(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
06 September 2019 (06.09.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/169409 A1

(51) International Patent Classification:
C12N 15/113 (2010.01) *C12N 15/82* (2006.01)
C12N 15/63 (2006.01)

(21) International Application Number:
PCT/US2019/020621

(22) International Filing Date:
04 March 2019 (04.03.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/638,010 02 March 2018 (02.03.2018) US

(71) Applicant: **ARIZONA BOARD OF REGENTS ON BEHALF OF ARIZONA STATE UNIVERSITY** [US/US]; SkySong, 1475 N. Scottsdale Road, Suite 200, Scottsdale, Arizona 85257 (US).

(72) Inventors: **MASON, Hugh**; 16833 S. 24th Place, Phoenix, Arizona 85048 (US). **DIAMOS, Andrew**; 1265 E. University Drive, Tempe, Arizona 85281 (US). **ROSENTHAL, Sun Hee**; 1475 N. Scottsdale, Suite 200, Scottsdale, Arizona 85257 (US).

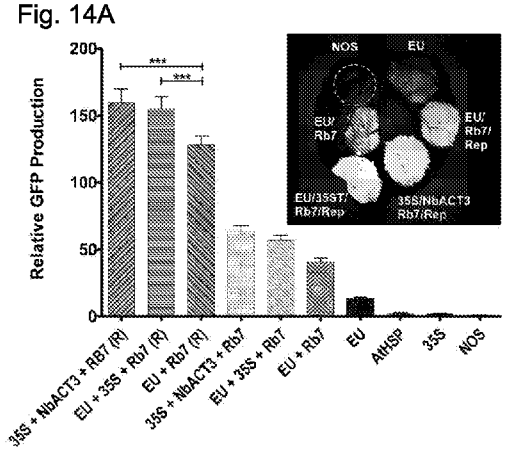
(74) Agent: **FULLER, Rodney J.**; BOOTH UDALL FULLER, PLC, 1255 W. Rio Salado Parkway, Suite 215, Tempe, Arizona 85281 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,

(54) Title: REPLICATING AND NON-REPLICATING VECTORS FOR RECOMBINANT PROTEIN PRODUCTION IN PLANTS AND METHOD OF USE THEREOF

Fig. 14A



(57) Abstract: The present disclosure relates to plant-based recombinant protein production systems and their methods of production and use. The plant-based recombinant protein production system is a vector comprising a 5' UTR and a 3' UTR, wherein the 3' UTR comprises at least one terminator selected from the group consisting of: EU, IEU, NbACT3, NbACT617, NbACT567, Pin2, BDB501, BDB282, NbHSP, NbHSPb, Rep, RbcS, SIR, SIR 5'/3', SIR 3', AtHSP, 35S, RepA, NOS, TMV, TNVD, PEMV, and BYDV. In certain implementations, the vector comprises two terminators in the 3' UTR, where the two terminators are fused to form a double terminator. For example, the double terminator comprises two members selected from the group consisting of: EU, IEU, NbACT3, NbACT617, NbACT567, Pin2, BDB501, BDB282, NbHSP, NbHSPb, Rep, RbcS, SIR, SIR 5'/3', SIR 3', AtHSP, 35S, RepA, NOS, TMV, TNVD, PEMV, and BYDV. In some aspects, the vector further comprises a chromatin scaffold/matrix attachment region (MAR) that is downstream of the terminators.

WO 2019/169409 A1

[Continued on next page]

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

Replicating and Non-Replicating Vectors for Recombinant Protein Production in Plants and Method of Use Thereof

Cross Reference to Related Applications

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/638,010, filed March 2, 2018, and incorporates the disclosure of the provisional application by reference thereto.

Incorporation-By-Reference of Material Electronically Filed

[0002] Incorporated by reference in its entirety herein is a computer-readable nucleotide/amino acid sequence listing submitted concurrently herewith and identified as follows: One 104,506 byte ASCII (text) file named "SeqList" created on March 1, 2019.

Technical Field

[0003] The disclosure relates to plant-based recombinant protein production systems and their methods of production and use.

Background

[0004] Plant-based recombinant protein production systems are have emerged as promising alternatives to traditional mammalian and microbial cell culture systems due to unique advantages of lower costs, high scalability, and improved safety (Chen and Davis 2016; Kamarova et al., 2010). Case studies have shown the potential for large cost reductions in capital investment and the cost of goods for plant-made therapeutics compared to conventional methods (Tusé et al., 2014; Nandi et al., 2016). The capacity for these systems to rapidly and safely produce therapeutics has been demonstrated by two success stories: the FDA approval of an enzyme replacement therapy for Gaucher's disease, which became the first plant-made therapeutic (Zimran et al., 2011; Fox 2012); and the monoclonal antibody therapy ZMapp given during the 2014 Ebola outbreak, which was shown to protect against lethal virus challenge (Lyon et al., 2014; Qui et al., 2014). Many strategies for improving protein production in plants have been explored, such as viral expression systems, subcellular targeting, agrobacterium strain, expression host, promoters, introns, and 5' untranslated regions (UTR). However, another key

component in many of these systems is the gene terminator and surrounding regions, which have not been systematically optimized.

Summary

[0005] The disclosure relates to plant-based recombinant protein production systems. In one aspect, the plant-based recombinant protein production system is a plant expression vector comprising at least one expression cassette. In some aspects, the disclosure relates to 3' UTRs that can be used in the expression cassettes disclosed herein.

[0006] The at least one expression cassette comprises a 5' UTR and a 3' UTR, wherein the 3' UTR comprises a first terminator; and a second terminator, a chromatin scaffold/matrix attachment region (MAR), or both. In some embodiments, the first terminator and the second terminator form a double terminator. In some embodiments, the 3' UTR further comprises MAR. In some aspects, MAR is downstream of the double terminator, while in other aspects, MAR is downstream of the first terminator. In certain implementations, the double terminator increases protein expression from the expression cassette.

[0007] In some embodiments, the first terminator is intronless tobacco extension terminator (EU) and the second terminator is selected from the group consisting of: *Nicotiana benthamiana* actin 3' UTR (NbACT3), p19 suppressor of RNA silencing from tomato bushy stunt virus (P19), *N. benthamiana* 18.8 kDa class II heat shock protein 3' UTR (NbHSP), short intergenic region of bean yellow dwarf virus (SIR), agrobacterium nopaline synthase 3' UTR (NOS), cauliflower mosaic virus 35S 3' UTR (35S), tobacco mosaic virus 3' UTR (TMV), BDB501 (bean dwarf mosaic virus DNA B nuclear shuttle protein 3' UTR, the intergenic region, the 3' end of the movement protein, and additional 200 nt downstream of the movement protein sequence), tobacco necrosis virus-D 3' UTR (TNVD), pea enation mosaic virus 3' UTR (PEMV), and barley yellow dwarf virus 3' UTR (BYDV). In some aspects, EU is upstream of the second terminator. Where the second terminator is 35S, 35S is upstream of EU in some embodiments.

[0008] In some embodiments, the first terminator is intron-containing tobacco extension terminator (IEU) and the second terminator is selected from the group consisting of: SIR, 35S, and long intergenic region from bean yellow dwarf virus (LIR). In some aspects, IEU is upstream of the second terminator.

[0009] In some embodiments, the at least one expression cassette comprises a 5' UTR and a 3' UTR, wherein the 3' UTR comprises a first terminator and MAR. In some aspects, the 3' UTR comprises EU and the MAR is selected from the group consisting of: Rb7 and TM6. In other aspects, the 3' UTR comprises IEU and the MAR is selected from the group consisting of: Rb7 and TM6. In certain embodiments, the 3' UTR of the at least one expression cassette comprises the first terminator, the second terminator, and MAR. In one embodiment, the 3' UTR comprises IEU, 35S, and Rb7, wherein IEU is upstream of 35S. In other embodiments, the 3' UTR comprises EU. In one aspect, such 3' UTR comprises EU, 35S, and Rb7, wherein EU is downstream or upstream of 35S. In another aspect, such 3' UTR comprises EU, NbACT3, and Rb7, wherein EU is upstream of NbACT3. In still another aspect, such 3' UTR comprises EU, BD501, and Rb7, wherein EU is upstream of BD501. In yet another aspect, such 3' UTR comprises EU, *A. thaliana* heat shock protein 3' UTR (AtHSP), and Rb7, wherein EU is downstream of AtHSP. In another aspect, such 3' UTR comprises EU, 35S, and TM6, wherein EU is upstream of 35S.

[0010] In another embodiment, the plant expression vector comprises an expression cassette with 3' UTR comprising at least one terminator selected from the group consisting of: EU, IEU, NbACT3, NbACT617 (downstream 617-nt region of NbACT3), NbACT567 (downstream 567nt of NbACT3), Pin2, BDB501, BDB282 (282 nucleotides comprising bean dwarf mosaic virus DNA B nuclear shuttle protein 3' UTR, the intergenic region, and the 3' end of the movement protein), NbHSP, NbHSPb (NbHSP missing 75 nt from 5' end), bean dwarf mosaic virus rep gene 3' UTR (Rep), pea rubisco small subunit 3' UTR (RbcS), SIR, SIR 5'/3' (SIR with additional sequences both upstream and downstream), SIR 3' (SIR with its additional downstream viral sequence), AtHSP, 35S, bean dwarf mosaic virus repA gene 3' UTR (RepA), NOS, TMV, TNVD, PEMV, and BYDV. In some aspects, the 3' UTR comprises at least one terminator selected from the group consisting of: NbACT3, NbACT617, NbACT567, Pin2, BDB501, BDB282, NbHSP, NbHSPb, Rep, RbcS, SIR, SIR 5'/3', SIR 3', AtHSP, and RepA. In some implementations, the 3' UTR comprises a double terminator, wherein the double terminator is a fusion of two members selected from the group consisting of: EU, IEU, NbACT3, NbACT617, NbACT567, Pin2, BDB501, BDB282, NbHSP, NbHSPb, Rep, RbcS, SIR, SIR 5'/3', SIR 3', AtHSP, 35S, RepA, NOS, TMV, TNVD, PEMV, and BYDV. For example, the 3' UTR comprises a double terminator, wherein the double terminator is a fusion of two members

selected from the group consisting of: EU, IEU, NbACT3, NbACT617, NbACT567, Pin2, BDB501, BDB282, NbHSP, NbHSPb, Rep, RbcS, SIR, SIR 5'/3', SIR 3', AtHSP, 35S, RepA, NOS, TMV, TNVD, PEMV, and BYDV. In some aspects, the 3' UTR comprises EU and a second terminator selected from the group consisting of: NbACT, P19, NbHSP, SIR, NOS, 35S, TMV, BDB501, TNVD, PEMV, and BYDV, wherein EU is upstream of the second terminator in some embodiments. In other aspects, the 3' UTR comprises 35S and a second terminator selected from the group consisting of: NbACT3, NOS, EU, NbHSP, Pin2, and BDB501, wherein in 35S is upstream of the second terminator in some embodiments. In some embodiments, the 3' UTR comprises 35S and NOS, wherein NOS is upstream of 35S. in some aspects, the 3' UTR comprises NbHSP and a second terminator selected from the group consisting of: NbACT3, NOS, and Pin2, wherein NbHSP is upstream of the second terminator in some embodiments.

[0011] In some implementations of the plant expression vector, the 3' UTR further comprises a chromatin scaffold/matrix attachment region (MAR) that is downstream of the terminators. In certain embodiments, the MAR is Rb7 or TM6. In some embodiments, the 3' UTR comprises Rb7 downstream of EU, IEU, AtHSp, 35S, BDB501, NbHSP, NOS, or NbACT3. In other embodiments, the 3' UTR comprises TM6 downstream of IEU, 35S, or NbACT3. In some aspects, the 3' UTR comprises RB7 downstream of a double terminator selected from the group consisting of: 35S+NbACT3, EU+35S, EU+NbACT3, NbHSP+NbACT3, 35S+EU, AtHSP+NOS, 35S+NOS, EU+BDB501, AtHSP+NbHSP, NbHSP+NOS, AtHSP+EU, NbHSP+Pin2, and IEU+35S. In other aspects, the 3' UTR comprises TM6 downstream of a double terminator selected from the group consisting of: EU+35S, 35S+NOS, NbHSP+NOS, and NbHSP+Pin2.

[0012] The disclosure also relates to the method of using the aforementioned plant-based recombinant protein production systems. In one implementation the vector described above are introduced into a plant or plant part. In some aspects, the plant is tobacco or lettuce or the plant part is from tobacco or lettuce. The some implementations, the vector transforms the plant or plant part using agrobacterium, for example, *Agrobacterium tumefaciens*.

Brief Description of the Drawings

[0013] Figs. 1A and 1B, in accordance with certain embodiments, depict schematics of the vector and constructs described herein. Fig. 1A shows a generalized schematic representation of

the T-DNA regions of the vectors encoding Norwalk virus capsid protein (NVCP), green fluorescent protein (GFP), or β-glucuronidase (GUS). Different terminators are placed downstream of gene of interest. As used in Fig. 1A, P35S/TEV5' refers to the CaMV 35S promoter with tobacco etch virus 5' UTR. NOS refers to the *Agrobacterium* nopaline synthase gene 3' element. VSP refers to the soybean vspB gene 3' element. 35S refers to the CaMV 35S terminator. EU refers to the intronless tobacco extensin gene terminator. LB and RB refer to the left and right borders, respectively, of the T-DNA region. NPTII refers to the expression cassette encoding nptII gene for kanamycin resistance. Fig. 1B depicts constructs containing various lengths of the extension (Ext) terminator used in the examples (GenBank accession no. D13951, 1-731 nt). In the top portion, the vertical black bars indicate the putative near upstream elements (NUEs). The positions of intron and polypurine sequence (PPS) are indicated with arrows. The bottom portion show the regions of Ext terminator fused downstream of the GFP gene in deletion constructs and are aligned with the map in the top portion. The numbers in parentheses indicate the portion of the Ext terminator used for each construct. As used in Figs. 1B and also in Figs 2A-9B (unless specified otherwise), IEU refers to the portion of the Ext terminator that contains the intron while EU refers to the portion of the Ext terminator that lack the intron.

[0014] Figs. 2A and 2B, in accordance with certain embodiments, depict the influence of the Ext intron (nt 24-249) on gene expression with either the Ext or NOS terminator. Fig. 2A shows GFP expression as determined by fluorometric analysis from three independently infiltrated samples and normalized by total soluble protein as measured by Bradford assay. Fig. 2B shows NVCP expression as measured by ELISA from six independent samples. White or filled bars represent constructs with or without intron, respectively. Data shown are means ± S.D. P-values were determined by student's t test.

[0015] Figs. 3A-3C, in accordance with certain embodiments, show that the Ext terminator increases transient transgene expression in *Nicotiana benthamiana* leaves. Fig. 3A depicts fluorometric analysis of GFP expression from constructs using different terminators. Infiltrated leaf extracts were analyzed by spectrofluorimetry and normalized by total soluble protein as measured by Bradford assay. The inset image is exemplary visualization of GFP expression. Infiltrated leaf was examined at 2 days post infiltration (DPI) under UV illumination (365 nm) generated by a B-100AP lamp (UVP). Fig. 3B shows NVCP expression from constructs with

different terminators. Fig. 3C shows GUS expression from constructs using different terminators. Data shown are means ± S.D. from six independently infiltrated samples.

[0016] Figs. 4A and 4B, in accordance with certain embodiments, show that the Ext terminator increases mRNA accumulation. Figs 4A and 4B respectively depict the relative mRNA accumulation of GFP and NVCP. Each sample was measured in triplicate for each transcript of interest. Elongation factor (EF1a) transcript served as the internal control. Transcript levels of GFP, NVCP and EF1a were quantified using separate standard curves and then the mRNA copy numbers of GFP and NVCP were normalized against the mRNA copy numbers of EF1a. Data shown are means ± S.D. from three independently infiltrated samples.

[0017] Figs. 5A-5C, in accordance with certain embodiments, show that the Ext terminator is a highly efficient mediator of transcription termination. Fig. 5A depicts a portion of the T-DNA region of the constructs used in the study. The *Agrobacterium* nopaline synthase gene 3' element (NOS) or Ext terminator (indicated as EU in Figs. 5A-5C) is placed downstream of the *GFP* or *NVCP* gene. RB refers to the right border of the T-DNA region. The horizontal arrows indicate primer regions used for the transcription readthrough detection by RT-PCR. Agarose gel electrophoresis of RT-PCR product to check transcription readthrough using constructs with GFP gene (Fig. 5B) or NVCP gene (Fig. 5C). Template types are indicated. cDNA of wild-type leaf sample and H_2O were used for negative (-) control. A gene specific forward primer was used with different reverse sense primers, which bind to different regions downstream of the terminator, indicated on the left of upper portion of Fig. 5A. To check RNA quality, RT-PCR was performed on the same RNA samples with GFP (Fig. 5B) or NVCP (Fig. 5C) and EF1a (Figs. 5B and C) primer sets.

[0018] Fig. 6 identifies the poly(A) sites of the Ext terminator. The RNA samples from construct pPS-OGFPEU were decapped, circularized, reverse transcribed, PCR amplified, cloned, and sequenced. The cDNA sequence of the intronless Ext 3' UTR are shown (SEQ ID NO. 6 lacking all but two nucleotides from the pair of extraneous sequences). The putative NUEs are underlined and bold. The polypurine sequence (PPS) in the Ext terminator is shown in gray. The text "pA" is used to indicate the polyadenylation YA dinucleotide, followed by the number of sequenced clones in parentheses.

[0019] Fig. 7, in accordance with certain embodiments, show that influence of deletions in the Ext terminator on gene expression. Modified terminators were placed downstream of the

GFP gene. Agroinfiltrated *N. benthamiana* leaves were harvested at 3 DPI and extracts were analyzed by SDS-PAGE followed by observation under UV illumination (365 nm) and by Coomassie staining. GFP band intensity was quantified using ImageJ software, using native plant proteins as a loading control. Data shown are means ± standard error of three independently infiltrated leaves. All leaves were infiltrated with construct pEU as an internal control for leaf and plant variability.

[0020] Figs. 8A and 8B, in accordance with certain embodiments, show the effect of deletion or substitution of PPS on Ext terminator function. Fig. 8A depicts a diagram of Ext terminator regions used in PPS testing constructs. Modified terminators are placed downstream of GFP gene. The positions of the putative NUEs and PPS are indicated with vertical bars and arrow, respectively. Deleted (pEUd) or substituted (pEUs) regions are indicated with discontinuous or dotted lines, respectively. Fig. 8B depicts the GFP expression relative to the reference construct pNOS. Test constructs were infiltrated side-by-side with reference construct pNOS and leaf extract were analyzed by spectrofluorimetry. Values from fluorometric analyses were converted to μg GFP/mg TSP and then normalized against the GFP expression of pNOS (indicated by dotted line). Asterisks denote a statistically significant reduction in GFP expression compared to pEU (Student's t test, $P < 0.05$). Absolute mean value of pEU was 112.3 μg GFP/mg TSP. Data shown are means ± S.D. from four to five independently infiltrated samples.

[0021] Figs. 9A and 9B, in accordance with certain embodiments, show that the Ext terminator has strong activity in plants other than *N. benthamiana*. GFP constructs containing the Ext terminator with or without intron were agroinfiltrated into the leaves of lettuce (Fig. 9A) or tobacco (Fig. 9B). At 4 DPI, leaf tissue was harvested, and protein extracts were analyzed by SDS- PAGE followed by observation under UV illumination (365 nm) and by Coomassie staining. GFP band intensity was quantified using ImageJ software, using native plant proteins as a loading control. Data shown are means ± standard error of three independently infiltrated leaves. All leaves were infiltrated with construct pEU as an internal control for leaf and plant variability.

[0022] Figs 10A and 10B, in accordance with certain embodiments, show evaluations of a variety of 3' UTRs on GFP production. Fig. 10A depicts generalized schematic representation of the T-DNA region of the vectors used in this study. RB and LB refers to the right and left borders of the T-DNA region, respectively. NPTII refers to the kanamycin resistance cassette.

P35S refers to the 35S promoter from cauliflower mosaic virus. TMV refers to the 5' UTR from tobacco mosaic virus. The 3' UTR is either a single terminator, double terminator, matrix attachment region, or combination of these elements as described in each experiment. Fig. 10B depicts the relative GFP production of nonreplicating vectors containing various 3' UTRs, which is inserted downstream from the *GFP* gene. The nonreplicating vectors were agroinfiltrated into *N. benthamiana* leaves. Leaves were photographed at 5 DPI under UV illumination (365 nm). The inset photo depicts a representative infiltrated leaf. Agroinfiltrated leaves were harvested between 4-5 DPI and extracts were analyzed by SDS-PAGE followed by observation under UV illumination (365nm) and Coomassie staining. GFP band intensity was quantified using ImageJ software, using native plant protein bands as a loading control. Data shown are means ± standard error of 3-4 independently infiltrated leaves. EU refers to intronless tobacco extensin 3' UTR. As used in Figs. 10B-17B, IEU refers to intron-containing tobacco extensin 3' UTR; NbACT3 refers to the *N. benthamiana* actin 3' UTR; NbHSP refers to the *N. benthamiana* 18.8kDa class II heat shock protein 3' UTR; pinII refers to the potato proteinase inhibitor II 3' UTR; RbcS refers to the pea rubisco small subunit 3' UTR; SIR refers to the short intergenic region of bean yellow dwarf virus, where SIR 5'/3' includes additional upstream and downstream sequences and SIR 3' include only additional downstream sequences; BDB refers to the bean dwarf mosaic virus DNA B nuclear shuttle protein 3' UTR; Rep refers to the bean dwarf mosaic virus rep gene 3' UTR; RepA refers to the bean dwarf mosaic virus repA gene 3' UTR; AtHSP refers to the *Arabidopsis thaliana* heat shock protein 3' UTR; 35S refers to the cauliflower mosaic virus 35S 3' UTR; and NOS refers to the *agrobacterium* nopaline synthase 3' UTR.

[0023] Fig. 11, in accordance with some embodiments, shows that double terminators strongly enhance GFP gene expression. Nonreplicating vectors containing different double terminators downstream from the GFP gene were agroinfiltrated into *N. benthamiana* leaves and analyzed for GFP production at 5 DPI. The filled bars indicate double terminators. Data shown are means ± standard error of 3-4 independently infiltrated leaves. The abbreviations refer to the same as those in Fig. 10B. BYDV refers to the barley yellow dwarf virus 3' UTR. PEMV refers to the pea enation mosaic virus 3' UTR. As used in Fig. 11, TNVD refers to the tobacco necrosis virus-D 3' UTR; TMV refers to the tobacco mosaic virus 3' UTR; and LIR refers to the long intergenic region from bean yellow dwarf virus.

[0024] Fig. 12, in accordance with some embodiments, show that chromatin scaffold/matrix attachment regions (MAR) strongly enhance GFP expression. Nonreplicating GFP vectors containing either the tobacco Rb7 or tobacco TM6 MAR sequences inserted 3' of the gene terminator were agroinfiltrated into the leaves of *N. benthamiana* and evaluated for GFP production at 5 DPI. Data shown are means ± standard error of 3-4 independently infiltrated leaves. "EU + Control" indicates DNA sequence obtained from an inverted region of the Norwalk virus capsid protein coding sequence was inserted 3' of the EU gene terminator in place of the Rb7 MAR.

[0025] Fig. 13, in accordance with some embodiments, show that combined 3' UTRs strongly enhance GFP expression. Nonreplicating GFP vectors with combined terminators were created, agroinfiltrated into the leaves of *N. benthamiana*, and evaluated for GFP production at 5 DPI. Data shown are means ± standard error of 3-4 independently infiltrated leaves. White bars indicate double terminators combined with Rb7 MAR; gray bars indicate double terminators combined with TM6 MAR; and black bars indicate single terminators combined with Rb7 MAR.

[0026] Figs. 14A and 14B, in accordance with some embodiments, show evaluations of a variety of combined 3' UTRs in replicating vectors on GFP expression. Replicating vectors containing elements of bean yellow dwarf virus (Diamos et al, 2016) were constructed with various combined 3' flanking regions, agroinfiltrated into the leaves of *N. benthamiana*, and evaluated for GFP production. Fig. 14A shows the mean relative GFP expression ± standard error of 3-4 independently infiltrated leaves. "(R)" indicates replicating geminiviral vector. Fig. 14B depicts SDS-PAGE gels showing GFP expression with the indicated constructs. RbcL refers to the large subunit of Rubisco.

[0027] Fig. 15, in accordance with some embodiments, compares the relative DsR production of the identified 3' UTRs. Nonreplicating vectors were constructed with single, double, or MAR-containing terminators downstream from the DsRed gene and agroinfiltrated into the leaves of *N. benthamiana*. DsRed production was evaluated at 5 DPI by SDS-PAGE and UV fluorescence. Data shown are means ± standard error of 3-4 independently infiltrated leaves.

[0028] Figs. 16A and 16B, in accordance with some embodiments, compares GFP expression of the a variety of 3' UTRs in tobacco and lettuce. Nonreplicating vectors were constructed with single, double, or MAR-containing terminators downstream from the GFP gene, and agroinfiltrated into the leaves of either tobacco (Fig. 16A) or lettuce (Fig. 16B) plants. GFP

production was evaluated at 5 DPI by UV fluorescence and SDS-PAGE. Data shown are means ± standard error of 3-4 independently infiltrated leaves.

[0029] Figs. 17A and 17B characterizes Rb7 and TM6 MARs. Fig. 17A depicts photographs of leaves agroinfiltrated with nonreplicating GFP vectors containing either the tobacco Rb7 or tobacco TM6 MAR sequences. The photographs were taken after 5 DPI under UV illumination (365 nm). In Fig. 17A, "No 3' Elements" refers to vectors that contain no terminator or MAR; "Rb7 Only" refers to vectors that contain only the Rb7 MAR with no terminator; "5' Rb7" refers vectors that contain the Rb7 MAR inserted 5' of the promoter; and "3' Rb7," "Rb7," or "TM6," identifies the MAR that was inserted 3' of the gene terminator in the vector. Fig. 17B depicts a schematic of deletion mutants of the Rb7 MAR. GFP expression was measured by SDS-PAGE using ImageJ to quantify band intensity. Relative expression is given compared to terminator EU lacking the Rb7 MAR.

[0030] Figs. 18A-D show that double terminators enhance VLP expression in a replicating vector derived from bean yellow dwarf virus. Fig. 18A depicts the generalized schematic of the replicating vector derived from bean yellow dwarf virus. The abbreviations used in Fig. 18A refers to the left and right T-DNA borders from Agrobacterium as RB and LB; the nopaline synthase 3' UTR from Agrobacterium as NOS3'; the p19 suppressor of RNA silencing from tomato bushy stunt virus as P19; the 35S promoter from cauliflower mosaic virus for P35S; the long intergenic region from bean yellow dwarf virus as LIR; the 5' UTR from *N. benthamiana* psaK gene as NbPsaK 5'; the target gene as indicated in each experiment as VLP; various single or double terminators as indicated in each experiment as T1/T2; the matrix attachment region from the tobacco Rb7 gene as Rb7 MAR; the short intergenic region from bean yellow dwarf virus for SIR; and the rep/repA genes from bean yellow dwarf virus Rep/RepA. Figs. 18B and 18D respectively depicts ELISA results measuring GII.4 norovirus (B) or GI norovirus (D) capsid production in the *N. benthamiana* leaves were agroinfiltrated with the replicating vector derived from bean yellow dwarf virus. Plant-produced GI or GII norovirus capsid served as standard. Data shown are means ± standard deviation from three independently infiltrated samples. Two asterisks (**) indicate $p < 0.05$, and three asterisks (***) indicate $p < 0.01$ as compared to the unmodified vector by student's t-test. Fig. 18C shows reducing SDS gel western of double terminator constructs with or without the addition of 0.1% Triton X-100 detergent (Triton). Probe is polyclonal rabbit anti-GII. The abbreviations as used in Figs 18B-18C refer to

intron-containing tobacco extensin terminator as IEU; IEU fused to the 35S terminator from cauliflower mosaic virus as IEU-35S; intronless tobacco extensin terminator fused to the putative ACT3-like terminator from *N. benthamiana* as EU-NbACT3; and 35S terminator fused to NbACT3 terminator as 35S-NbACT3.

[0031] Figs. 19-21 depict schematics of three exemplary plasmids comprising, in its expression cassettes, the 3' UTR comprising a double terminator and MAR.

[0032] Fig. 19 depict the schematic for a 14115 bp plasmid (pBYR2eAK2Mc-GFP, sequence set forth in SEQ ID NO. 115) comprising in the 3' region of the expression cassette a double terminator of EU fused upstream of NbACT3 and the Rb7 MAR downstream of the double terminator.

[0033] Fig. 20 depicts the schematic for a 17111 bp plasmid (pBY11HA-GFP, sequence set forth in SEQ ID NO. 116), which comprises two expression cassettes. The upstream expression cassette comprises in the 3' region a double terminator of EU fused upstream of NbHSP (shown as HSP20 3' in the figure) and the Rb7 MAR downstream of the double terminator. The downstream expression cassette comprises in the 3' region a double terminator of EU fused upstream of NbACT3 and the Rb7 MAR downstream of the double terminator.

[0034] Fig. 21 depicts the schematic for a 19330 bp plasmid (pBY!11-h6D8M2, sequence set forth in SEQ ID NO. 117), which comprises expression cassettes for a modified mAb, where the upstream cassette codes the heavy chain fused to influenza antigen M2e and the second cassette codes the corresponding light chain. The 3' region of the upstream expression cassette comprises a double terminator of EU fused upstream of NbACT3 and the RB7 MAR downstream of the double terminator. The 3' region of the downstream expression cassette comprises a double terminator of EU fused upstream of NbHSP (shown as HSP20 3' in the figure) and the Rb7 MAR downstream of the double terminator.

[0035] Figs 22A-22B depicts expression results of the plasmids of Figs. 19-21. Fig. 22A shows that relative percentage of the total soluble protein produced from leaves infiltrated with the plasmid with Fig. 21 that is the modified antibody encoded by the two expression cassettes. Although typical extraction protocol did not appear to yield much of the antibody in the total soluble proteins, the combination of urea and Triton X-100 resulted in over 40% of the total soluble protein being the modified antibody encoded in the plasmid. Fig. 22B depicts an exemplary gel that corresponds with the results of Fig. 22A. Fig. 22C compares fluorescence

generated from the plasmid of Fig. 19 with plasmid that comprises a single terminator (pBYR2eK2Mc-GFP, labeled as 2eK2Mc in the figure), which shows the double terminator construct produces (pBYR2eAK2Mc-GFP, labeled in the figure as 2eAK2Mc) brighter fluorescence than the single terminator construct.

Detailed Description

[0036] Detailed aspects and applications of the disclosure are described below in the following drawings and detailed description of the technology. Unless specifically noted, it is intended that the words and phrases in the specification and the claims be given their plain, ordinary, and accustomed meaning to those of ordinary skill in the applicable arts.

[0037] In the following description, and for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the various aspects of the disclosure. It will be understood, however, by those skilled in the relevant arts, that embodiments of the technology disclosed herein may be practiced without these specific details. It should be noted that there are many different and alternative configurations, devices and technologies to which the disclosed technologies may be applied. The full scope of the technology disclosed herein is not limited to the examples that are described below.

[0038] The singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “a step” includes reference to one or more of such steps.

[0039] As used herein, the term “expression cassette” refers to a distinct component of vector DNA, which contains gene sequences and regulatory sequences to be expressed by the transfected cell. An expression cassette comprises three components: a promoter sequence (part of the 5’ untranslated region, 5’ UTR), an open reading frame, and a 3’ untranslated region (3’ UTR). In some aspects, the regulatory sequences are found in the 5’ UTR and the 3’ UTR.

[0040] As used herein, the term “terminator” refers to a DNA sequence that causes the dissociation of RNA polymerase from DNA and hence terminates transcription of DNA into mRNA. Accordingly, while the term encompasses terminator sequences of known genes, the term also encompasses other sequences that perform the same function, for example, sequences around the short intergenic region of bean yellow dwarf virus.

[0041] The disclosure relates to 3' untranslated regions (UTRs), which in an expression cassette encoding a protein increases the expression level of the protein, and vectors for recombinant protein production in plants that utilize in at least one of its expression cassettes the 3' UTR disclosed herein. In some aspects, the plant expression vector is a replicating vector, for example a geminivirus vector. In other aspects, the plant expression vector is a non-replicating vector.

[0042] The plant expression vector described herein comprise at least one expression cassette, wherein the 3' UTR of the expression cassette comprises a single terminator or a double terminator. As used herein, a single terminator refers to a terminator element that contains one set of terminator sequences. As used herein, a double terminator refers to a terminator element that contains one set of terminator sequences fused with another set of terminator sequences. In some aspects, the expression cassette further comprises a chromatin scaffold/matrix attachment region (MAR). The MAR is downstream of the single terminator of the double terminator.

[0043] The vectors described herein results an increase in protein production (for example, as determined by the reporter gene GFP) compared to vectors using the most widely used terminators in the past 30 years, which include nopaline synthase (NOS) and octopine synthase (OCS) terminators from *Agrobacterium tumefaciens*, the 35S terminator from cauliflower mosaic virus (MacFarlane et al.,1992; Ellis et al., 1987; Pietrzak et al., 1986), and the terminator of soybean vegetative storage protein (VSP). In some embodiments, the increase in recombinant protein production is more than 5-fold, 7-fold, 10-fold, 15-fold, 20-fold, 25-fold, 30-fold, 40-fold, 50-fold, 60-fold, 100-fold, or 150-fold. In some aspects, the increased recombinant protein production is due increased stability of the transcripts. The benefits of the vectors described in herein is seen in a variety of plants (including, for example, tobacco and lettuce) and with a variety of recombinant proteins.

[0044] The 3' UTR regions that provide enhanced production of the recombinant protein are the extensin 3' UTR (also referenced herein as the extensin terminator), *Nicotiana benthamiana* actin 3' UTR (NbACT3), potato proteinase inhibitor II 3' UTR (Pin2), bean dwarf mosaic virus DNA B nuclear shuttle protein 3' UTR (BDB), *N. benthamiana* 18.8 kDa class II heat shock protein 3' UTR (NbHSP), bean dwarf mosaic virus rep gene 3' UTR (Rep), pea rubisco small subunit 3' UTR (RbcS), short intergenic region of bean yellow dwarf virus (SIR), *A. thaliana* heat shock protein 3' UTR (AtHSP), cauliflower mosaic virus 35S 3' UTR (35S), bean dwarf

mosaic virus repA gene 3' UTR (RepA), and agrobacterium nopaline synthase 3' UTR (NOS). The sequences of these 3'UTR are well-known in the art. In some implementations, the oligonucleotide sequences of these 3' UTRs for the synthesis of the vectors described herein are produced in the methods described in the Examples.

[0045] In some aspects, the nucleic acid sequence of the extensin terminator selected from the terminator sequences of the extensin gene in *Nicotiana tabacum*, *Nicotiana tomentosiformis*, *Nicotiana plumbaginifolia*, *Nicotinana attenuata*, *Nicotinana sylvestris*, *Nicotiana benthamiana*, *Solanum tuberosum*, *Solanum lycopersicum*, *Solanum pennellii*, *Capsicum annuum*, and *Arabidopsis thaliana*, the sequences of which are determinable from GenBank or the Sol Genomics Network. The nucleic acid sequence of the extension terminator comprises a polypurine sequence, an atypical near upstream element (NUE), an alternative polyA site, a far upstream element (FUE)-like region, a major NUE, and a major polyA region, and in certain embodiments, the nucleic acid sequence has at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79% identity to the sequence of the tobacco (*N. tabacum*) extension terminator. In some embodiments, the nucleic acid sequence of the extension terminator is that of the tobacco extensin gene. In certain embodiments, the portion of the extensin 3' UTR in the disclosed vector lacks the intron. In a particular embodiment, the 3' UTR region of the vector comprises an intronless tobacco extensin terminator (EU). Thus in some aspects, the nucleic acid sequence of EU spans nt 2764-3126 of the complete *N. tabcacum* gene for extensin (GenBank D13951.1). In certain other embodiments, the disclosed vector comprises intron-containing extensin terminator. Thus in some aspects, the 3' UTR region of the vector comprises an intron-containing tobacco extensin terminator (IEU). In such embodiments, the nucleic acid sequence of IEU spans nt 2396-3126 of the complete *N. tabcacum* gene for extensin (GenBank D13951.1).

[0046] In some aspects, the nucleic acid sequence of NbACT3 comprises nt 1460-1853 of actin gene (Gene ID Niben101Scf00096g04015.1). In some aspects, the nucleic acid sequence of NbACT3 comprises nt 33-1023 of the sequence set forth in SEQ ID NO. 8. In some aspects, the *N. benthamiana* actin 3' UTR is not the entirety of the 3' UTR, but only the downstream 617-nt region of NbACT3 (NbACT617). In such embodiments, the nucleic acid sequence of NbACT617 comprises nt 606-1023 of the sequence set forth in SEQ ID NO. 8. In other aspects,

the *N. benthamiana* actin 3' UTR is not the entirety of the 3' UTR, but only the downstream 567-nt region of NbACT3 (NbACT567).

[0047] In some embodiments, the nucleic acid sequence of Pin2 spans nt 1507-1914 of the potato gene for proteinase inhibitor II (GenBank: X04118.1). In some aspects, the sequence of pinII is obtained from pHB114 (Richter et al., 2000) by SacI-EcoRI digestion.

[0048] In some embodiments, the nucleic acid sequence of BDB comprises the 3' end of the nuclear shuttle protein, the intergenic region, the 3' end of the movement protein, and additional 200 nt downstream of the movement protein sequence (BDB501), which spans nt 1213-1713 of bean dwarf mosaic virus segment DNA-B (GenBank: M88180.1). In some embodiments, the nucleic acid sequence of BDB comprises only the 282 nucleotides that include the 3' end of the nuclear shuttle protein, the intergenic region, and the 3' end of the movement protein (BDB282).

[0049] In some embodiments, the nucleic acid sequence of NbHSP comprises the complement to nt 988867-989307 of the sequence of Gene ID Niben101Scf04040. In some aspects, the nucleic acid sequence of NbHSP spans nt 33-424, nt 33-447, nt 33-421, nt 33-453, nt 45-424, nt 45-447, nt 45-421, or nt 45-453 of the sequence set forth in SEQ ID NO. 7. In one embodiment, the nucleic acid sequence spanning nt 45-421 of the sequence set forth in SEQ ID NO. 7 is NbHSP. In embodiments, the nucleic acid sequence of NbHSPb comprises the complement to nt 988942-989307 of the sequence of Gene ID Niben101Scf04040. In some aspects, the nucleic acid sequence spanning nt 45-372 of the sequence set forth in SEQ ID NO. 7 is NbHSPb.

[0050] In some embodiments, the nucleic acid sequence of Rep comprises a sequence with at least 95%, preferably 99%, sequence identity to the complement of nt 859-1522 of bean yellow dwarf virus putative genes V1, V2, C1, C1:C2 (GenBank: Y11023.2). In some aspects, the sequence of Rep is set forth in SEQ ID NO. 14.

[0051] In some embodiments, the nucleic acid sequence of rbcS comprises a sequence that is complementary to the sequence spanning nt 6-648 of transient gene expression vector pUCPMA-M24 (GenBank: KT388099.1). In some aspects, the sequence of rbcS is obtained from pRTL2-GUS (Carrington et al., 1999) by SacI-EcoRI digestion.

[0052] In some embodiments, the 3' UTR comprises SIR, SIR with its additional downstream viral sequence (SIR 3'), or SIR with additional sequences both upstream and downstream (SIR 5'/3'). In some aspects, the nucleic acid sequence of SIR5'/3' comprises a

sequence with at least 95%, preferably 99%, sequence identity to nt 730-1966 of bean yellow dwarf virus putative genes V1, V2, C1, C1:C2 (GenBank: Y11023.2). In some embodiments, the sequence of SIR 5'/3' is set forth in SEQ ID NO. 11. In some aspects, the nucleic acid sequence of SIR 3' comprises a sequence with at least 95%, preferably 99%, sequence identity to nt 1155-1966 of bean yellow dwarf virus putative genes V1, V2, C1, C1:C2 (GenBank: Y11023.2). In some embodiments, the sequence of SIR 3' is set forth in nt 7-818 of SEQ ID NO. 10. In aspects, the nucleic acid sequence of SIR comprises nt 1122-1326 of bean yellow dwarf virus putative genes V1, V2, C1, C1:C2 (GenBank: Y11023.2). In some embodiments, the nucleic acid sequence of SIR is set forth in nt 4-208 of SEQ ID NO. 9.

[0053] In some embodiments, the nucleic acid sequence of AtHSP comprises nt 1-250 of the partial sequence of the *A. thaliana* heat shock protein 18.3 gene (GenBank KP008108.1). In some aspects, the nucleic acid sequence of AtHSP spans nt 7-257 of SEQ ID NO. 13.

[0054] In some embodiments, the nucleic acid sequence of 35S comprises a sequence spanning nt 3511-3722 of plant transformation vector pSITEII-8C1 (GenBank: GU734659.1). In some aspects, the sequence of 35S is set forth in nt 7-218 of SEQ ID NO. 2. In some aspects, the sequence of 35S is the sequence of the amplication of pRTL2-GUS (Carrington et al 1991) using the primers 35STm-1 (SEQ ID NO. 26) and 35STm-2 (SEQ ID NO. 27).

[0055] In some embodiments, the nucleic acid sequence of RepA comprises the complementary sequence to nt 859-1311 of bean yellow dwarf virus putative genes V1, V2, C1, C1:C2 (GenBank: Y11023.2). In some aspects, the nucleic acid sequence of RepA is set forth in nt 6-458 of SEQ ID NO. 15.

[0056] In some embodiments, the nucleic acid sequence of NOS comprises nt 22206-22271 of the T-DNA region of cloning vector pSLJ8313 (GenBank: Y18556.1). In some aspects, the sequence of NOS is that of the fragment obtained from pHB103 (Richter et al., 2000) by SacI-EcoRI digestion. In some aspects, the nucleic acid sequence of NOS is set forth in nt 6-261 of SEQ ID NO. 1.

[0057] In some embodiments, the 3' UTR region comprises at least one member from the group consisting of: EU, IEU, NbACT3, NbACT617, NbACT567, Pin2, BDB501, BDB282, NbHSP, NbHSPb, Rep, RbcS, SIR, SIR 5'/3', SIR 3', AtHSP, 35S, RepA, and NOS. In certain embodiments, the 3' UTR region of the vector consists of a terminator selected from the group consisting of: EU, NbACT3, Pin2, BDB501, NbHSP, Rep, RbcS, NbACT617, SIR 5'/3',

NbACT567, NbHSPb, and AtHSP. In some implementations, the 3' UTR region of the vector consists of a terminator selected from the group consisting of: EU, NbACT3, Pin2, BDB501, NbHSP, Rep, and RbcS.

[0058] In some aspects, the 3' UTR comprises two terminators, which produces a double terminator. The double terminator may be a repeat of same terminator or a combination of different terminators (for example, a fusion of two different terminators). In some embodiments, the double terminator consists of EU with NbACT, P19, NbHSP, SIR, NOS, 35S, tobacco mosaic virus 3' UTR (TMV), BDB501, tobacco necrosis virus-D 3' UTR (TNVD), pea enation mosaic virus 3' UTR (PEMV), or barley yellow dwarf virus 3' UTR (BYDV). In some aspects, the aforementioned pair of terminators are arranged where EU is arranged upstream of the other terminator, which is denoted as EU+NbACT, EU+P19, EU+NbHSP, EU+SIR, EU+NOS, EU+35S, EU+TMV, EU+BDB501, EU+TNVD, EU+PEMV, or EU+BYDV. In some embodiments, the double terminator consists of 35S with NbACT3, NOS, EU, NbHSP, Pin2, or BDB501. In some aspects, the aforementioned pair of terminators are arranged where 35S is arranged upstream of the other terminator, which is denoted as 35S+NbACT3, 35S+NOS, 35S+EU, 35S+NbHSP, 35S+Pin2, or 35S+BDB501. In some embodiments, the double terminator consists of IEU with SIR, 35S, or long intergenic region from bean yellow dwarf virus (LIR). In some aspects, the aforementioned pair of terminators are arranged where IEU is arranged upstream of the other terminator, which are denoted as IEU+SIR, IEU+35S, or IEU+LIR. In some embodiments, the double terminator consists of NbHSP with NbACT3, NOS, or Pin2. In some aspects, the aforementioned pair of terminators are arranged where NbHSP is upstream of the other terminator, which is denoted as NbHSP+NbACt3, NbHSP+NOS, or NbHSP+Pin2. In some embodiments, the double terminator consists of NOS with 35S, where NOS is arranged upstream of 35S (NOS+35S).

[0059] As used herein, the term "P19" refers to the P19 suppressor of RNAi silencing. An exemplary vector backbone that comprises P19 is pEAQ-HT (see Sainsbury et al., 2009).

[0060] In accordance with certain embodiments, the nucleic acid sequence of TMV spans nt 489-693 of the tobacco mosaic virus isolate TMV-JGL coat protein gene (GenBank: KJ624633.1). In some aspects, the nucleic acid sequence of TMV is set forth in nt 7-211 of SEQ ID NO. 21.

[0061] In accordance with certain embodiments, the nucleic acid sequence of TNVD has at least 85% identity, preferably 87% identity, to the sequence spanning nt 3457-3673 of the complete genome of tobacco necrosis virus D genome RNA (GenBank: D00942.1). In other embodiments, the nucleic acid sequence of TNVD has at least 90%, preferably 93%, sequence identity with nt 3460-3673 of tobacco necrosis virus-D genome (GenBank: U62546.1). In some embodiments, the nucleic acid sequence of TNVD comprises the sequence set forth in nt 29-222 of SEQ ID NO. 19.

[0062] In accordance with certain embodiments, the nucleic acid sequence of PEMV has at least 95%, preferably 98%, sequence identity with nt 3550-4250 of the pea enation mosaic virus-2 strain UK RNA-dependent RNA-polymerase, hypothetical protein, phloem RNA movement protein, and cell-to-cell RNA movement protein genes (GenBank: AY714213.1). In some aspects, the nucleic acid sequence of PEMV is set forth in nt 1-703 of SEQ ID NO. 20.

[0063] In accordance with certain embodiments, the nucleic acid sequence of BYDV has at least 95%, preferably 99%, sequence identity with nt 4807-5677 of barley yellow dwarf virus - PAV genomic RNA (GenBank: X07653.1). In some aspects, the nucleic acid sequence of BYDV is set forth in nt 5-875 of SEQ ID NO. 18.

[0064] In another embodiment, the vector further comprises at a chromatin scaffold/matrix attachment region (MAR) downstream of the region comprising the at least one terminator. In a preferred embodiment, the MAR is the Rb7 MAR (GenBank: U67619.1) or the TM6 enhancer region (GenBank: KC5555564.1). As used herein, the term “Rb7” refers to a sequence comprising the sequence of GenBank ID U67619.1or set forth in nt 7-1174 of SEQ ID NO. 16. As used herein, the term “TM6” refers to a sequence comprising the sequence of GenBank ID KC5555564.1 or set forth in nt 10-1202 of SEQ ID NO. 17. Accordingly, in some implementations, the vector comprises the terminator EU in combination with Rb7, the terminator IEU with Rb7 or TM6, the terminator AtHSp with Rb7, the terminator 35S with Rb7 or TM6, the terminator BDB501 with Rb7, the terminator NbHSP with Rb7, the terminator NOS with Rb7, or the terminator NbACT3 with Rb7 or TM6.

[0065] In certain embodiments, the vector comprises a double terminator and a MAR, wherein the MAR is downstream of the double terminators. In some implementations, the MAR is Rb7, and it is downstream of the double terminators 35S+NbACT3, EU+35S, EU+NbACT3, NbHSP+NbACT3, 35S+EU, AtHSP+NOS, 35S+NOS, EU+BDB501, AtHSP+NbHSP,

NbHSP+NOS, AtHSP+EU, NbHSP+Pin2, or IEU+35S. In other implementations, the MAR is TM6, it is downstream of the double terminators EU+35S, 35S+NOS, NbHSP+NOS, or NbHSP+Pin2.

[0066] The disclosure is also related to oligonucleotides for the production of disclosed vectors. SEQ ID NOs.1-21 provides the nucleic acid sequences for incorporating the aforementioned 3' UTRs into vectors. The nucleic acid sequence of the template for incorporating NOS is set forth in SEQ ID NO. 1. The nucleic acid sequence of the template for incorporating 35S is set forth in SEQ ID NO. 2. The nucleic acid sequence of the template for incorporating pinII is set forth in SEQ ID NO. 3. The nucleic acid sequence of the template for rbcS is set forth in SEQ ID NO. 4. The nucleic acid sequence of the template for incorporating IEU is set forth in SEQ ID NO. 5. The nucleic acid sequence of the template for incorporating EU is set forth in SEQ ID NO. 6. The nucleic acid sequence of the template for incorporating NbHSP is set forth in SEQ ID NO. 7. The nucleic acid sequence of the template for incorporating NbACT3 is set forth in SEQ ID NO. 8. The nucleic acid sequence of the template for incorporating SIR is set forth in SEQ ID NO. 9. The nucleic acid sequence of the template for incorporating SIR 3' is set forth in SEQ ID NO. 10. The nucleic acid sequence of the template for incorporating SIR 5'/3' is set forth in SEQ ID NO. 11. The nucleic acid sequence of the template for incorporating BDB501 is set form in SEQ ID NO. 12. The nucleic acid sequence of the template for incorporating AtHSP is set forth in SEQ ID NO. 13. The nucleic acid sequence of the template for incorporating Rep is set forth in SEQ ID NO. 14. The nucleic acid sequence of the template for incorporating RepA is set forth in SEQ ID NO. 15. The nucleic acid sequence of the template for incorporating Rb7 MAR is set forth in SEQ ID NO. 16. The nucleic acid sequence of the template for incorporating TM6 MAR is set forth in SEQ ID NO. 17. The nucleic acid sequence of the template for incorporating barley yellow dwarf virus's (BYDV's) 3' UTR is set forth in SEQ ID NO. 18. The nucleic acid sequence of the template for incorporating TNVD 3' UTR is set forth in SEQ ID NO. 19. The nucleic acid sequence of the template for incorporating PEMV 3' UTR is set forth in SEQ ID NO. 20. The nucleic acid sequence of the template for incorporating tobacco mosaic virus 3' UTR is set forth in SEQ ID NO. 21.

[0067] The disclosure is further related to methods of producing recombinant protein in a plant or plant part. In some aspects, the method produced at least 5-fold, 7-fold, 10-fold, 15-fold, 20-fold, 25-fold, 30-fold, or 40-fold yield of the recombinant protein than methods of the prior

art. The method comprises introducing a vector described above into the plant or plant part. In some implementations, the plant or plant part is transformed by the vector of the disclosure using an Agrobacterium, for example, *Agrobacterium tumefaciens*, or more specifically, *A. tumefaciens* GV3101. In one aspect, the plant or plant part is transformed by the vector of the disclosure using agroinfiltration. In one implementation, the plant is tobacco or tomato while the plant part is from a tobacco plant or tomato plant.

Illustrative, Non-Limiting Examples in Accordance with Certain Embodiments

[0068] The disclosure is further illustrated by the following examples that should not be construed as limiting. The contents of all references, patents, and published patent applications cited throughout this application, as well as the Figures, are incorporated herein by reference in their entirety for all purposes.

1. Tobacco Extensin (Ext) Terminator

a. Influence of the Ext intron on gene expression

[0069] The Ext terminator consists of 746 nt and contains an intron between nt 24 and 249. To characterize the activity of the Ext terminator with and without the intron, different forms were cloned into Agrobacterium T-DNA vectors. Intron-containing Ext terminator constructs were generated (Fig. 1B): the Ext terminator segments from nt 1-731 or nt 1-464, including the intron sequence (nt 24-249), were PCR amplified from tobacco genomic DNA and fused to the GFP gene under the control of the strong 35S promoter from cauliflower mosaic virus (pIEU or pIEU2, respectively). Similar constructs were created with the intron deleted, using nt 252-731 or 252-464 (pEU or pEU2, respectively). These constructs were then compared by Agrobacterium-mediated transient expression in *N. benthamiana* leaves. The intron-less constructs pEU and pEU2 showed a significantly higher level of GFP expression compared to the intron-containing construct pIEU and pIEU2 (~3-fold, Fig. 2A). To test whether the effects of removing the Ext intron were gene-specific, the GFP gene was replaced with the NVCP gene. NVCP is a candidate vaccine antigen for the protection against Norwalk virus infections, which cause epidemic acute gastroenteritis in humans. The NVCP expression of the intron-containing construct, pIEU or pIEU2, was also lower than that of the intron-less constructs, pEU or pEU2 (~29-75% activity remained, Fig. 2B), though the magnitude of the inhibitory effect of the intron

was not as great with pIEU. From the intron-containing constructs, either with GFP or NVCP gene, efficient splicing was confirmed by RT-PCR as no detectable unspliced product was found in ethidium bromide-stained agarose gel electrophoresis. The intron also greatly reduced expression of DsRed. To determine whether the inhibitory effect of the intron was species-specific, pEU and pIEU were agroinfiltrated into tobacco and lettuce leaves. In agreement with our results obtained with *N. benthamiana*, the presence of the intron substantially reduced gene expression in both species, with 34% activity remaining in lettuce and 14% remaining in tobacco (Fig. 9). These data indicate that the Ext intron may have a deleterious effect on gene expression, in agreement with previously reported findings for introns inserted into 3' UTRs (Kertész 2006). However, we cannot exclude the possibility that the effect was due to the incorporation of the 23 nt (1-23 nt upstream of the intron) in the intron-containing constructs.

[0070] It has been reported that the effect of intron insertion changes in a context-dependent manner (Kertész 2006). The effect of the Ext intron was also tested in context of the NOS terminator: PCR-amplified Ext intron (1-251 nt) was fused to the 5' end of the NOS terminator preceded by the GFP or NVCP gene. Unexpectedly, the addition of the Ext intron to NOS terminator caused slight but statistically insignificant increases in GFP or NVCP expression by 51% and 34%, respectively (Figs. 2A and 2B, pNOS vs. pINOS). Taken together, these data indicate that the Ext intron has context-dependent effects on transient transgene expression.

b. Ext terminator increases transient transgene expression

[0071] We evaluated the effects of the tobacco Ext terminator on transient transgene expression in comparison to other widely used terminators, including NOS, CaMV 35S, and soybean vegetative storage protein (VSP). For this comparison, we placed the intron-less Ext terminator (nt 252-731) and the other terminators downstream of the GFP gene, driven by the CaMV 35S promoter with the tobacco etch virus (TEV) 5' UTR (Fig. 1A). The resulting constructs were introduced into *N. benthamiana* leaves by agroinfiltration, and at 2 DPI the level of GFP expression was analyzed. Expression peaked between 2-3 DPI and declined thereafter, likely due to gene silencing (data not shown). The Ext construct (pEU) yielded the highest GFP expression level, at 13.5-fold, 11.9-fold, and 2.8-fold higher than those with the NOS, VSP, and 35S constructs (pNOS, pVSP, and p35S), respectively (Fig. 3A). To test whether the enhancing effect of the intron-less Ext terminator is gene specific, we replaced the GFP gene with the

NVCP gene and compared the level of NVCP expression. Results using NVCP were similar to those with GFP, suggesting that the enhanced transgene expression by the Ext terminator is not gene specific (Fig. 3B). To determine whether the Ext terminator also functions without TEV 5' UTR, we directly fused the 35S promoter to a GUS reporter gene, followed by the various terminators (Fig. 1A). We found that the Ext terminator increased expression of GUS without TEV 5' UTR (Fig. 3C), suggesting that the ability of the Ext terminator to increase expression is independent of the specific 5' UTR and transgene. To assess whether these results are generalizable to other plant species, we tested the extensin, NOS, and 35S terminators in tobacco and lettuce. In agreement with our results in *N. benthamiana*, the intronless extensin terminator strongly enhanced transgene expression compared to either NOS or 35S, but the magnitude of the enhancing effect was slightly reduced in lettuce (Fig. 9). These results show that the intronless extensin terminator functions in a variety of different species and gene contexts as a potent enhancer of transgene expression compared to other commonly used terminators.

c. Ext terminator increases mRNA accumulation

[0072] The 3' UTR influences the fate of mRNA through a complex interplay of multiple nuclear and cytoplasmic processes, including polyadenylation, transcript termination, transcript reinitiation, nuclear export, and translatability, as well as by avoiding deleterious interactions with RNA silencing and mRNA decay pathways. The upregulated transgene expression mediated by the intronless Ext terminator could be caused by an increase in either mRNA level or translational efficiency. To investigate whether the Ext terminator affects mRNA accumulation, the levels of accumulated transgene mRNAs were compared. Construct pEU produced approximately 20-fold increase in GFP mRNA accumulation compared to construct pNOS (Fig. 4A). Consistently, NVCP mRNA accumulation using construct pEU was approximately 10-fold higher compared to that of construct pNOS (Fig. 4B). These results indicate that the enhanced transgene expression mediated by the Ext terminator is due at least in part to increased mRNA accumulation.

[0073] We also compared the levels of mRNA accumulation between intronless and intron-containing constructs (pEU vs. pIEU). The use of the Ext intron caused a 40-50% decrease in mRNA accumulation (Figs. 4A and 4B), which was consistent with the protein expression data (Figs. 2A and 2B).

d. Ext terminator mediates efficient transcription termination.

[0074] The stability of mRNA is greatly affected by the efficiency of transcription termination and mRNA 3' end processing. Improperly terminated and unpolyadenylated mRNA is targeted by RNA-dependent RNA polymerase 6 (RDR6)-mediated RNA silencing (Luo and Chen 2007) and long 3' UTRs are subject to the nonsense-mediated decay pathway (Kertész 2006). To determine if the Ext terminator increases the efficiency of transcription termination, we tested for the presence of read-through transcripts from Ext and NOS terminator constructs using RT-PCR with random primed cDNA (Fig. 5A). A forward primer specific to either the GFP (Fig. 5A) or NVCP gene (Fig. 5C) (gfp-3f or sNV-3f) was paired with one of four antisense primers (RT-0, RT- 1, RT-2, and RT-3) specific to different downstream regions of the terminator. From the NOS terminator samples, specific RT-PCR bands of the same size as the positive controls were detected with the first three reverse primers, RT-0, RT-1, and RT-2, indicating that the NOS terminator generated readily detectable read-through transcripts when driven by the strong CaMV 35S promoter. In contrast, no detectable levels of specific bands were amplified in the Ext terminator samples with any primer set. These data demonstrate that read-through transcription is greatly reduced by the Ext terminator compared to the NOS terminator.

e. Polyadenylation Sites of the Extensin Terminator

[0075] Positions of five putative NUEs and polypurine sequence (PPS) of the Ext terminator are indicated in Figure 6. To characterize the location and features of the poly(A) sites of the Ext terminator, transgene transcripts were analyzed by circularized RT-PCR, a method which can detect the native position and length of poly(A) tails (Slomovic and Schuster 2013). GFP transcripts were decapped, circularized, RT-PCR amplified using primers specific for circularized templates, cloned, and sequenced. The data show that the Ext terminator contains a region of two nearby poly(A) sites (12 out of 14 sequences, 86%) at a YA dinucleotide either 13 or 25 nt downstream from the fourth putative NUE, a canonical AAUAAA motif (Fig. 6). Of the 12 polyadenylated transcripts detected in this region, 9 occurred at nt 527, while the other three occurred at nt 515. As multiple cleavage sites are typically found in the same area downstream from a single NUE, these data are consistent with the hypothesis that the fourth NUE directs

most of the polyadenylation in the Ext terminator (Rothnie et al 2001). In support of the functional relevance of this region, homologous extensin 3' UTRs from plant species with available sequences (obtained from GenBank of the Sol Genomics Network) were aligned using the Clustal Omega Program. The extensin terminators of *N. tabacum*, *N. tomentosiformis, N. plumbaginifolia*, *N. attenuate*, *N. sylvestris*, *N. benthamiana*, *S. tuberosum*, *S. lycopersicum*, *S. pennellii*, *C. annuum*, and *A. thaliana* were compared, and the alignment revealed that the fourth NUE is highly conserved.

[0076] Alternative polyadenylation sites have been found in more than 50% of all eukaryotic genes (Tian and Manley 2013). An infrequently utilized alternative poly(A) site (14%, 2 out of 14) was found at position 424, between the third and fourth putative NUEs. This site was also highly conserved among related extensin terminators. As NUEs have been shown typically to reside 10-40 nucleotides upstream from the polyadenylation site (Loke et al 2005), and the closest typical NUE is 70 nt upstream from the alternative polyadenylation site, our results suggest an atypical NUE may direct polyadenylation at this site, consistent with its infrequent utilization. We found no evidence that the other putative NUEs were functionally active, though we cannot exclude the possibility that they are infrequently utilized.

f. Deletion analysis of the Ext terminator

[0077] Using a series of deletion constructs depicted in Fig. 1B, we investigated whether specific regions of the Ext terminator controlled its function by measuring transient expression of GFP (Fig. 7). Compared to the full-length construct pEU (nt 252-731), construct pEU2 (nt 252-464) removed the major polyadenylation site identified by cRT-PCR. Deletion construct pEU2 exhibited a substantially lower level of GFP gene expression compared to pEU (11-14% activity remained). These data indicate that the 89-bp sequence (nt 465-553) is essential to yield maximum gene expression. This result is consistent with the polyadenylation data that the fourth putative NUE directs most of the polyadenylation of the Ext terminator. The residual activity of this deletion construct may result from the infrequently utilized alternative polyadenylation site, which remained intact in this construct. To further confirm the role of the fourth NUE in the function of the Ext terminator, we created construct pNUE4a, which substitutes the A-rich region "AATAAACTAA" hypothesized to serve as the NUE, with "TCGTAGCTCT." This substitution greatly reduced GFP expression to a similar level compared to pEU2 (Fig. 7).

[0078] The deletion construct pEU1 (nt 252-553) retains all five NUEs, but removes the area immediately downstream from the putative major cleavage site, including a U-rich region typical of CEs. Compared to the full-length construct pEU, expression of pEU1 was substantially reduced (~36% activity remained), suggesting that this region is essential for optimum efficiency of the extensin terminator.

[0079] Deletion construct pEU3 (nt 465-731) was designed to include the fourth and fifth NUEs, but removes their upstream regions. Construct pEU3 showed no detectable level of GFP expression, indicating that the identified major polyA site, its putative NUE, and its downstream sequence alone were not sufficient to support gene expression. Inspection of this sequence reveals an UG-rich region, typical of plant FUEs that is well conserved among other plant extensin terminator sequences. This region also contains the alternative polyadenylation site. Therefore, both polyadenylation sites may have been disrupted by this deletion. To further confirm this result, construct pEU4 (nt 349-731) was tested. pEU4 contains the fourth and fifth NUEs, but additionally contains the larger FUE-like region. While this construct had restored activity similar to pEU2, it was still drastically reduced from the full-length construct, indicating that the first 97 nucleotides of the intronless extensin terminator are also essential for high levels of gene expression. Collectively, these data indicate that multiple regions of the Ext terminator are essential for its function, including canonical FUE, NUE, and CE regions, but also a region containing an atypical polypurine-rich region sequence ~175 nt upstream from the major polyadenylation sites. To determine whether this region contributed to the enhancement provided by the extensin terminator, we deleted or substituted the polypurine sequence and tested its effects on GFP expression (Figs. 8A and 8B). For the substitution, some A residues were replaced with T, while maintaining the segment length. Upon deletion or substitution, the levels of GFP expression were substantially decreased (36-40% activity remained), yet they were still 4.7-5.2-fold higher than with the construct using the NOS terminator. These data indicate that the polypurine sequence is also an important element contributing to the enhancing effect of the Ext terminator.

g. Discussion

[0080] Pre-mRNA 3' end processing is an essential step in the expression of genes and is greatly affected by transcription terminators. Hence, careful selection of an optimal terminator is

important for high yield recombinant protein production. Extensins are highly abundant components of the plant cell wall. Additionally, a genome-wide analysis of mRNA stability in *A. thaliana* found that extensin mRNAs were often highly resistant to degradation (Narsai et al 2007). These factors, as well as the unusual presence of an intron, led us to evaluate the tobacco Ext terminator for its potential to enhance transgene expression. When the native intron was removed, the Ext terminator produced higher levels (up to 13.5-fold) of transient transgene expression from three transgenes, GFP, NVCP, and GUS, compared to other commonly used gene terminators. Interestingly, the three transgenes showed similar expression level patterns in coordination with the tested terminators; i.e., all transgenes showed the highest expression level with the Ext terminator, followed in order by the 35S, VSP, and NOS terminators (Figs. 3A-3C).

[0081] However, the magnitude of the difference varied between different transgenes, suggesting that if gene-specific effects exist, their effects are small compared to the intrinsic efficiency of each terminator. The NOS terminator is known to contain a cryptic poly(A) site that is only functional if a heterologous FUE is present upstream, which could be provided by some transgenes but not others (Sanfaçon et al 1994). In agreement with our results in *N. benthamiana*, we found that the intronless extensin terminator strongly outperformed the 35S and NOS terminators in both its native tobacco and in lettuce, although again the magnitude of the enhancement varied. We did not directly address the effect of the 5' UTR on the efficiency of the Ext terminator, but the enhancing effect of the Ext terminator was not obviously affected by the 5' UTR, considering that the GFP and NVCP expressing vectors contained the TEV 5' UTR, while the GUS constructs did not. Additionally, we have found the Ext terminator to support very high levels of transgene production when used in conjunction with diverse plant, viral, and human 5' UTRs (Diamos et al 2016). These results demonstrate that the intronless extensin terminator is a highly effective terminator in multiple plant species and transgene contexts.

[0082] In addition to terminators, introns can also significantly affect gene expression. For example, some introns boost expression by containing enhancer elements or through a poorly defined process termed intron-mediated enhancement (IME) (Rethmeier et al 1997; Rose 2008). Conversely, some introns are required for tissue specific or developmentally restricted gene expression: introns in the *Arabidopsis agamous* (AG) gene and the Seedstick (STK) gene restrict AG and STK expression to specific tissues within the flower (Sieburth and Meyerowitz 1997;

Kooiker et al 2005); and an intron in the Arabidopsis floral repressor Flowering Locus C (FLC) gene reduces FLC expression in response to vernalization (Sheldon et al 2002). It has been observed that the expression of Ext is regulated tissue-specifically and developmentally and is induced by various stress conditions (Showalter et al 2010; Hirsinger et al 1997). The presence of an intron in the 3' UTR of the Ext gene and its regulated expression might imply the involvement of the intron in the regulated Ext expression. Additionally, plant 3' UTRs containing introns have been shown to activate the nonsense-mediated decay pathways, sometimes resulting in decreased mRNA accumulation (Kertész 2006). We showed that the Ext intron might have regulatory function on transient transgene expression in a context-dependent manner.

[0083] Specifically, the intron produced a deleterious effect (up to 70% decrease) when present in the Ext terminator but showed a slight increase when used in combination with NOS terminator (Figs. 2A-2B). Consistent with these data, Menossi et al. (2003) reported the tissue dependent regulatory function of Maize HRGP gene 3' UTR: it increased GUS expression in fast-growing coleoptiles, but decreased GUS expression in maize suspension cells. Additionally, the extent to which nonsense-mediated decay is activated by the presence of an intron in the 3' UTR varies in a context-dependent manner (Kertész 2006). Further studies of the Ext intron may provide valuable insight into the intron-mediated regulation of gene expression.

[0084] We showed that the Ext 3' UTR substantially reduces the level of readthrough transcripts compared to other terminators (Figs. 5A-5C). Although limited transcription readthrough (~1% of transcripts) is a normal phenomenon of plant gene expression (Xing et al 2010), there are strong data indicating that readthrough transcripts have negative effects on gene expression.

[0085] For example, readthrough transcripts can trigger RDR6-mediated RNA silencing, which leads to a cascade of mRNA degradation (Luo and Chen 2007). Best studied in this regard is the GUS transgene expression in A. thaliana. A GUS transgene without a terminator produced readthrough mRNA and consistent RDR6-dependent RNA silencing. However, this phenomenon was resolved when two 3' terminators were placed 3' of the GUS transgene: the level of readthrough transcripts and GUS-specific small interfering RNA were decreased, resulting in higher GUS expression. Enhancement of transient gene expression from minimal cassettes using a double terminator was also suggested to be correlated with the reduction in aberrant RNA

formation and therefore prevention of the triggering of post transcriptional gene silencing (PTGS) via the RDR6 pathway (Beyene G et al 2011). This mechanism was further characterized by Baeg et al (2017), showing that aberrant mRNAs lacking polyA tails are specifically selected for amplification by RDR6.

[0086] Mapendano et al. (2010) demonstrated another possible mechanism whereby readthrough transcripts mediated down regulation of gene expression. When the poly(A) signal was mutated, RNA polymerase II (RNAPII) complexes were found in the readthrough region of the mutated gene. This readthrough RNAPII, engaged on the mutated transcription units, sequesters the transcription initiation/elongation factors, and leading to their depletion at the promoter, thereby preventing transcript reinitiation. Based on these references and our data showing that the Ext terminator decreases the level of readthrough transcripts and increases mRNA accumulation, we hypothesize that the Ext terminator increases transgene expression at least in part by preventing the activation of gene silencing by RDR6 and/or by stimulating continued transcription.

[0087] The Ext 3' UTR contains five putative NUEs (Fig. 6). Sequence analyses of the 3' end of transgene transcripts revealed that most transcripts (86%) are polyadenylated at one of two YA dinucleotides shortly after the fourth NUE, which consists of a canonical AAUAAA motif. The fourth NUE resides 25 nt upstream from the cleavage site, which is typical of reported plant NUEs. While the AAUAAA NUE motif is very common in mammals and is still the single most abundant NUE motif identified in plants, it has been found to direct polyadenylation in only 10% of Arabidopsis transcripts (Loke et al 2005). Phylogenetic analysis showed that the general structure of the Ext terminator is found among many *Nicotiana* species, but in particular the region between the intron and the major polyA site is well conserved in *S. tuberosum* (79% identity), *C. annum* (76% identity), *S. lycopersicum* (73% identity), and *S. pennellii* (72% identity). Alignment of this region revealed that the AAUAAA motif was strongly conserved, supporting our data indicating that it functions as the dominant polyA site. In addition to this site, we found that a minority of Ext terminator transcripts were polyadenylated > 60 nt downstream of the third putative NUE (also AAUAAA). However, the typical positioning of a NUE is 10-40 nt upstream from the poly(A) site, suggesting that an atypical NUE may instead direct its polyadenylation. Plant NUEs are generally A/AU-rich, but they tolerate significant degeneracy (Mogen et al 1991; Loke et al 2005). Thus, an unusual motif, such as the "AAGUUA" located 23

nt upstream from the infrequently utilized poly(A) site, may inefficiently direct 3'-end processing in a minority of Ext transcripts. This alternative polyadenylation site was also strongly conserved among related genes, suggesting alternative polyadenylation may be a general feature of plant extensins.

[0088] We performed a deletion analysis to identify which regions of the Ext terminator are necessary for its function (Fig. 7). The importance of the identified major polyadenylation site was supported by deletion construct pEU2, which showed a drastic, but not complete, reduction of Ext activity when the region containing the fourth NUE was removed. However, pEU2 retained the infrequently used polyadenylation site at nt 424 and its upstream sequence, which may explain its residual activity despite loss of the dominant polyadenylation site. These findings suggest that although an alternative polyadenylation site exists in the Ext terminator, it cannot compensate for the loss of the dominant polyadenylation site.

[0089] Plant polyadenylation requires the coordinated action of multiple regions upstream and downstream from the cleavage and polyadenylation site. Construct pEU1, which deleted the region downstream from the polyadenylation site, had greatly reduced expression (Fig. 7), in agreement with previous studies showing that U-rich regions downstream from the cleavage site are important for proper terminator function (Dong et al 2007). pEU3 included the dominant polyadenylation sites and their respective NUE, and yet had no detectable activity. Inspection of the upstream sequence removed in construct pEU3 revealed a UG-rich region located ~60-80 nt upstream from the cleavage site, typical of plant FUEs (Loke et al 2005), which was generally conserved across other homologous Ext genes. While including this putative FUE in pEU4 restored some of the activity of the terminator (~16%), our data clearly show that a region ~175 nt upstream from the identified polyA sites is also essential for optimum functioning of the extensin terminator.

[0090] The extensin terminator contains an unusual 45 nt polypurine sequence (PPS) in this upstream region that strongly affects its function. Deletion or substitution of the PPS caused a large ~60% decrease in GFP expression (Fig. 8b). Despite the enhancement provided by the PPS, the deletion construct lacking the PPS still outperformed all other tested terminators, indicating that the PPS alone does not account for the remarkable enhancement provided by the Ext terminator. The PPS contains a potential NUE, however no poly(A) sites were found in proximity to the PPS, suggesting that this NUE is not active. Additionally, the PPS shares no

similarity to other known terminator elements, suggesting that its enhancing effect is due to an as of yet unidentified mechanism. The PPS is present at varying lengths in many *Nicotiana* species' Ext genes, however it is substantially shorter in the Ext genes from other sequenced plant species. Additionally, even among Nicotiana extensin terminators, the PPS is particularly long in tobacco. Further work is needed to determine whether the PPS is functional in Ext genes from other plant species, and whether the length of the PPS is an important contributing factor to its enhancing activity.

[0091] In conclusion, we find that the production of recombinant proteins in plants is enhanced by use of the Ext terminator, which increases transgene expression up to 13.5-fold compared to commonly used terminators when its naturally occurring intron is removed. The increased gene expression was associated with a reduction in readthrough transcription, and an increase in mRNA accumulation. The high activity of the intronless Ext terminator requires a NUE region consisting of a canonical AAUAAA motif 25 nt upstream from the dominant poly(A) site and surrounding CE, as well as a FUE-like UG-rich region located 60-80 nt upstream of the cleavage site, as well as an unusual polypurine rich region ~200n t upstream of the cleavage site. An infrequently utilized alternative polyadenylation site was also identified. This work highlights the importance of the terminator in controlling gene expression, and we anticipate that the enhancing effect of the Ext terminator will be broadly applicable to plant-based recombinant protein expression systems.

2. Other terminators and chimeric 3' flanking regions

a. Evaluation of Diverse Terminators on GFP Production

[0092] To systematically evaluate diverse terminators, we constructed expression vectors using 20 different terminators from plant and viral sources, placed 3' of a GFP reporter gene, which was driven by the strong 35S promoter and tobacco mosaic virus 5' UTR (Fig. 10A). These constructs were delivered to *N. benthamiana* leaves by agroinfiltration and evaluated for GFP production. To minimize leaf-to-leaf variation, each leaf was also infiltrated with a GFP vector containing the intronless tobacco extensin terminator (EU) as an internal control, which we previously found to be a potent enhancer of gene expression (Diamos et al., 2016; Rosenthal et al., 2018)

[0093] While much previous work used the NOS and 35S terminators, the *A. thaliana* 18.2 kDa heat shock protein terminator (AtHSP) was reported to enhance transgene production compared to the NOS terminator (Nagaya et al., 2010). In agreement with these results, we found that the AtHSP terminator provided a 2.5-fold increase in GFP production compared to the NOS terminator (Fig. 10B). Also consistent with previous work in transgenic potato (Richter et al., 2000), the potato pinII 3' UTR provided a very strong 8.5-fold increase compared to the NOS terminator (Fig. 10B). The rubisco small subunit (rbcS) 3' UTR from pea showed a 5.4-fold enhancement. These data demonstrate that many terminators from diverse plant species have high activity in *N. benthamiana*.

[0094] To identify new candidates, a genome-wide study of mRNA stability levels in *A. thaliana* (Narsai et al., 2007) was used to locate genes with potentially stability-enhancing 3' UTRs. We identified a *N. benthamiana* homolog of the *A. thaliana* 17.6 kDa class II heat shock protein (At5g12020). The 3' flanking region from this homolog (referred to as NbHSP) was highly active, increasing reporter gene expression by 6.3-fold compared to the NOS terminator, more than doubling the enhancement provided by the AtHSP terminator (Fig. 10B). We also identified an N. benthamiana homolog of A. thaliana actin 7 (At5g09810), referred to as NbACT3. While the downstream 617-nt region of NbACT3 enhanced expression by 3.9-fold compared to NOS, extending the 3' UTR to include more downstream sequence (1044 nt) resulted in a large 8.9-fold enhancement (NbACT3, Fig. 10B).

[0095] Many of the most highly active genetic elements in recombinant protein production systems are derived from viral sources. Therefore, we investigated the potential of viral terminators to enhance gene expression. The downstream short intergenic region (SIR) from the coat protein gene of bean yellow dwarf virus (BeYDV) showed no intrinsic terminator function by itself, or when additional downstream viral sequence (SIR 3') was included (Fig. 10B). However, when an upstream region (SIR 5'/3') from the bean yellow dwarf virus coat protein coding sequence was also included, it was highly functional, providing GFP production 3-fold greater than the NOS terminator (Fig. 10B). These data suggest that upstream elements present in the BeYDV coat protein gene are required for proper 3' end processing. The downstream sequences from the BeYDV rep and repA genes were also found to be highly active, providing a 5.6-fold and 2-fold respective enhancement. To test sequences from other geminiviruses, the terminators of bean dwarf mosaic virus (BDMV) genes were also investigated. A 282 nt

sequence including the 3' end of the nuclear shuttle protein, the intergenic region, as well as the 3' end of the movement protein (BDB282) performed similarly to the NOS terminator. However, when an additional 200 nt of the downstream movement protein sequence was included (BDB501), it provided a 6.4-fold enhancement. A construct containing the BDMV coat protein downstream sequence alone was not functional, again suggesting that necessary signals may also reside in the gene coding sequence upstream from the terminator.

[0096] Taken together, these results show that many 3' UTRs from diverse sources exceed the enhancement provided by the commonly used NOS or 35S terminators, at least in a transient expression system in *N. benthamiana* leaves. Consistent with our previous work, the EU terminator outperformed the other 19 3' UTRs tested, providing a 13.6-fold increase compared to the NOS terminator, indicating that it is a uniquely potent enhancer of gene expression.

b. Combined Gene Terminators Strongly Enhance GFP Production

[0097] A double terminator consisting of the 35S terminator fused to the NOS terminator greatly enhanced protein production in various plant species compared to either terminator alone (Beyene et al., 2011). To investigate the potential for tandem terminators to synergistically enhance recombinant protein production, we tested combinations of those previously tested in Fig. 1B. We found an 18.4-fold enhancement by the 35S-NOS double terminator compared to the NOS terminator alone (Fig. 11), exceeding the highest production by the best single terminator. Interestingly, reversing the position of the two terminators (NOS-35S) provided a much lower 11.2-fold enhancement, but still greatly exceeded the GFP production obtained with either terminator alone.

[0098] Fusion of 35S with pinII, NbHSP, and BDB501 3' regions all substantially enhanced protein production compared to either terminator alone. However, despite the individual superiority of each of these terminators compare to NOS, when paired with 35S, none exceeded the GFP production of 35S-NOS (Fig. 11, compare 35S-Pin2, 35S-NbHSP, 35S-BDB501, and 35S-NOS). In contrast, fusion of the individually strong NbACT3 terminator to the 3' end of either the 35S or NbHSP terminators resulted in a potent enhancement of GFP production, exceeding the production of 35S-NOS (Fig. 11).

[0099] As the extensin EU terminator was the best individual terminator identified, we evaluated its potential combined with other terminators. Addition of either the NbHSP, NOS, or

35S terminators to the 3' end of the EU terminator nearly doubled the GFP production provided by EU alone, exceeding the gene expression provided by 35S-NOS. The two best individual terminators, EU and NbACT3, when combined, exceeded all other combinations, providing a remarkable 37.7-fold increase compared to NOS alone (Fig. 11, EU + NbACT3). Interestingly, although the 35S terminator performed best when placed 5' of the NOS terminator, the opposite was found when paired with EU: the enhancement provided by 35S-EU was significantly lower than EU-35S (Fig. 11). Furthermore, addition of BDB501 to EU resulted in a slight decrease in expression (Fig. 11). Therefore, these results indicate that terminators placed in tandem interact either synergistically or antagonistically, in a context-dependent manner.

[00100] Previously, we found that the 5' and 3' UTRs from the RNA viruses barley yellow dwarf virus (BYDV) and pea enation mosaic virus (PEMV) severely inhibited expression in N. benthamiana leaves using a replicating system containing the extensin terminator (Diamos et al., 2016). A non-replicating expression system based on the 5' and 3' UTRs from cowpea mosaic virus (CMPV) was reported to enhance gene expression largely due to incorporation of the viral 3' UTR before the NOS terminator (Sainsbury and Lomonossoff, 2008; Meshcheriakova et al., 2014). In this study, we evaluated virus-derived 3' UTRs in nonreplicating vectors. Similar to our results with replicating vectors, we found that the 3' UTRs from PEMV, BYDV, and tobacco necrosis virus D strongly inhibited gene expression when inserted downstream from the EU terminator, and the TMV 3' UTR had a negligible effect on gene expression (Fig. 11). In agreement with the results of Sainsbury et al., pEAQ-HT-GFP, which contains the cowpea mosaic virus 5' and 3' UTRs, provided a 17.1-fold increase compared to NOS alone (Fig. 12). However, this vector also contains the P19 suppressor of RNAi silencing. Coinfiltration of EU with P19 provided a 29.9-fold increase compared to NOS (Fig. 11, compare EU and EU + P19). The 3' UTRs derived from DNA viruses also performed very well. While the BeYDV SIR showed no terminator function by itself, addition of the SIR to the 3' end of EU nearly doubled its GFP production. Interestingly, although the intron-containing extensin terminator (IEU) performed very poorly on its own compared to the intronless version (EU), addition of the BeYDV SIR completely negated the detrimental effect of the intron (Fig. 10B, compare IEU + SIR and EU + SIR). Addition of the 35S terminator to IEU also greatly enhanced expression, however in this case the total yield was lower than the comparable vector with the intron removed (Fig. 11, compare IEU + 35S and EU + 35S). These results indicate that viral 3'

flanking regions, have potential to strongly increase gene expression when inserted downstream from the gene terminator.

c. Matrix Attachment Regions Are Potent Enhancers of Transient Expression

[00101] While MAR has been widely used in transgenic expression systems, there are few reports of their use in transient expression systems. We found that the tobacco Rb7 MAR strongly enhanced transient expression in a replicating geminiviral transient expression system when placed downstream from the gene terminator (Diamos et al., 2016). To more fully characterize the potential for MAR to function in transient expression systems, the tobacco Rb7 and TM6 MAR were inserted into nonreplicating GFP expression vectors in combination with 8 different gene terminators.

[00102] Insertion of the Rb7 MAR downstream from the EU terminator resulted in a striking 3-fold enhancement of GFP production (40-fold compared to NOS alone), exceeding the best double terminator configuration (Fig. 12, compare EU and EU + Rb7). Interestingly, similar to the SIR, the Rb7 MAR also dramatically improved production of the otherwise weak IEU terminator, increasing expression 19-fold, bringing it nearly on par with the EU-Rb7 vector (Fig. 12, compare IEU, IEU + Rb7, and EU + Rb7 and Fig. 17A). To verify that the observed enhancement was unique to the Rb7 MAR, a control DNA sequence of similar size, derived from a synthetic norovirus capsid protein coding sequence, was instead inserted downstream of GFP, and found to provide no significant difference in GFP production (Fig. 12, compare EU and EU + Control). Inspection of the AT-rich Rb7 MAR sequence reveals many terminator-like elements, however in the absence of a terminator no detectable GFP activity was produced using the MAR as the sole 3' region, indicating that it does not act as a fully functional terminator (Fig. 17A). Consistent with our findings in replicating systems, positioning of the Rb7 MAR 5' of the promoter had no effect on gene expression (Fig. 17A).

[00103] We further found that Rb7 MAR provided a large enhancement when used in conjunction with the 35S (13.8-fold), AtHSP (13.6-fold), NOS (12-fold), BDB501 (3.6-fold), or NbHSP (2-fold) terminators (Fig. 12). In the absence of the Rb7 MAR, the NbHSP terminator provided nearly 3-fold more expression than the 35S terminator. However, upon addition of the MAR, the 35S/Rb7 combination provided double the expression of NbHSP/Rb7 (Fig. 12).

Interestingly, while the Rb7 MAR substantially enhanced 7 of the 8 terminators tested, the NbACT3 terminator was unaffected by addition of the Rb7 MAR.

[00104] The tobacco TM6 MAR reportedly exceeded the enhancing effect of the Rb7 MAR in transgenic tobacco (Ji et al. 2013). To test the TM6 MAR in our transient expression system, the full sequence was cloned from tobacco plants and inserted in place of the Rb7 MAR. The TM6 MAR enhanced GFP production when paired with the 35S, NOS, or IEU terminators, but not with the NbACT3, similar to our findings for the Rb7 MAR (Fig. 12, purple bars). However, the magnitude of the enhancement provided by the TM6 MAR was significantly less than that of the Rb7 MAR in all combinations tested (Fig. 12 and Fig. 17A).

[00105] Using deletion studies, we investigated which regions of the 1193-bp Rb7 MAR were responsible for the observed enhancement. Deletion of nucleotides 144-1193 or 437-1193 eliminated the enhancing effect of the Rb7 MAR, however deletion of nucleotides 1-144, 144-437, 1-437, 421-730, or 1-730 did not impair MAR activity (Fig. 17B). In fact, a small but repeatable increase in the enhancement provided by the Rb7 MAR was observed upon deletion of nucleotides 1-730 (Fig. 17B). These data indicate that a relatively short region at the 3' end of the Rb7 MAR is responsible for all of the observed enhancement in this system.

d. Synergistic Enhancement of Combined 3' Flanking Regions

[00106] We investigated the potential for double terminators and the Rb7 or TM6 MAR to further increase gene expression when used in combination. Addition of the Rb7 MAR to the EU-35S double terminator significantly increased the expression provided by the double terminator alone (2.4-fold), and by either individual terminator with or without Rb7 MAR (Fig. 13, compare EU + Rb7, 35S + Rb7, EU + 35S + Rb7 and Fig. 3 EU + 35S). This represents a 56.7-fold total increase compared to the NOS terminator. However, the fold- increase provided by addition of the Rb7 MAR to EU-35S was smaller than the increase provided by addition of Rb7 to either EU or 35S alone (Fig. 17A). Interestingly, despite the failure of the Rb7 MAR to enhance expression of the NbACT3 terminator by itself, the 35S-NbACT3 double terminator was further enhanced by 2.4-fold upon addition of the Rb7 MAR, more than doubling the expression provided by 35S-Rb7 (Fig. 13). EU-NbACT3, the highest expressing double terminator, was also improved by 1.5-fold upon addition of the Rb7 MAR.

[00107] While the 35S and NOS terminators were substantially improved when combined in either orientation (Fig. 12), and while addition of the Rb7 MAR to the 35S-NOS double terminator provided a further 1.9-fold increase in expression (compare Fig. 13 35S + NOS + Rb7 to Fig. 12 35S + NOS), the total yield of 35S-NOS-Rb7 was no better than the single terminator construct 35S-Rb7 (Fig. 12, compare 35S + NOS + Rb7 and 35S + Rb7). Similarly, although we observed a large synergy between the Rb7 MAR and the individual terminators AtHSP, EU, BDB501, and 35S, and although the expression provided by the double terminators AtHSP-EU, 35S-EU, and EU-BDB501 was enhanced by addition of the Rb7, all performed worse than EU-Rb7. Furthermore, one double terminator combination was not improved at all by the presence of a MAR. Addition of the Rb7 MAR to NbHSP-Pin2 resulted in a small 1.1-fold increase, which was not statistically significant (compare Fig. 12 NbHSP + Pin2 + Rb7 to Fig. 11 NbHSP + Pin2).

e. Evaluation of Combined 3' Flanking Regions in a Replicating System

[00108] Previously, we reported a plant transient expression system based on the geminivirus bean yellow dwarf virus, which enhances gene expression by increasing accumulation of DNA copies of the gene of interest (Huang et al. 2009). We found that expression was substantially increased by insertion of the extensin terminator and the Rb7 MAR, among other modifications (Diamos et al., 2016). To evaluate the potential of combined 3' UTRs to function in this system, several of the best performing 3' UTR combinations were cloned into geminiviral vectors expressing GFP. The geminiviral vector containing EU-Rb7 enhanced expression 3.1-fold more than the nonreplicating vector (Fig. 14A). Similarly, the two best nonreplicating vectors containing EU-35S-Rb7 and 35S-NbACT3-Rb7 were further enhanced 2.7-fold and 2.5-fold, respectively, when placed in geminiviral vectors (Fig. 14A). Similar to their non-replicating counterparts, in geminiviral vectors the 35S-EU-Rb7 and 35S-NbACT3-Rb7 combinations increased GFP production by up to 20% compared to a replicating vector containing only EU-Rb7 (Fig. 14A). The total GFP expression of the best construct is estimated at up to 50% total soluble protein or 3-5 g per kg leaf fresh weight (Fig. 14B).

f. Gene-Specific and Plant-Specific Activity of Single and Combined 3' Flanking Regions

[00109] To determine whether the identified terminators performed similarly with a reporter gene other than GFP, vectors containing a variety of individual or combined terminators were constructed with the DsRed gene replacing the GFP gene. DsRed shares no sequence homology with GFP. For single terminators, the extensin terminator provided the highest level of gene expression (Fig. 15), in agreement with our data with GUS or Norwalk virus capsid protein(Rosenthal et al., 2018) . Other terminators performed similarly as when paired with GFP, however some small differences were observed, such as the increased activity of NbHSP and 35S with DsRed compared to NOS (Fig. 15). The IEU-35S double terminator substantially exceeded the DsRed production of EU alone, however the 35S-NOS double terminator performed substantially worse with DsRed than with GFP (Fig. 15). Use of the Rb7 MAR strongly enhanced DsRed expression with most terminators, and the enhancement provided by Rb7 exceeded that of TM6 in the one case tested (Fig. 15). Interestingly, while 35S-Rb7 performed very well with GFP, it did not with DsRed.

[00110] The functionality of genetic elements often varies among species. To assess the generality of these results in other plant systems, a subset of 3' UTRs were tested in tobacco (N. tabacum) and lettuce. Similar to our results in N. benthamiana, in both tobacco and lettuce plants GFP gene expression with EU was >10-fold higher than with NOS, and EU exceeded all other single terminators tested (Fig. 16). While EU performed substantially better than all other single terminators in its native tobacco, the AtHSP, NbHSP, and NbACT3 performed nearly as well as EU in lettuce (Fig. 16). Addition of the Rb7 MAR strongly enhanced GFP expression when added to EU in tobacco and lettuce, and the 35S-NbACT3-Rb7 combination, which performed very strongly in N. benthamiana, also did so in tobacco and lettuce (Fig. 16). Despite its strong performance in N. benthamiana, IEU-35S-Rb7, which contains the extensin intron, had much lower expression in tobacco (Fig. 16). Similarly, NbHSP-NbACT3 with or without Rb7 MAR performed substantially worse in tobacco than in N. benthamiana (Fig. 14 and Fig. 16).

g. Discussion

[00111] To more broadly assess the potential of 3' flanking regions to enhance gene expression in plant systems, we systematically compared a diverse set of terminators from

various plant and viral sources. Narsai et al. (2007) reported a genome-wide analysis of mRNA stability in A. thaliana, showing that characteristic 3' UTR motifs are enriched in long-lived or short-lived transcripts. To rationally derive putative terminator candidates with potential to enhance gene expression, we identified *N. benthamiana* homologs of two highly stable *A. thaliana* transcripts: an 18.8 kDa class II heat shock protein gene, and an actin-like gene. Both terminators outperformed all of those frequently used previously (Fig. 10B). Since we tested only two terminators identified in this manner, we suspect that other candidates can be discovered. Overall, we found that 12 terminators exceeded the performance of 35S or NOS.

[00112] Highly expressed genes from strong promoters are targets for RNA silencing (Que et al. 1997; Schubert et al. 2004), mediated by RDR6 (Béclin et al. 2002). Luo and Chen (2007) demonstrated that improperly terminated mRNAs result in RDR6-mediated silencing of the transgene, and that the use of a 35S-NOS double terminator reduced this effect while enhancing GUS expression is transgenic *A. thaliana*. Beyene et al. (2007) also found a large enhancing effect of a 35S-NOS double terminator in several plant species. While the intronless EU terminator by itself was significantly better than all 19 other individual terminators tested, we identified 8 double terminators that significantly exceeded the performance of EU alone, 7 of which significantly outperformed the 35S-NOS double terminator (Fig. 11). We found that nearly every combination outperformed either individual terminator alone, showing that tandem-linked terminators have excellent potential to enhance gene expression. Interestingly, in both tested cases, reversal of the position of the two terminators resulted in a substantial difference in expression, indicating that the observed enhancement does not arise entirely from the individual action of the two terminators, but rather on a synergistic interaction between the two terminators, which depends in part on their relative position. Furthermore, when positioned upstream from NOS, the 35S terminator performed strongly; however, the opposite was observed when paired with EU: expression was enhanced by 50% when 35S was instead positioned downstream from EU. These results indicate that the optimal position for a given terminator depends on the individual terminator with which it is paired. Further work is needed to study the individual mechanisms underlying these differences.

[00113] The 3' flanking regions from RNA viruses contain many mechanisms to enhance mRNA stability or increase translation (Fan et al., 2012; Simon and Miller 2013). However, when expressed in the plant nucleus, these 3' regions may contain cryptic splice sites and other

detrimental sequences. Most of the RNA virus-derived 3' flanking regions we tested were poorly functional when transiently expressed in *N. benthamiana* leaves, except for those derived from cowpea mosaic virus. The 5' and 3' UTRs from cowpea mosaic virus were reported to be potent enhancers of protein expression (Sainsbury and Lomonossoff, 2008; Meshcheriakova et al., 2014). In general agreement, we found pEAQ-HT-GFP, which contains the NOS terminator and the cowpea mosaic virus UTRs, enhanced GFP expression 17.1-fold compared to NOS alone. However, the cowpea mosaic virus vector pEAQ-HT-GFP also contains the P19 suppressor of RNA silencing, which likely enhances RNA stability (Sainsbury et al., 2009), making direct comparisons to other 3' UTRs difficult. While pEAQ-HT-GFP provided 20% more GFP than the extensin terminator alone, it provided 40% less when extensin was also supplemented with P19 (Fig. 11). We also found that other highly expressing terminator combinations identified here were also strongly enhanced by addition of P19 (data not shown). As pEAQ-HT-GFP contains the relatively weak NOS terminator, we suspect the cowpea mosaic virus flanking regions may perform better with a stronger terminator, unless some particular synergy exists with NOS.

[00114] Flanking regions derived from nucleus-adapted DNA viruses, such as the geminiviruses, were found to be potent enhancers of gene expression, especially when used in conjunction with a functional terminator. The short 200 bp SIR from bean yellow dwarf virus showed no terminator activity by itself, but it was found to strongly increase gene expression when used in conjunction with the extensin terminator, on par with the best double terminator combinations tested. However, extending the SIR to include upstream and downstream coding sequence from the BeYDV coat and rep proteins showed that it also has strong terminator function on its own. Similar results were obtained with 3' UTRs obtained from bean dwarf mosaic virus. These results highlight the influence of the upstream gene coding sequence on 3' UTR function. Further work is needed to better characterize the enhancing potential of geminiviral 3' UTRs, and to determine whether the observed enhancing effect of the SIR is terminator-specific.

[00115] Though MAR was previously used in transgenic systems, we found the tobacco Rb7 MAR substantially improved gene expression using a geminiviral transient expression system (Diamos et al. 2016). Only a small percentage of T-DNA delivered by agrobacterium undergoes chromosomal integration, while the majority is transiently transcribed in the nucleus. It has been shown that the agrobacterium proteins VirE2, which coats the T-DNA, and VirD2, which

attaches to 5' end of the T-DNA and mediates nuclear entry, both associate with cellular histones (Lacroix et al. 2008; van Heusden et al. 2015). As MAR is thought to influence chromatin structure, the association of T-DNA with histones suggests a possible mechanism by which MAR function in vectors delivered by agroinfiltration. Here, we find that both the tobacco Rb7 and TM6 elements greatly enhance transient gene expression in agroinfiltrated leaf tissue (Fig. 12). The effect of the Rb7 MAR varies in a promoter-dependent manner (Mankin et al. 2003). Similarly, we find that the effect of the MAR also varies in a terminator-dependent manner. EU was the strongest individual terminator and EU-Rb7 was the strongest MAR combination, exceeding the best double terminator by over 50%. However, while NbACT3 was the second strongest individual terminator, NbACT3-Rb7 was the lowest expressing MAR combination. Interestingly, we saw a similar effect when NbACT3 was paired with the TM6 MAR, suggesting that some enhancing activity present in both MAR is not active when paired with the NbACT3 region. All other terminators greatly benefited from addition of the Rb7 or TM6 MARs, although the magnitude of the enhancement varied in a manner that was not correlated to the individual expression level mediated by each terminator alone.

[00116] Ji et al. (2013) found that the TM6 MAR enhanced GUS expression at a level greater than the Rb7 MAR in transgenic tobacco. However, we consistently found that the Rb7 MAR increased transient expression more than the TM6 MAR. This observed discrepancy could be due to different expression systems, or different reporter genes. We found that the entire enhancing activity of the Rb7 MAR resides in a 463 bp region at its 3' end. Although a detailed characterization of the functional regions of the Rb7 MAR has not been reported, the region we found to be dispensable includes several AT-rich regions, a matrix attachment recognition sequence motif, and a topoisomerase II binding site, all of which were previously suspected to play a role in MAR function (Allen et al. 1996). Ji et al. (2013) found that deletion of similar MAR elements substantially reduced the enhancing effect of TM6. Additionally, it has been reported that the TM2 MAR functions best when placed 5' of the gene of interest (Zhang et al. 2009), whereas we found no effect of 5' insertion of the Rb7 MAR. As MAR are thought to contain multiple active regions responsible for their enhancing function, there may be differences in the key functional regions of the Rb7, TM2, and TM6 MARs, making direct comparison difficult. Alternatively, while Rb7 and TM6 are both clearly active in our transient expression

system, the mechanisms by which expression is enhanced may differ between transient and transgenic systems. Further studies are needed to resolve these discrepancies.

[00117] Previously, we found that combining optimized 5' UTRs and the Rb7 MAR resulted in a synergistic enhancement of gene expression (Diamos et al. 2016). Other studies obtained favorable results by duplicating or combining highly functional genetic elements, such by tandem-linking TM2 MAR (Zhang et al. 2002) or combining the 5' UTR from alcohol hydrogenase and the AtHSP terminator (Limkul 2015). Here, we find that combining double terminators with the Rb7 MAR enhanced gene expression more than either component by itself in some, but not most, cases (Fig. 13). Particularly high synergy was observed between the Rb7 MAR and the EU-35S, 35S-NbACT3, and EU-NbACT3 double terminators, reaching very high expression levels of up to 60-fold more than the NOS terminator alone.

[00118] We saw variable effects when combining double terminators with MAR. While 35S-NOS was a relatively strong double terminator, it had little synergy when combined with Rb7. Similarly, while the AtHSP terminator had high synergy with the Rb7 MAR, double terminators containing AtHSP did not improve expression compared to AtHSP-Rb7 alone (Fig. 13). As all AtHSP double terminators tested in this study had AtHSP positioned as the upstream terminator, it is possible reversing terminator positions may result in better performance. Notably, EU-35S-Rb7 was one of the best combinations tested, but the reversed 35S-EU-Rb7 had a 50% reduction in expression (Fig. 13).

[00119] We have created a replicating transient expression system based on the geminivirus bean yellow dwarf virus, which amplifies the gene of interest to high copy number in the plant nucleus (Huang et al. 2009, Huang et al. 2010). By incorporating optimized 5' and 3' UTRs with other modifications, we have used this system to produce vaccine antigens and pharmaceutical proteins at levels greater than or similar to the highest levels reported in plant-based systems (Diamos et al. 2016). Here, we find that gene expression with the double terminator and MAR constructs 35S-NbACT3-Rb7 and EU-35S-Rb7 is improved by ~2.5-fold when placed in a replicating vector, a 20% increase compared to the best replicating construct containing only a single terminator and MAR (Fig. 14A). This represents a more than 150-fold increase compared to the original NOS vector alone, providing an estimate yield of around 40-60% total soluble protein or 4-5mg recombinant protein per kg of leaf tissue (Fig. 14B), which appears to approach the theoretical limit achievable in a plant-based system.

[00120] The upstream gene coding sequence has been shown to interact with the 3' UTR. The NOS terminator contains a cryptic polyadenylation site that requires an upstream element to be present for its function (Sanfaçon et al., 1991; Sanfaçon and Hohn, 1990). We found that the intergenic regions of bean yellow dwarf virus and bean dwarf mosaic virus both require upstream coat protein coding sequence for terminator function (Fig. 10B). These results indicate that 3' UTRs may perform differently in the context of different genes of interest. Using DsRed as an alternative reporter gene to GFP with no shared homology, we found that most single or combined 3' UTRs performed similarly relative to one another, with a few notable exceptions. The 35S-NOS and 35S-Rb7 3' UTRs both performed substantially worse with DsRed than they did with GFP, however both still performed better than 35S alone. Despite these combinations performing worse, the 35S terminator alone performed better with DsRed than with GFP, and combinations with the 35S terminator positioned as the second terminator were still highly functional (Fig. 15).

[00121] Lettuce has been shown to be a promising plant system capable of rapidly producing recombinant proteins (Lai et al., 2012; Chen et al., 2016). To further investigate the generality of our results, we also tested a variety of 3' UTRs in tobacco and lettuce. As with *N. benthamiana*, EU was the best individual terminator in lettuce. Further, combined terminators containing the Rb7 MAR substantially outperformed any individual terminator tested (Figs. 16A and 16B). However, a few of the terminator combinations that performed very well in *N. benthamiana* performed relatively poorly in either lettuce or tobacco. As we did not test every identified combination with multiple genes or other plant systems, it is likely that other gene-specific or plant-specific effects exist that we did not discover here. Overall, our data suggests that the optimal terminator for a given system must be determined empirically. However, the potent enhancing effect of the intronless EU terminator has been demonstrated with GFP, DsRed, GUS, and Norwalk virus capsid protein, in *N. benthamiana*, lettuce, and tobacco (Fig. 10B, Fig. 15, Fig. 17A, Fig. 17B, and Rosenthal et al., 2018), indicating that the terminator is intrinsically highly active in many gene contexts. Additionally, our results clearly demonstrate that combining terminators is a highly effective strategy to improve gene expression in a variety of systems.

[00122] In conclusion, we have identified a diverse set of gene terminator regions that greatly exceed the gene expression provided by the most commonly used terminators in *N. benthamiana*,

tobacco, and lettuce leaves. The intronless tobacco extensin terminator is a uniquely potent enhancer of gene expression. In nearly every case tested, double terminators outperformed either individual terminator alone, often exceeding the gene expression of the best individual terminators by more than 2-fold. We find that MAR, especially the 3' end of the Rb7 MAR, are strong enhancers of transient gene expression, and when combined with double terminators, synergistically enhance expression. Incorporating these combined terminators into a replicating geminiviral expression system has allowed us to produce recombinant proteins comparable to the highest levels ever reported in a plant-based system. The 3' UTR combinations identified here have broad potential to improve other DNA-based plant expression systems.

3. Double Terminators Enhance VLP Expression

[00123] Attempts to express the GII.4 norovirus capsid using the TMV-based magnICON system resulted in the rapid onset of cell death, and correspondingly low VLP yields of 0.3 mg per gram leaf fresh weight (LFW) (Mathew et al. 2014). The total soluble protein extractable from one gram of *N. benthamiana* leaf issue is 8-10 mg. Previously, we have reported expression of GI Norwalk capsid protein (NVCP) at 1.8 mg per gram leaf fresh weight (LFW) using bean yellow dwarf virus vectors, or ~20% of total soluble protein. To attempt production of GII.4 norovirus VLPs using optimized vectors designed to reduce cell death, bean yellow dwarf virus vectors (Fig. 18A) containing the Rb7 matrix attachment region, and the NbPsaK 5' UTR with or without reduced Rep/RepA expression (Diamos et al. 2016, and data to be published elsewhere) were infiltrated using agrobacterium strain EHA105 at an OD600 of 0.2 into the leaves of Nicotiana benthamiana and monitored for cell death. An unmodified vector without these optimizations was used as a control.

[00124] Combined gene terminators, which we have found to be potent enhancers of plant transient gene expression, were tested in BeYDV vectors expressing GII.4 norovirus capsids. The gene terminators tested IEU, IEU+35S (indicated in Figs. 18B and 18C as IEU-35S), EU+NbACT3 (indicated in Figs. 18B and 18C as EU-NbACT3), and 35S+NbACT3 (indicated in Figs. 18B and 18C as 35S-NbACT3). By both western blot and ELISA, the EU+NbACT3 and 35S+NbACT3 double terminators enhanced expression compared to IEU or IEU+35S. The best double terminator, EU+NbACT3, provided a 64% increase in expression, producing up to 1.1 mg/g LFW of GII.4 norovirus capsids (Figs. 18B and 18C). Addition of 0.1% Triton X-100

detergent enhanced norovirus capsid recovery (Fig. 18C). Production of GI NVCP was also tested with the IEU, IEU+35S, and 35S+NbACT3 double terminators. Like GII.4 VLP, 35S+NbACT3 increased expression 30% compared to IEU, resulting in a yield of 2.3 mg/g NVCP, or up to 28% total soluble protein.

[00125] Using the BeYDV plant expression system, we have produced norovirus VLPs at 2-3 times the highest levels reported in plant-based systems. Modified BeYDV vectors allowed high-level production of GII.4 norovirus VLPs without eliciting the plant cell death response. By optimizing extraction conditions, we have achieved > 90% purity of VLPs with no losses in yield, allowing production of milligram quantities of VLPs from a single plant leaf.

4. Exemplary plant expression vectors

[00126] pBYR2eAK2Mc-GP (See Fig. 19 and SEQ ID NO. 115) is a BeYDV replicon that lies between the 2 long intergenic regions (indicated as LIR in Fig. 20), bounded by AscI and AgeI sites. The replicon contains one expression cassette with the CaMV 35S dual-enhancer promoter (indicated as P 35Sx2e in Fig. 19) followed by the *N. benthamiana* PsaK2 gene 5' UTR. The expression cassette follows in tandem P 35Sx2e and PsaK2 5' UTR followed by the GFP coding sequence, which is flanked by unique XbaI and SacI sites. Just downstream lie a chimeric terminator (also referred to herein as a double terminator) comprises the intronless tobacco extensin terminator (indicated as EU 3' in Fig. 19) and *N. benthamiana* actin 3 terminator (NbAct3). The double terminator is followed by the tobacco Rb7 matrix attachment region (indicated as Rb7 MARc delta 144-437 in Fig. 19). Downstream of the Rb7 MARc lies the BeYDV short intergenic region (SIR), which also contains polyadenylation signals. The Rep/RepA (C1/C2) gene in the inverse orientation is driven by a BeYDV promoter in the downstream LIR. The expression cassette for the siRNA binding protein p19 (from tomato bushy stunt virus) is located between the PvuI and AscI sites (~0-45° on Fig. 19). The p19 coding sequence is driven by a single enhancer 35S promoter and the tobacco mosaic virus (TMVΩ) 5' UTR. The potato pinII gene 3' region (Pin2) serves as the terminator for the p19 expression cassette. The T-DNA region (DNA that is transferred into plant cells by *A. tumefaciens*) is delineated by the left border and right border segments. As shown in Fig. 22C, this construct with a double terminator downstream of the GFP coding sequence resulted in brighter fluorescence than a corresponding construct with a single terminator (pBYR2eK2Mc-GFP).

[00127] pBY11HA-GFP (see Fig. 20 and SEQ ID NO. 116) contains a BeYDV replicon that lies between the 2 long intergenic regions (indicated as LIR in Fig. 20), bounded by AscI and AgeI sites. The replicon contains two expression cassettes having the CaMV 35S dual-enhancer promoter (indicated as P 35Sx2e in Fig. 20) followed by the *N. benthamiana* PsaK2 gene 5' UTR. The first cassette has 2 BsaI sites downstream of the 5' UTR; when cleaved by BsaI, the vector ends have 5' single strand protrusions (cohesive ends) 5'-CTAG (upstream) and 5'-AGCT (downstream), that permit insertion of a coding sequence. The BsaI sites are followed by a chimeric terminator comprising the intronless tobacco extensin terminator (indicated as Ext 3' UTR in Fig. 20) and *N. benthamiana* HSP20 gene terminator (also referred to herein as *N. benthamiana* 18.8 kDa class II heat shock protein 3' UTR or NbHSP), which is followed by the tobacco Rb7 matrix attachment region (indicated as Rb7 MARc in Fig. 20). Downstream of the first Rb7 MARc lies the BeYDV short intergenic region (SIR), which also contains polyadenylation signals. The second expression cassette follows in tandem (P 35Sx2e, PsaK2 5' UTR), followed by the GFP coding sequence, which is flanked by unique XbaI and SacI sites. Just downstream lie a chimeric terminator comprising the intronless tobacco extensin terminator (indicated as EU 3' in Fig. 20) and *N. benthamiana* actin 3 terminator (NbAct3). Another MARc and SIR segment follow the terminators. The Rep/RepA (C1/C2) gene in the inverse orientation is driven by a BeYDV promoter in the downstream LIR. An expression cassette for the siRNA binding protein p19 (from tomato bushy stunt virus) is located between the PvuI and AscI sites (~0-45° on Fig. 20). The p19 coding sequence is driven by a single enhancer 35S promoter and the tobacco mosaic virus (TMVΩ) 5' UTR. The potato pinII gene 3' region (Pin2) serves as the terminator for the p19 expression cassette. The T-DNA region (DNA that is transferred into plant cells by *A. tumefaciens*) is delineated by the left border and right border segments. The pBY11HA-GFP vector is optimized for the co-expression of two proteins and has unique restriction sites for ease of inserting two coding sequences.

[00128] pBY!11-h6D8M2e (see Fig. 21 and SEQ ID NO. 117) is similar to the dual-cassette vector pBY11HA-GFP, but has expression cassettes for a modified mAb (heavy chain fused to influenza antigen M2e in first cassette; light chain in second cassette). Figs. 22A and 22B depict expression data for this vector, namely the expression of a recombinant immune complex (RIC). RIC is inherently insoluble due to the immune complex, but by modifying the extract protocol, RIC can be at least 40% of total soluble protein extracted from the transformed plant. Thus, the

dual expression cassette vector containing dual chimeric terminators on both cassettes performs very well.

5. Experimental Procedures

a. Vector Construction

[00129] For the sequence positions for the Ext terminator, the position of the U of the Ext stop codon is designated as -3. The initial construct was pBY027-IEU for the experiments resulting in Figs. 1A-9, which contain a fragment spanning 1-731 nt from a *Nicotiana tabacum* genomic DNA (GenBank accession D13951) cloned in pBY027 (Mor et al 2003) using the SacI and EcoRI sites. The fragment was amplified by PCR using primers Ext-1 and Ext-2. The sequences of all primers used in experiments resulting in Figs. 1A-9 are listed in Table 1. The intron-less form of Ext terminator (spanning 252-731 nt) was amplified by PCR from pBY027-IEU using primers Ext-3 and Ext-2 and cloned into the SacI and EcoRI sites of pBY027 to produce pBY027-EU. A XhoI-EcoRI fragment of pBY027-EU was excised and replaced the corresponding fragment in pPS1 (Huang and Mason 2004) to generate pEU (Fig. 1). The various forms of Ext terminator were similarly cloned using primer pairs of Ext-1 and Ext-2 (pIEU), Ext-3 and Ext-4 (pEU1), Ext-3 and Ext-6 (pEU2), Ext-5 and Ext-2 (pEU3), and Ext-1 and Ext-6 (pIEU2). The control plasmids pNOS, p35S, and pVSP were generated as follows. The NOS terminator was excised from construct pIBT211.1 (Richter et al 2000) and cloned into pBY027 to yield pBY027-NOS using the SacI/EcoRI sites. The 35S terminator was amplified by PCR from pRTL2-GUS (Carrington et al 1991) using the primers 35STm-1 and 35STm-2 and cloned into pBY027 opened with SacI/EcoRI to give pBY027-35S. The XhoI-EcoRI fragments of pBY027-NOS, pBY027-35S, and pBY027 (contains VSP terminator) were excised and replaced the corresponding fragment in pPS1 to produce pNOS, p35S, and pVSP. A fragment that contains the intron of Ext terminator was amplified by PCR from pIEU using primers Ext-1 and Ext-8 and cloned into pNOS opened with SacI to give pINOS. A set of constructs encoding Norwalk virus capsid protein (NVCP) with various terminators was obtained by replacing the XhoI/SacI fragment containing the green fluorescent protein (GFP) gene from the constructs pEU, pEU1, pEU2, pEU3, pIEU, pIEU2, p35S, pVSP, pNOS, and pINOS with the NVCP coding region excised from the plasmid pSNV210 (Zhang et al 2006) with the XhoI and SacI sites. Another set of constructs containing the β-glucuronidase gene (with intron) was obtained by

replacing the GFP gene of constructs pEU, p35S, pVSP, and pNOS with the PCR fragment amplified from pGPTVK-GI (Collens et al 2007) using primers GUS-1 and GUS-2 and digested with XhoI and SacI.

[00130] For experiments resulting in Figs. 10A-17B, an agrobacterium T-DNA binary vector containing the 35S promoter, tobacco mosaic virus 5' UTR, GFP gene, and full-length tobacco extensin terminator was constructed by 3-fragment ligation: the vector backbone from pPS1 (Huang and Mason, 2004) was obtained by XhoI-EcoRI digestion; the TMV 5' UTR-GFP fragment was excised from pBYR2e-GFP (Diamos et al., 2016) by XhoI-SacI digestion; and the intronless extensin terminator was excised from pBY-GFP212 (Diamos et al., 2016). The resulting vector, pPS-OGFP-EU, was used to construct single and double terminator constructs. The DsRed gene was amplified from pBYDsRed (Huang et al., 2010) with primers DsR-Xba-F and VspHT, digested XbaI-SacI, and inserted into pPS-OGFP-EU to create pPS-ODsR-EU. Terminators were amplified by PCR using primers (Table 1) designed to insert SacI and EcoRI restriction sites at its 5' and 3' ends, respectively, digested SacI-EcoRI, and ligated into pPS-OGFP digested likewise. The pinII, NOS, and rbcS 3' regions were obtained from pHB114 (Richter et al., 2000), pHB103 (Richter et al., 2000), and pRTL2-GUS (Carrington et al. 1991) respectively by SacI-EcoRI digestion and cloned into pPS-OGFP digested likewise.

[00131] Specifically, The NbHSP (homolog of At5g12020) and NbACT3 (homolog of At5g09810) terminators were identified using the Sol Genomics Network *N. benthamiana* draft genome (Fernandez-Pozo et al., 2015). Primers (Table 1) specific for NbHSP and NbACT3 designed to introduce SacI and EcoRI sites were used to amplify the downstream segments of each gene from N. benthamiana genomic DNA. The PCR products were digested SacI-EcoRI and inserted into pPS-OGFP-EU digested likewise.

[00132] For double terminator constructs, the upstream segment was amplified by PCR using primers (Table S1) designed to insert a SacI site at the 5' end, and BsaI site at the 3' end. The downstream segment was amplified with a BsaI site at the 5' end designed to generate compatible overhang with the upstream BsaI site, and an EcoRI site at the 3' end. The final construct was assembled by 3-fragment ligation: pPS-OGFP-EU digested SacI-EcoRI, the upstream segment digested SacI-BsaI, and the downstream segment digested BsaI-EcoRI.

[00133] For MAR constructs, the tobacco Rb7 or TM6 MAR was inserted downstream from the terminator. First, the Rb7 MAR was inserted into pPS-OGFP-EU by three fragment ligation:

pPS-OGFP-EU was digested PvuI-SphI to obtain the vector fragment; pPS-OGFP-EU was digested PvuI-EcoRI to obtain the GFP cassette; and pBYR2e-MRtxGM (Diamos et al., 2016) was digested EcoRI-SphI to obtain the Rb7 MAR fragment. The resulting vector was digested KpnI-AgeI, the ends were blunted with Klenow fragment DNA polymerase, and the vector fragment was self-ligated to yield pPS-OGFPM-EU. The TM6 MAR (genbank accession KC555564) was PCR amplified from tobacco genomic DNA using primers TM6-EcoRI-F and TM6-KpnI-R (Ji et al., 2013), digested EcoRI-KpnI, and inserted into pUC19. The EcoRI-AvrII fragment containing the TM6 MAR was then excised and inserted into pPS-OGFPM-EU digested likewise to yield pPS-OGFPT-EU. Single or double terminators were inserted into pPS-OGFPM-EU or pPS-OGFPT-EU by SacI-EcoRI digestion as described for pPS-OGFP-EU. For Rb7 deletion mutants, native restriction sites were used as shown in Figs. 14A and 14B. After digestion with each respective enzyme, the ends were blunted with Klenow fragment DNA polymerase, and self-ligated.

Table 1. Primers Used

Name/Seq ID No.		Sequence (5' to 3')
CAA-2	22	GTTGTTGTTGTTGTTTTGTGATGTTTGAACGATCGGGG
CAA-3	23	CAACAACAACAACAACAAAACATTGGCAATAAAGTTTCTTAAGATTG
CAA-4	24	TGTTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTTGTG
CAA-5	25	CACAAAACAACAACAACAACAACAACAACAACAACAACAAAACA
35STm-1	26	GTGAGCTCGTCCGCAAAAATCACCAG
35STm-2	27	CAGAATTCGTCACTGGATTTTGGTTTTAGG
DT-2	28	CGCTGTGCTGCCGTGCCTTTTTTTTTTTTTTTTT
DT-4	29	CGCTGTGCTGCCGTGCCTT
Ed-2	30	GGCTTTATTATTTAATTGTCTCTCGAAAATGG
Ed-3	31	TCGAGAGACAATTAAATAATAAAGCCCCACAG
EFf	32	CTGGTGGTTTTGAAGCTGGTA
EFp	33	FAM-CACGCATTGCTTGCTTTCACCC-TAMRA
EFr	34	GGTGGTAGCATCCATCTTGTT
Et-2	35	ATCATCATCATCATCAATTCACATTGTCTCTCGAAAATGGAAAAAG
Et-3	36	GATGATGATGATGATGAATTGTGTAAATAATAAAGCCCCACAGG
Ext-1	37	GTGAGCTCGAAGTGACATCACAAAGTTGAAG
Ext-2	38	CAGAATTCGTCATAACTGTAGAAATGATTCC
Ext-3	39	GTGAGCTCAAAGCAGAATGCTGAGCTA
Ext-4	40	CAGAATTCGGCTAGCTAGAGAGAGGAGGAGAAT

Name/Seq ID No.		Sequence (5' to 3')
Ext-6	41	CAGAATTCGAAACTGAACAAAACATACACAATGACAG
Ext-8	42	TGCGAGCTCCTGCATAAGAATATACATTGTGTG
gfp-3f	43	ATGGTCCTGCTGGAGTTCGTGACC
gfp-f	44	GTCCAGGAGCGCACCATCTTCT
gfp-p	45	FAM-CGGGTCTTGTAGTTGCCGTCGTCCTTG-TAMRA
gfp-r	46	GATGCCCTTCAGCTCGATGCGGTT
GFP-F1	47	GTCCTGCTGGAGTTCGTGACC
GFP-F2	48	GCATGGACGAGCTGTACAAGTAAGAGC
GFP-R1	49	CGTTTACGTCGCCGTCCAGC
GFP-R2	50	CTCGCCCTTGCTCACCATTGTTC
GUS-1	51	GTCTCGAGAACAATGTTACGTCCTGTAGAAACC
GUS-2	52	ACGAGCTCTCATTGTTTGCCTCCCTGC
NE-2	53	TTCTTCTTCTTCTTCTTTTCTCATGTTTGAACGATCGGGG
NE-3	54	GAAGAAGAAGAAGAAGAAAAGATTGGCAATAAAGTTTCTTAAGATTG
NE-4	55	TTCTTCTTCTTCTTCTTTTCTCAGGATTCAATCTTAAGAAAC
NE-5	56	GAAGAAGAAGAAGAAGAAAAGAGGTTGCCGGTCTTGCGATG
NET-2	57	ATCATCATCATCATCAATTCACATGTTTGAACGATCGGGG
NET-3	58	GATGATGATGATGATGAATTGTTTGGCAATAAAGTTTCTTAAGATTG
NOSREco	59	CTCGAATTCCCGATCTAGTAAC
RS-2	60	CAGCCCTACAACTATATCCGATGATGTTTGAACGATCGGGG
RS-3	61	GAGTAAATATTACGAGTCGAGGTTGGCAATAAAGTTTCTTAAGATTG
RS-4	62	CCTCGACTCGTAATATTTACTCAGCCCTACAACTATATCCGATG
RS-5	63	CATCGGATATAGTTGTAGGGCTGAGTAAATATTACGAGTCGAGG
RT-0	64	AACGAT TGAAGGAGCCACTCA
RT-1	65	ACTGAAGGCGGGAAACGAC
RT-2	66	GTGGTTGGCATGCACATACAA
RT-3	67	ATTTGCCTGCGCACCTGT
sNV-3f	68	AATGGTGCTAGCAGCGGTCCAC
sNV-f	69	CAATGTCCAAAGTGTGCAATTCC
sNV-p	70	FAM-TGATGGAAGACTTGTTGGCACC-TAMRA
sNV-r	71	CCTCTTATCTTGGCAACATGTGAC
VE-5	72	GAGAAAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAAGA
VE-6	73	TCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTTTCTC
Vet-5	74	GTGAATTGATGATGATGATGATGATGATGATGATGAATTGT
Vet-6	75	ACAATTCATCATCATCATCATCATCATCATCATCAATTCAC
TM6-EcoRI-F	76	TCCGAATTCTAATATTTAGAAATTTAATTAACATAACCAAGG
TM6-KpnI-R	77	CTGGTACCGACATCCTAGGTTCAATCAAAT

Name/Seq ID No.		Sequence (5' to 3')
DsR-Xba-F	78	GAGTCTAGAACATGGTGCGCTCCTCC
VspHT	79	TGAATAGTGCATATCAGCATACCTTA
EU-Bsa-F	80	AGGGTCTCGGCTCAAAGCAGAATGCTG
EU-Bsa-R	81	AAGGTCTCGGAGCGTCATAACTGTAGAAATGATTCC
BDB-Bsa-F	82	GGGTCTCGGCTCTGACAACATCAGCAAG
NbHSP-Bsa-F	83	AGGGTCTCGGCTCACTGAGGAAATATATAGACAAATTAAG
35S-Bsa-R	84	AAGGTCTCGGAGCGTCACTGGATTTTGGTTTTAGG
NbHSP-Bsa-R	85	AAGGTCTCAGAGCTCCCAAAGGAAACTATGTGTAC
NOS-Bsa-R	86	AGGGTCTCGGCTCAGATCGTTCAAACATTTG
Pin2-Bsa-F	87	GAGGTCTCAGCTCGTACCCTGCAATGTGACC
NbACT-Bsa-F	88	AGGGTCTCGGCTCATACAGCATTCCCA
BDB501-Sac-F	89	CCGAGCTCTGACAACATCAGCAAGAACG
BDB501-Eco-R	90	AAGAATTCAAAGGAAACCCATAAGATGCG
NbHSP-Sac-F	91	TCGAGCTCACTGAGGAAATATATAGACAAATTAAGTTTGGTTCTATG
NbHSP-Eco-R	92	GTGAATTCGCTCCCAAAGGAAACTATGTGTACTTC
NbACT-Sac-F	93	GCGAGCTCATACAGCATTCCCAGAAAGAGAAAC
NbACT-Eco-R	94	TAGAATTCATGCTAGCTTGTTTACACCTCG
BDA375-Sac-F	95	GAGAGCTCGGAGAACGCCTTATTATTGTATATGGC
BDA375-Eco-R	96	AAGAATTCGCTCATCACTGCACTTCAAGC
AtHSP-Sac-F	97	TAGAGCTCATATGAAGATGAAGATGAAATATTTGGTGTG
AtHSP-Eco-R	98	ATGAATTCCTTATCTTTAATCATATTCCATAGTCCATACC
AtHSP-Bsa-R	99	AAGGTCTCGGAGCCTTATCTTTAATCATATTCCATAGTCCATACC
Rep-Sac-F	100	AGCGAGCTCTAATAGGTTGCCAGTCTGATTTC
Rep-Eco-R	101	CTAGAATTCTTGCCATCGTTTTGTGG
RepA-Sac-F	102	TCGGAGCTCTGAACGTGCCTCTCCTC
SIR-Sac-F	103	AAGGAGCTCTAAAATGATTATTTTATGAATATATTTCATTGTGC
NbACT617-EcoR	104	AATGAATTCGAACCCCAATTACTGGAGC
35S-Bsa-F	105	GCGGTCTCGGCATGGTGGAGCACGA

Name/Seq ID No.		Sequence (5' to 3')
NOS-Bsa-F	106	AGGGTCTCGGCTCAGATCGTTCAAACATTTG
TNVD3-Bsr-F	107	ATTGTACAAGTAATTGCTTTCATAGATCCGTCTTCC
TNVD3-Sac-R	108	TAGAGCTCGGGTTCCTAGAGAGATCTCTAGG
TMV3-Bsr-F	109	TATGTACAAGTAAGGTAGTCAAGATGCATAATAAATAACGGATTGTG
TMV3-Sac-R	110	TAGAGCTCTGGGCCCCTACCGGGGGTAA
PEMV3-Bsr-F	111	ATTGTACAAGTAAGGCTTCGCTTCCCGCC
BYDV3-Kpn-F	112	AAGGTACCAGTGAAGACAACACC
BYDV3-Sac-R	113	ATGAGCTCGGGTTGCCGAACTGC

b. DNA constructs for polypurine sequence (PPS) mutants

[00134] PPS mutant vectors were constructed using standard overlapping PCR and molecular cloning methods. For pEUd, primer sets Ext-3/Ed-2 and Ed-3/Ext-6 were used for initial amplification in separate PCR reactions using pBY027.IEU as template. The resulting PCR fragments were mixed and amplified using primers Ext-3 and Ext-6, complementary to the ends of the two initial fragments. The resulting PCR product was inserted into pBY027 using SacI/EcoRI sites and then subcloned into pPS1 via XhoI/EcoRI. For pEUs, similar steps were performed to amplify two overlapping DNA fragments in separate PCRs. The 5' fragment was obtained by double round PCR: in the first round, primers Ext-1 and Et-2 were used with pBY027.IEF as template and the resulting fragment was used for a second round of amplification with primers Ext-1 and Vet-6. The 3' fragment was also obtained by double round PCR: in the first round, primers Et-3 and Ext-2 were used with pBY027.EF as template and the resulting fragment was used for a second round of amplification with primers Vet-5 and Ext-2. The resulting 5' and 3' fragments were mixed and amplified using primers Ext-3 and Ext-2. The final PCR product was inserted into pBY027 using SacI/EcoRI sites and then subcloned via XhoI/EcoRI into pPS1.

c. Agroinfiltration procedure

[00135] Binary vectors were separately introduced into Agrobacterium tumefaciens LBA4404 (for Figs. 1A-9) or GV3101 (for Figs. 10A-17B) by electroporation. The resulting strains were

verified by PCR, grown overnight at 30°C, and used to infiltrate leaves. For the experiments of Figs. 1A-9, the leaves were from 6- to 8-week-old growth chamber-grown *N. benthamiana*, tobacco (*N. tabacum* cv. 81V9), or lettuce (*Lactuca sativa* cv. Lolla Rosa) plants. For the experiments of Figs. 10A-17B, the leaves were from 5- to 6-week-old *Nicotiana benthamiana*, tobacco (*N. tabacum*), or lettuce (*Lactuca sativa* "Black Seeded Simpson") maintained at 23-25°C. Briefly, the bacteria were pelleted by centrifugation for 5 min at 5,000*g* and then resuspended in infiltration buffer (10 mM 2-(N-morpholino)ethanesulfonic acid (MES), pH 5.5 and 10 mM $MgSO_4$) to OD600 = 0.2. The resulting bacterial suspensions were infiltrated using a syringe without needle into fully expanded leaves through a small puncture (Huang and Mason 2004). Plant tissue was harvested 5 days post infiltration (DPI). Leaves producing GFP were photographed under UV illumination generated by a B-100AP lamp (UVP, Upland, CA).

d. GFP assay

[00136] Total protein was extracted from leaf samples harvested at 2-3 DPI with extraction buffer (25 mM sodium phosphate buffer, pH 6.6, 100 mM NaCl, 1 mM EDTA, 0.05% Triton X-100, 50 mM sodium ascorbate, and 10 µg/ml leupeptin) using a FastPrep machine (Bio101). Cleared supernatants were obtained by centrifugation at 13,000g for 10 min. The protein concentration from the leaf samples was determined using Bradford reagent (Bio-Rad®) with bovine serum albumin (BSA) as the reference standard. The GFP fluorescence intensity was examined on a microplate reader (Molecular Device Co, Spectra Max M2). GFP samples were prepared by serial 2-fold dilution with phosphate buffered saline (PBS, 137 mM NaCl, 2.6 mM KCl, 10 mM Na2HPO4, and 1.8 mM KH2PO4, pH 7.4) and 50 µl of each sample was added to black-wall 96-well plates (Corning), in duplicate. The excitation and emission wavelengths were 485 nm and 538 nm, respectively. All measurements were performed at room temperature and the reading of negative control (extract of uninfiltrated plant leaf) was subtracted before graphing. *E. coli* expressed GFP was used to generate standard curve. His-tagged GFP gene was cloned into pET28 expression vector (Invitrogen), introduced into *E. coli* strain BL21(DE3), and IPTG-induced GFP was purified using TALON His-Tag purification resin (Clontech®). GFP production was also analyzed by SDS-PAGE. Briefly, clarified plant protein extracts from 3 DPI were mixed with sample buffer (50 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 0.02 % bromophenol blue) and separated on 4-15% polyacrylamide gels (Bio-Rad®). The GFP band

was visualized under UV illumination (365 nm), and the band intensity was quantified using ImageJ software.

e. NVCP ELISA

[00137] Total leaf protein extracts prepared and normalized as described above were assayed by NVCP sandwich ELISA (Mason et al 1996). Briefly, a rabbit polyclonal anti-NVCP and a guinea pig polyclonal anti-NVCP were used as capture and detection antibodies, respectively. Insect cell-derived recombinant NVCP (Jiang et al 1992) was used as the reference standard.

f. GUS activity assay

[00138] GUS activity was measured as described by Francis and Spiker (2005). Briefly, total leaf protein was extracted from 100mg leaf tissue at 2 DPI with 1 ml GUS Extraction Buffer [150 mM sodium phosphate pH 7.0, 10mM EDTA, 10mM β-mercaptoethanol, 0.1% Triton X-100, 0.1% sarcosyl, 10 μg/ml leupeptin] using a FastPrep machine (Bio101). The cleared extract (10 μl) was incubated with 130 μl assay buffer [GUS Extraction Buffer containing 1.2 mM 4-methyl- umbelliferyl-β-D-glucuronide (MUG) (Sigma(R))] in a dark 37°C incubator for 20 minutes. The reaction was stopped by transferring 10 μl of the reaction to 190 μl stop buffer [200 mM sodium carbonate] in a black wall 96-well plate. Fluorescence was measured on a SpectraMax M2 microplate reader (Molecular Devices) at 460 nm when excited at 355nm. A standard curve was generated using 50, 25, 5, 2.5, 0.5, 0.25, and 0 μM 4-methylumbelliferone (MU) on every plate. Fluorescence values were converted to moles of MU/minute, and then standardized by protein concentration determined by Bio-Rad RC DC protein assay kit according to the kit instructions, using BSA as the reference standard.

g. RNA extraction and Quantitative PCR

[00139] Total RNA was purified from infiltrated tobacco leaves at 2 DPI using Plant RNA Reagent (Invitrogen), and residual DNA was removed by DNAfree system (Ambion). First-strand cDNA was synthesized from 1μg of total RNA and oligo dT22 primer using the Superscript III First-Strand Synthesis System (Invitrogen) according to the manufacturer's protocol. Real-time PCR for each transcript of interest was performed on an IQ5 Real-Time PCR Detection System (Bio- Rad) using gene specific primers (gfp-f and gfp-r for GFP; sNV-f and

sNV-r for NVCP) and custom-made Taqman FAM/MGB probes (GFP-p and sNV-p, Integrated DNA Technologies). Each sample was measured in triplicate for each transcript of interest and an internal reference gene. Elongation factor (EF1a) transcripts served as internal control (using primers EFf and EFr and probe EFp, Integrated DNA Technologies). Transcript levels of GFP, NVCP and EF1a were quantified using separate standard curves prepared with plasmids pBY027, pSNV210, and pCR4-Topo-EF1a, respectively. The construct pCR4-Topo-EF1a was obtained as follows. RT-PCR was performed with cDNA from wild-type *N. benthamiana* RNA using EF1f and EF1r primers, which amplify 119bp (nt 167-285) fragment of EF1a gene (accession number AY206004). The amplified product was cloned into a PCR cloning vector, pCR4-Topo (Invitrogen). The relative quantification of GFP and NVCP transcripts was normalized against EF1a transcripts.

h. Sequence Alignment of Ext Genes

[00140] The nucleotide sequences of plant Ext gene terminators were obtained from GenBank. The *N. benthamiana* Ext nucleotide sequence was obtained from the Sol Genomics Network (Bombarely et al 2012). Alignment was performed using the online program Clustal Omega.

i. Read-through PCR

[00141] First-strand cDNA was synthesized from 1 µg of total RNA, described above, with a random primer. PCRs were performed to evaluate the transcription read-through of GFP and NVCP transcripts with four sets of primers. For GFP transcripts, a forward primer (gfp-3f), specific to GFP gene, was paired with one of four reverse primers RT-0, RT-1, RT-2, and RT-3, positioned at 42, 156, 291, and 389, respectively, downstream of the EcoRI site on the expression vectors. For NVCP transcripts, a NVCP specific sense primer sNV-3f was paired with the same four antisense primers used for GFP transcripts test.

j. Analysis of transcript 3' ends by circularized RT-PCR

[00142] Circularized RT-PCR was performed as described, with modification (Slomovic and Schuster 2013). Total RNA was purified from leaves of *N. benthamiana* at 4 days post infiltration using RNeasy Plant Mini Kit (Qiagen®). Prior to circularization, 5 µg total RNA was

decapped with 25 U of RNA 5' pyrophosphohydrolase (NEB) and 40 U of RNAseOUT RNA inhibitor (Invitrogen) in a total volume of 50 µl buffer (20 mM Tris-HCl, pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100) for 1 hour at 37°C. The reaction was stopped by addition of 1 µl of 500 mM EDTA, and heated to 65°C for 5 minutes. Decapped RNA was purified using a spin column (Qiagen®). RNA was then circularized with 10 U of T4 RNA Ligase (NEB) and 40 U RNAseOUT in a total volume of 20 µl buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 10% PEG8000, 50 µM ATP) for 1.5 hours at 25°C. The reaction was terminated by boiling for 2 minutes. First-strand cDNA was synthesized using Superscript III First-Strand Synthesis System (Invitrogen®) according to the manufacturer's protocol from purified circularized RNA using the primer GFP-R1, which binds near the 5' end of the GFP mRNA in the reverse direction. PCRs were performed by standard procedures with the cDNA obtained. Divergent primers were used for amplification of the fused 3' and 5' ends of the GFP mRNA: the forward primer, GFP-F1, which binds the 3' end of the GFP gene, and the reverse primer GFP-R2, which binds to the 5' end of the GFP gene, were designed to specifically amplify a cDNA product derived from a circularized transcript. A second round of PCR using primers GFP-R2 and GFP-F2 to increase specificity was also performed as suggested (Slomovic and Schuster 2013). The PCR products were cloned using native SacI-XhoI sites present in the mRNA, and 14 positive clones were randomly selected and sequenced.

k. Protein extraction and fluorescence analysis

[00143] Total protein extract was obtained by homogenizing agroinfiltrated leaf samples with 1:5 (*w:v*) ice cold extraction buffer (25 mM sodium phosphate, pH 7.4, 100 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 10 mg/mL sodium ascorbate, 0.3 mg/mL PMSF) using a Bullet Blender machine (Next Advance, Averill Park, NY) following the manufacturer's instruction. To enhance solubility, the homogenized samples were end-over-end mixed at room temperature or 4°C for 30 min. The crude plant extract was clarified by centrifugation at 13,000*g* for 10 min at 4°C. Protein concentration of clarified leaf extracts was measured using a Bradford protein assay kit (Bio-Rad®) with bovine serum albumin as standard. For SDS-PAGE, clarified plant proteins extract were mixed with sample buffer containing a final concentration of 50 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 0.02 % bromophenol blue, and separated on 4-15% polyacrylamide gels (Bio-Rad®). For GFP or DsRed fluorescence, PAGE gels were visualized under UV

illumination (365 nm) and stained with Coomassie stain (Bio-Rad®) following the manufacturer's instructions. The fluorescent band corresponding to GFP or DsRed was analyzed using ImageJ software to quantify the band intensity using native plant protein bands as an internal loading control.

l. SDS-PAGE and Western Blot

[00144] Clarified plant protein extract was mixed with sample buffer (50 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 0.5M DTT, 0.02 % bromophenol blue), boiled for 10 min, and then separated on 4-15% polyacrylamide gels (Bio-Rad). Polyacrylamide gels were either transferred to a PVDF membrane or stained with Coomassie stain (Bio-Rad) following the manufacturer's instructions. For GII.4 norovirus capsid detection, the protein transferred membranes were blocked with 5% dry milk in PBST (PBS with 0.05% Tween-20) for 1 h at 37˚C and probed in succession with polyclonal rabbit-anti GII.4 [44] diluted 1:5000 in 1% PBSTM followed by goat anti-rabbit IgG antibody-horseradish peroxidase conjugate (Sigma) diluted 1:5000 in 1% PBSTM. Bound antibody was detected with ECL reagent (Amersham).

m. Protein Quantification by ELISA

[00145] Norovirus capsid production was analyzed by sandwich ELISA. A rabbit polyclonal anti-GI or anti-GII antibody was bound to 96-well high-binding polystyrene plates (Corning), and the plates were blocked with 5% nonfat dry milk in PBST. After washing the wells with PBST (PBS with 0.05% Tween 20), the plant extracts were added and incubated. The bound norovirus capsids were detected by incubation with guinea pig polyclonal anti-GI or anti-GII antibody followed by goat anti-guinea pig IgG antibody-horseradish peroxidase conjugate (Sigma). The plate was developed with TMB substrate (Pierce) and the absorbance was read at 450nm. Plant-produced GI or GII capsids were used as the reference standard (Kentucky Bio Processing).

References Cited

- Ali S, Taylor WC (2001) The 3' non-coding region of a C4 photosynthesis gene increases transgene expression when combined with heterologous promoters Plant Mol Biol 46:325-333
- Allen, E., and Howell, M. D. (2010). miRNAs in the biogenesis of trans-acting siRNAs in higher plants. Semin. Cell Dev. Biol. 21, 798–804. doi:10.1016/j.semcdb.2010.03.008.
- Allen, G. C., Hall, G., Michalowski, S., Newman, W., Spiker, S., Weissinger, A. K., et al. (1996). High-Level Transgene Expression in Plant Cells : Effects of a Strong Scaffold Attachment Region from Tobacco. Plant Cell 8, 899–913. doi:10.1105/tpc.8.5.899.Arntzen CJ (2008) Plant science. Using tobacco to treat cancer Science 321:1052-1053 doi:10.1126/science.1163420
- Andersen, P. K., Lykke-Andersen, S., and Jensen, T. H. (2012). Promoter-proximal polyadenylation sites reduce transcription activity. Genes Dev. 26, 2169–2179. doi:10.1101/gad.189126.112.
- Baeg K, Iwakawa H, Tomari Y (2017) The poly(A) tail blocks RDR6 from converting self mRNAs into substrates for gene silencing. Nat Plants 3:17036. doi: 10.1038/nplants.2017.36
- Béclin, C., Boutet, S., Waterhouse, P., and Vaucheret, H. (2002). A branched pathway for transgene-induced RNA silencing in plants. Curr. Biol. 12, 684–688. doi:10.1016/S0960-9822(02)00792-3.
- Beyene G, Buenrostro-Nava MT, Damaj MB, Gao S-J, Molina J, Mirkov TE (2010) Unprecedented enhancement of transient gene expression from minimal cassettes using a double terminator Plant Cell Reports 30:13-25 doi:10.1007/s00299-010-0936-3
- Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB (2012) A Draft Genome Sequence of Nicotiana benthamianato Enhance Molecular Plant- Microbe Biology Research. Mol Plant-Microbe Interact 25:1523–1530. doi: 10.1094/mpmi-06-12-0148-ta
- Calikowski, T. T., Meulia, T., and Meier, I. (2003). A proteomic study of the arabidopsis nuclear matrix. J. Cell. Biochem. 90, 361–78. doi:10.1002/jcb.10624.
- Carrington JC, Freed DD, Leinicke AJ (1991) Bipartite Signal Sequence Mediates Nuclear Translocation of the Plant Potyviral NIa Protein The Plant Cell 3:953 doi:10.2307/3869157

- Chen, Q., and Davis, K. R. (2016). The potential of plants as a system for the development and production of human biologics. F1000Research 5, 912. doi:10.12688/f1000research. 8010.1.
- Chen, Q., Dent, M., Hurtado, J., Stahnke, J., McNulty, A., Leuzinger, K., et al. (2016). Transient Protein Expression by Agroinfiltration in Lettuce. Methods Mol. Biol. 1385, 55–67. doi:10.1007/978-1-4939-3289-4_4.
- Chung BYW, Simons C, Firth AE, Brown CM, Hellens RP (2006) BMC Genomics 7:120 doi:10.1186/1471-2164-7-120
- Collens JI, Mason HS, Curtis WR (2007) Agrobacterium-mediated viral vector-amplified transient gene expression in Nicotiana glutinosa plant tissue culture Biotechnol Prog 23:570-576 doi:10.1021/bp060342u
- Collens JI, Mason HS, Curtis WR (2008) Agrobacterium-Mediated Viral Vector-Amplified Transient Gene Expression in Nicotiana glutinosa Plant Tissue Culture Biotechnology Progress 23:570-576 doi:10.1021/bp060342u
- D'Aoust, M.-A., Couture, M. M.-J., Charland, N., Trépanier, S., Landry, N., Ors, F., et al. (2010). The production of hemagglutinin-based virus-like particles in plants: a rapid, efficient and safe response to pandemic influenza. Plant Biotechnol. J. 8, 607–19. doi:10.1111/j.1467-7652.2009.00496.x.
- Diamos AG, Rosenthal SH, Mason HS (2016) 5′ and 3′ Untranslated Regions Strongly Enhance Performance of Geminiviral Replicons in Nicotiana benthamiana Leaves Frontiers in Plant Science 7 doi:10.3389/fpls.2016.00200
- Dong H et al. (2007) An exploration of 3′-end processing signals and their tissue distribution in Oryza sativa Gene 389:107-113 doi:10.1016/j.gene.2006.10.015
- Fan, Q., Treder, K., and Miller, W. A. (2012). Untranslated regions of diverse plant viral RNAs vary greatly in translation enhancement efficiency. BMC Biotechnol. 12, 22. doi:10.1186/1472-6750-12-22.
- Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., et al. (2015). The Sol Genomics Network (SGN)--from genotype to phenotype to breeding. Nucleic Acids Res. 43, D1036-41. doi:10.1093/nar/gku1195.

- Fox, J. L. (2012). First plant-made biologic approved. Nat. Biotechnol. 30, 472–472. doi:10.1038/nbt0612-472.Francis KE, Spiker S (2004) Identification of Arabidopsis thaliana transformants without selection reveals a high occurrence of silenced T-DNA integrations The Plant Journal 41:464-477 doi:10.1111/j.1365-313x.2004.02312.x
- Gilmartin GM (2005) Eukaryotic mRNA 3' processing: a common means to different ends Genes & Development 19:2517-2521 doi:10.1101/gad.1378105
- Gleba YY, Tusé D, Giritch A (2014) Plant viral vectors for delivery by Agrobacterium. In: Current Topics in Microbiology and Immunology. pp 155–192Gray NK, Hentze MW (1994) Regulation of protein synthesis by mRNA structure Molecular Biology Reports 19:195-200 doi:10.1007/bf00986961
- Halweg, C. (2005). The Rb7 Matrix Attachment Region Increases the Likelihood and Magnitude of Transgene Expression in Tobacco Cells: A Flow Cytometric Study. Plant Cell Online 17, 418–429. doi:10.1105/tpc.104.028100.
- Hiatt, A., Bohorova, N., Bohorov, O., Goodman, C., Kim, D., Pauly, M. H., et al. (2014). Glycan variants of a respiratory syncytial virus antibody with enhanced effector function and in vivo efficacy. Proc. Natl. Acad. Sci. 111, 5992–5997. doi:10.1073/pnas.1402458111.
- Hori K, Watanabe Y (2007) Context analysis of termination codons in mRNA that are recognized by plant NMD. Plant Cell Physiol 48:1072–8. doi:10.1093/pcp/pcm075
- Hirsinger C, Parmentier Y, Durr A, Fleck J, Jamet E (1997) Plant Molecular Biology 33:279-289 doi:10.1023/a:1005738815383
- Huang Z, Mason HS (2004) Conformational analysis of hepatitis B surface antigen fusions in an Agrobacterium-mediated transient expression system Plant Biotechnology Journal 2:241-249 doi:10.1111/j.1467-7652.2004.00068.x
- Huang, Z., Chen, Q., Hjelm, B., Arntzen, C., and Mason, H. (2009). A DNA replicon system for rapid high-level production of virus-like particles in plants. Biotechnol. Bioeng. 103, 706–714. doi:10.1002/bit.22299.
- Huang, Z., Phoolcharoen, W., Lai, H., Piensook, K., Cardineau, G., Zeitlin, L., et al. (2010). High-level rapid production of full-size monoclonal antibodies in plants by a single-vector DNA replicon system. Biotechnol. Bioeng. 106, 9–17. doi:10.1002/bit.22652.
- Hunt AG (2008) Messenger RNA 3’ end formation in plants. Curr. Top. Microbiol. Immunol. 326:151–177 doi:10.1007/978-3-540-76776-3_9

- Ingelbrecht LW, Herman LMF, Dekeyser RA, Montagu MC Van, Depicker AG (1989) Different 3' End Regions Strongly lnfluence the Level of Gene Expression in Plant Cells. Plant Cell 1:671–680 doi:10.1105/tpc.1.7.671
- Ji G, Zheng J, Shen Y, Wu X, Jiang R, Lin Y, Loke JC, Davis KM, Reese GJ, Li QQ (2007) Predictive modeling of plant messenger RNA polyadenylation sites. BMC Bioinformatics 8:43. doi: 10.1186/1471-2105-8-43
- Ji, L., Xu, R., Lu, L., Zhang, J., Yang, G., Huang, J., et al. (2013). TM6, a novel nuclear matrix attachment region, enhances its flanking gene expression through influencing their chromatin structure. Mol. Cells 36, 127–137. doi:10.1007/s10059-013-0092-z.
- Jiang X, Wang M, Graham DY, Estes MK (1992) Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein J Virol 66:6527-6532
- Kato H, Xie G, Sato Y, Imai R (2010) Isolation of Anther-specific Gene Promoters Suitable for Transgene Expression in Rice. Plant Mol Biol Report 28:381–387. doi: 10.1007/s11105-009-0162-8
- Kertész S, Kerényi Z, Mérai Z, Bartos I, Pálfy T, Barta E, Silhavy D (2006) Both introns and long 3′-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants Nucleic Acids Research 34:6147-6157 doi:10.1093/nar/gkl737
- Komarova, T. V, Baschieri, S., Donini, M., Marusic, C., Benvenuto, E., and Dorokhov, Y. L. (2010). Transient expression systems for plant-derived biopharmaceuticals. Expert Rev. Vaccines 9, 859–876. doi:10.1586/erv.10.85.
- Kooiker M, Airoldi CA, Losa A, Manzotti PS, Finzi L, Kater MM, Colombo L (2005) BASIC PENTACYSTEINE1, a GA binding protein that induces conformational changes in the regulatory region of the homeotic Arabidopsis gene SEEDSTICK. Plant Cell 17:722–729. doi: 10.1105/tpc.104.030130
- Lacroix, B., Loyter, A., and Citovsky, V. (2008). Association of the Agrobacterium T-DNA-protein complex with plant nucleosomes. Proc. Natl. Acad. Sci. 105, 15429–15434. doi:10.1073/pnas.0805641105.
- Lai, H., He, J., Engle, M., Diamond, M. S., and Chen, Q. (2012). Robust production of virus-like particles and monoclonal antibodies with geminiviral replicon vectors in lettuce. Plant Biotechnol. J. 10, 95–104. doi:10.1111/j.1467-7652.2011.00649.x.

- Libri D, Dower K, Boulay J, Thomsen R, Rosbash M, Jensen TH (2002) Interactions between mRNA export commitment, 3`-end quality control, and nuclear degradation. Mol Cell Biol 22:8254–8266. doi: 10.1128/MCB.22.23.8254- 8266.2002
- Liebich, I., Bode, J., Reuter, I., and Wingender, E. (2002). Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). Nucleic Acids Res. 30, 3433–3442. doi:10.1093/nar/gkf446.
- Limkul, J., Misaki, R., Kato, K., and Fujiyama, K. (2015). The combination of plant translational enhancers and terminator increase the expression of human glucocerebrosidase in Nicotiana benthamiana plants. Plant Sci. 240, 41–49. doi:10.1016/j.plantsci.2015.08.018.
- Loke JC (2005) Compilation of mRNA Polyadenylation Signals in Arabidopsis Revealed a New Signal Element and Potential Secondary Structures. Plant Physiol 138:1457–1468. doi: 10.1104/pp.105.060541
- Lomonossoff GP, D'Aoust M-A (2016) Plant-produced biopharmaceuticals: A case of technical developments driving clinical deployment. Science 353:1237–40. doi: 10.1126/science.aaf6638
- Luo Z, Chen Z (2007) Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis. Plant Cell 19:943–58. doi: 10.1105/tpc.106.045724
- Lyon, G. M., Mehta, A. K., Varkey, J. B., Brantly, K., Plyler, L., McElroy, A. K., et al. (2014). Clinical Care of Two Patients with Ebola Virus Disease in the United States. N. Engl. J. Med. 371, 2402–2409. doi:10.1056/nejmoa1409838.
- Macfarlane, S. A., Gilmer, D., and Davies, J. W. (1992). Efficient inoculation with CaMV 35 S promoter-driven DNA clones of the tobravirus PEBV. Virology 187, 829–831. doi:10.1016/0042-6822(92)90488-b.
- Mandel CR, Bai Y, Tong L (2007) Protein factors in pre-mRNA 3′-end processing Cellular and Molecular Life Sciences 65:1099-1122 doi:10.1007/s00018-007- 7474-3
- Mankin, S. L., Allen, G. C., Phelan, T., Spiker, S., and Thompson, W. F. (2003). Elevation of transgene expression level by flanking matrix attachment regions (MAR) is promoter dependent: a study of the interactions of six promoters with the RB7 3' MAR. Transgenic Res 12, 3–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12650520.

- Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH (2010) Crosstalk between mRNA 3′ End Processing and Transcription Initiation Molecular Cell 40:410-422 doi:10.1016/j.molcel.2010.10.012
- Mason HS, Ball JM, Shi JJ, Jiang X, Estes MK, Arntzen CJ (1996) Expression of Norwalk virus capsid protein in transgenic tobacco and potato and its oral immunogenicity in mice Proceedings of the National Academy of Sciences 93:5335-5340 doi:10.1073/pnas.93.11.5335
- Mason HS, Richter LJ, Thanavala Y, Arntzen CJ (2000) Nature Biotechnology 18:1167-1171 doi:10.1038/81153
- Mathew LG, Maloney B, Takeda N, Mason HS (2011) Spurious polyadenylation of Norovirus Narita 104 capsid protein mRNA in transgenic plants Plant Molecular Biology 75:263-275 doi:10.1007/s11103-010-9725-1
- Menossi M, Rabaneda F, Puigdomenech P, Martinez-Izquierdo JA (2003) Analysis of regulatory elements of the promoter and the 3' untranslated region of the maize Hrgp gene coding for a cell wall protein Plant Cell Rep 21:916-923 doi:10.1007/s00299-003-0602-0
- Meshcheriakova, Y. A., Saxena, P., and Lomonossoff, G. P. (2014). Fine-tuning levels of heterologous gene expression in plants by orthogonal variation of the untranslated regions of a nonreplicating transient expression system. Plant Biotechnol. J. 12, 718–727. doi:10.1111/pbi.12175.
- Millevoi S, Vagner S (2009) Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation Nucleic Acids Research 38:2757-2774 doi:10.1093/nar/gkp1176
- Mogen BD, MacDonald MH, Graybosch R, Hunt AG (1990) Upstream Sequences Other than AAUAAA Are Required for Efficient Messenger RNA 3'-End Formation in Plants The Plant Cell 2:1261 doi:10.2307/3869344
- Mlynárová, L., Hricová, A., Loonen, A., and Nap, J.-P. (2003). The presence of a chromatin boundary appears to shield a transgene in tobacco from RNA silencing. Plant Cell 15, 2203–17. doi:10.1105/tpc.012070.
- Moore MJ, Proudfoot NJ (2009) Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation Cell 136:688-700 doi:10.1016/j.cell.2009.02.001

- Mor TS, Moon Y-S, Palmer KE, Mason HS (2002) Geminivirus vectors for high-level expression of foreign proteins in plant cells Biotechnology and Bioengineering 81:430-437 doi:10.1002/bit.10483
- Nagaya S, Kawamura K, Shinmyo A, Kato K (2009) The HSP Terminator of Arabidopsis thaliana Increases Gene Expression in Plant Cells Plant and Cell Physiology 51:328-332 doi:10.1093/pcp/pcp188
- Nandi, S., Kwong, A. T., Holtz, B. R., Erwin, R. L., Marcel, S., and McDonald, K. A. (2016). Techno-economic analysis of a transient plant-based platform for monoclonal antibody production. MAbs 8, 1456–1466. doi:10.1080/19420862.2016.1227901.
- Narsai R, Howell KA, Millar AH, O'Toole N, Small I, Whelan J (2007) Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in Arabidopsis thaliana. PLANT CELL ONLINE 19:3418–3436. doi: 10.1105/tpc.107.055046
- Olinger, G. G., Pettitt, J., Kim, D., Working, C., Bohorov, O., Bratcher, B., et al. (2012). Delayed treatment of Ebola virus infection with plant-derived monoclonal antibodies provides protection in rhesus macaques. Proc. Natl. Acad. Sci. U. S. A. 109, 18030–5. doi:10.1073/pnas.1213709109.
- Pietrzak, M., Shillito, R. D., Hohn, T., and Potrykus, I. (1986). Expression in plants of two bacterial antibiotic resistance genes after protoplast transformation with a new plant expression vector. Nucleic Acids Res. 14, 5857–5868. doi:10.1093/nar/14.14.5857.
- Qin, C., Shi, N., Gu, M., Zhang, H., Li, B., Shen, J., et al. (2012). Involvement of RDR6 in short-range intercellular RNA silencing in Nicotiana benthamiana. Sci. Rep. 2, 467. doi:10.1038/srep00467.
- Qiu, X., Wong, G., Audet, J., Bello, A., Fernando, L., Alimonti, J. B., et al. (2014). Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. Nature 514, 47–53. doi:10.1038/nature13777.
- Qu X, Lykke-Andersen S, Nasser T, Saguez C, Bertrand E, Jensen TH, Moore C (2009) Assembly of an Export-Competent mRNP Is Needed for Efficient Release of the 3'-End Processing Complex after Polyadenylation Molecular and Cellular Biology 29:5327-5338 doi:10.1128/mcb.00468-09
- Que, Q., Wang, H. Y., English, J. J., and Jorgensen, R. A. (1997). The Frequency and Degree of Cosuppression by Sense Chalcone Synthase Transgenes Are Dependent on Transgene

Promoter Strength and Are Reduced by Premature Nonsense Codons in the Transgene Coding Sequence. Plant Cell 9, 1357–1368. doi:10.1105/tpc.9.8.1357.

- Rethmeier N, Seurinck J, Van Montagu M, Cornelissen M (1997) Intron-mediated enhancement of transgene expression in maize is a nuclear, gene-dependent process The Plant Journal 12:895-899 doi:10.1046/j.1365-313x.1997.12040895.x
- Richter, L. J., Thanavala, Y., Arntzen, C. J., and Mason, H. S. (2000). Production of hepatitis B surface antigen in transgenic plants for oral immunization. Nat. Biotechnol. 18, 1167–71. doi:10.1038/81153.
- Rose AB (2008) Intron-Mediated Regulation of Gene Expression. Springer Berlin Heidelberg. doi:10.1007/978-3-540-76776-3_15
- Rosenthal, S. H., Diamos, A. G., and Mason, H. S. (2018). An intronless form of the tobacco extensin gene terminator strongly enhances transient gene expression in plant leaves. Plant Mol. Biol. doi:10.1007/s11103-018-0708-y.
- Rothnie HM (1996) Plant mRNA 3?-end formation Plant Molecular Biology 32:43-61 doi:10.1007/bf00039376
- Rothnie HM, Chen G, Futterer J, Hohn T (2001) Polyadenylation in Rice Tungro Bacilliform Virus: cis-Acting Signals and Regulation Journal of Virology 75:4184- 4194 doi:10.1128/jvi.75.9.4184-4194.2001
- Sainsbury, F., and Lomonossoff, G. P. (2008). Extremely high-level and rapid transient protein production in plants without the use of viral replication. Plant Physiol. 148, 1212–8. doi:10.1104/pp.108.126284.
- Sainsbury, F., Thuenemann, E. C., and Lomonossoff, G. P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol. J. 7, 682–93. doi:10.1111/j.1467-7652.2009.00434.x.
- Sanfacon H, Brodmann P, Hohn T (1991) A dissection of the cauliflower mosaic virus polyadenylation signal Genes & Development 5:141-149 doi:10.1101/gad.5.1.141
- Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., and Schmidt, R. (2004). Silencing in Arabidopsis T-DNA transformants: the predominant role of a gene-specific RNA sensing mechanism versus position effects. Plant Cell 16, 2561–72. doi:10.1105/tpc.104.024547.

- Sharma AK, Sharma MK (2009) Plants as bioreactors: Recent developments and emerging opportunities Biotechnology Advances 27:811-832 doi:10.1016/j.biotechadv.2009.06.004
- Sheldon CC, Conn AB, Dennis ES, Peacock WJ (2002) Different Regulatory Regions Are Required for the Vernalization-Induced Repression of FLOWERING LOCUS C and for the Epigenetic Maintenance of Repression. Plant Cell Online 14:2527–2537. doi: 10.1105/tpc.004564
- Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR (2010) A Bioinformatics Approach to the Identification, Classification, and Analysis of Hydroxyproline-Rich Glycoproteins PLANT PHYSIOLOGY 153:485-513 doi:10.1104/pp.110.156554
- Sieburth LE, Meyerowitz EM (1997) Molecular Dissection of the AGAMOUS Control Region Shows That cis Elements for Spatial Regulation Are Located Intragenically The Plant Cell 9:355 doi:10.2307/3870487
- Simon, A. E., and Miller, W. A. (2013). 3' cap-independent translation enhancers of plant viruses. Annu. Rev. Microbiol. 67, 21–42. doi:10.1146/annurev-micro-092412-155609.
- Singh, G., Pratt, G., Yeo, G. W., and Moore, M. J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu. Rev. Biochem. 84, 325–354. doi:10.1146/annurev-biochem-080111-092106.
- Slomovic S, Schuster G (2013) Circularized RT-PCR (cRT-PCR). Elsevier. doi:10.1016/b978-0-12-420037-1.00013-0
- Strasser, R., Altmann, F., and Steinkellner, H. (2014). Controlled glycosylation of plant-produced recombinant proteins. Curr. Opin. Biotechnol. 30, 95–100. doi:10.1016/j.copbio.2014.06.008.
- Streatfield SJ (2007) Approaches to achieve high-level heterologous protein production in plants Plant Biotechnology Journal 5:2-15 doi:10.1111/j.1467- 7652.2006.00216.x
- Tan-Wong SM, Wijayatilake HD, Proudfoot NJ (2009) Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex Genes & Development 23:2610-2624 doi:10.1101/gad.1823209
- Tan-Wong SM et al. (2012) Gene Loops Enhance Transcriptional Directionality Science 338:671-675 doi:10.1126/science.1224350

- Thanavala Y, Huang Z, Mason HS (2006) Plant–derived vaccines: a look back at the highlights and a view to the challenges on the road ahead Expert Review of Vaccines 5:249-260 doi:10.1586/14760584.5.2.249
- Tian B, Manley JL (2013) Alternative cleavage and polyadenylation: the long and short of it Trends in Biochemical Sciences 38:312-320 doi:10.1016/j.tibs.2013.03.005
- Tokuhisa, J. G., Singh, K., Dennis, E. S., and Peacock, W. J. (1990). A DNA-Binding Protein Factor Recognizes Two Binding Domains within the Octopine Synthase Enhancer Element. Plant Cell 2, 215. doi:10.2307/3869136.
- Tora L, Dantonel J-C, Murthy KGK, Manley JL (1997) Nature 389:399-402 doi: 10.1038/38763
- Tusé, D., Tu, T., and McDonald, K. A. (2014). Manufacturing Economics of Plant-Made Biologics: Case Studies in Therapeutic and Industrial Enzymes. Biomed Res. Int. 2014, 1–16. doi:10.1155/2014/256135.
- Wigington, C. P., Williams, K. R., Meers, M. P., Bassell, G. J., and Corbett, A. H. (2014). Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions. Wiley Interdiscip. Rev. RNA 5, 601–622. doi:10.1002/wrna.1233.
- Wolterink-van Loo, S., Escamilla Ayala, A. A., Hooykaas, P. J. J., and van Heusden, G. P. H. (2015). Interaction of the Agrobacterium tumefaciens virulence protein VirD2 with histones. Microbiology 161, 401–10. doi:10.1099/mic.0.083410-0.
- Xing A, Moon BP, Mills KM, Falco SC, Li Z (2010) Revealing frequent alternative polyadenylation and widespread low-level transcription read-through of novel plant transcription terminators Plant Biotechnology Journal 8:772-782 doi:10.1111/j.1467-7652.2010.00504.x
- Zhang X, Mason H (2006) Bean Yellow Dwarf Virus replicons for high-level transgene expression in transgenic plants and cell cultures Biotechnology and Bioengineering 93:271-279 doi:10.1002/bit.20695
- Zhang, J., Lu, L., Ji, L., Yang, G., and Zheng, C. (2009). Functional characterization of a tobacco matrix attachment region-mediated enhancement of transgene expression. Transgenic Res. 18, 377–385. doi:10.1007/s11248-008-9230-3.

- Zhao H, Xing D, Li QQ (2009) Unique features of plant cleavage and polyadenylation specificity factor revealed by proteomic studies. Plant Physiol 151:1546–56. doi: 10.1104/pp.109.142729
- Zhao J, Hyman L, Moore C (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis Microbiol Mol Biol Rev 63:405-445
- Zhao, C.-P., Guo, X., Chen, S.-J., Li, C.-Z., Yang, Y., Zhang, J.-H., et al. (2017). Matrix attachment region combinations increase transgene expression in transfected Chinese hamster ovary cells. Sci. Rep. 7, 42805. doi:10.1038/srep42805.
- Zimran, A., Brill-Almon, E., Chertkoff, R., Petakov, M., Blanco-Favela, F., Munoz, E. T., et al. (2011). Pivotal trial with plant cell-expressed recombinant glucocerebrosidase, taliglucerase alfa, a novel enzyme replacement therapy for Gaucher disease. Blood 118, 5767–5773. doi:10.1182/blood-2011-07-366955.

Claims

We claim:

1. A plant expression vector comprising an expression cassette, wherein the expression cassette comprises a 5' UTR and a 3' UTR, wherein the 3' UTR comprises:

a first terminator; and

a second terminator, a chromatin scaffold/matrix attachment region (MAR), or both.

2. The plant expression vector of claim 1, wherein the first terminator and second terminator form a double terminator.

3. The plant expression vector of claim 2, wherein the 3' UTR further comprises MAR downstream of the double terminator.

4. The plant expression vector of claim 1, wherein the 3' UTR comprises MAR.

5. The plant expression vector of claim 4, wherein the double terminator increases protein expression from the expression cassette.

6. The plant expression vector of claim 1, wherein the first terminator is intronless tobacco extension terminator (EU) and the second terminator is selected from the group consisting of: *Nicotiana benthamiana* actin 3' UTR (NbACT3), p19 suppressor of RNA silencing from tomato bushy stunt virus (P19), *N. benthamiana* 18.8 kDa class II heat shock protein 3' UTR (NbHSP), short intergenic region of bean yellow dwarf virus (SIR), agrobacterium nopaline synthase 3' UTR (NOS), cauliflower mosaic virus 35S 3' UTR (35S), tobacco mosaic virus 3' UTR (TMV), BDB501 (bean dwarf mosaic virus DNA B nuclear shuttle protein 3' UTR, the intergenic region, the 3' end of the movement protein, and additional 200 nt downstream of the movement protein sequence), tobacco necrosis virus-D 3' UTR (TNVD), pea enation mosaic virus 3' UTR (PEMV), and barley yellow dwarf virus 3' UTR (BYDV).

7. The plant expression vector of claim 6, wherein EU is upstream of the second terminator.

8. The plant expression vector of claim 6, wherein the 3' UTR comprises EU and 35S, wherein 35S is upstream of EU.

9. The plant expression vector of claim 1, wherein the first terminator is intron-containing tobacco extension terminator (IEU) and the second terminator is selected from the group consisting of: SIR, 35S, and long intergenic region from bean yellow dwarf virus (LIR).

10. The plant expression vector of claim 9, wherein IEU is upstream of the second terminator.

11. The plant expression vector of claim 1, wherein the 3' UTR comprises EU and MAR, wherein MAR is selected from the group consisting of: Rb7 and TM6.

12. The plant expression vector of claim 1, wherein the 3' UTR comprises IEU and MAR, wherein MAR is selected from the group consisting of: Rb7 and TM6.

13. The plant expression vector of claim 1, wherein the 3' UTR comprises the first terminator, the second terminator, and MAR.

14. The plant expression vector of claim 13, wherein the 3' UTR comprises the combination of the first terminator, the second terminator, and MAR, wherein the first terminator, the second terminator and MAR are selected from the group consisting of:

EU, 35S, and Rb7, wherein EU is downstream or upstream of 35S;

EU, NbACT3, and Rb7, wherein EU is upstream of NbACT3;

EU, BD501, and Rb7, wherein EU is upstream of BD501;

EU, *A. thaliana* heat shock protein 3' UTR (AtHSP), and Rb7, wherein EU is downstream of AtHSP;

EU, 35S, and TM6, wherein EU is upstream of 35S; and

IEU, 35S, and Rb7, wherein IEU is upstream of 35S.

15. The plant expression vector of claim 14, wherein the sequence of the plant expression vector is set forth in SEQ ID NO. 115, SEQ ID NO. 116, or SEQ ID NO. 117.

16. A plant expression vector comprising an expression cassette, wherein the expression cassette comprises a 5' UTR and a 3' UTR, wherein the 3' UTR comprises at least one terminator selected from the group consisting of: EU, IEU, NbACT3, NbACT617 (downstream 617-nt region of NbACT3), NbACT567 (downstream 567nt of NbACT3), Pin2, BDB501, BDB282 (282 nucleotides comprising bean dwarf mosaic virus DNA B nuclear shuttle protein 3' UTR, the intergenic region, and the 3' end of the movement protein), NbHSP, NbHSPb (NbHSP missing 75 nt from 5' end), bean dwarf mosaic virus rep gene 3' UTR (Rep), pea rubisco small subunit 3' UTR (RbcS), SIR, SIR 5'/3' (SIR with additional sequences both upstream and downstream), SIR 3' (SIR with its additional downstream viral sequence), AtHSP, 35S, bean dwarf mosaic virus repA gene 3' UTR (RepA), NOS, TMV, TNVD, PEMV, and BYDV.

17. The plant expression vector of claim 16, wherein the 3' UTR comprises at least one terminator selected from the group consisting of: NbACT3, NbACT617, NbACT567, Pin2, BDB501, BDB282, NbHSP, NbHSPb, Rep, RbcS, SIR, SIR 5'/3', SIR 3', AtHSP, and RepA.

18. The plant expression vector of claim 16, wherein the 3' UTR comprises a double terminator, wherein the double terminator is a fusion of two members selected from the group consisting of: EU, IEU, NbACT3, NbACT617, NbACT567, Pin2, BDB501, BDB282, NbHSP, NbHSPb, Rep, RbcS, SIR, SIR 5'/3', SIR 3', AtHSP, 35S, RepA, NOS, TMV, TNVD, PEMV, and BYDV.

19. The plant expression vector of claim 16, wherein the 3' UTR comprises EU and a second terminator selected from the group consisting of: NbACT, P19, NbHSP, SIR, NOS, 35S, TMV, BDB501, TNVD, PEMV, and BYDV.

20. The plant expression vector of claim 19, wherein EU is upstream of the second terminator.

21. The plant expression vector of claim 20, wherein the sequence of the plant expression vector is set forth in SEQ ID NO. 115, SEQ ID NO. 116, or SEQ ID NO. 117.

22. The plant expression vector of claim 19, wherein the 3' UTR comprises 35S and a second terminator selected from the group consisting of: NbACT3, NOS, EU, NbHSP, Pin2, and BDB501.

23. The plant expression vector of claim 21, wherein 35S is upstream of the second terminator.

24. The plant expression vector of claim 21, wherein NOS is upstream of 35S.

25. The plant expression vector of claim 16, wherein the 3' UTR comprises IEU and a second terminator selected from the group consisting of: SIR, 35S, and LIR.

26. The plant expression vector of claim 24, wherein IEU is upstream of the second terminator.

27. The plant expression vector of claim 16, wherein the 3' UTR comprises NbHSP and a second terminator selected from the group consisting of: NbACT3, NOS, and Pin2.

28. The plant expression vector of claim 26, wherein NbHSP is upstream of the second terminator.

29. The plant expression vector of any one of claims 16-28, wherein the 3' UTR further comprises MAR, wherein MAR is downstream of the at least one terminator or the second terminator wherein the second terminator is the downstream terminator.

30. The plant expression vector of claim 29, wherein the MAR is selected from the group consisting of: Rb7 and TM6.

31. The plant expression vector of claim 30, wherein the 3' UTR comprises Rb7 downstream of EU, IEU, AtHSp, 35S, BDB501, NbHSP, NOS, or NbACT3.

32. The plant expression vector of claim 30, wherein the 3' UTR comprises TM6 downstream of IEU, 35S, or NbACT3.

33. The plant expression vector of claim 30, wherein the 3' UTR comprises RB7 downstream of a double terminator selected from the group consisting of: 35S+NbACT3, EU+35S, EU+NbACT3, NbHSP+NbACT3, 35S+EU, AtHSP+NOS, 35S+NOS, EU+BDB501, AtHSP+NbHSP, NbHSP+NOS, AtHSP+EU, NbHSP+Pin2, and IEU+35S.

34. The plant expression vector of claim 40, wherein the 3' UTR comprises TM6 downstream of a double terminator selected from the group consisting of: EU+35S, 35S+NOS, NbHSP+NOS, and NbHSP+Pin2.

35. A plant expression vector comprising an expression cassette, wherein the expression cassette comprises a 5' UTR and a 3' UTR, wherein the 3' UTR comprises at least one terminator and a chromatin scaffold/matrix attachment region (MAR) selected from the group consisting of: Rb7 and TM6.

36. The plant expression vector of claim 35, wherein MAR is Rb7, the at least one terminator is selected from the group consisting of: EU, IEU, AtHSP, 35S, BDB501, NbHSP, NOS, and NbACT3.

37. The plant expression vector of claim 35, wherein MAR is Rb7, the at least one terminator comprises two terminators selected from the group consisting of: EU, IEU, AtHSP, 35S, BDB501, NbHSP, NOS, NbACT3, and Pin2.

38. The plant expression vector of claim 35, wherein the MAR is TM6, the at least one terminator is selected from the group consisting of: IEU, 35S, BDB501, NOS, and NbACT3.

39. The plant expression vector of claim 35, wherein the MAR is TM6 and the at least one terminator comprises two terminators selected from the group consisting of: EU, IEU, 35S, NbHSP, NOS, and Pin2.

40. The plant expression vector of any one of claims 1-39, wherein the vector is a non-replicating vector.

41. The plant expression vector of any one of claims 1-39, wherein the vector is a replicating vector.

42. The plant expression vector of claim 41, wherein the replicating vector is a geminiviral vector.

43. The plant expression vector of any one of claims 1-42, wherein the expression cassette comprises a promoter selected from 35S promoter, tobacco mosaic virus 5' UTR, or a combination thereof.

44. The plant expression vector of claim 43, wherein the promoter is in the 5' UTR.

45. A method of producing recombinant protein in a plant or plant part, the method comprising introducing the plant expression vector of any one of claims 1-44 into the plant or plant part.

46. The method of claim 45, wherein the plant or plant part is transformed by the plant expression vector of any one of claims 1-44 using an Agrobacterium.

47. The method of claim 46, wherein the Agrobacterium is *Agrobacterium tumefaciens*.

48. The method of claim 46, wherein the *A. tumefaciens* GV3101.

49. The method of any one of claims 45-48, wherein the plant or plant part is transformed by the vector of any one of claims 1-44 via agroinfiltration.

50. The method of any one of claim 45-49, wherein the plant is tobacco or lettuce or the plant part is from tobacco or lettuce.

51. A 3' UTR comprising a first terminator and a MAR.

52. The 3' UTR of claim 51, wherein the first terminator is EU and MAR is selected from the group consisting of: Rb7 and TM6.

53. The 3' UTR of claim 51, wherein the first terminator is IEU and MAR is selected from the group consisting of: Rb7 and TM6.

54. A 3' UTR comprising a first terminator and a second terminator, wherein the first terminator and second terminator form a double terminator.

55. The 3' UTR of claim 54, wherein the first terminator is EU and the second terminator is selected from the group consisting of: NbACT3, p19 suppressor of RNA silencing from tomato P19, NbHSP, SIR, NOS, 35S, TMV, BDB501, TNVD, PEMV, and BYDV.

56. The 3' UTR of claim 55, wherein EU is upstream of the second terminator.

57. The 3' UTR of claim 54 comprising EU and 35S, wherein 35S is upstream of EU.

58. The 3' UTR of claim 54, wherein the first terminator is IEU and the second terminator is selected from the group consisting of: SIR, 35S, and LIR.

59. The 3' UTR of claim 58, wherein IEU is upstream of the second terminator.

60. The 3' UTR of any one of claims 54-59 further comprising a MAR.

61. The 3' UTR of claim 60, wherein MAR is downstream of the double terminator.

62. The 3' UTR of claim 61, comprises a combination of the first terminator, the second terminator and MAR selected from the group consisting of:

EU, 35S, and Rb7, wherein EU is downstream or upstream of 35S;

EU, NbACT3, and Rb7, wherein EU is upstream of NbACT3;

EU, BD501, and Rb7, wherein EU is upstream of BD501;

EU, AtHSP, and Rb7, wherein EU is downstream of AtHSP;

EU, 35S, and TM6, wherein EU is upstream of 35S; and

IEU, 35S, and Rb7, wherein IEU is upstream of 35S.

Fig. 1A

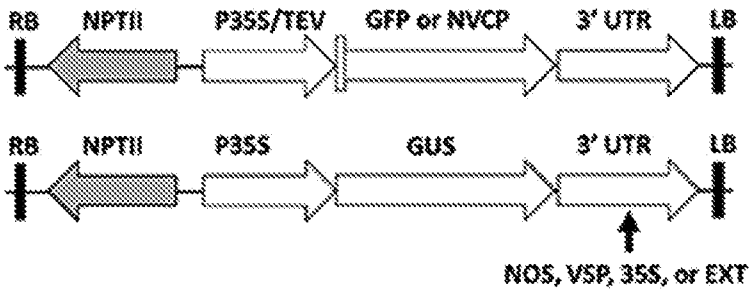


Fig. 1B

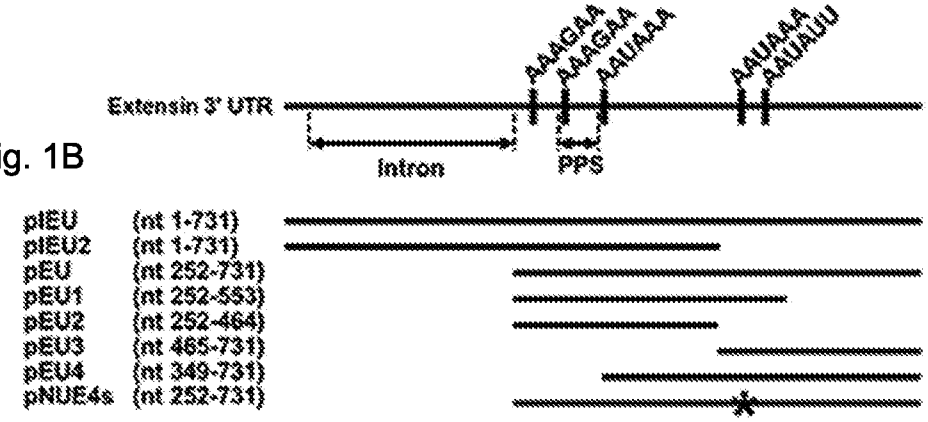


Fig. 2A

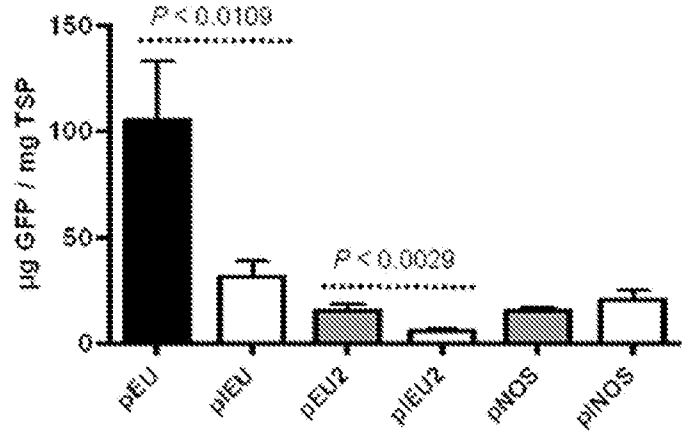
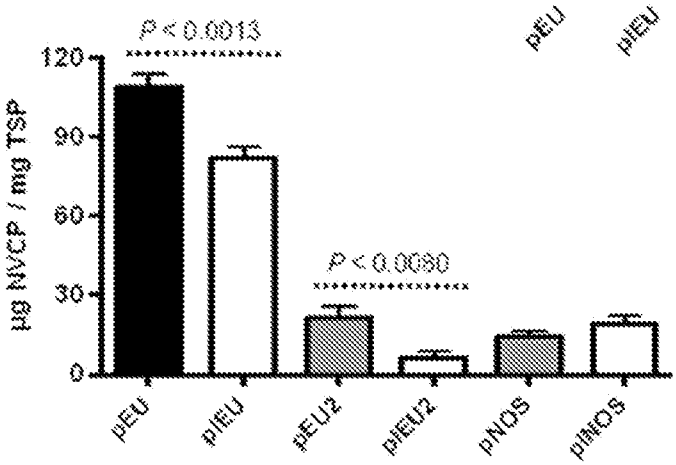
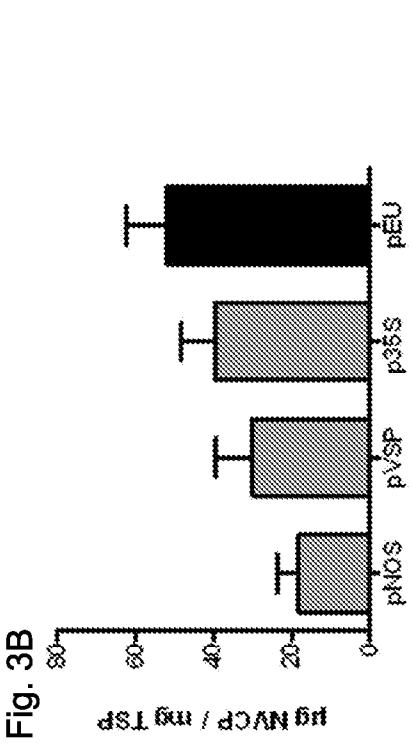
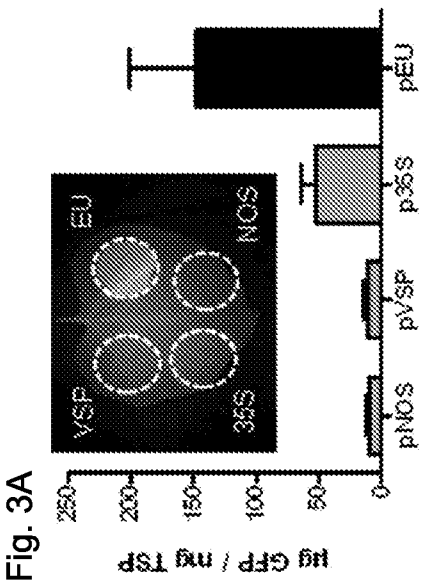
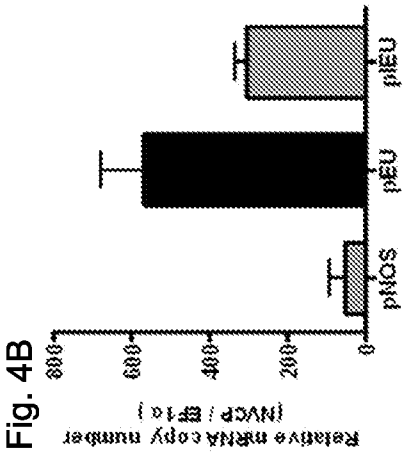


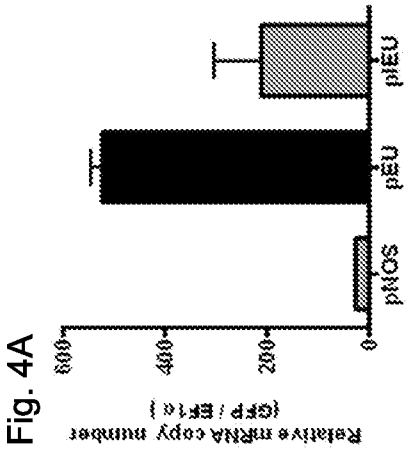
Fig. 2B




Fig. 3B
μg NVCP / mg TSP
0
20
40
60
80
pNOS
pVSP
p35S
pEU


Fig. 3A
μg GFP / mg TSP
0
50
100
150
200
250
pNOS
pVSP
p35S
pEU
VSP
EU
35S
NOS


Fig. 4B
Relative mRNA copy number
(NVCP / EF1α)
0
200
400
600
800
pNOS
pEU
pIEU


Fig. 4A
Relative mRNA copy number
(GFP / EF1α)
0
200
400
600
pNOS
pEU
pIEU

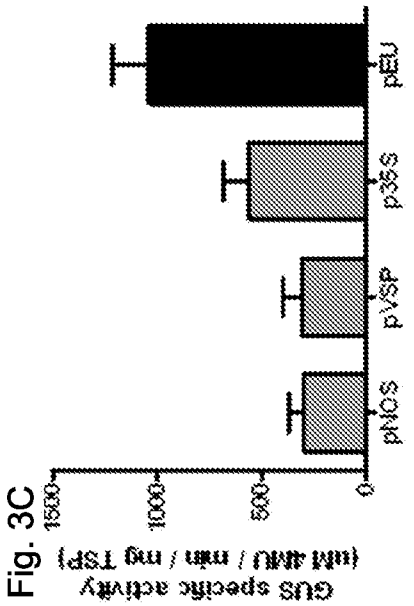

Fig. 3C
GUS specific activity
(uM 4MU / min / mg TSP)
0
500
1000
1500
pNOS
pVSP
p35S
pEU

Fig. 5A

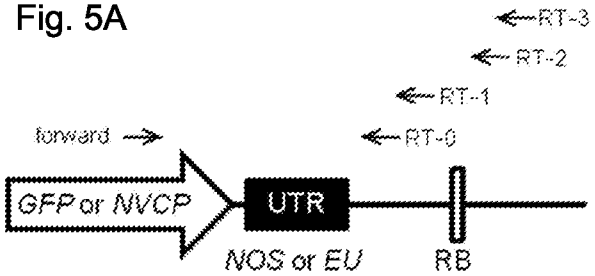


Fig. 5B

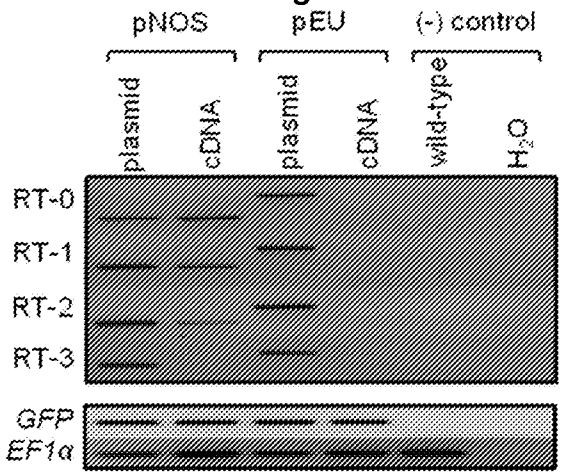


Fig. 5C

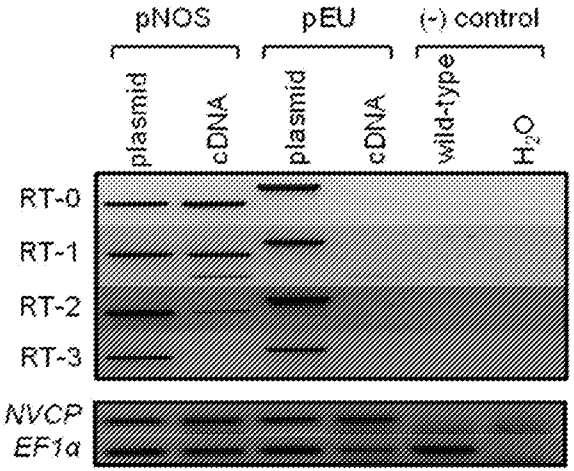


Fig. 6

```
                                    NUE1
252 ~ caaagcagaatgctgagctaaaagaaaaggcttttttccatttttcgagagac

                  NUE2    Polypurine Sequence (PPS)
302 ~ aatgagaaaaagaaggaagaagaagaagaagaagaagaagaaaagagta

          NUE3
352 ~ aataataaagccccacaggaggcgaagttcttgtagctccatgttatcta

                                pA(2)
402 ~ agttattgatattgtttgccctatattttatttctgtcattgtgtatgtt

                                                    NUE4
452 ~ ttgttcagtttcgatctccttgcaaaatgcagagattatgagatgaataa

                                    NUE5
                        pA(3)       pA(9)
502 ~ actaagttatattattatacgtgtttaatattctcctcctctctctagcta
552 ~ gccttttgttttctcttttttcttatttgatttttctttaaatcaatccatt
602 ~ ttaggagagggccaggaagtgatccagcaaaacatgaagattagaaagaaa
652 ~ cttccctctttttttttcctgaaaacaatttaacgtcgagatttatctctt
702 ~ tttgtaatggaatcatttctacagttatgacg
```

Fig. 7

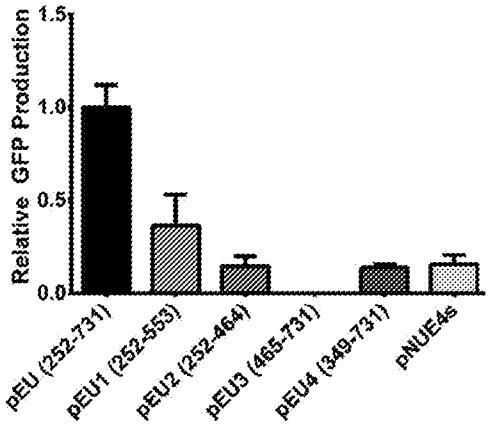


Fig. 8A

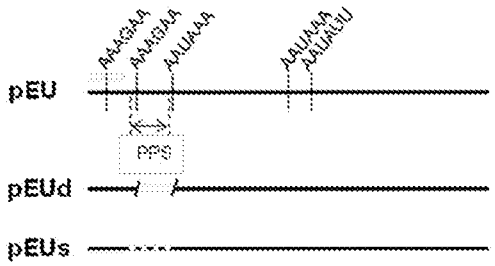


Fig. 8B

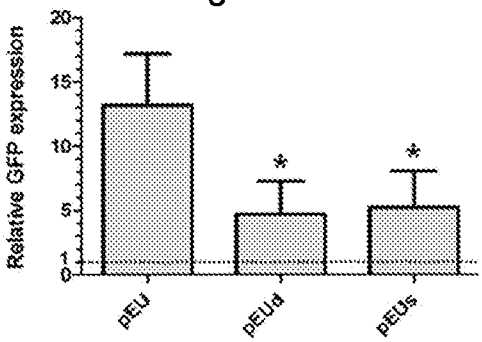


Fig. 9A

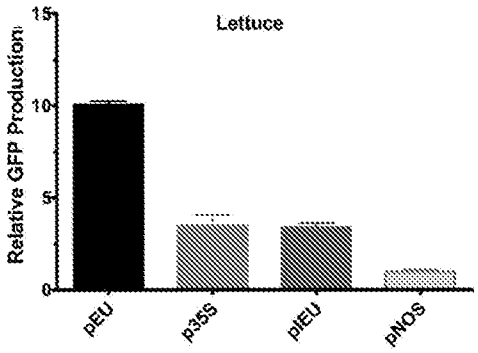


Fig. 9B

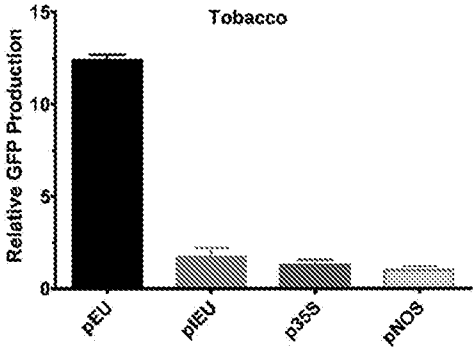


Fig. 10A

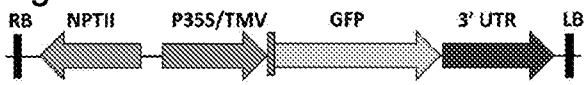


Fig. 10B

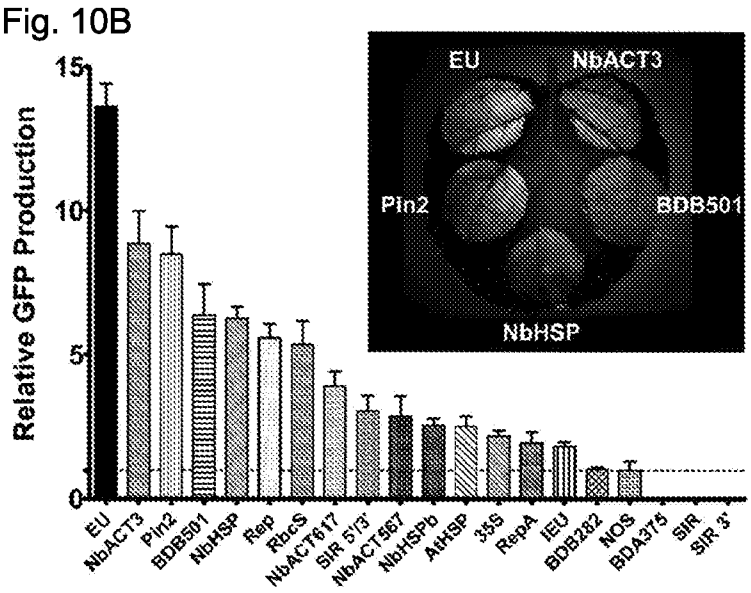


Fig. 11

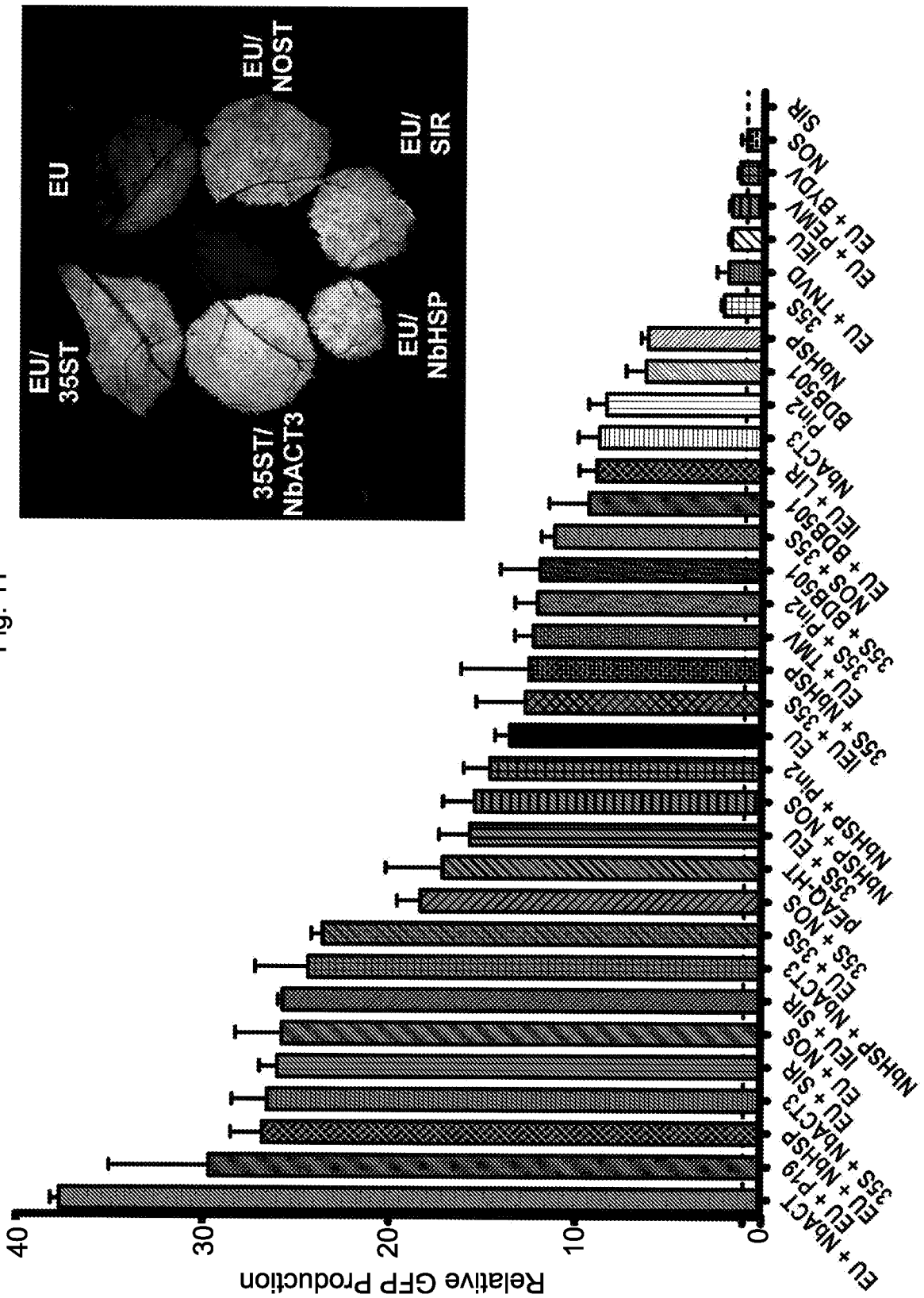

EU/
35ST
EU
EU/
NOST
EU/
SIR
35ST/
NbACT3
EU/
NbHSP
Relative GFP Production
0
10
20
30
40
EU + NbACT
EU + P19
EU + NbHSP
35S + NbACT3
EU + SIR
EU + NOS
NbHSP + NbACT3
EU + 35S
35S + NOS
PEAQ-HT
35S + EU
NbHSP + NOS
NbHSP + pin2
EU + 35S
35S + NbHSP
EU + TMV
35S + pin2
NOS + 35S
EU + BDB501
EU + LIR
NbACT3
pin2
BDB501
NbHSP
35S
EU + TNVD
EU + PEMV
EU + BYDV
NOS SIR

Fig. 12

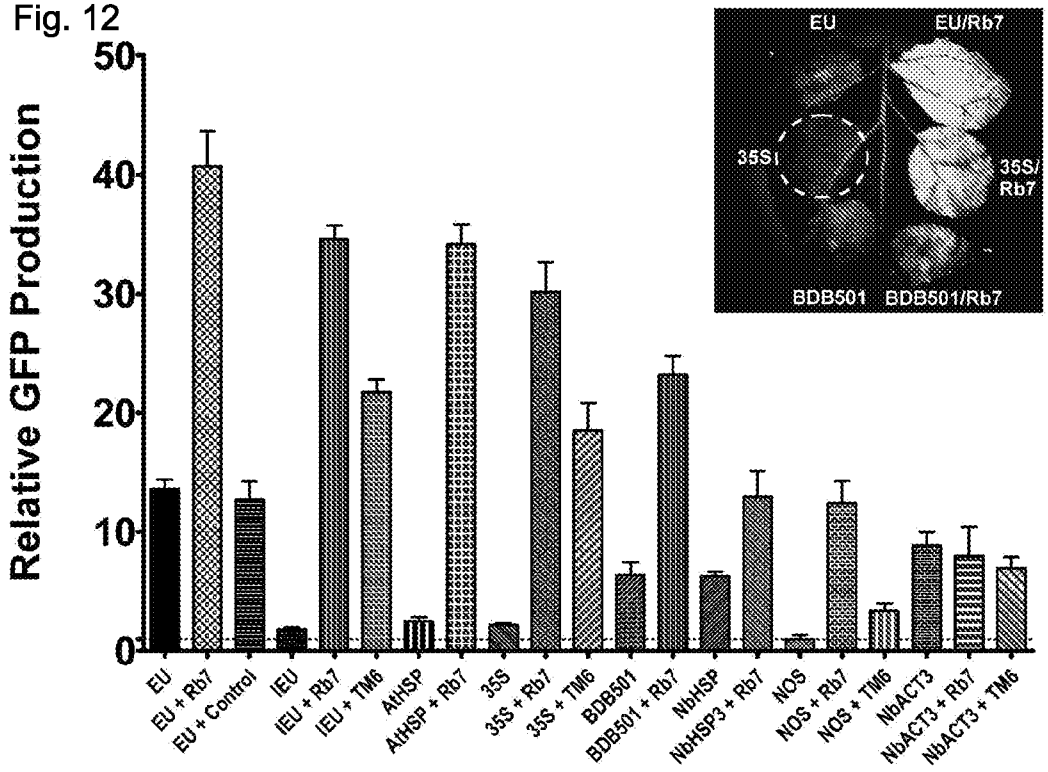

Relative GFP Production
0
10
20
30
40
50
EU
EU + Rb7
EU + Control
IEU
IEU + Rb7
IEU + TM6
AtHSP
AtHSP + Rb7
35S
35S + Rb7
35S + TM6
BDB501
BDB501 + Rb7
NbHSP
NbHSP3 + Rb7
NOS
NOS + Rb7
NOS + TM6
NbACT3
NbACT3 + Rb7
NbACT3 + TM6
EU
EU/Rb7
35S
35S/ Rb7
BDB501
BDB501/Rb7

Fig. 13

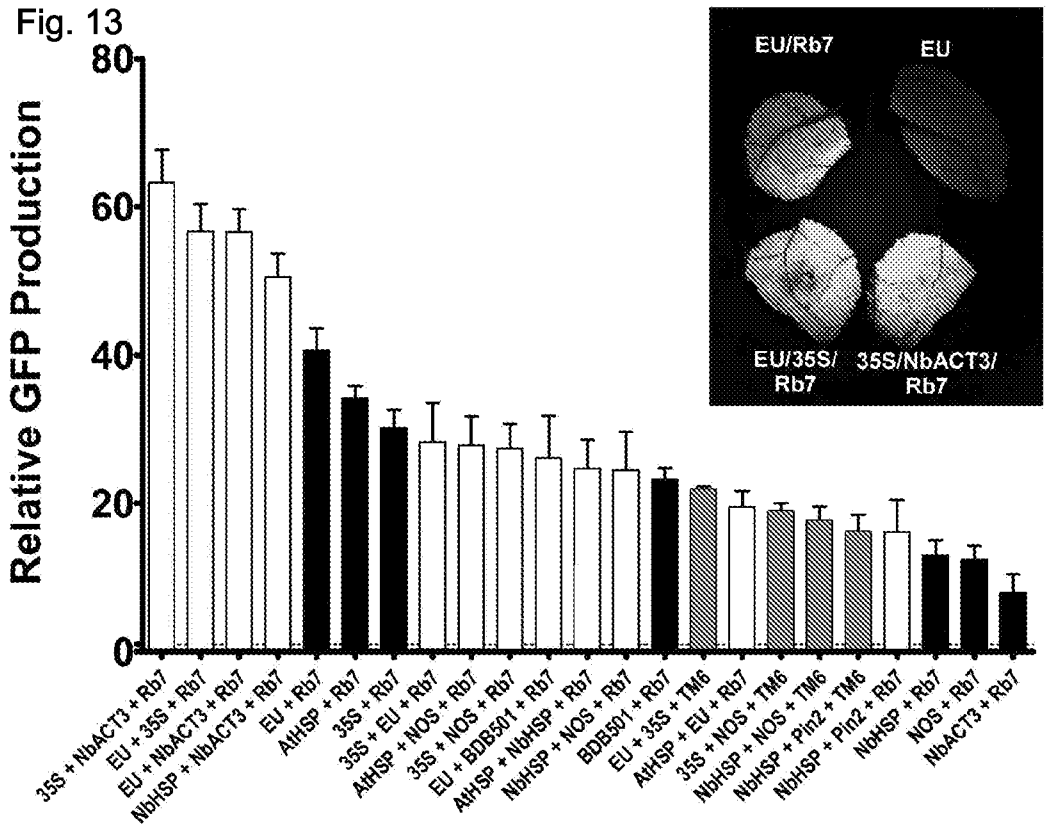

Relative GFP Production
0
20
40
60
80
35S + NbACT3 + Rb7
EU + 35S + Rb7
EU + NbACT3 + Rb7
NbHSP + NbACT3 + Rb7
EU + Rb7
AtHSP + Rb7
35S + Rb7
35S + EU + Rb7
AtHSP + NOS + Rb7
35S + NOS + Rb7
EU + BDB501 + Rb7
AtHSP + NbHSP + Rb7
NbHSP + NOS + Rb7
BDB501 + Rb7
EU + 35S + TM6
AtHSP + EU + Rb7
35S + NOS + TM6
NbHSP + NOS + TM6
NbHSP + Pin2 + TM6
NbHSP + Pin2 + Rb7
NbHSP + Rb7
NOS + Rb7
NbACT3 + Rb7
EU/Rb7
EU
EU/35S/ Rb7
35S/NbACT3/ Rb7

Fig. 14A

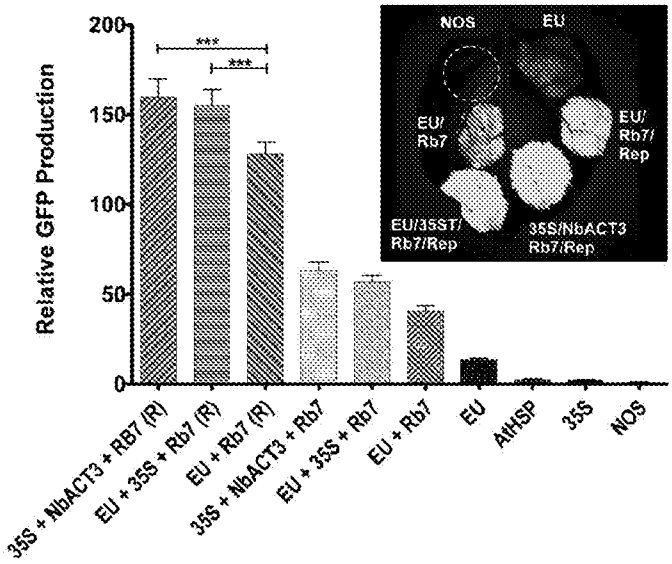


Fig. 14B

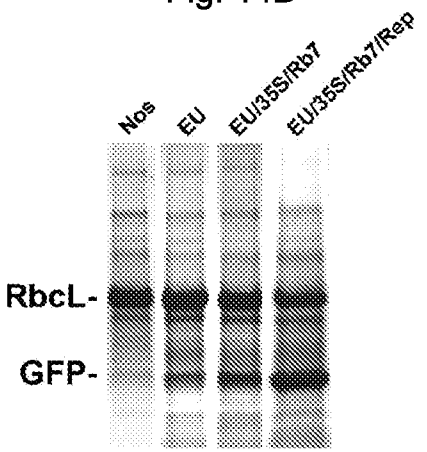


Fig. 15

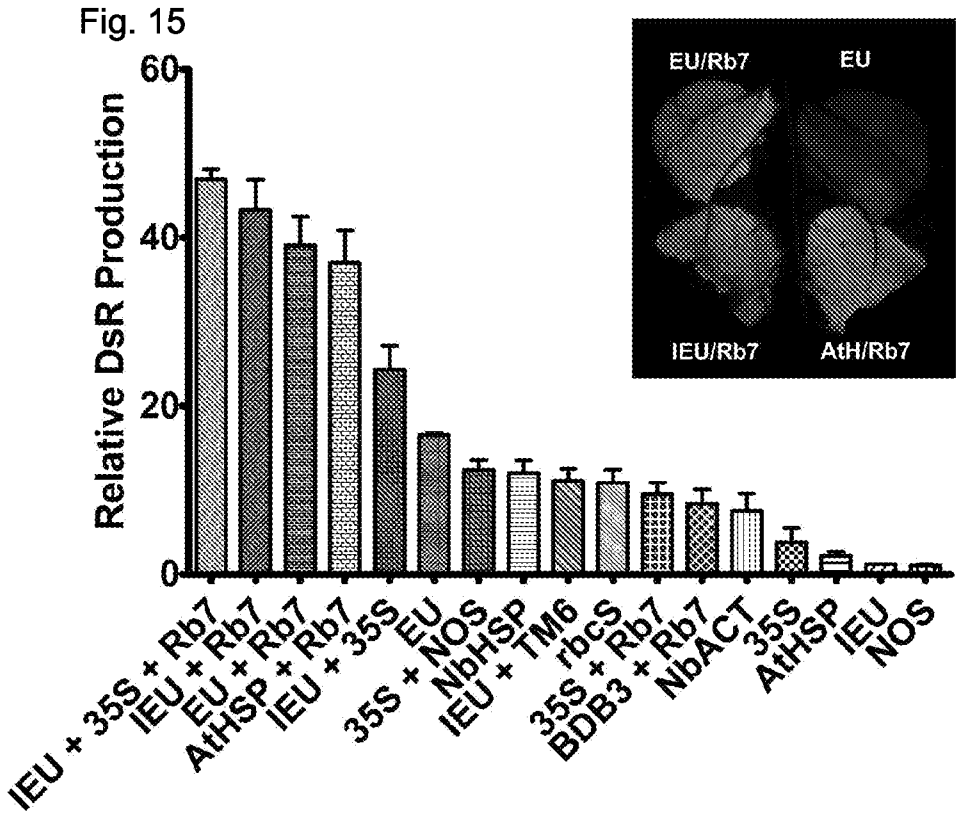


Fig. 16A

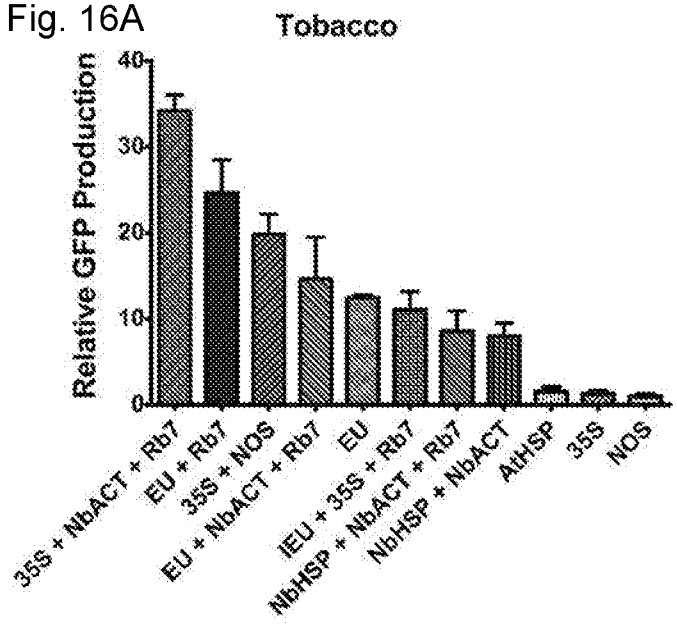


Fig. 16B

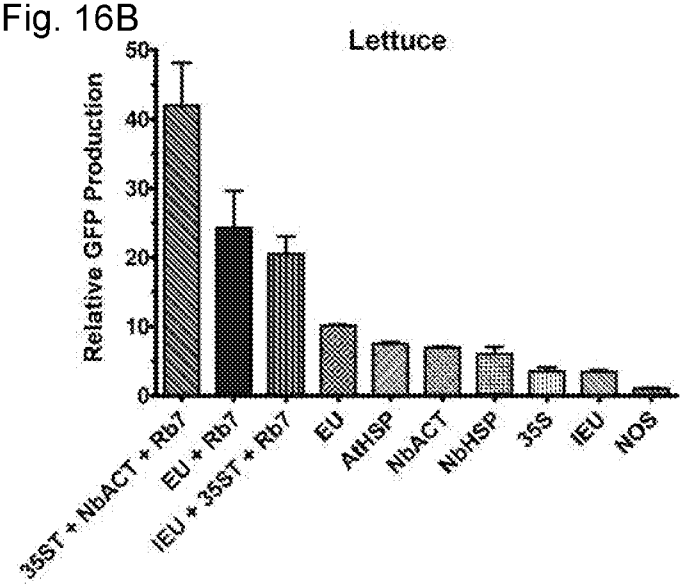


Fig. 17A

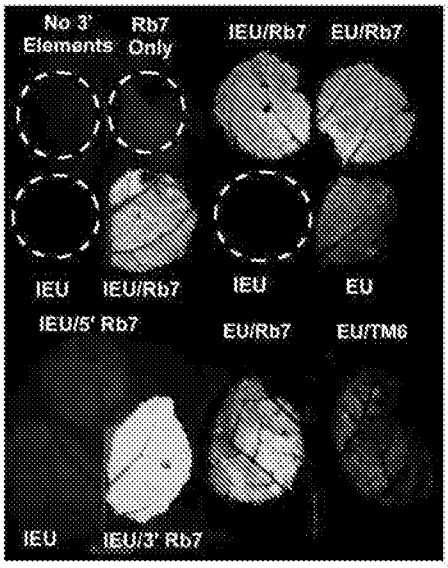
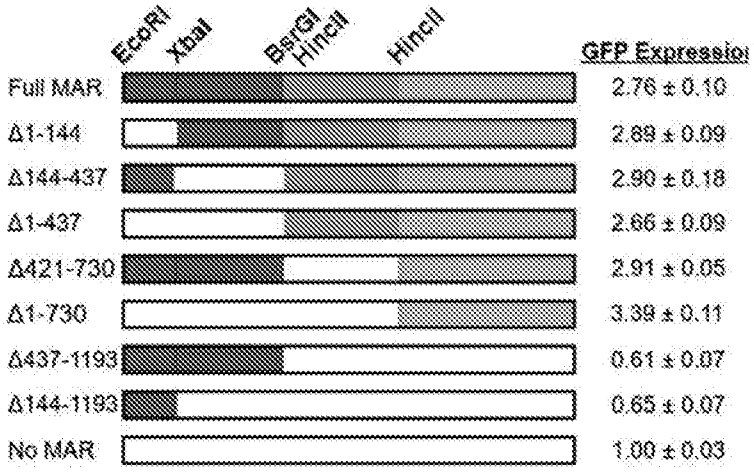


Fig. 17B



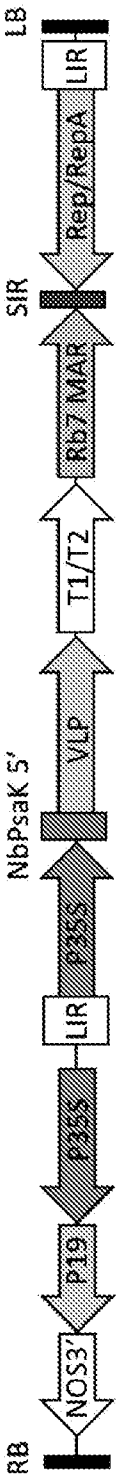


Fig. 18A

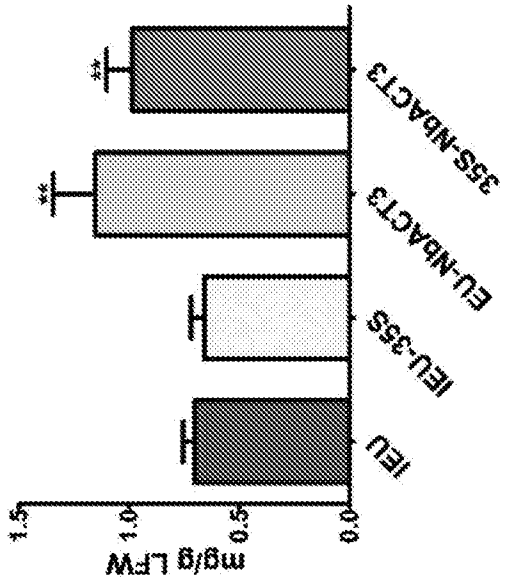


Fig. 18B

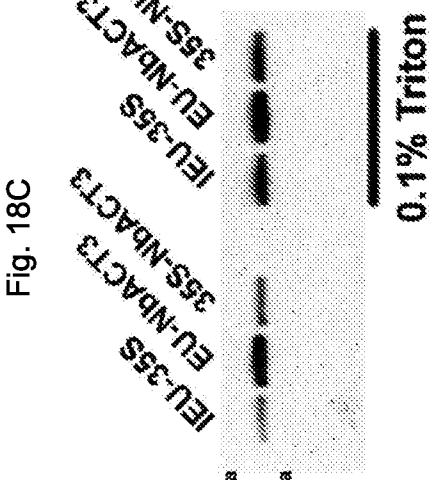


Fig. 18C

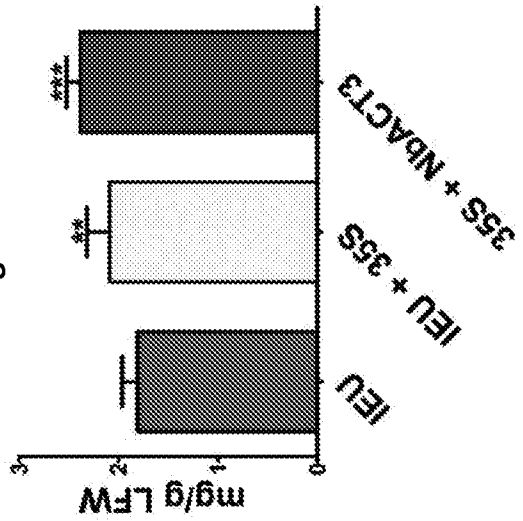


Fig. 18D

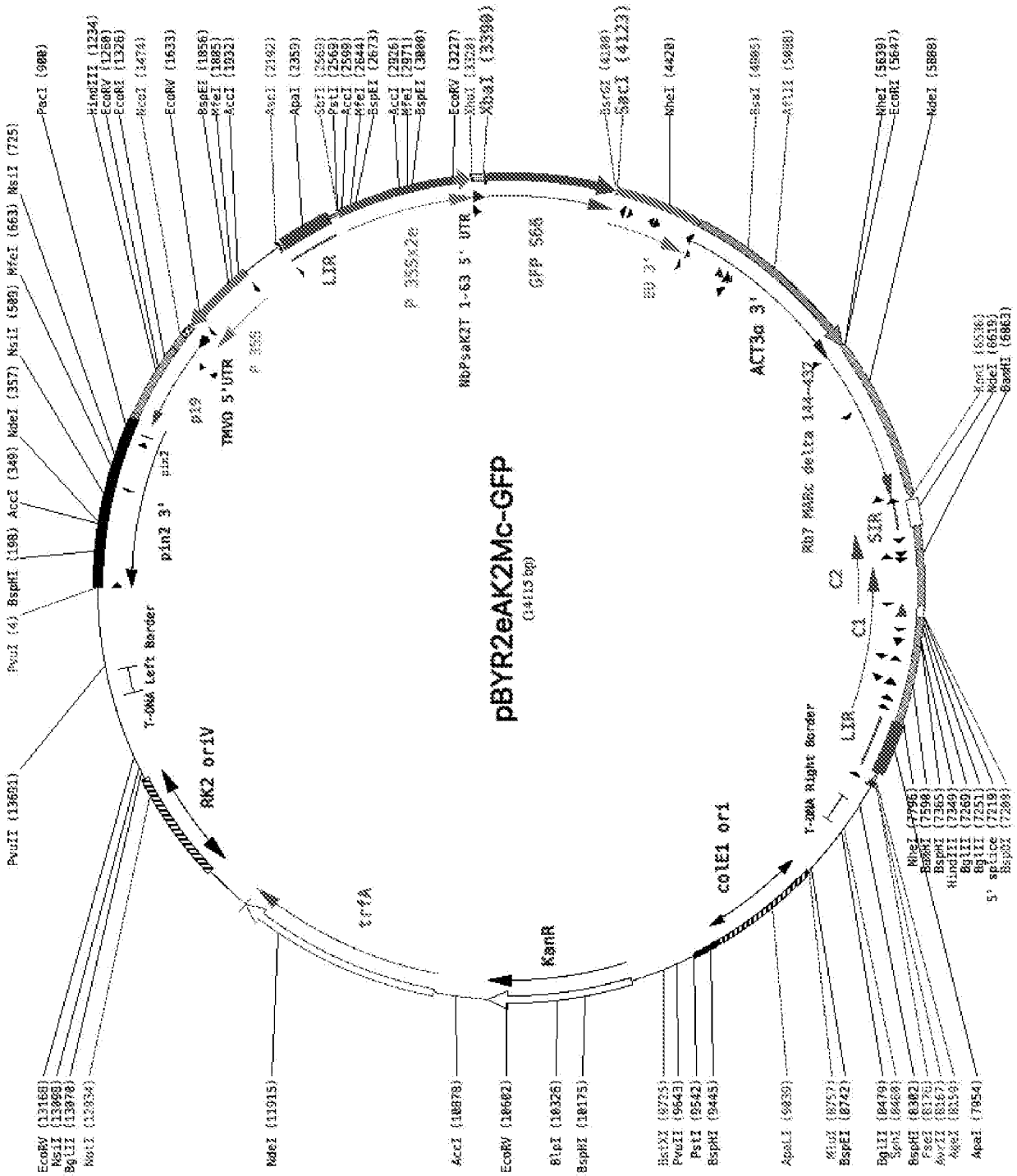

pBYR2eAK2Mc-GFP
(14113 bp)
RK2 oriV
T-DNA Left Border
pin2 3'
TMV5'UTR
LIR
GFP S65t
ACT3a 3'
T-DNA Right Border
colE1 ori
KanR
trfA
SIR
C1
C2
C3

Fig. 19

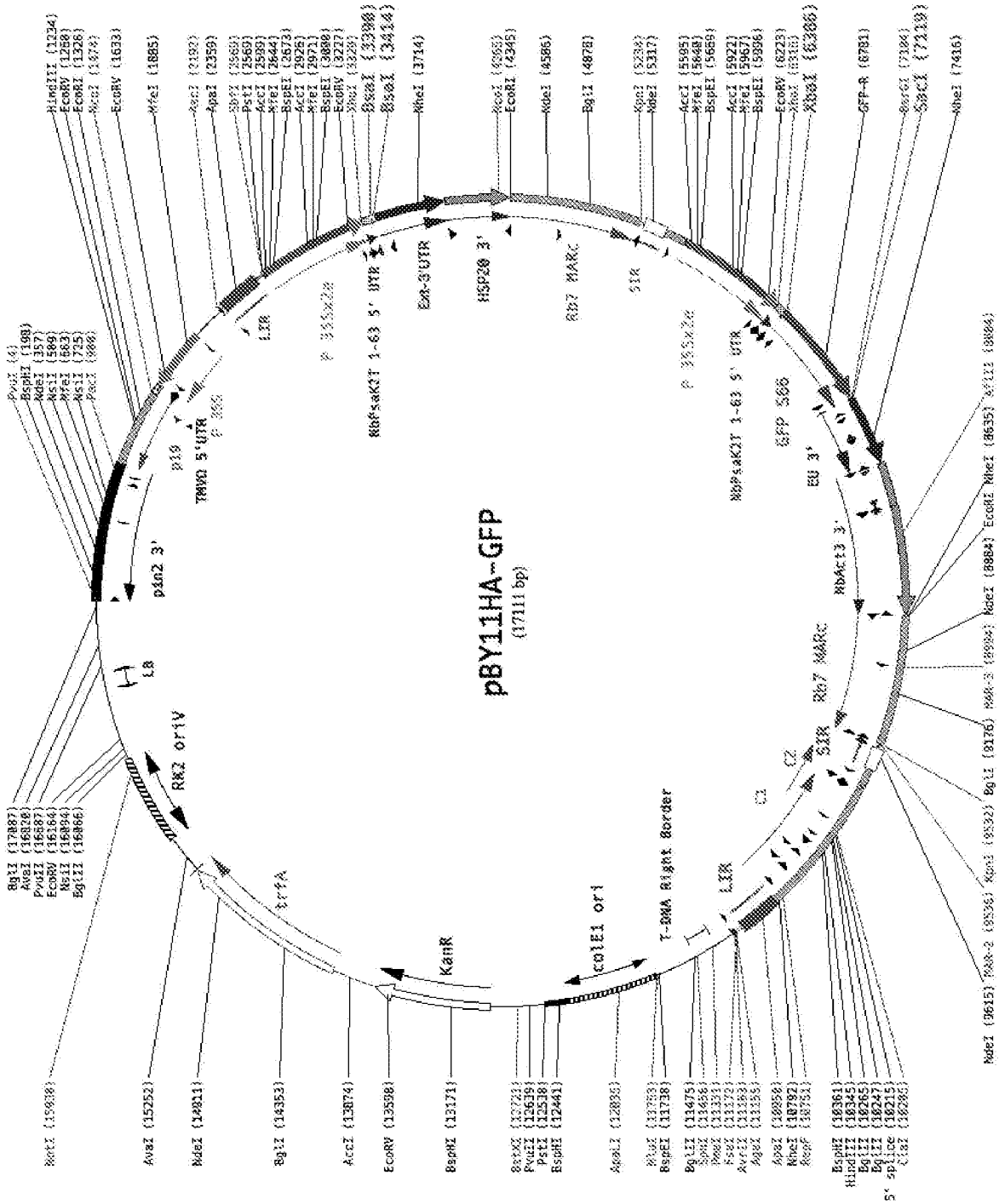

pBY11HA-GFP
(17111 bp)
RK2 oriV
KanR
colE1 ori
T-DNA Right Border
LB
RB
pAn2 3'
TMV 5'UTR
KbPsaK2T 1-63 5' UTR
Ext-3'UTR
HSP26 3'
Rb7 MARc
SIR
GFP S65
E8 3'
NbAct3 3'
GFP-R (6781)

Fig. 20

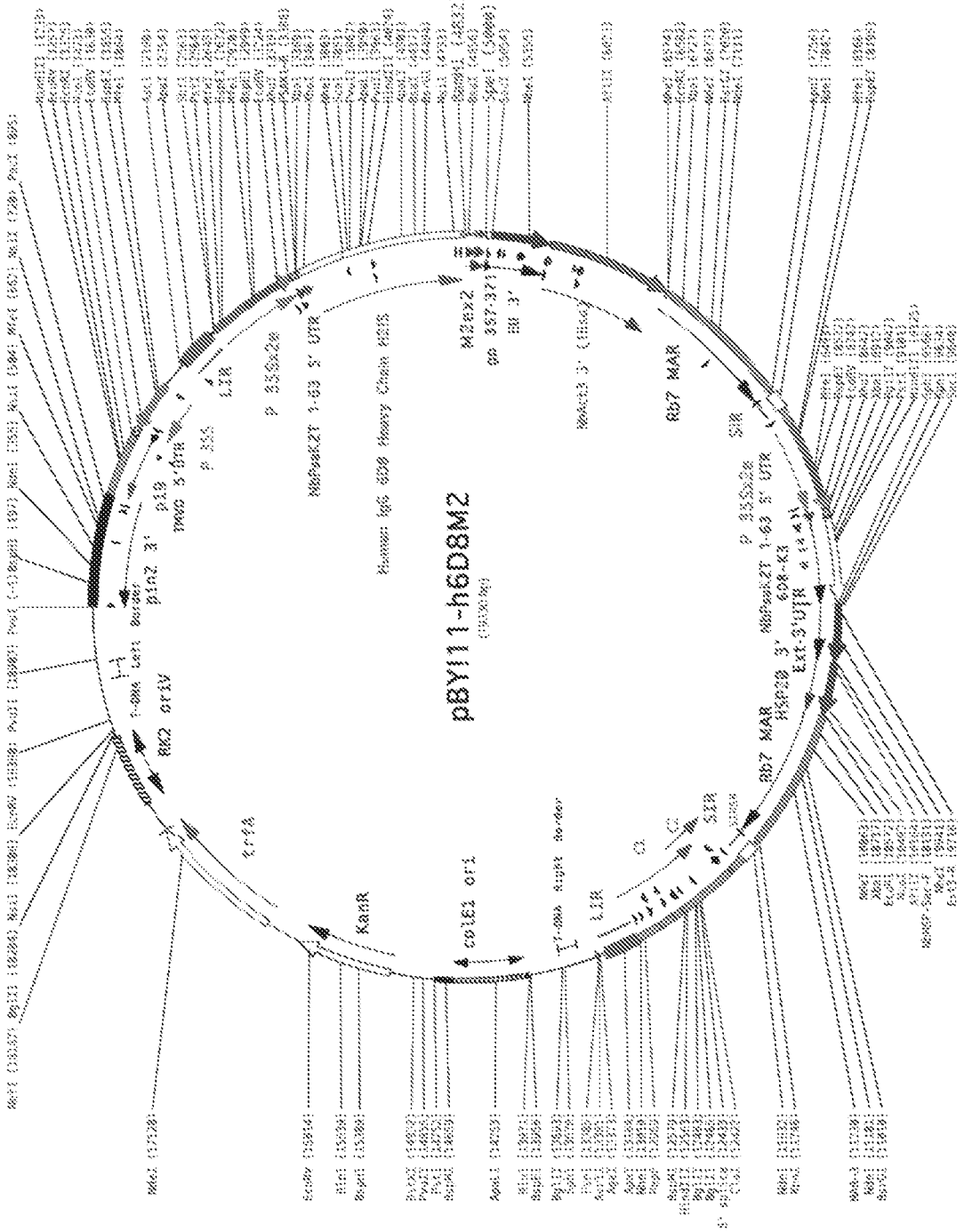

p8Y111-h6D8M2
Human IgG 6D8 Heavy Chain HSS
MBPmaK2T 1-83 5' UTR
P 35S
LIR
M13ori2
RB7 MAR
SIR
P 35S-2x
MBPmaK2T 1-83 5' UTR
608-K3
Ext-5'UTR
RB7 MAR
RSPCB 3'
RK2 oriV
ColE1 ori
KanR
ErrA
T1-DNA Right Border
CI
LIR
SIR

Fig. 21

Fig. 22A

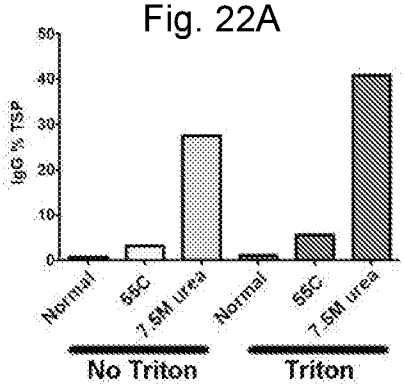
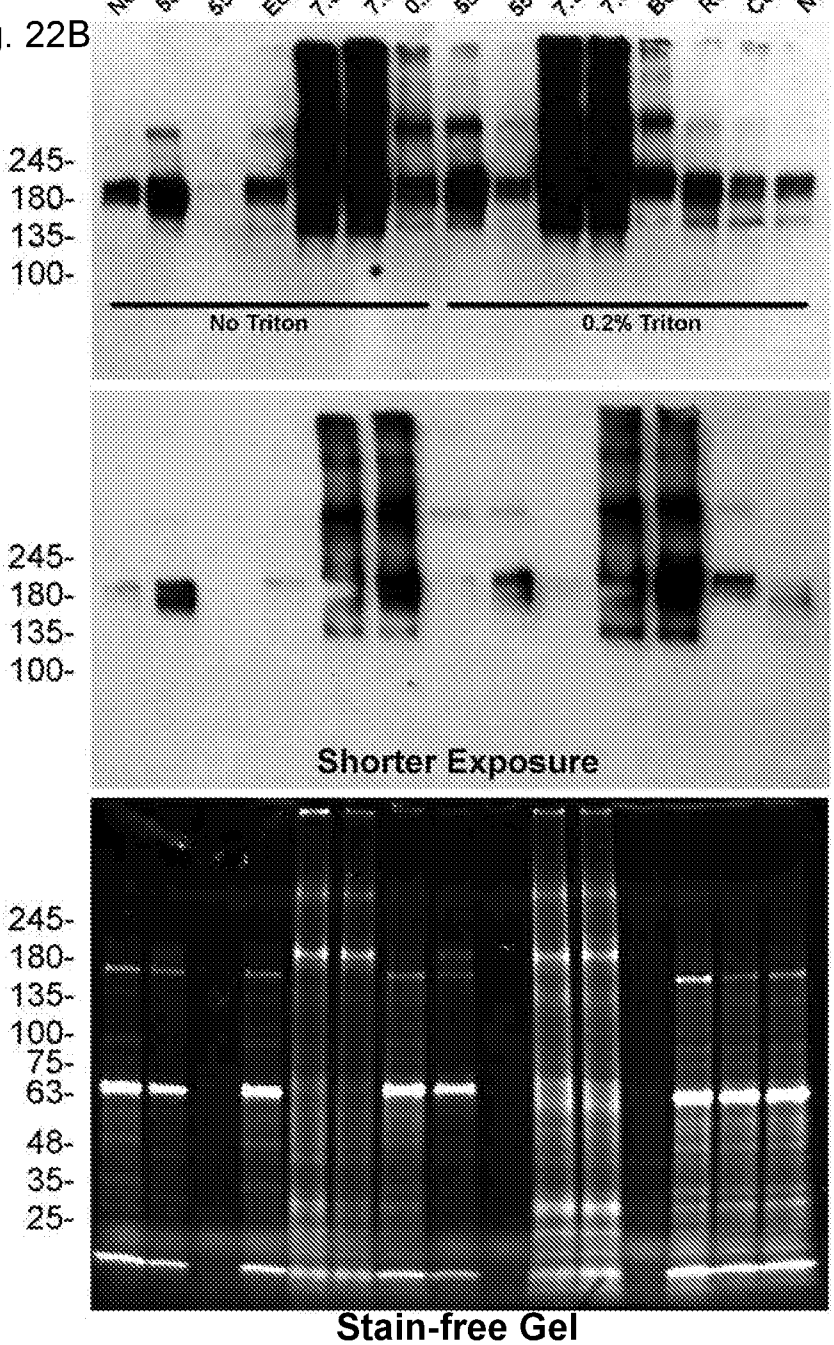


Fig. 22C



Fig. 22B



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/020621

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NOs: 115-117 were searched.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/020621

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 40-50
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/020621

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 15/113; C12N 15/63; C12N 15/82 (2019.01)

CPC - C12N 15/113; C12N 15/82; C12N 15/8216; C12N 15/8218; C12N 15/8257; C12N 2830/36 (2019.05)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 435/468; 435/69.1; 435/320.1; 536/24.1; 800/298 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DIAMOS, A.G. "Chapter 3: Chimeric 3' flanking regions strongly enhance gene expression in plants" Optimization of a Viral System to Produce Vaccines and other Biopharmaceuticals in Plants, Doctoral Thesis, Arizona State University, 31 December 2017 (31.12.2017), Pgs. 86-132. Retrieved from the Internet: <https://repository.asu.edu/attachments/194162/content/Diamos_asu_0010E_17585.pdf> on 20 May 2019 (20.05.2019). entire document	1-14, 16-20, 22, 25, 27, 29-39, 51-62
P, X	DIAMOS et al. "Chimeric 3' flanking regions strongly enhance gene expression in plants," Plant Biotechnol J, 21 May 2018 (21.05.2018), Vol. 16, Pgs. 1971-1982. entire document	1-39, 51-62
A	BEYENE et al. "Unprecedented enhancement of transient gene expression from minimal cassettes using a double terminator," Plant Cell Rep, 22 October 2010 (22.10.2010), Vol. 30, Pgs. 13-25. entire document	1-39, 51-62
A	US 2014/0130205 A1 (BHYRI et al) 08 May 2014 (08.05.2014) entire document	1-39, 51-62
A	DIAMOS et al. "5' and 3' Untranslated Regions Strongly Enhance Performance of Geminiviral Replicons in Nicotiana benthamiana Leaves," Front Plant Sci, 24 February 2016 (24.02.2016), Vol. 7, No. 200, Pgs. 1-15. entire document	1-39, 51-62
A	US 2017/0081676 A1 (DOW AGROSCIENCES LLC) 23 March 2017 (23.03.2017) entire document	1-39, 51-62

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 May 2019	29 MAY 2019
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/020621

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HALWEG et al. "The Rb7 matrix attachment region increases the likelihood and magnitude of transgene expression in tobacco cells: a flow cytometric study," Plant Cell, 19 January 2005 (19.01.2005), Vol. 17, Pgs. 418-429. entire document	1-39, 51-62
A	JI et al. "TM6, a novel nuclear matrix attachment region, enhances its flanking gene expression through influencing their chromatin structure," Mol Cells, 12 July 2013 (12.07.2013), Vol. 36, Pgs. 127-137. entire document	1-39, 51-62
A	US 2015/0368660 A1 (MASON et al) 24 December 2015 (24.12.2015) entire document	1-39, 51-62
A	ROSENTHAL et al. "An intronless form of the tobacco extensin gene terminator strongly enhances transient gene expression in plant leaves," Plant Mol Biol, 10 February 2018 (10.02.2018), Vol. 96, Pgs. 429-443. entire document	1-39, 51-62