spr-F	5′-CTATCCTTCGCAAGACCCTTC-3′
Amp-R	5′-ATAATACCGCGCCACATAGC-3′
pBR322ori-F	5′-GGGAAACGCCTGGTATCTTT-3′

HindIII and SbfI restriction sites were incorporated into the insertion sequence by means of PCR. Amplification involved primers selected for the borders of RcU6(n)-2xBso31I-gRNAgene constructs and containing corresponding restriction sites in the 5′-region (Table 5). PCR products were purified, cloned in the pAL2-T AT vector, and sequenced to verify the absence of synthesis errors. Once the plasmid pRGE31 and the pAL2-T plasmids featuring RcU6(n)-2xBso31I-gRNAgene structures were processed with HindIII and SbfI restriction endonucleases, the products were separated by electrophoresis, and target fragments were cut out. Purified fragments were ligated and used to transform competent *Escherichia coli* cells. Clones were selected by PCR with primers selected for the borders of RcU6(n)-2xBso31I-gRNAgene constructs. Plasmids extracted from the selected clones were tested with HindIII (linearizes the target plasmid) and Bso31I (does not break down the target plasmid), then sequenced using the insertion primers (SbfI-reverse) and primers for the target regions: Cas9 gene (35Spr-F), the ampicillin resistance gene (Amp-R), ori (pBR322ori-F) (see Table 5).

The created vector constructs were further compared against each other and against vectors containing the *Arabidopsis U6* gene in terms of effectiveness. Such experiments seem advisable, since, for example, potatoes have been shown to produce twice as many mutant forms where the gRNA gene promoter, AtU6 (*U6* promoter of *Arabidopsis thaliana*, GenBank accession no. X52527.1) was substituted with StU6 (*U6* promoter of *Solanum tuberosum*, GenBank accession no. Z17290.1) [9].

The same conclusions were made by Long et al. [11] when they researched into the optimization of CRISPR/Cas9 for editing the cotton genome. gRNA expression grew by a factor of 6-7 once the endogenous cotton promoter, GhU6.3 (*U6* promoter of *Gossypium hirsutum*) replaced AtU6-29 (*U6* promoter *Arabidopsis thaliana*) in the CRISPR/Cas9 vector structure. As a result, the mutagenic effectiveness of CRISPR/Cas9 rose by a factor of 4-6 [11]. In turn, Sun et al. [10] obtained similar results in their comparison of endogenous and exogenous promoters in CRISPR/Cas9 cassettes for soy genome editing. gRNA expression under GmU6 (U6 promoter of *Glycine max*) was twice as high as under an exogenous promoter, AtU6-26 (*U6* promoter of *Arabidopsis thaliana*). The effectiveness of gene editing varied within 14.7% to 20.2% in the first case and within 3.2% to 9.7% in the second case [10]. These experiments were carried out on non-related plant species from different families; their results clearly show the benefits of using endogenous promoters in CRISPR/Cas9 vectors. However, it is important to compare not only endogenous vs. exogenous promoters but also endogenous promoters against each other. This research identified various castor bean *U6* promoters with USE and TATA regulatory elements which are necessary for normal functioning; it would be logical to compare them against each other in editing experiments on castor bean, as endogenous promoters, as was reported earlier, might vary in the effectiveness of controlling gene expression [27-30].

Thus, the castor bean genome contains six *U6* promoters with USE and TATA elements. Sequencing these promoters in Zanzibar Green and Gibzonskaya varieties showed them to be highly conservative. The promoters are medium-homologous to each other and to the promoters of other species. The promoters were found to contain USE elements; in three promoters, those matched the consensus sequence of the USE element in *Arabidopsis*; in three others, they did not. CRISPR/Cas9 based on the identified castor bean *U6* promoters will constitute a valuable tool for further attempts to edit the genome of this species and to evaluate the effectiveness of various promoters with respect to gRNA synthesis.

REFERENCES

1. Moshkin V.A. *Kleshchevina* [Castor beans]. Moscow, 1980 (in Russ.).
2. Aznaur'yan M.P., Eloshvili N.G., Nazarov N.V., Yakubovich H.H., Meshcheryakov C.B., Girsin Yu.A. *Izvestiya vuzov, Pishchevaya tekhnologiya*, 1993, 1-2(212-213): 87-89 (in Russ.).
3. Naughton F.C. Production, chemistry and commercial applications of various chemicals from castor oil. *Journal of the American Oil Chemists' Society*, 1974, 51(3): 65-71 (doi: 10.1007/BF00000015).
4. Shul'pekova Yu.O. *Prilozhenie RMZH «Bolezni organov pishchevareniya»*, 2006, 8(2): 90-96 (in Russ.).
5. Anandan S., Kumar G.K.A., Ghosh J., Ramachandra K.S. Effect of different physical and chemical treatments on detoxification of ricin in castor cake. *Animal Feed Science and Technology*, 2005, 120(1-2): 159-168 (doi: 10.1016/j.anifeedsci.2004.10.002).
6. Gusynin I.A. *Toksikologiya yadovitykh rastenii* [Toxicology of poisonous plants]. Moscow, 1962 (in Russ.).
7. Ishino Y., Krupovic M., Forterre P. History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *Journal of Bacteriology*, 2018, 200(7): e00580-17 (doi: 10.1128/JB.00580-17).
8. Zlobin N.E., Ternovoi V.V., Grebenkina N.A., Taranov V.V. *Vavilovskii zhurnal genetiki i selektsii*, 2017, 21(1): 104-111 (doi: 10.18699/VJ17.228) (in Russ.).
9. Andersson M., Turesson H., Nicolia A., Fält A.S., Samuelsson M., Hofvander P. Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Reports*, 2017, 36(1): 117-128 (doi: 10.1007/s00299-016-2062-3).
10. Sun X., Hu Z., Chen R., Jiang Q., Song G., Zhang H., Xi Y. Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Scientific Reports*, 2015, 5: 10342 (doi: 10.1038/srep10342).
11. Long L., Guo D.D., Gao W., Yang W.W., Hou L.P., Ma X.N., Miao Y.C., Botella J.R., Song C.P. Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods*, 2018, 14: 85 (doi: 10.1186/s13007-018-0353-0).
12. Yan P.U., Chao L.I.U., JiYang L.I., TaShi A., Yan H.U., XiaoDong L.I.U. Different SlU6 promoters cloning and establishment of CRISPR/Cas9 mediated gene editing system in tomato. *Scientia Agricultura Sinica*, 51(2): 315-326.
13. Bernard G., Gagneul D., Dos Santos H.A., Etienne A., Hilbert J.-L., Rambaud C. Efficient genome editing using CRISPR/Cas9 technology in chicory. *International Journal of Molecular Sciences*, 2019, 20(5): 1155 (doi: 10.3390/ijms20051155).
14. Marshallsay C., Kiss T., Filipowicz W. Amplification of plant U3 and U6 snRNA gene sequences using primers specific for an upstream promoter element and conserved intragenic regions. *Nucleic Acids Research*, 1990, 18(12): 3459-3466 (doi: 10.1093/nar/18.12.3459).
15. Chan A.P., Crabtree J., Zhao Q., Lorenzi H., Orvis J., Puiu D., Melake-Berhan A., Jones K.M., Redman J., Chen G., Cahoon E.B., Gedil M., Stanke M., Haas B.J., Wortman J.R., Fraser-Liggett C.M., Ravel J., Rabinowicz P.D. Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology*, 2010, 28: 951-956 (doi: 10.1038/nbt.1674).
16. Horvath D.P., Patel S., Doğramaci M., Chao W.S., Anderson J.V., Foley M.E., Scheffler B., Lazo G., Dorn K., Yan C., Childers A., Schatz M., Marcus S. Gene space and transcriptome assemblies of leafy spurge (*Euphorbia esula*) identify promoter sequences, repetitive elements, high-quality markers, and a full-length chloroplast genome. *Weed Science*, 2018, 66(3): 355-367 (doi: 10.1017/wsc.2018.2).
17. Rahman A.Y., Usharraj A.O., Misra B.B., Thottathil G.P., Jayasekaran K., Feng Y., Hou S., Ong S.Y., Ng F.L., Lee L.S., Tan H.S., Sakaff M.K., Teh B.S., Khoo B.F., Badai S.S., Aziz N.A., Yuryev A., Knudsen B., Dionne-Laporte A., Mchunu N.P., Yu Q., Langston B.J., Freitas T.A., Young A.G., Chen R., Wang L., Najimudin N., Saito J.A., Alam M. Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics*, 2013, 14: 75 (doi: 10.1186/1471-2164-14-75).
18. Wang W., Feng B., Xiao J., Xia Z., Zhou X., Li P., Zhang W., Wang Y., Møller B.L., Zhang P., Luo M.C., Xiao G., Liu J., Yang J., Chen S., Rabinowicz P.D., Chen X., Zhang H.B., Ceballos H., Lou Q., Zou M., Carvalho L.J., Zeng C., Xia J., Sun S., Fu Y., Wang H., Lu C., Ruan M., Zhou S., Wu Z., Liu H., Kannangara R.M., Jørgensen K., Neale R.L., Bonde M., Heinz N., Zhu W., Wang S., Zhang Y., Pan K., Wen M., Ma P.A., Li Z., Hu M., Liao W., Hu W., Zhang S., Pei J., Guo A., Guo J., Zhang J., Zhang Z., Ye J., Ou W., Ma Y., Liu X., Tallon L.J., Galens K., Ott S., Huang J., Xue J., An F., Yao Q., Lu X., Fregene M., López-Lavalle L.A., Wu J., You F.M., Chen M., Hu S., Wu G., Zhong S., Ling P., Chen Y., Wang Q., Liu G., Liu B., Li K., Peng M. Cassava genome from a wild ancestor to cultivated varieties. *Nature Communications*, 2014, 5: 5110 (doi: 10.1038/ncomms6110).
19. Wu P., Zhou C., Cheng S., Wu Z., Lu W., Han J., Chen Y., Chen Y., Ni P., Wang Y., Xu X., Huang Y., Song C., Wang Z., Shi N., Zhang X., Fang X., Yang Q., Jiang H., Chen Y., Li M., Wang Y., Chen F., Wang J., Wu G. Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. *Plant Journal*, 2015, 81(5): 810-821 (doi: 10.1111/tpj.12761).

20. Halling K.C., Halling A.C., Murray E.E., Ladin B.F., Houston L.L., Weaver R.F. Genomic cloning and characterization of a ricin gene from *Ricinus communis*. *Nucleic Acids Research*, 1985, 13(22): 8019-8033 (doi: 10.1093/nar/13.22.8019).
21. Tregear J.W., Roberts L.M. The lectin gene family of *Ricinus communis*: cloning of a functional ricin gene and three lectin pseudogenes. *Plant Molecular Biology*, 1992, 18(3): 515-525 (doi: 10.1007/BF00040667).
22. Ladin B.F., Murray E.E., Halling A.C., Halling K.C., Tilakaratne N., Long G.L., Houston L.L., Weaver R.F. Characterization of a cDNA encoding ricin E, a hybrid ricin-*Ricinus communis* agglutinin gene from the castor plant *Ricinus communis*. *Plant Molecular Biology*, 1987, 9(3): 287-295 (doi: 10.1007/BF00166464).
23. Alexandrov O. Study of the upstream ricin gene sequences in different castor (*Ricinus communis*) varieties as a preliminary step in CRISPR/Cas9 editing. *Research on Crops*, 2020, 21(2): 344-348 (doi: 10.31830/2348-7542.2020.058).
24. Aleksandrov O.S. *Aktual'naya biotekhnologiya*, 2018, 3(26): 161-162 (in Russ.).
25. Doyle J.J., Doyle J.L. Isolation of plant DNA from fresh tissue. *Focus*, 1990, 12(1): 13-15.
26. Razumova O.V., Alexandrov O.S., Divashuk M.G., Sukhorada T.I., Karlov G.I. Molecular cytogenetic analysis of monoecious hemp (*Cannabis sativa* L.) cultivars reveals its karyotype variations and sex chromosomes constitution. *Protoplasma*, 2016, 253(3): 895-901 (doi: 10.1007/s00709-015-0851-0).
27. Wang M.B., Helliwell C.A., Wu L.M., Waterhouse P.M., Peacock W.J., Dennis E.S. Hairpin RNAs derived from RNA polymerase II and polymerase III promoter-directed transgenes are processed differently in plants. *RNA*, 2008, 14(5): 903-913 (doi: 10.1261/rna.760908).
28. Mikami M., Toki S., Endo M. Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. *Plant Molecular Biology*, 2015, 88(6): 561-572 (doi: 10.1007/s11103-015-0342-x).
29. Domitrovich A.M., Kunkel G.R. Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. *Nucleic Acids Research*, 2003, 31(9): 2344-2352 (doi: 10.1093/nar/gkg331).
30. Wang P., Zhang J., Sun L., Ma Y., Xu J., Liang S., Deng J., Tan J., Zhang Q., Tu L., Daniell H., Jin S., Zhang X. High efficient multisites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system. *Plant Biotechnology Journal*, 2018, 16(1): 137-150 (doi: 10.1111/pbi.12755).