Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2945965 A1 2015/10/22

(21) 2 945 965

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2015/04/15
(87) **Date publication PCT/PCT Publication Date:** 2015/10/22
(85) **Entrée phase nationale/National Entry:** 2016/10/14
(86) **N° demande PCT/PCT Application No.:** US 2015/025887
(87) **N° publication PCT/PCT Publication No.:** 2015/160893
(30) **Priorité/Priority:** 2014/04/15 (US61/979,633)

(51) **Cl.Int./Int.Cl.** ***C12P 21/06*** (2006.01),
C12N 15/00 (2006.01), ***C12N 15/64*** (2006.01),
C12N 9/00 (2006.01)

(71) **Demandeur/Applicant:**
APPLIED GENETIC TECHNOLOGIES CORPORATION, US

(72) **Inventeurs/Inventors:**
YE, GUO-JIE, US;
LIU, JILIN, US

(74) **Agent:** BORDEN LADNER GERVAIS LLP

(54) **Titre : ACIDE NUCLEIQUE A CODONS OPTIMISES CODANT POUR UN REGULATEUR DE GTPASE DANS LA RETINITE PIGMENTAIRE (RPGR)**
(54) **Title: CODON OPTIMIZED NUCLEIC ACID ENCODING A RETINITIS PIGMENTOSA GTPASE REGULATOR (RPGR)**

(57) **Abrégé/Abstract:**
This invention relates generally to a codon optimized nucleic acid encoding a retinitis pigmentosa GTPase regulator (RPGR) protein. The nucleic acid has enhanced stability during plasmid production relative to a wildtype cDNA encoding the RPGR protein. The invention also relates to expression cassettes, vectors, and host cells comprising the codon optimized nucleic acid. Methods for preparing a recombinant adeno-associated (rAAV) expression vector comprising the codon optimized nucleic acid sequence are also provided. The nucleic acids, expression cassettes, vectors, and host cells provided may be useful in the large scale production of rAAV expression vectors for gene therapy applications.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 October 2015 (22.10.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/160893 A1

(51) International Patent Classification:
C12P 21/06 (2006.01) *C12N 15/00* (2006.01)
C12N 15/64 (2006.01) *C12N 9/00* (2006.01)

(21) International Application Number:
PCT/US2015/025887

(22) International Filing Date:
15 April 2015 (15.04.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/979,633 15 April 2014 (15.04.2014) US

(71) Applicant: **APPLIED GENETIC TECHNOLOGIES CORPORATION** [US/US]; 11801 Research Drive, Suite D, Alachua, FL 32615 (US).

(72) Inventor: **YE, Guo-Jie**; 2178 NW 101 Street, Gainesville, FL 32606 (US).

(74) Agents: **SPARKS, Jonathan, M.** et al.; McCarter & English, LLP, 265 Franklin Street, Boston, MA 02110 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, 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).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2015/160893 A1

(54) **Title**: CODON OPTIMIZED NUCLEIC ACID ENCODING A RETINITIS PIGMENTOSA GTPASE REGULATOR (RPGR)

(57) **Abstract**: This invention relates generally to a codon optimized nucleic acid encoding a retinitis pigmentosa GTPase regulator (RPGR) protein. The nucleic acid has enhanced stability during plasmid production relative to a wildtype cDNA encoding the RPGR protein. The invention also relates to expression cassettes, vectors, and host cells comprising the codon optimized nucleic acid. Methods for preparing a recombinant adeno-associated (rAAV) expression vector comprising the codon optimized nucleic acid sequence are also provided. The nucleic acids, expression cassettes, vectors, and host cells provided may be useful in the large scale production of rAAV expression vectors for gene therapy applications.

CODON OPTIMIZED NUCLEIC ACID ENCODING A RETINITIS PIGMENTOSA GTPASE REGULATOR (RPGR)

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 61/979,633, filed on April 15, 2014, the entire contents of which is expressly incorporated herein by reference.

FIELD OF THE INVENTION

This invention relates generally to codon optimized nucleic acid sequences encoding a human retinitis pigmentosa GTPase regulator (RPGR).

BACKGROUND OF THE INVENTION

Retinitis pigmentosa (RP) is an inherited degenerative disease of the retina that affects approximately one in 3,500 individuals, with an estimated 1.5 million patients worldwide. See Churchill et al., 2013, Invest. Ophthalmol. Vis. Sci. 54(2): 1411-1416. RP is caused by progressive loss of rod and cone photoreceptors, resulting in night blindness followed by loss of visual fields. The disease may result in legal or even complete blindness. Mutations in the retinitis pigmentosa GTPase regulator (RPGR) gene account for greater than 70% of the cases of human X-linked retinitis pigmentosa (XLRP), the most severe subtype of RP. See Beltran et al., 2012, PNAS 109(6): 2132-2137 and Bader et al., 2003, Invest. Ophthalmol. Vis. Sci. (44)4: 1458-1463.

Alternative splicing of the RPGR gene results in expression of multiple isoforms of the RPGR protein. The mRNA for isoform A contains all 19 exons of the gene, while the mRNA for isoform C contains exons 1 to 15 and a large part of intron 15. Intron 15 is a purine-rich region that contains highly repetitive sequences that code for glutamate and glycine repeats (EEEGEGEGE in human and EEGEGE in mouse), see Vervoort et al., Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. Nat Genet 2000;25:462-6. Isoform A is constitutively expressed in all tissues while isoform C, which is also referred to as “ORF15”, is the predominant form expressed in the connecting cilium of photoreceptor, see Hong et al., Invest Ophthalmol Vis Sci 2002;43:3373-82, and Hong et al., Invest Ophthalmol Vis Sci 2003;44:2413-21.

A total of 55% of RPGR-related XLRP is caused by mutations in ORF15, all of which result from deletions that lead to truncated proteins. Most of the other cases are caused by

mutations in exons 1-13, which can be either missense or nonsense mutations, with a small number caused by mutations in introns or large deletions. No cases have been identified due to mutations in exons 16 to 19.

Recent studies have demonstrated the potential of gene therapy approaches to treating XLRP caused by mutations in the RPGR gene. For example, Beltran et al. have shown that subretinal injections of adeno-assocatied virus (AAV) vectors expressing human RPGR increased rod and cone photoreceptor function in a canine model of XLRP.

However one of the challenges in large-scale production of AAV vectors for clinical use is that nucleic acid sequences encoding a protein of interest such as RPGR may be unstable, resulting in the accumulation of several mutations and deletions. For example, the RPGR gene contains a region of 1.2 kb called ORF15 near the 3' end of the cDNA that is highly repetitive and GA rich. This region is a mutation "hot spot" in population. This repetitive region is very unstable during cloning and vector preparation and clones obtained generally contain mutations and deletions. These mutations can potentially alter or eliminate RPGR protein function, limiting the use of this protein in gene therapy applications. Therefore a need exists to identify methods of stabilizing RPGR cDNAs during large-scale production of AAV vectors.

SUMMARY OF THE INVENTION

It has been surprisingly found that the nucleic acid sequence of SEQ ID NO: 1 encoding the human RPGR protein is stable in large scale production of AAV plasmid pTR-IRBP-RPGRsyn. This nucleic acid sequence was developed through codon optimization of the wild type RPGR cDNA. In one aspect, the present invention provides a polynucleotide comprising the nucleic acid sequence of SEQ ID NO: 1 encoding a human RPGR protein.

In one aspect, the invention features a polynucleotide comprising the nucleic acid sequence of SEQ ID NO: 1 encoding a human retinitis pigmentosa GTPase regulator (RPGR) protein.

In one embodiments, the invention features an expression cassette comprising the polynucleotide of the above aspect, and an expression control sequence operably linked and heterologous to the nucleic acid sequence.

In another embodiment, the invention features a vector comprising the polynucleotide of claim 1. In a further embodiment, the vector is a recombinant adeno-associated (rAAV) expression vector.

In another embodiment, the invention features a recombinant herpes simplex virus (rHSV) comprising the polynucleotide of any one of the above aspects.

In another embodiment, the invention features a host cell comprising the polynucleotide of any one of the above aspects. In a related embodiment, the host cell is a mammalian cell. In a further related embodiment, the host cell is a HeLa cell, a BHK21 cell or a Vero cell. In another further embodiment, the host cell is a V27 cell.

In another embodiment, the expression control sequence is a human interphotoreceptor retinoid-binding protein (IRBP) promoter. In a further related embodiment, the human IRBP promoter comprises a nucleic acid sequence having at least 95% sequence identity to the nucleic acid sequence of SEQ ID NO: 8 and directs preferential expression in rods and cones. In another further embodiment, the human IRBP promoter comprises the nucleic acid sequence of SEQ ID NO: 8.

In one embodiment, the polynucleotide comprises the nucleic acid sequence of SEQ ID NO: 7.

The invention also features in another embodiment a method of producing the rAAV expression vector of the above aspect, comprising (a) infecting a host cell with a recombinant herpes simplex virus (rHSV) comprising the nucleic acid sequence of SEQ ID NO: 1; (b) incubating the host cell; and (c) following incubation, collecting rAAV from the host cell of step (b).

In one embodiment, the host cell is a HeLa cell, a BHK21 cell or a Vero cell.

In another embodiment, the rHSV further comprises a human IRBP promoter operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a further embodiment, the human IRBP promoter comprises a nucleic acid sequence having at least 95% sequence identity to the nucleic acid sequence of SEQ ID NO: 8 and directs preferential expression in rods and cones. In a further related embodiment, the human IRBP promoter comprises the nucleic acid sequence of SEQ ID NO: 8.

In another embodiment, the rHSV comprises the nucleic acid sequence of SEQ ID NO: 7.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1B show a sequence alignment of codon optimized RPGR cDNA (RPGRsyn; SEQ ID NO: 1) and the wildtype RPGR cDNA (Genbank Accession No. NM_001034853; SEQ ID NO: 5).

Figure 2 shows a map of plasmid pUC57-RPGRsyn.

Figure 3 shows pUC57-RPGRsyn plasmid DNA clones N5 and N6 prepared by mini-prep and larger scale midi-prep (Midi) and digested with restriction enzymes NotI and PciI. Plasmid DNA from mini-preps was retransformed into SURE2 cells before larger scale production by midi-prep.

Figure 4 shows pUC57-RPGRsyn plasmid DNA from mini-preps (mini_2 and mini_3) digested with restriction enzymes NotI and PciI. Plasmid DNA was not detectable in larger scale midi preps (midi_2 and midi_3). Seeding culture was stored at 4°C overnight and used as the inoculant for larger scale plasmid production.

Figure 5 shows a map of AAV proviral plasmid pTR-IRBP-RPGRsyn

Figure 6 shows the restriction maps of pTR-IRBP-RPGRsyn plasmid DNA isolated from transformed bacteria after 4 rounds of serial overnight propagation, along with a control plasmid of pTR-IRBP-CNGB3co. Bacteria transformed with pTR-IRBP-RPGRsyn or pTR-IRBP/GNAT2-hCNGB3co plasmids were grown in medium at 37°C, overnight. In the next morning, plasmid DNA was purified from 1.5 mL of overnight culture, and the remaining culture was left at room temperature until late afternoon and then used to inoculate 2 mLs of culture medium (1:1000 dilution) for the 2nd round propagation. Same procedures were followed for the 3rd and 4th round of propagation. Plasmid DNA purified from each round were then analyzed by restriction digestion with Sma I to confirm the integrity of the ITR sequence of the plasmid. Restriction maps kept same for both pTR-IRBP-RPGRsyn and the control plasmid pTR-IRBP-CNGB3co, through the 3 rounds of propagation in bacteria. However, the yield was significantly decreased after 3rd round propagation and almost no plasmid restriction fragments were detected after 4th round propagation in bacteria.

Figure 7 shows the sequence alignment of the consensus sequence of contigs obtained from pTR-IRBP-RPGRsyn plasmid DNA to the reference pTR-IRBP-RPGRsyn sequence.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides a polynucleotide comprising the nucleic acid sequence of SEQ ID NO: 1 encoding a human retinitis pigmentosa GTPase regulator (RPGR) protein. The nucleic acid sequence has been codon optimized for enhanced stability during vector replication, and may be used, for example, for production of adeno-assocatied virus (AAV) vectors for gene therapy applications.

Nucleic acid sequences may be codon optimized to improve stability or heterologous expression in host cells without changing the encoded amino acid sequence. For example, codon optimization may be used to remove sequences that negatively impact gene

expression, transcript stability, protein expression or protein stability, such as transcription splice sites, DNA instability motifs, polyadenylation sites, secondary structure, AU-rich RNA elements, secondary ORFs, codon tandem repeats, or long range repeats. Codon optimization may also be used to adjust the G/C content of a sequence of interest.

A codon consists of a set of three nucleotides and encodes a specific amino acid or results in the termination of translation (i.e. stop codons). The genetic code is redundant in that multiple codons specify the same amino acid, i.e., there are a total of 61 codons encoding 20 amino acids. Codon optimization replaces codons present in a DNA sequence with preferred codons encoding the same amino acid, for example, codons preferred for mammalian expression. Thus, the amino acid sequence is not altered during the process. Codon optimization can be performed using gene optimization software. The codon optimized nucleotide sequence is translated and aligned to the original protein sequence to ensure that no changes were made to the amino acid sequence. For example, the nucleotide sequence of SEQ ID NO: 1 encoding human RPGR is a codon optimized version of the wild type human RPGR nucleotide sequence (Genbank Accession No. NM_001034853, SEQ ID NO: 5). Both SEQ ID NO: 1 and SEQ ID NO: 5 encode the same RPGR protein (SEQ ID NO: 6).

Methods of codon optimization are known in the art and are described, for example, in U.S. Application Publication No. 2008/0194511 and U.S. Pat. No. 6,114,148.

The nucleic acid sequences of the present invention can be made as synthetic sequences. Techniques for constructing synthetic nucleic acid sequences are known in the art, and synthetic gene sequences may be purchased from several companies, including DNA 2.0 (Menlo Park, Calif.) and GenScript USA Inc. (Piscataway, N.J.). Alternatively, codon changes can be introduced by standard molecular biology techniques such as site-specific in vitro mutagenesis, PCR, or any other genetic engineering methods known in art which are suitable for specifically changing a nucleic acid sequence. In vitro mutagenesis protocols are described, for example, in In Vitro Mutagenesis Protocols, Braman, ed., 2002, Humana Press, and in Sankaranarayanan, Protocols in Mutagenesis, 2001, Elsevier Science Ltd.

The human RPGR gene is located in chromosomal region Xp21.1 and spans 172 kb. Shu et al., 2012, Invest. Ophthalmol. Vis. Sci. 53(7): 3951-3958. There are multiple alternatively spliced transcripts, all of which encode an amino (N)-terminal RCC1-like (RCCL) domain. The RCCL domain is structurally similar to the RCC1 protein, a guanine nucleotide exchange factor for the small guanosine triphosphate–binding protein, Ran. The RPGR gene contains 19 exons (RPGRex1–19), encoding a predicted 90 kDa protein. Exons

2 to 11 encode the RCCL domain, whereas exons 12 to 19 encode a carboxyl (C)-terminal domain rich in acidic residues and ending in an isoprenylation anchorage signal. Mutations found in RPGRex1–19 account for 15% to 20% of XLRP patients, and subsequent studies revealed many more disease-causing mutations within one or more transcripts containing an alternatively spliced C-terminal exon called ORF15 (RPGRORF15). A high frequency of microdeletions, frameshift, and premature stop mutations are found within the ORF15.

In one embodiment, the RPGR cDNA used for codon optimization is the full-length human RPGRORF15 clone, variant C, Genbank Accession No. NM_001034853 (SEQ ID NO: 5). See Vervoort et al., 2000, Nat Genet 25: 462-466. This clone contains exons 1-ORF15 and was generated using three-way ligation by step-wise amplifying exons 1-part of 15b (nucleotides 169–1990) from human lymphocytes and 1991–3627 from human genomic DNA. See Beltran et al., 2012, PNAS 109(6): 2132-2137.

RPGR is widely expressed and shows a complex expression pattern. See Shu et al., cited above. RPGR transcripts are detected in different tissues, including brain, eye, kidney, lung, and testis in several different species. RPGR protein is detected in retina, trachea, brain, and testis. In human, mouse, and bovine retina, RPGR mainly localizes to photoreceptor connecting cilia, but expression has also been reported in outer segments in some species. RPGR is expressed in the transitional zone of motile cilia and within human and monkey cochlea.

The invention also provides an expression cassette comprising the nucleic acid sequence of SEQ ID NO: 1 and an expression control sequence operably linked and heterologous to the nucleic acid sequence. The term "expression control sequence" refers to any genetic element (e.g., polynucleotide sequence) that can exert a regulatory effect on the replication or expression (transcription or translation) of the nucleic acid sequence. Common expression control sequences include promoters, polyadenylation signals, transcription termination sequences, upstream regulatory domains, origins of replication, internal ribosome entry sites (IRES), and enhancers.

An expression control sequence is operably linked with a nucleic acid sequence when the expression control sequence is placed in a functional relationship with the second nucleic acid sequence. For example, a promoter is operably linked to a coding sequence if the promoter affects the expression of the coding sequence. The term operably linked encompasses, for example, an arrangement of an expression control sequence with the nucleic acid sequence to be expressed and optionally further expression control sequences,

such as a terminator or enhancer, such that each of the expression control sequences can allow, modify, facilitate or otherwise influence expression of the nucleic acid sequence.

The term "heterologous" refers to nucleic acid or amino acid sequences that are obtained or derived from different source organisms or from different genes or proteins within the same source organism. For example, an expression control sequence that is not a native expression control sequence of the human RPGR gene is considered to be heterologous to the human RPGR gene. In certain embodiments, the expression control sequence is a promoter that is heterologous to the RPGR gene.

In a preferred embodiment, the expression control sequence is a human interphotoreceptor retinoid-binding protein (IRBP) promoter. IRBP is a large glycoprotein that is expressed only in the photoreceptor cells of the retina and to a much lesser extent in pinealocytes in the pineal gland in the brain. See Al-Ubaidi et al., 1992, J Cell Biology, 119(6) 1681-1687. The IRBP promoter region is well characterized. For example, Albini et al. (1990, Nucleic Acids Research 18(17): 5181-5187) describe a nucleotide sequence of the human IRBP promoter region (Genbank Accession No. X53044) containing 2818 bp of the 5' untranscribed region (SEQ ID NO: 2). Beltran et al. (cited above) demonstrated that a 235 bp fragment of the human IRBP promoter directed GFP expression in both rods and cones of normal canine retina in a dose- and time-dependent manner. A 1.3 kb fragment of the 5' untranslated region of the human IRBP gene (SEQ ID NO: 3) directed expression of a bacterial reporter gene (chloramphenicol acetyltransferase, CAT) specifically to photoreceptor cells in transgenic mice. See Al Ubaidi et al. 1992, J Cell Biology 119: 1681-1687. Nested deletion analysis of a 1783 bp fragment of the mouse IRBP 5' flanking region indicated that high promoter activity was maintained with a fragment consisting of 70 bp 5' to the transcription start site (SEQ ID NO: 4), but that elements upstream of this 70 bp fragment are required for complete tissue-specific regulation. See Boatright, et al., 1997, Molecular Vision 3: 15.

In a preferred embodiment, the human IRBP promoter comprises a nucleic acid sequence having at least 95%, 96%, 97%, 98% or 99% sequence identity to SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 8. In a further preferred embodiment, the human IRBP promoter comprises SEQ ID NO: 8.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked, e.g., a plasmid. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. An "rAAV vector" is a recombinant vector that includes nucleic acid

sequences derived from adeno-associated virus (AAV). Recombinant AAV is produced in vitro by introduction of gene constructs into cells known as producer cells. Recombinant AAV has been studied extensively as a vehicle for gene therapy and for its potential applicability as a treatment for human diseases based on genetic defects. At the clinical level, the rAAV vector has been used in human clinical trials to deliver the cftr gene to cystic fibrosis patients and the Factor IX gene to hemophilia patients (Flotte, et al., 1998, Methods Enzymol 292:717–732; and Wagner et al., 1998, Lancet 351:1702–1703). Systems for production of rAAV employ three elements: 1) a gene cassette containing the gene of interest, 2) a gene cassette containing AAV rep and cap genes and 3) a source of "helper" virus proteins. Methods of producing rAAV are known in the art and are described, for example, in U.S. Pat. No. 7,091,029.

Production of rAAV vectors for gene therapy is carried out in vitro, using suitable producer cell lines. A preferred cell line is 293, but production of rAAV can be achieved using other cell lines, including but not limited to human or monkey cell lines such as Vero, WI 38 and HeLa, and rodent cells, such as BHK cells, e.g. BHK21.

In particular embodiments, the rAAV comprises the nucleic acid sequence of SEQ ID NO: 1 encoding the human RPGR protein. The rAAV may further comprise one or more expression control sequences operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a preferred embodiment, the expression control sequence is a human IRBP promoter. In a further preferred embodiment, the human IRBP promoter comprises a nucleic acid sequence having at least 95%, 96%, 97%, 98% or 99% sequence identity to the nucleic acid sequence of SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 8 and directs preferential expression in rods and cones. In a particularly preferred embodiment, the human IRBP promoter comprises SEQ ID NO: 8.

In certain embodiments, the rAAV further comprises an SV40 poly A tail, an SV40 splice donor/splice acceptor (SD/SA) sequence, and a Kozak sequence, each operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a preferred embodiment, the rAAV comprises the nucleic acid sequence of SEQ ID NO: 7.

One strategy for delivering all of the required elements for rAAV production to the producer cell line involves transfecting the cells with plasmids containing gene cassettes encoding the necessary gene products, as well as infection of the cells with the helper virus Ad to provide the helper functions. This system employs plasmids with two different gene cassettes. The first is a proviral plasmid encoding the recombinant DNA to be packaged as rAAV. The second is a plasmid encoding the rep and cap genes. Other DNA viruses, such as

Herpes simplex virus type 1 (HSV-1) can be used instead of Ad to provide helper virus gene products needed for rAAV production (Conway et al., 1999, Gene Ther. 6:973–985).

Another strategy for rAAV production is based on the use of two or more recombinant rHSV-1 viruses to simultaneously co-infect producer cells with all of the components necessary for producing rAAV. This strategy employs at least two different forms of rHSV, each containing a different gene cassette. In addition to supplying the necessary helper functions, each of these rHSV viruses is engineered to deliver different AAV (and other) genes to the producer cells upon infection. The two rHSV forms are referred to as the "rHSV/rc virus" and the "rHSV expression virus." The rHSV/rc virus contains a gene cassette in which the rep and cap genes from AAV are inserted into the HSV genome. The rep genes are responsible for replication and packaging of the rAAV genome in host cells infected with AAV. The cap genes encode proteins that comprise the capsid of the rAAV produced by the infected cells.

The second recombinant HSV is an "rHSV expression virus." A usual element of an rAAV production system is an expression cassette containing transgene DNA sequences encoding a gene(s) of interest, such as the RPGR gene, along with promoter elements necessary for expression of the gene. In particular embodiments, the rHSV comprises the nucleic acid sequence of SEQ ID NO: 1 encoding the human RPGR protein. Expression vectors engineered for rAAV production are generally constructed with the gene of interest inserted between two AAV-2 inverted terminal repeats (ITRs). The ITRs are responsible for the ability of native AAV to insert its DNA into the genome of host cells upon infection or otherwise persist in the infected cells. The expression cassette is incorporated into the rHSV expression virus described above. This second rHSV virus is used for simultaneous co-infection of the cells along with the rHSV-1/rc virus.

The terms "recombinant HSV," "rHSV," "rHSV vector," and "rHSV expression vector" refer to isolated, genetically modified forms of herpes simplex virus (HSV) containing heterologous genes incorporated into the viral genome. Methods for production of rHSV are known in the art and are described, for example, by Conway et al. (1999, Gene Ther. 6:973–985); Conway et al. (1997, J Virol 71: 8780-8789) and U.S. Pat. No. 7,037,723.

In particular embodiments, the rHSV comprises the nucleic acid sequence of SEQ ID NO: 1 encoding the human RPGR protein. The rHSV may further comprise one or more expression control sequences for regulating expression of the nucleic acid sequence of SEQ ID NO: 1, wherein the expression control sequence is operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a preferred embodiment, the expression control sequence is a

human IRBP promoter that is operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a further preferred embodiment, the human IRBP promoter comprises a nucleic acid sequence having at least 95%, 96%, 97%, 98% or 99% sequence identity to the nucleic acid sequence of SEQ ID SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 8 and directs preferential expression in rods and cones. In a particularly preferred embodiment, the human IRBP promoter comprises SEQ ID NO: 8.

In certain embodiments of the aforementioned methods, the rHSV further comprises an SV40 poly A tail, an SV40 splice donor/splice acceptor (SD/SA) sequence, and a Kozak sequence, each operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a preferred embodiment, the rHSV comprises the nucleic acid sequence of SEQ ID NO: 7.

The invention also provides a method of producing an rAAV expression vector comprising a polynucleotide comprising the nucleic acid sequence of SEQ ID NO: 1 encoding a human RPGR protein. In one embodiment, the method comprises (a) infecting a host cell with a recombinant herpes simplex virus (rHSV) comprising the nucleic acid sequence of SEQ ID NO: 1; (b) incubating the host cell; and (c) following incubation, collecting rAAV from the host cell of step (b).

Methods of producing rAAV expression vectors by infecting a host cell with an rHSV are known in the art and are described for example in U.S. Pat. No. 7,091,029. For example, in one embodiment, the host cells are infected with rHSV by diluting the virus in growth medium such as DMEM and adding the virus to flasks containing the host cells. The host cells may be incubated with the virus for various intervals, for example, 22, 26, 30, 34, or 46 hours. Following the incubation interval, the virus-infected cells may be harvested by pelleting, followed by resuspension in DMEM. Cell-associated rAAV may be collected from the host cells by lysis of the cells using standard techniques involving three rounds of freezing and thawing (See Conway et al., 1999, cited above).

In particular embodiments, the host cell used for producing an rAAV expression vector in the aforementioned methods is a HeLa cell, a BHK21 cell or a Vero cell.

The rHSV used in the aforementioned method may further comprise one or more expression control sequences for regulating expression of the nucleic acid sequence of SEQ ID NO: 1 that is operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a preferred embodiment, the expression control sequence is a human IRBP promoter that is operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a further preferred embodiment, the human IRBP promoter comprises a nucleic acid sequence having at least 95%, 96%, 97%, 98% or 99% sequence identity to the nucleic acid sequence of SEQ ID NO:

2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 8 and directs preferential expression in rods and cones. In a particularly preferred embodiment, the human IRBP promoter comprises SEQ ID NO: 8.

In certain embodiments of the aforementioned methods, the rHSV further comprises an SV40 poly A tail, an SV40 splice donor/splice acceptor (SD/SA) sequence, and a Kozak sequence, each operably linked to the nucleic acid sequence of SEQ ID NO: 1. In a preferred embodiment, the rHSV comprises the nucleic acid sequence of SEQ ID NO: 7.

Description of Sequences

SEQ ID NO:	**Description**
1	Codon modified RPGR cDNA
2	Human IRBP promoter, 2818 bp. Albini et al., 1990, Nuc Acid Res 18: 5181-5187). SEQ ID NO: 2 comprises SEQ ID NO: 3 and 4.
3	Human IRBP promoter, 1326 bp. Al Ubaidi et al. 1992, J Cell Biology 119: 1681-1687
4	Mouse IRBP core promoter region, 70 bp. Boatright, et al., 1997, Molecular Vision 3: 15.
5	Wildtype RPGR cDNA, Genbank Accession No. NM_001034853
6	Wildtype RPGR amino acid sequence
7	3871 bp synthesized sequence comprising SEQ ID NO: 1, an SV40 poly A tail, the SV40 SD/SA sequence, Kozak sequence, and restriction sites
8	Human IRBP promoter, 234 bp fragment used in the RPGRsyn expression cassette

The following examples serve to illustrate certain embodiments and aspects of the present invention and are not to be construed as limiting the scope of the invention.

EXAMPLES

Example 1

Codon optimization of the RPGR gene and evaluation of plasmid stability

A wildtype RPGR cDNA in an AAV plasmid used for AAV manufacturing was found to contain several mutations and deletions in the region from nt 2461 to nt 3057. There were a total of 42 bp accumulated deletions or substitutions across this region. The plasmid clone was found to be stable during plasmid propagation in bacteria, and no sequence changes were found in the AAV vector.

A 3459 bp coding sequence of the RPGR gene, variant C (SEQ ID NO: 5) was codon-optimized at Genscript, Inc. for mammalian expression. Codon optimization was used both to select codons of high frequency in mammals and to alter GC content to enhance stability, and to reduce the repetitive nature of the gene. The codon optimized version of the RPGR coding sequence (RPGRsyn; SEQ ID NO: 1) shares 72.1% sequence identity with the original gene (SEQ ID NO: 5). See Fig. 1. The codon optimized gene encodes the same polypeptide as the original gene, i.e. the polypeptide of SEQ ID NO: 6. The RPGRsyn gene was synthesized at GenScript along with an SV40 poly A tail, the SV40 SD/SA sequence, a Kozak sequence, and restriction sites for cloning purposes. The entire 3871 bp synthesized sequence is provided as SEQ ID NO: 7.

A map of the plasmid containing RPGRsyn (pUC57-RPGRsyn) is shown in Fig. 2. This plasmid was able to propagate stably in bacteria in small scale plasmid production. This plasmid also maintained its stability in larger scale production after being retransformed into SURE2 cells, a bacteria strain used for cloning of the AAV plasmid. See Fig. 3. Clone N5 of the pUC57-RPGRsyn plasmid DNA produced in large scale production was confirmed to be identical to the original plasmid by DNA sequencing. The plasmid yield could range from very low yield to none at all if the seeding culture was stored at 4°C overnight and used as the inoculant for large scale plasmid production. See Fig. 4. The RPGRsyn cDNA was then released from pUC57-RPGRsyn plasmid and inserted into a pTR containing plasmid to generate the AAV proviral plasmid pTR-IRBP-RPGRsyn (Figure 5). pTR-IRBP-RPGRsyn contains inverted terminal repeats (ITR) of AAV2 and IRBP promoter. Large scale production of the plasmid confirmed to be 100% correct upon DNA sequencing (Figure 6). To further confirm the stability of pTR-IRBP-RPGRsyn, bacteria transformed with pTR-IRBP-RPGRsyn or pTR-IRBP/GNAT2-hCNGB3co plasmids were grown in medium at

37°C, overnight. In the next morning, plasmid DNA was purified from 1.5 mL of overnight culture, and the remaining culture was left at room temperature until late afternoon and then used to inoculate 2 mLs of culture medium (1:1000 dilution) for the 2nd round propagation. Same procedures were followed for the 3rd and 4th round propagation. Plasmid DNA purified from each round was then analyzed by restriction digestion with Sma I to confirm the integrity of the ITR sequence of the plasmid. As shown in Fig 6, the yield of pTR-IRBP-RPGRsyn declined during the serial passages; however, the same pattern is observed for pTR-IRBP-CNGB3co, a plasmid that contains the stable hCNGB3 cDNA. Therefore, the decline of plasmid yield is related to bacteria itself or other features such as TR, but not to the RPGRsyn. Also noted in Fig 7, the 4.2kb band containing RPGRsyn has been stable over the passages (it will become loose or smear if unstable).

Example 2

Construction of AAV plasmids and evaluation in bacteria

An AAV plasmid (pTR-IRBP-RPGRsyn) comprising an RPGRsyn expressing cassette comprising the IRBP promoter (234 bp), the RPGRsyn cDNA (SEQ ID NO: 1), and an SV40 polyA signal sequence is constructed. This IRBP fragment is contained within the 235 bp fragment used by Beltran et al. in the canine model (See Beltran et al., 2012, PNAS 109(6): 2132-2137). After construction of pTR-IRBP-RPGRsyn, the plasmid is tested for stability in bacteria using the methods described in Example 1.

Once the stability of pTR-IRBP-RPGRsyn is confirmed, an HSV recombination plasmid comprising the IRBP-RPGRsyn expression cassette (pHSV106-IRBP-RPGRsyn) is constructed. pHSV106-IRBP-RPGRsyn is used for construction of HSV-IRBP-RPGRsyn helper vector for large scale production of the AAV vector AAV-IRBP-RPGRsyn. The rHSV helper viruses are propagated in mammalian cells (V27, an ICP27-complementing Vero cell line). RPGRsyn cDNA is more stable in mammalian cells than in bacteria. This increased stability will eliminate the need for large-scale production of an AAV proviral plasmid containing the RPGRsyn cDNA, which is a reagent required for rAAV production by plasmid transfection methods.

APPENDIX

SEQUENCES

SEQ ID NO: 1

atgagagagccagaggagctgatgccagatagcggagcagtgtttaccttcggaaagtccaagttcgcagagaat
aacccaggaaagttctggtttaaaaacgacgtgcccgtccacctgtcttgtggcgatgagcatagtgccgtggtc
actgggaacaataagctgtatatgttcgggtccaacaattggggacagctggggctgggatccaaatctgctatc
tctaagccaacctgcgtgaaggcactgaaacccgagaaggtcaaactggccgcttgtggcagaaaccacactctg
gtgagcaccgagggcgggaatgtctatgccaccggaggcaacaatgagggacagctgggactgggggacactgag
gaaaggaatacctttcacgtgatctccttctttacatctgagcataagatcaagcagctgagcgccggctccaac
acatctgcagccctgactgaggacgggcgcctgttcatgtggggagataattcagagggccagattgggctgaaa
aacgtgagcaacgtgtgcgtgcctcagcaggtgaccatcggaaagccagtcagttggatttcatgtggctactat
catagcgccttcgtgaccacagatggcgagctgtacgtctttggggagcccgaaaacggaaaactgggcctgcct
aaccagctgctgggcaatcaccggacaccccagctggtgtccgagatccctgaaaaagtgatccaggtcgcctgc
gggggagagcatacagtggtcctgactgagaatgccgtgtacaccttcggactgggccagtttggccagctgggg
ctgggaaccttcctgtttgagacatccgaaccaaaagtgatcgagaacattcgcgaccagactatcagctacatt
tcctgcggagagaatcacaccgcactgatcacagacattggcctgatgtataccttggcgatgggcggcacggg
aagctgggactgggcctggagaacttcactaatcacttcatccccaccctgtgctctaacttcctgcggttcatc
gtgaaactggtcgcttgcggcgggtgtcacatggtggtcttcgctgcacctcataggggcgtggctaaggagatc
gaatttgacgagattaacgatacatgcctgagcgtggcaactttcctgccatacagctccctgacttctggcaat
gtgctgcagagaaccctgagtgcaaggatgcggagaagggagagggaacgctctcctgacagtttctcaatgcga
cgaaccctgccacctatcgaggggacactgggactgagtgcctgcttcctgcctaactcagtgtttccacgatgt
agcgagcggaatctgcaggagtctgtcctgagtgagcaggatctgatgcagccagaggaacccgactacctgctg
gatgagatgaccaaggaggccgaaatcgacaactctagtacagtggagtccctgggcgagactaccgatatcctg
aatatgacacacattatgtcactgaacagcaatgagaagagtctgaaactgtcaccagtgcagaagcagaagaaa
cagcagactattggcgagctgactcaggacaccgccctgacagagaacgacgatagcgatgagtatgaggaaatg
tccgagatgaaggaaggcaaagcttgtaagcagcatgtgagtcaggggatcttcatgacacagccagccacaact
attgaggctttttcagacgaggaagtggagatccccgaggaaaaagagggcgcagaagattccaaggggaatgga
attgaggaacaggaggtggaagccaacgaggaaaatgtgaaagtccacggaggcaggaaggagaaaacagaaatc
ctgtctgacgatctgactgacaaggccgaggtgtccgaaggcaaggcaaaatctgtcggagaggcagaagacgga
ccagagggacgaggggatggaacctgcgaggaaggctcaagcggggctgagcattggcaggacgaggaacgagag
aagggcgaaaaggataaaggccgcggggagatggaacgacctggagagggcgaaaaagagctggcagagaaggag
gaatggaagaaaagggacggcgaggaacaggagcagaaagaaagggagcagggccaccagaaggagcgcaaccag
gagatggaagagggcggcgaggaagagcatggcgagggagaagaggaagagggcgatagagaagaggaagaggaa
aaagaaggcgaagggaaggaggaaggagagggcgaggaagtggaaggcgagagggaaaaggaggaaggagaacgg
aagaaagaggaaagagccggcaaagaggaaaagggcgaggaagagggcgatcagggcgaaggcgaggaggaagag
accgagggccgcggggaagagaaagaggagggaggagaggtggagggcggagaggtcgaagagggaaagggcgag
cgcgaagaggaagaggaagagggcgagggcgaggaagaagagggcgagggggaagaagaggagggagagggcgaa
gaggaagagggggagggaaagggcgaagaggaaggagaggaagggggagggagaggaagagggggaggagggcgag
ggggaaggcgaggaggaagaaggagagggggaaggcgaagaggaaggcgagggggaaggagaggaggaagaagggg
gaaggcgaaggcgaagaggagggagaaggagagggggaggaagaggaaggagaagggaagggcgaggaggaaggc
gaagagggagagggggaaggcgaggaagaggaaggcgagggcgaaggagaggacggcgagggcgagggagaagag
gaggaaggggaatgggaaggcgaagaagaggaaggcgaaggcgaaggcgaagaagagggcgaaggggagggcgag
gagggcgaaggcgaaggggaggaagaggaaggcgaaggagaaggcgaggaagaagagggagaggaggaaggcgag
gaggaaggagagggggaggaggagggagaaggcgagggcgaagaagaagaagagggagaagtggagggcgaagtc
gaggggggaggagggagaaggggaaggggaggaagaagagggcgaagaagaaggcgaggaaagagaaaaagaggga
gaaggcgaggaaaaccggagaaatagggaagaggaggaagaggaagagggaaagtaccaggagacaggcgaagag
gaaaacgagcggcaggatggcgaggaatataagaaagtgagcaagatcaaaggatccgtcaagtacggcaagcac
aaaacctatcagaagaaaagcgtgaccaacacacaggggaatggaaaagagcagcgaagtaaaatgcctgtgcag
tcaaaacggctgctgaagaatggcccaagcgggtctaaaaaattctggaacaatgtcctgccacactatctggaa
ctgaagtaa

SEQ ID NO: 2

gctccttcctgtactgcccagctccgcttgctccctgaccatccctgcagcagccctgatgtgtcattgtccccc
tcttaacctgcgctgcagtgctgcagggctgggctctggagctgggtctggtcatttctccttagatatgtagag
gcccaggaaaggtttggagcctaagaagccctaggactccaggtctccagggcagccccagcctcttggaatgac
tttccctaataccacaggggtgttctaatcccaggcagacccaagctgcccctcaccaactcctacgtcctcaac

ttcctttcataacttctaggatggaaacacctaatcctccagcaatactgaggcttttctccttattctgttttc
ccttttgaagaagccaaggctcagagcagtcgagtcacctaatcatggtctcatgtcgcctgatcaaggtctcat
gtcaccttatcaagatctcacccactcacctattcagttctcaccagttcagttcaggatggcttctaagctacc
ctgcacagctctgcccacaggacatttgtataagtgaggggtgcaggccttccagccccctccaactccaaaac
tcagccccaagatcaagtggactctctgaacccaccctggccctacagttgtcagggtctggatgggaagatgt
agagctctcggctttcactctggggacttacccagaacatattctcctcatgagctaaggaggctggctgccatc
ttcctacatccccccacggcctgggggcaaggacaccctggccccctggagtctggagaactctgaggacagaac
ttgctcttccacctgcttgggccttacccacaggagaagcactgcttctctacccatgccccatccaactcaggc
accccagggacttgcaacagtctgatttttctcacgtccttcttaaggctctgggctagccacacaaatcaaat
cccagtgataggtccagacaatcctatcctgaaactacatcttagtaagactccagggaatcctttccccaaaga
cagtcttactcctgttctcccccccaagcctttctgggccagaagctttgcctggactcaagcaatggcagacaag
tgccctctgaggacacggaagtgcatgctcagaactgtgattctccaagtggaggcagaggagaaggcccaggct
tcccagcagggctaaggatatgcaaggagtgcattcatccggaggtgttggcagcatcccagccccacccccattc
tcatcgtaaatcaggctcacttccattggctgcatacggtggagtgatgtgaccatatgtcacttgagcattaca
caaatcctaatgagctaaaaatatgtttgttttagctaattgacctctttggccttcataaagcagttggtaaac
atcctcagataatgatttccaaagagcagattgtgggtctcagctgtgcagagaaagcccacgtccctgagacca
ccttctccagctgcctactgaggcacacaggggcgcctgcctgctgcccgctcagccaaggcggtgttgctggag
ccagcttgggacagctctcccaacgctctgccctggccttgcgaccactctctgggccgtagttgtctgtctgtt
aagtgaggaaagtgcccatctccagaggcattcagcggcaaagcagggcttccaggttccgaccccatagcagga
cttcttggatttctacagccagtcagttgcaagcagcacccatattatttctataagaagtggcaggagctggga
tctgaagagttcagcagtctacctttccctgtttcttgtgctttatgcagtcaggaggaatgatctggattccat
gtgaagcctgggaccacggagacccaagacttcctgcttgattctccctgcgaactgcaggctgtgggctgagcc
ttcaagaagcaggagtcccctctagccattaactctcagagctaacctcatttgaatgggaacactagtcctgtg
atgtctggaaggtgggcgcctctacactccacaccctacatggtggtccagacacatcattcccagcattagaaa
gctgtaggggacccgttctgttccctggaggcattaaagggacatagaaataaatctcaagctctgaggctgat
gccagcctcagactcagcctctgcactgtatgggccaattgtagccccaaggacttcttcttgctgcaccccta
tctgtccacacctaaaacgatgggcttctatttagttacagaactctctggcctgttttgttttgctttgctttg
ttttgttttgttttttgttttttgttttttagctatgaaacagaggtaatatctaatacagataacttaccag
taatgagtgcttcctacttactgggtactgggaagaagtgctttacacatattttctcatttaatctacacaata
agtaattaagacatttccctgaggccacgggagagacagtggcagaacagttctccaaggaggacttgcaagtta
ataactggactttgcaaggctctggtggaaactgtcagcttgtaaaggatggagcacagtgtctggcatgtagca
ggaactaaaataatggcagtgattaatgttatgatatgcagacacaacacagcaagataagatgcaatgtacctt
ctgggtcaaaccaccctggccactcctccccgatacccagggttgatgtgcttgaattagacaggattaaaggct
tactggagctggaagccttgccccaactcaggagtttagcccc

SEQ ID NO: 3

ctgcctactgaggcacacaggggcgcctgcctgctgcccgctcagccaaggcggtgttgctggagccagcttggg
acagctctcccaacgctctgccctggccttgcgaccactctctgggccgtagttgtctgtctgttaagtgaggaa
agtgcccatctccagaggcattcagcggcaaagcagggcttccaggttccgaccccatagcaggacttcttggat
ttctacagccagtcagttgcaagcagcacccatattatttctataagaagtggcaggagctgggatctgaagagt
tcagcagtctacctttccctgtttcttgtgctttatgcagtcaggaggaatgatctggattccatgtgaagcctg
ggaccacggagacccaagacttcctgcttgattctccctgcgaactgcaggctgtgggctgagccttcaagaagc
aggagtcccctctagccattaactctcagagctaacctcatttgaatgggaacactagtcctgtgatgtctggaa
ggtgggcgcctctacactccacaccctacatggtggtccagacacatcattcccagcattagaaagctgtagggg
gacccgttctgttccctggaggcattaaagggacatagaaataaatctcaagctctgaggctgatgccagcctca
gactcagcctctgcactgtatgggccaattgtagccccaaggacttcttcttgctgcaccccctatctgtccaca
cctaaaacgatgggcttctatttagttacagaactctctggcctgttttgttttgctttgctttgttttgttttg
ttttttgttttttgttttttagctatgaaacagaggtaatatctaatacagataacttaccagtaatgagtgc
ttcctacttactgggtactgggaagaagtgctttacacatattttctcatttaatctacacaataagtaattaag
acatttccctgaggccacgggagagacagtggcagaacagttctccaaggaggacttgcaagttaataactggac
tttgcaaggctctggtggaaactgtcagcttgtaaaggatggagcacagtgtctggcatgtagcaggaactaaaa
taatggcagtgattaatgttatgatatgcagacacaacacagcaagataagatgcaatgtaccttctgggtcaaa
ccaccctggccactcctccccgatacccagggttgatgtgcttgaattagacaggattaaaggcttactggagct
ggaagccttgccccaactcaggagtttagccccagaccttctgtccaccag

SEQ ID NO: 4

gcttgaattagacaggattaaaggcttactggagctggaagccttgccccaactcaggagtttagcccca

SEQ ID NO: 5

atgagggagccg

atgagggagccggaagagctgatgcccgattcgggtgctgtgtttacatttgggaaaagtaaatttgctgaaaat
aatcccggtaaattctggtttaaaaatgatgtccctgtacatctttcatgtggagatgaacattctgctgttgtt
accggaaataataaactttacatgtttggcagtaacaactggggtcagttaggattaggatcaaagtcagccatc
agcaagccaacatgtgtcaaagctctaaaacctgaaaaagtgaaattagctgcctgtggaaggaaccacaccctg
gtgtcaacagaaggaggcaatgtatatgcaactggtggaaataatgaaggacagttggggcttggtgacaccgaa
gaaagaaacacttttcatgtaattagctttttacatccgagcataagattaagcagctgtctgctggatctaat
acttcagctgccctaactgaggatggaagacttttatgtggggtgacaattccgaagggcaaattggtttaaaa
aatgtaagtaatgtctgtgtccctcagcaagtgaccattgggaaacctgtctcctggatctcttgtggatattac
cattcagcttttgtaacaacagatggtgagctatatgtgtttggagaacctgagaatgggaagttaggtcttccc
aatcagctcctgggcaatcacagaacaccccagctggtgtctgaaattccggagaaggtgatccaagtagcctgt
ggtggagagcatactgtggttctcacggagaatgctgtgtatacctttgggctgggacaatttggtcagctgggt
cttggcacttttcttttgaaacttcagaacccaaagtcattgagaatattagggatcaaacaataagttatatt
tcttgtggagaaaatcacacagctttgataacagatatcggccttatgtatacttttggagatggtcgccacgga
aaattaggacttggactggagaattttaccaatcacttcattcctactttgtgctctaatttttgaggtttata
gttaaattggttgcttgtggtggatgtcacatggtagttttgctgctcctcatcgtggtgtggcaaaagaaatt
gaattcgatgaaataaatgatacttgcttatctgtggcgacttttctgccgtatagcagtttaacctcaggaaat
gtactgcagaggactctatcagcacgtatgcggcgaagagagagggagaggtctccagattctttttcaatgagg
agaacactacctccaatagaagggactcttggcctttctgcttgttttctccccaattcagtctttccacgatgt
tctgagagaaacctccaagagagtgtcttatctgaacaggacctcatgcagccagaggaaccagattatttgcta
gatgaaatgaccaaagaagcagagatagataattcttcaactgtagaaagccttggagaaactactgatatctta
aacatgacacacatcatgagcctgaattccaatgaaaagtcattaaaattatcaccagttcagaaacaaaagaaa
caacaaacaattggggaactgacgcaggatacagctcttactgaaaacgatgatagtgatgaatatgaagaaatg
tcagaaatgaaagaagggaaagcatgtaaacaacatgtgtcacaagggattttcatgacgcagccagctacgact
atcgaagcattttcagatgaggaagtagagatcccagaggagaaggaaggagcagaggattcaaaaggaaatgga
atagaggagcaagaggtagaagcaaatgaggaaaatgtgaaggtgcatggaggaagaaaggagaaaacagagatc
ctatcagatgaccttacagacaaagcagaggtgagtgaaggcaaggcaaaatcagtgggagaagcagaggatggg
cctgaaggtagaggggatggaacctgtgaggaaggtagttcaggagcagaacactggcaagatgaggagagggag
aaggggagaaagacaagggtagaggagaaatggagaggccaggagagggagagaaggaactagcagagaaggaa
gaatggaagaagagggatggggaagagcaggagcaaaaggagagggagcagggccatcagaaggaaagaaaccaa
gagatggaggagggaggggaggaggagcatggagaaggagaagaagaggagggagacagagaagaggaagaagag
aaggagggagaagggaaagaggaaggagaaggggaagaagtggagggagaacgtgaaaaggaggaaggagagagg
aaaaaggaggaaagagcggggaaggaggagaaaggagaggaagaaggagaccaaggagaggggggaagaggaggaa
acagaggggagaggggaggaaaaagaggagggaggggaagtagagggaggggaagtagaggaggggaaaggagag
agggaagaggaagaggaggagggtgaggggaagaggaggaaggggaggggaagaggaggaaggggaggggggaa
gaggaggaaggagaagggaaaggggaggaagaaggggaagaaggagaaggggaggaagaaggggaggaaggagaa
ggggaggggaagaggaggaaggagaaggggagggagaagaggaaggagaaggggaggagaagaggaggaaggga
gaaggggagggagaagaggaaggagaaggggagggagaagaggaggaaggagaagggaaaggggaggaggaagga
gaggaaggagaagggaggggaagaggaggaaggagaagggaggggaagaggaaggagaaggggaggggaagaag
gaggaaggagaatgggaggggaagaggaggaaggagaaggggaggggaagaggaaggagaaggggaaggggag
gaaggagaaggggaggggaagaggaggaaggagaaggggaggggaagaggaaggagaaggggaaggggaggtg
gaaggggaggaaggagaggggaaggagaggaagaggaaggagaggaggaaggagaagaaagggaaaaggagggg
gaaggagaagaaaacaggaggaacagagaagaggaggaggaagaagaggggaagtatcaggagacaggcgaagaa
gagaatgaaaggcaggatggagaggagtacaaaaaagtgagcaaaataaaaggatctgtgaaatatggcaaacat
aaaacatatcaaaaaaagtcagttactaacacacagggaaatgggaaagagcagaggtccaaaatgccagtccag
tcaaaacgacttttaaaaaacgggccatcaggttccaaaaagttctggaataatgtattaccacattacttggaa
ttgaagtaa

SEQ ID NO: 6

MREPEELMPDSGAVFTFGKSKFAENNPGKFWFKNDVPVHLSCGDEHSAVVTGNNKLYMFGSNNWGQLGLGSKSAI
SKPTCVKALKPEKVKLAACGRNHTLVSTEGGNVYATGGNNEGQLGLGDTEERNTFHVISFFTSEHKIKQLSAGSN
TSAALTEDGRLFMWGDNSEGQIGLKNVSNVCVPQQVTIGKPVSWISCGYYHSAFVTTDGELYVFGEPENGKLGLP
NQLLGNHRTPQLVSEIPEKVIQVACGGEHTVVLTENAVYTFGLGQFGQLGLGTFLFETSEPKVIENIRDQTISYI
SCGENHTALITDIGLMYTFGDGRHGKLGLGLENFTNHFIPTLCSNFLRFIVKLVACGGCHMVVFAAPHRGVAKEI
EFDEINDTCLSVATFLPYSSLTSGNVLQRTLSARMRRRERERSPDSFSMRRTLPPIEGTLGLSACFLPNSVFPRC

SERNLQESVLSEQDLMQPEEPDYLLDEMTKEAEIDNSSTVESLGETTDILNMTHIMSLNSNEKSLKLSPVQKQKK
QQTIGELTQDTALTENDDSDEYEEMSEMKEGKACKQHVSQGIFMTQPATTIEAFSDEEVEIPEEKEGAEDSKGNG
IEEQEVEANEENVKVHGGRKEKTEILSDDLTDKAEVSEGKAKSVGEAEDGPEGRGDGTCEEGSSGAEHWQDEERE
KGEKDKGRGEMERPGEGEKELAEKEEWKKRDGEEQEQKEREQGHQKERNQEMEEGGEEEHGEGEEEEGDREEEEE
KEGEGKEEGEGEEVEGEREKEEGERKKEERAGKEEKGEEEGDQGEGEEEETEGRGEEKEEGGEVEGGEVEEGKGE
REEEEEEGEGEEEEGEGEEEEGEGEEEEGEGKGEEEGEEGEGEEEGEEGEGEGEEEEGEGEGEEEGEGEGEEEEG
EGEGEEEGEGEGEEEEGEGKGEEEGEEGEGEGEEEEGEGEGEDGEGEGEEEEGEWEGEEEEGEGEGEEEGEGEGE
EGEGEGEEEEGEGEGEEEEGEEEGEEEGEGEEEGEGEGEEEEEGEVEGEVEGEEGEGEGEEEEGEEEGEEREKEG
EGEENRRNREEEEEEEGKYQETGEEENERQDGEEYKKVSKIKGSVKYGKHKTYQKKSVTNTQGNGKEQRSKMPVQ
SKRLLKNGPSGSKKFWNNVLPHYLELK

SEQ ID NO: 7

tctagactcgaggaactgaaaaaccagaaagttaactggtaagtttagtcttttgtcttttatttcaggtcccg
gatccggtggtggtgcaaatcaaagaactgctcctcagtggatgttgcctttacttctaggcctgtacggaagtg
ttacttctgctctaaaagctgcggaattgtacccGCGGCCGCgccaccATGAGAGAGCCAGAGGAGCTGATGCCA
GATAGCGGAGCAGTGTTTACCTTCGGAAAGTCCAAGTTCGCAGAGAATAACCCAGGAAAGTTCTGGTTTAAAAAC
GACGTGCCCGTCCACCTGTCTTGTGGCGATGAGCATAGTGCCGTGGTCACTGGGAACAATAAGCTGTATATGTTC
GGGTCCAACAATTGGGGACAGCTGGGGCTGGGATCCAAATCTGCTATCTCTAAGCCAACCTGCGTGAAGGCACTG
AAACCCGAGAAGGTCAAACTGGCCGCTTGTGGCAGAAACCACACTCTGGTGAGCACCGAGGGCGGGAATGTCTAT
GCCACCGGAGGCAACAATGAGGGACAGCTGGGACTGGGGGACACTGAGGAAAGGAATACCTTTCACGTGATCTCC
TTCTTTACATCTGAGCATAAGATCAAGCAGCTGAGCGCCGGCTCCAACACATCTGCAGCCCTGACTGAGGACGGG
CGCCTGTTCATGTGGGGAGATAATTCAGAGGGCCAGATTGGGCTGAAAAACGTGAGCAACGTGTGCGTGCCTCAG
CAGGTGACCATCGGAAAGCCAGTCAGTTGGATTTCATGTGGCTACTATCATAGCGCCTTCGTGACCACAGATGGC
GAGCTGTACGTCTTTGGGGAGCCCGAAAACGGAAAACTGGGCCTGCCTAACCAGCTGCTGGGCAATCACCGGACA
CCCCAGCTGGTGTCCGAGATCCCTGAAAAAGTGATCCAGGTCGCCTGCGGGGGAGAGCATACAGTGGTCCTGACT
GAGAATGCCGTGTACACCTTCGGACTGGGCCAGTTTGGCCAGCTGGGGCTGGGAACCTTCCTGTTTGAGACATCC
GAACCAAAAGTGATCGAGAACATTCGCGACCAGACTATCAGCTACATTTCCTGCGGAGAGAATCACACCGCACTG
ATCACAGACATTGGCCTGATGTATACCTTTGGCGATGGGCGGCACGGGAAGCTGGGACTGGGCCTGGAGAACTTC
ACTAATCACTTCATCCCCACCCTGTGCTCTAACTTCCTGCGGTTCATCGTGAAACTGGTCGCTTGCGGCGGGTGT
CACATGGTGGTCTTCGCTGCACCTCATAGGGGCGTGGCTAAGGAGATCGAATTTGACGAGATTAACGATACATGC
CTGAGCGTGGCAACTTTCCTGCCATACAGCTCCCTGACTTCTGGCAATGTGCTGCAGAGAACCCTGAGTGCAAGG
ATGCGGAGAAGGGAGAGGGAACGCTCTCCTGACAGTTTCTCAATGCGACGAACCCTGCCACCTATCGAGGGGACA
CTGGGACTGAGTGCCTGCTTCCTGCCTAACTCAGTGTTTCCACGATGTAGCGAGCGGAATCTGCAGGAGTCTGTC
CTGAGTGAGCAGGATCTGATGCAGCCAGAGGAACCCGACTACCTGCTGGATGAGATGACCAAGGAGGCCGAAATC
GACAACTCTAGTACAGTGGAGTCCCTGGGCGAGACTACCGATATCCTGAATATGACACACATTATGTCACTGAAC
AGCAATGAGAAGAGTCTGAAACTGTCACCAGTGCAGAAGCAGAAGAAACAGCAGACTATTGGCGAGCTGACTCAG
GACACCGCCCTGACAGAGAACGACGATAGCGATGAGTATGAGGAAATGTCCGAGATGAAGGAAGGCAAAGCTTGT
AAGCAGCATGTGAGTCAGGGGATCTTCATGACACAGCCAGCCACAACTATTGAGGCTTTTTCAGACGAGGAAGTG
GAGATCCCCGAGGAAAAAGAGGGCGCAGAAGATTCCAAGGGGAATGGAATTGAGGAACAGGAGGTGGAAGCCAAC
GAGGAAAATGTGAAAGTCCACGGAGGCAGGAAGGAGAAAACAGAAATCCTGTCTGACGATCTGACTGACAAGGCC
GAGGTGTCCGAAGGCAAGGCAAAATCTGTCGGAGAGGCAGAAGACGGACCAGAGGGACGAGGGGATGGAACCTGC
GAGGAAGGCTCAAGCGGGGCTGAGCATTGGCAGGACGAGGAACGAGAGAAGGGCGAAAAGGATAAAGGCCGCGGG
GAGATGGAACGACCTGGAGAGGGCGAAAAAGAGCTGGCAGAGAAGGAGGAATGGAAGAAAAGGGACGGCGAGGAA
CAGGAGCAGAAAGAAAGGGAGCAGGGCCACCAGAAGGAGCGCAACCAGGAGATGGAAGAGGGCGGCGAGGAAGAG

CATGGCGAGGGAGAAGAGGAAGAGGGCGATAGAGAAGAGGAAGAGGAAAAAGAAGGCGAAGGGAAGGAGGAAGGA
GAGGGCGAGGAAGTGGAAGGCGAGAGGGAAAAGGAGGAAGGAGAACGGAAGAAAGAGGAAAGAGCCGGCAAAGAG
GAAAAGGGCGAGGAAGAGGGCGATCAGGGCGAAGGCGAGGAGGAAGAGACCGAGGGCCGCGGGGAAGAGAAAGAG
GAGGGAGGAGAGGTGGAGGGCGGAGAGGTCGAAGAGGGAAAGGGCGAGCGCGAAGAGGAAGAGGAAGAGGGCGAG
GGCGAGGAAGAAGAGGGCGAGGGGGAAGAAGAGGAGGGAGAGGGCGAAGAGGAAGAGGGGGAGGGAAAGGGCGAA
GAGGAAGGAGAGGAAGGGGAGGGAGAGGAAGAGGGGGAGGAGGGCGAGGGGGAAGGCGAGGAGGAAGAAGGAGAG
GGGGAAGGCGAAGAGGAAGGCGAGGGGGAAGGAGAGGAGGAAGAAGGGGAAGGCGAAGGCGAAGAGGAGGGAGAA
GGAGAGGGGGAGGAAGAGGAAGGAGAAGGGAAGGGCGAGGAGGAAGGCGAAGAGGGAGAGGGGGAAGGCGAGGAA
GAGGAAGGCGAGGGCGAAGGAGAGGACGGCGAGGGCGAGGGAGAAGAGGAGGAAGGGGAATGGGAAGGCGAAGAA
GAGGAAGGCGAAGGCGAAGGCGAAGAAGAGGGCGAAGGGGAGGGCGAGGAGGGCGAAGGCGAAGGGGAGGAAGAG
GAAGGCGAAGGAGAAGGCGAGGAAGAAGAGGGAGAGGAGGAAGGCGAGGAGGAAGGAGAGGGGGAGGAGGAGGGA
GAAGGCGAGGGCGAAGAAGAAGAAGAGGGAGAAGTGGAGGGCGAAGTCGAGGGGGAGGAGGGAGAAGGGGAAGGG
GAGGAAGAAGAGGGCGAAGAAGAAGGCGAGGAAAGAGAAAAAGAGGGAGAAGGCGAGGAAAACCGGAGAAATAGG
GAAGAGGAGGAAGAGGAAGAGGGAAAGTACCAGGAGACAGGCGAAGAGGAAAACGAGCGGCAGGATGGCGAGGAA
TATAAGAAAGTGAGCAAGATCAAAGGATCCGTCAAGTACGGCAAGCACAAAACCTATCAGAAGAAAAGCGTGACC
AACACACAGGGGAATGGAAAAGAGCAGCGAAGTAAAATGCCTGTGCAGTCAAAACGGCTGCTGAAGAATGGCCCA
AGCGGGTCTAAAAAATTCTGGAACAATGTCCTGCCACACTATCTGGAACTGAAGTAAgcggccgcgcggatccag
acatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtgaaa
tttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcatt
ttatgtttcaggttcaggggaggtgtgggaggtttttagcatgc

SEQ ID NO: 8

Agcacagtgtctggcatgtagcaggaactaaaataatggcagtgattaatgttatgatatgcagacacaacacag
caagataagatgcaatgtaccttctgggtcaaaccaccctggccactcctccccgatacccagggttgatgtgct
tgaattagacaggattaaaggcttactggagctggaagccttgccccaactcaggagtttagccccagaccttct
gtccaccagc

CLAIMS

1. A polynucleotide comprising the nucleic acid sequence of SEQ ID NO: 1 encoding a human retinitis pigmentosa GTPase regulator (RPGR) protein.

2. An expression cassette comprising the polynucleotide of claim 1 and an expression control sequence operably linked and heterologous to the nucleic acid sequence.

3. A vector comprising the polynucleotide of claim 1.

4. The vector of claim 3, wherein the vector is a recombinant adeno-associated (rAAV) expression vector.

5. A recombinant herpes simplex virus (rHSV) comprising the polynucleotide of claim 1.

6. A host cell comprising the polynucleotide of claim 1.

7. The host cell of claim 6, wherein the host cell is a mammalian cell.

8. The host cell of claim 6, wherein the host cell is a HeLa cell, a BHK21 cell or a Vero cell.

9. The host cell of claim 6, wherein the host cell is a V27 cell.

10. The expression cassette of claim 2, wherein the expression control sequence is a human interphotoreceptor retinoid-binding protein (IRBP) promoter.

11. The expression cassette of claim 10, wherein the human IRBP promoter comprises a nucleic acid sequence having at least 95% sequence identity to the nucleic acid sequence of SEQ ID NO: 8 and directs preferential expression in rods and cones.

12. The expression cassette of claim 10, wherein the human IRBP promoter comprises the nucleic acid sequence of SEQ ID NO: 8.

13. The polynucleotide of claim 1, wherein the polynucleotide comprises the nucleic acid sequence of SEQ ID NO: 7.

14. A method of producing the rAAV expression vector of claim 4, comprising

(a) infecting a host cell with a recombinant herpes simplex virus (rHSV) comprising the nucleic acid sequence of SEQ ID NO: 1;

(b) incubating the host cell; and

(c) following incubation, collecting rAAV from the host cell of step (b).

15. The method of claim 14, wherein the host cell is a HeLa cell, a BHK21 cell or a Vero cell.

16. The method of claim 14, wherein the rHSV further comprises a human IRBP promoter operably linked to the nucleic acid sequence of SEQ ID NO: 1.

17. The method of claim 16, wherein the human IRBP promoter comprises a nucleic acid sequence having at least 95% sequence identity to the nucleic acid sequence of SEQ ID NO: 8 and directs preferential expression in rods and cones.

18. The method of claim 16, wherein the human IRBP promoter comprises the nucleic acid sequence of SEQ ID NO: 8.

19. The method of claim 14, wherein the rHSV comprises the nucleic acid sequence of SEQ ID NO: 7.

```
                          1                                                                  80
SEQ ID NO: 5      (1)     ATGAGGGAGCCGGAAGAGCTGATGCCCGATTCGGGTGCTGTGTTTACATTTGGGAAAAGTAAATTTGCTGAAAATAATCC
SEQ ID NO: 1      (1)     ATGAGAGAGCCAGAGGAGCTGATGCCAGATAGCGGAGCAGTGTTTACCTTCGGAAAGTCCAAGTTCGCAGAGAATAACCC
                          81                                                                 160
SEQ ID NO: 5     (81)     CGGTAAATTCTGGTTTAAAAATGATGTCCCTGTACATCTTTCATGTGGAGATGAACATTCTGCTGTTGTTACCGGAAATA
SEQ ID NO: 1     (81)     AGGAAAGTTCTGGTTTAAAAACGACGTGCCCGTCCACCTGTCTTGTGGCGATGAGCATAGTGCCGTGGTCACTGGGAACA
                         161                                                                 240
SEQ ID NO: 5    (161)     ATAAACTTTACATGTTTGGCAGTAACAACTGGGGTCAGTTAGGATTAGGATCAAAGTCAGCCATCAGCAAGCCAACATGT
SEQ ID NO: 1    (161)     ATAAGCTGTATATGTTCGGGTCCAACAATTGGGGACAGCTGGGGCTGGGATCCAAATCTGCTATCTCTAAGCCAACCTGC
                         241                                                                 320
SEQ ID NO: 5    (241)     GTCAAAGCTCTAAAACCTGAAAAAGTGAAATTAGCTGCCTGTGGAAGGAACCACACCCTGGTGTCAACAGAAGGAGGCAA
SEQ ID NO: 1    (241)     GTGAAGGCACTGAAACCCGAGAAGGTCAAACTGGCCGCTTGTGGCAGAAACCACACTCTGGTGAGCACCGAGGGCGGGAA
                         321                                                                 400
SEQ ID NO: 5    (321)     TGTATATGCAACTGGTGGAAATAATGAAGGACAGTTGGGGCTTGGTGACACCGAAGAAAGAAACACTTTTCATGTAATTA
SEQ ID NO: 1    (321)     TGTCTATGCCACCGGAGGCAACAATGAGGGACAGCTGGGACTGGGGGGACACTGAGGAAAGGAATACCTTTCACGTGATCT
                         401                                                                 480
SEQ ID NO: 5    (401)     GCTTTTTTACATCCGAGCATAAGATTAAGCAGCTGTCTGCTGGATCTAATACTTCAGCTGCCCTAACTGAGGATGGAAGA
SEQ ID NO: 1    (401)     CCTTCTTTACATCTGAGCATAAGATCAAGCAGCTGAGCGCCGGCTCCAACACATCTGCAGCCCTGACTGAGGACGGGCGC
                         481                                                                 560
SEQ ID NO: 5    (481)     CTTTTTATGTGGGGTGACAATTCCGAAGGGCAAATTGGTTTAAAAAATGTAAGTAATGTCTGTGTCCCTCAGCAAGTGAC
SEQ ID NO: 1    (481)     CTGTTCATGTGGGGAGATAATTCAGAGGGCCAGATTGGGCTGAAAAACGTGAGCAACGTGTGCGTGCCTCAGCAGGTGAC
                         561                                                                 640
SEQ ID NO: 5    (561)     CATTGGGAAACCTGTCTCCTGGATCTCTTGTGGATATTACCATTCAGCTTTTGTAACAACAGATGGTGAGCTATATGTGT
SEQ ID NO: 1    (561)     CATCGGAAAGCCAGTCAGTTGGATTTCATGTGGCTACTATCATAGCGCCTTCGTGACCACAGATGGCGAGCTGTACGTCT
                         641                                                                 720
SEQ ID NO: 5    (641)     TTGGAGAACCTGAGAATGGGAAGTTAGGTCTTCCCAATCAGCTCCTGGGCAATCACAGAACACCCCAGCTGGTGTCTGAA
SEQ ID NO: 1    (641)     TTGGGGAGCCCGAAAACGGAAAACTGGGCCTGCCTAACCAGCTGCTGGGCAATCACCGGACACCCCAGCTGGTGTCCGAG
                         721                                                                 800
SEQ ID NO: 5    (721)     ATTCCGGAGAAGGTGATCCAAGTAGCCTGTGGTGGAGAGCATACTGTGGTTCTCACGGAGAATGCTGTGTATACCTTTGG
SEQ ID NO: 1    (721)     ATCCCTGAAAAAGTGATCCAGGTCGCCTGCGGGGGAGAGCATACAGTGGTCCTGACTGAGAATGCCGTGTACACCTTCGG
                         801                                                                 880
SEQ ID NO: 5    (801)     GCTGGGACAATTTGGTCAGCTGGGTCTTGGCACTTTTCTTTTTGAAACTTCAGAACCCAAAGTCATTGAGAATATTAGGG
SEQ ID NO: 1    (801)     ACTGGGCCAGTTTGGCCAGCTGGGGCTGGGAACCTTCCTGTTTGAGACATCCGAACCAAAAGTGATCGAGAACATTCGCG
                         881                                                                 960
SEQ ID NO: 5    (881)     ATCAAACAATAAGTTATATTTCTTGTGGAGAAAATCACACAGCTTTGATAACAGATATCGGCCTTATGTATACTTTTGGA
SEQ ID NO: 1    (881)     ACCAGACTATCAGCTACATTTCCTGCGGAGAGAATCACACCGCACTGATCACAGACATTGGCCTGATGTATACCTTTGGC
```

FIG. 1A

CA 02945965 2016-10-14

```
                        961                                                                1040
SEQ ID NO: 5    (961)   GATGGTCGCCACGGAAAATTAGGACTTGGACTGGAGAATTTTACCAATCACTTCATTCCTACTTTGTGCTCTAATTTTTT
SEQ ID NO: 1    (961)   GATGGGCGGCACGGGAAGCTGGGACTGGGCCTGGAGAACTTCACTAATCACTTCATCCCCACCCTGTGCTCTAACTTCCT
                        1041                                                               1120
SEQ ID NO: 5   (1041)   GAGGTTTATAGTTAAATTGGTTGCTTGTGGTGGATGTCACATGGTAGTTTTTGCTGCTCCTCATCGTGGTGTGGCAAAAG
SEQ ID NO: 1   (1041)   GCGGTTCATCGTGAAACTGGTCGCTTGCGGCGGGTGTCACATGGTGGTCTTCGCTGCACCTCATAGGGGCGTGGCTAAGG
                        1121                                                               1200
SEQ ID NO: 5   (1121)   AAATTGAATTCGATGAAATAAATGATACTTGCTTATCTGTGGCGACTTTTCTGCCGTATAGCAGTTTAACCTCAGGAAAT
SEQ ID NO: 1   (1121)   AGATCGAATTTGACGAGATTAACGATACATGCCTGAGCGTGGCAACTTTCCTGCCATACAGCTCCCTGACTTCTGGCAAT
                        1201                                                               1280
SEQ ID NO: 5   (1201)   GTACTGCAGAGGACTCTATCAGCACGTATGCGGCGAAGAGAGAGGGAGAGGTCTCCAGATTCTTTTTCAATGAGGAGAAC
SEQ ID NO: 1   (1201)   GTGCTGCAGAGAACCCTGAGTGCAAGGATGCGGAGAAGGGAGAGGGAACGCTCTCCTGACAGTTTCTCAATGCGACGAAC
                        1281                                                               1360
SEQ ID NO: 5   (1281)   ACTACCTCCAATAGAAGGGACTCTTGGCCTTTCTGCTTGTTTTCTCCCCAATTCAGTCTTTCCACGATGTTCTGAGAGAA
SEQ ID NO: 1   (1281)   CCTGCCACCTATCGAGGGGACACTGGGACTGAGTGCCTGCTTCCTGCCTAACTCAGTGTTTCCACGATGTAGCGAGCGGA
                        1361                                                               1440
SEQ ID NO: 5   (1361)   ACCTCCAAGAGAGTGTCTTATCTGAACAGGACCTCATGCAGCCAGAGGAACCAGATTATTTGCTAGATGAAATGACCAAA
SEQ ID NO: 1   (1361)   ATCTGCAGGAGTCTGTCCTGAGTGAGCAGGATCTGATGCAGCCAGAGGAACCCGACTACCTGCTGGATGAGATGACCAAG
                        1441                                                               1520
SEQ ID NO: 5   (1441)   GAAGCAGAGATAGATAATTCTTCAACTGTAGAAAGCCTTGGAGAAACTACTGATATCTTAAACATGACACACATCATGAG
SEQ ID NO: 1   (1441)   GAGGCCGAAATCGACAACTCTAGTACAGTGGAGTCCCTGGGCGAGACTACCGATATCCTGAATATGACACACATTATGTC
                        1521                                                               1600
SEQ ID NO: 5   (1521)   CCTGAATTCCAATGAAAAGTCATTAAAATTATCACCAGTTCAGAAACAAAAGAAACAACAAACAATTGGGGAACTGACGC
SEQ ID NO: 1   (1521)   ACTGAACAGCAATGAGAAGAGTCTGAAACTGTCACCAGTGCAGAAGCAGAAGAAACAGCAGACTATTGGCGAGCTGACTC
                        1601                                                               1680
SEQ ID NO: 5   (1601)   AGGATACAGCTCTTACTGAAAACGATGATAGTGATGAATATGAAGAAATGTCAGAAATGAAAGAAGGGAAAGCATGTAAA
SEQ ID NO: 1   (1601)   AGGACACCGCCCTGACAGAGAACGACGATAGCGATGAGTATGAGGAAATGTCCGAGATGAAGGAAGGCAAAGCTTGTAAG
                        1681                                                               1760
SEQ ID NO: 5   (1681)   CAACATGTGTCACAAGGGATTTTCATGACGCAGCCAGCTACGACTATCGAAGCATTTTCAGATGAGGAAGTAGAGATCCC
SEQ ID NO: 1   (1681)   CAGCATGTGAGTCAGGGGATCTTCATGACACAGCCAGCCACAACTATTGAGGCTTTTTCAGACGAGGAAGTGGAGATCCC
                        1761                                                               1840
SEQ ID NO: 5   (1761)   AGAGGAGAAGGAAGGAGCAGAGGATTCAAAAGGAAATGGAATAGAGGAGCAAGAGGTAGAAGCAAATGAGGAAAATGTGA
SEQ ID NO: 1   (1761)   CGAGGAAAAAGAGGGCGCAGAAGATTCCAAGGGGAATGGAATTGAGGAACAGGAGGTGGAAGCCAACGAGGAAAATGTGA
                        1841                                                               1920
```

FIG. 1A (Continued)

```
SEQ ID NO: 5   (1841) AGGTGCATGGAGGAAGAAAGGAGAAAACAGAGATCCTATCAGATGACCTTACAGACAAAGCAGAGGTGAGTGAAGGCAAG
SEQ ID NO: 1   (1841) AAGTCCACGGAGGCAGGAAGGAGAAAACAGAAATCCTGTCTGACGATCTGACTGACAAGGCCGAGGTGTCCGAAGGCAAG
                      1921                                                                        2000
SEQ ID NO: 5   (1921) GCAAAATCAGTGGGAGAAGCAGAGGATGGGCCTGAAGGTAGAGGGGATGGAACCTGTGAGGAAGGTAGTTCAGGAGCAGA
SEQ ID NO: 1   (1921) GCAAAATCTGTCGGAGAGGCAGAAGACGGACCAGAGGGACGAGGGGATGGAACCTGCGAGGAAGGCTCAAGCGGGGCTGA
                      2001                                                                        2080
SEQ ID NO: 5   (2001) ACACTGGCAAGATGAGGAGAGGGAGAAGGGGGAGAAAGACAAGGGTAGAGGAGAAATGGAGAGGCCAGGAGAGGGAGAGA
SEQ ID NO: 1   (2001) GCATTGGCAGGACGAGGAACGAGAGAAGGGCGAAAAGGATAAAGGCCGCGGGGAGATGGAACGACCTGGAGAGGGCGAAA
                      2081                                                                        2160
SEQ ID NO: 5   (2081) AGGAACTAGCAGAGAAGGAAGAATGGAAGAAGAGGGATGGGGAAGAGCAGGAGCAAAAGGAGAGGGAGCAGGGCCATCAG
SEQ ID NO: 1   (2081) AAGAGCTGGCAGAGAAGGAGGAATGGAAGAAAAGGGACGGCGAGGAACAGGAGCAGAAAGAAAGGGAGCAGGGCCACCAG
                      2161                                                                        2240
SEQ ID NO: 5   (2161) AAGGAAAGAAACCAAGAGATGGAGGAGGGAGGGGAGGAGGAGCATGGAGAAGGAGAAGAAGAGGAGGGAGACAGAGAAGA
SEQ ID NO: 1   (2161) AAGGAGCGCAACCAGGAGATGGAAGAGGGCGGCGAGGAAGAGCATGGCGAGGGAGAAGAGGAAGAGGGCGATAGAGAAGA
                      2241                                                                        2320
SEQ ID NO: 5   (2241) GGAAGAAGAGAAGGAGGGAGAAGGGAAAGAGGAAGGAGAAGGGGAAGAAGTGGAGGGAGAACGTGAAAAGGAGGAAGGAG
SEQ ID NO: 1   (2241) GGAAGAGGAAAAAGAAGGCGAAGGGAAGGAGGAAGGAGAGGGCGAGGAAGTGGAAGGCGAGAGGGAAAAGGAGGAAGGAG
                      2321                                                                        2400
SEQ ID NO: 5   (2321) AGAGGAAAAAGGAGGAAAGAGCGGGGAAGGAGGAGAAAGGAGAGGAAGAAGGAGACCAAGGAGAGGGGGAAGAGGAGGAA
SEQ ID NO: 1   (2321) AACGGAAGAAAGAGGAAAGAGCCGGCAAAGAGGAAAAGGGCGAGGAAGAGGGCGATCAGGGCGAAGGCGAGGAGGAAGAG
                      2401                                                                        2480
SEQ ID NO: 5   (2401) ACAGAGGGGAGAGGGGAGGAAAAAGAGGAGGGAGGGGAAGTAGAGGGAGGGGAAGTAGAGGAGGGGAAAGGAGAGAGGGA
SEQ ID NO: 1   (2401) ACCGAGGGCCGCGGGGAAGAGAAAGAGGAGGGAGGAGAGGTGGAGGGCGGAGAGGTCGAAGAGGGAAAGGGCGAGCGCGA
                      2481                                                                        2560
SEQ ID NO: 5   (2481) AGAGGAAGAGGAGGAGGGTGAGGGGGAAGAGGAGGAAGGGGAGGGGGAAGAGGAGGAAGGGGAGGGGGAAGAGGAGGAAG
SEQ ID NO: 1   (2481) AGAGGAAGAGGAAGAGGGCGAGGGCGAGGAAGAAGAGGGCGAGGGGGAAGAAGAGGAGGGAGAGGGCGAAGAGGAAGAGG
                      2561                                                                        2640
SEQ ID NO: 5   (2561) GAGAAGGGAAAGGGGAGGAAGAAGGGGAAGAAGGAGAAGGGGAGGAAGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAG
SEQ ID NO: 1   (2561) GGGAGGGAAAGGGCGAAGAGGAAGGAGAGGAAGGGGAGGGAGAGGAAGAGGGGGAGGAGGGCGAGGGGGAAGGCGAGGAG
                      2641                                                                        2720
SEQ ID NO: 5   (2641) GAGGAAGGAGAAGGGGAGGGAGAAGAGGAAGGAGAAGGGGAGGGAGAAGAGGAGGAAGGAGAAGGGGAGGGAGAAGAGGA
SEQ ID NO: 1   (2641) GAAGAAGGAGAGGGGGAAGGCGAAGAGGAAGGCGAGGGGGAAGGAGAGGAGGAAGAAGGGGAAGGCGAAGGCGAAGAGGA
```

FIG. 1B

```
                        2721                                                                2800
SEQ ID NO: 5   (2721) AGGAGAAGGGGAGGGAGAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAGGAAGGAGAGGAAGGAGAAGGGGAGGGGGAAG
SEQ ID NO: 1   (2721) GGGAGAAGGAGAGGGGGAGGAAGAGGAAGGAGAAGGGAAGGGCGAGGAGGAAGGCGAAGAGGGAGAGGGGGAAGGCGAGG
                        2801                                                                2880
SEQ ID NO: 5   (2801) AGGAGGAAGGAGAAGGGGAAGGGGAGGATGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAATGGGAGGGGGAAGAGGAG
SEQ ID NO: 1   (2801) AAGAGGAAGGCGAGGGCGAAGGAGAGGACGGCGAGGGCGAGGGAGAAGAGGAGGAAGGGGAATGGGAAGGCGAAGAAGAG
                        2881                                                                2960
SEQ ID NO: 5   (2881) GAAGGAGAAGGGGAGGGGGAAGAGGAAGGAGAAGGGGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGA
SEQ ID NO: 1   (2881) GAAGGCGAAGGCGAAGGCGAAGAAGAGGGCGAAGGGGAGGGCGAGGAGGGCGAAGGCGAAGGGGAGGAAGAGGAAGGCGA
                        2961                                                                3040
SEQ ID NO: 5   (2961) AGGGGAGGGGGAAGAGGAGGAAGGGGAAGAAGAAGGGGAGGAAGAAGGAGAGGGAGAGGAAGAAGGGGAGGGAGAAGGGG
SEQ ID NO: 1   (2961) AGGAGAAGGCGAGGAAGAAGAGGGAGAGGAGGAAGGCGAGGAGGAAGGAGAGGGGGAGGAGGAGGGAGAAGGCGAGGGCG
                        3041                                                                3120
SEQ ID NO: 5   (3041) AGGAAGAAGAGGAAGGGGAAGTGGAAGGGGAGGTGGAAGGGGAGGAAGGAGAGGGGGAAGGAGAGGAAGAGGAAGGAGAG
SEQ ID NO: 1   (3041) AAGAAGAAGAAGAGGGAGAAGTGGAGGGCGAAGTCGAGGGGGAGGAGGGAGAAGGGGAAGGGGAGGAAGAAGAGGGCGAA
                        3121                                                                3200
SEQ ID NO: 5   (3121) GAGGAAGGAGAAGAAAGGGAAAAGGAGGGGGAAGGAGAAGAAAACAGGAGGAACAGAGAAGAGGAGGAGGAAGAAGAGGG
SEQ ID NO: 1   (3121) GAAGAAGGCGAGGAAAGAGAAAAAGAGGGAGAAGGCGAGGAAAACCGGAGAAATAGGGAAGAGGAGGAAGAGGAAGAGGG
                        3201                                                                3280
SEQ ID NO: 5   (3201) GAAGTATCAGGAGACAGGCGAAGAAGAGAATGAAAGGCAGGATGGAGAGGAGTACAAAAAAGTGAGCAAAATAAAAGGAT
SEQ ID NO: 1   (3201) AAAGTACCAGGAGACAGGCGAAGAGGAAAACGAGCGGCAGGATGGCGAGGAATATAAGAAAGTGAGCAAGATCAAAGGAT
                        3281                                                                3360
SEQ ID NO: 5   (3281) CTGTGAAATATGGCAAACATAAAACATATCAAAAAAAGTCAGTTACTAACACACAGGGAAATGGGAAAGAGCAGAGGTCC
SEQ ID NO: 1   (3281) CCGTCAAGTACGGCAAGCACAAAACCTATCAGAAGAAAAGCGTGACCAACACACAGGGGAATGGAAAAGAGCAGCGAAGT
                        3361                                                                3440
SEQ ID NO: 5   (3361) AAAATGCCAGTCCAGTCAAAACGACTTTTAAAAAACGGGCCATCAGGTTCCAAAAAGTTCTGGAATAATGTATTACCACA
SEQ ID NO: 1   (3361) AAAATGCCTGTGCAGTCAAAACGGCTGCTGAAGAATGGCCCAAGCGGGTCTAAAAAATTCTGGAACAATGTCCTGCCACA
                        3441           3459
SEQ ID NO: 5   (3441) TTACTTGGAATTGAAGTAA
SEQ ID NO: 1   (3441) CTATCTGGAACTGAAGTAA
```

FIG. 1B (Continued)

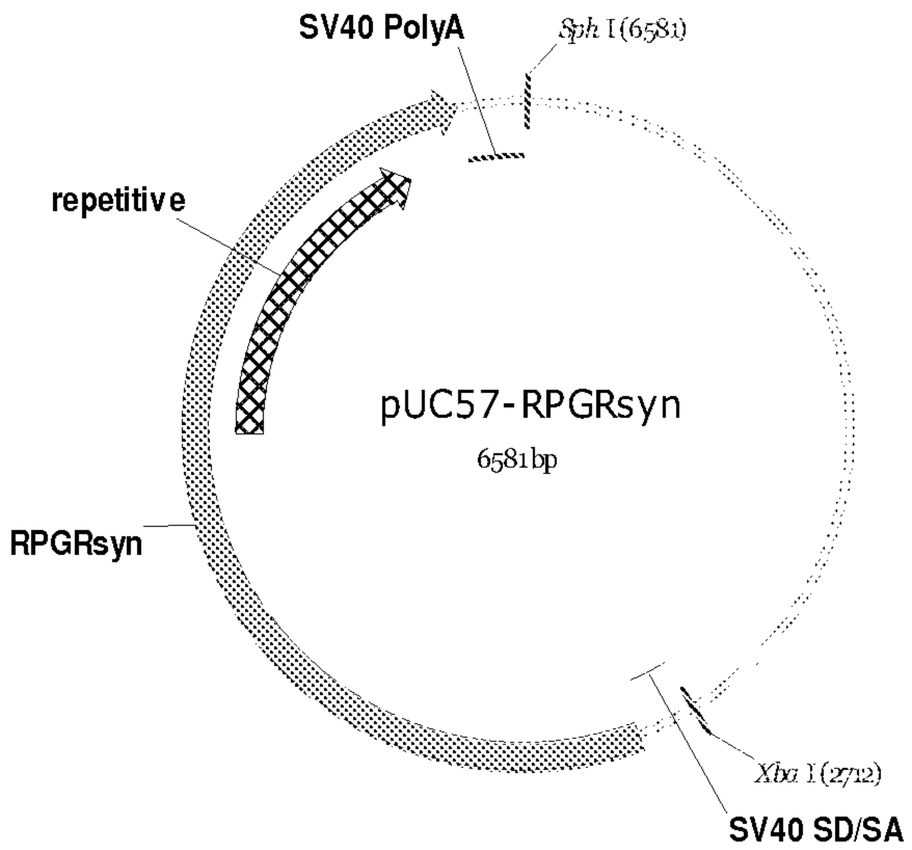
SV40 PolyA
Sph I (6581)
repetitive
pUC57-RPGRsyn
6581bp
RPGRsyn
Xba I (2712)
SV40 SD/SA

FIG. 2

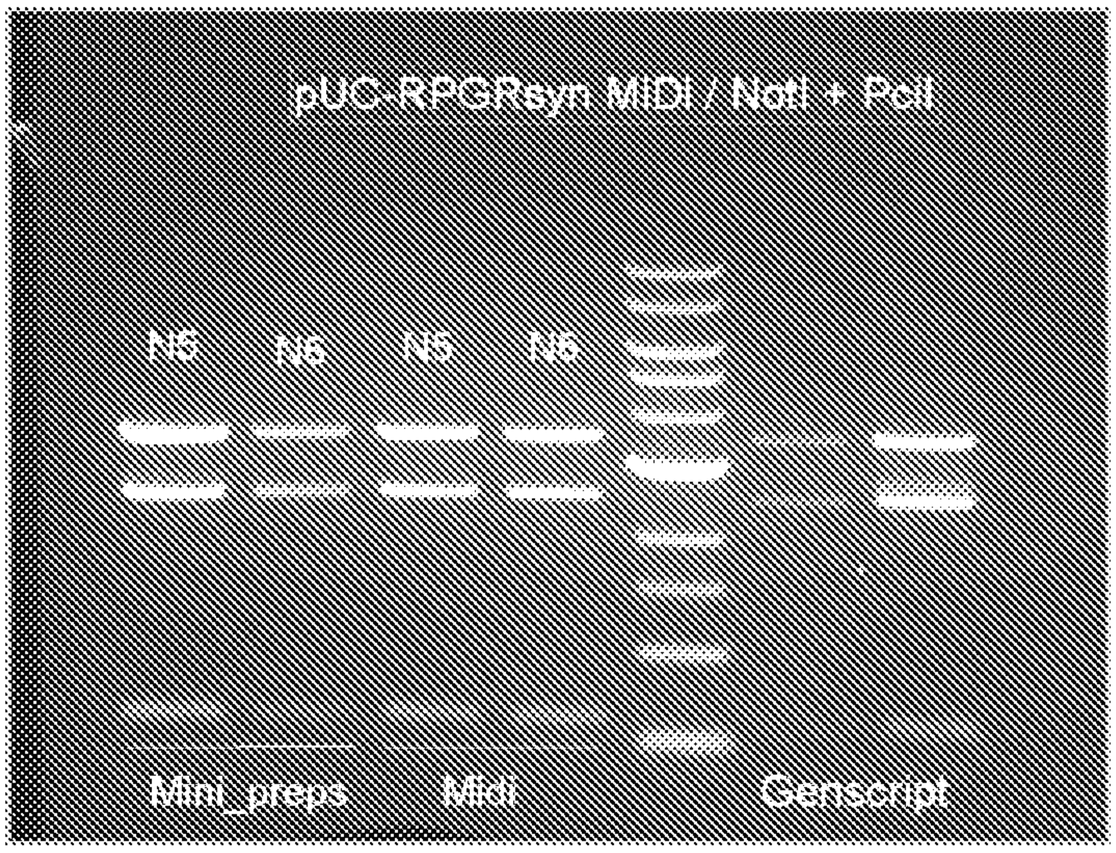
pUC-RPGRsyn MIDI / NotI + PciI
N5
N6
N5
N6
Mini_preps
Midi
Genscript

FIG. 3

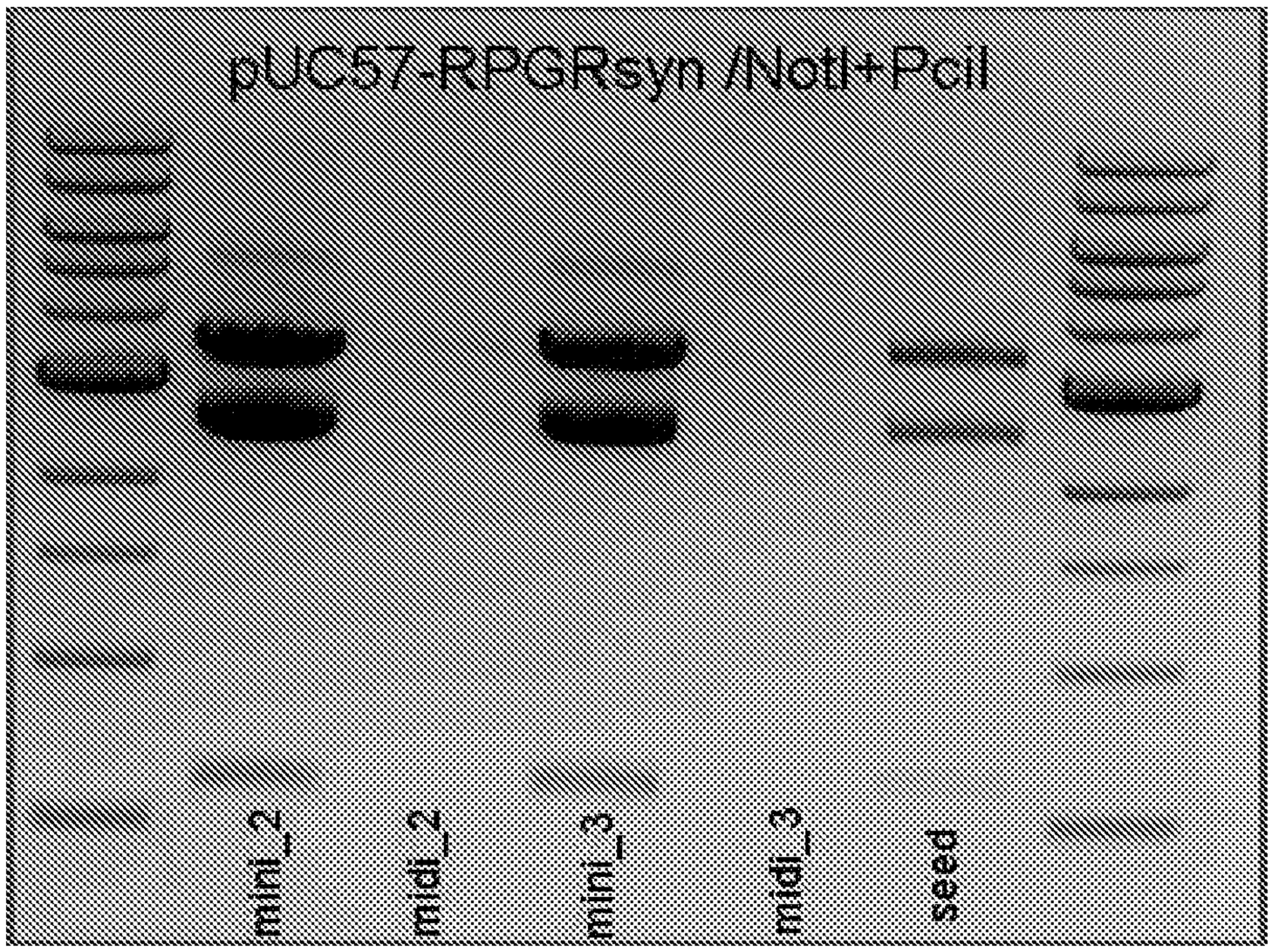
pUC57-RPGRsyn /NotI+PciI
mini_2
midi_2
mini_3
midi_3
seed

FIG. 4

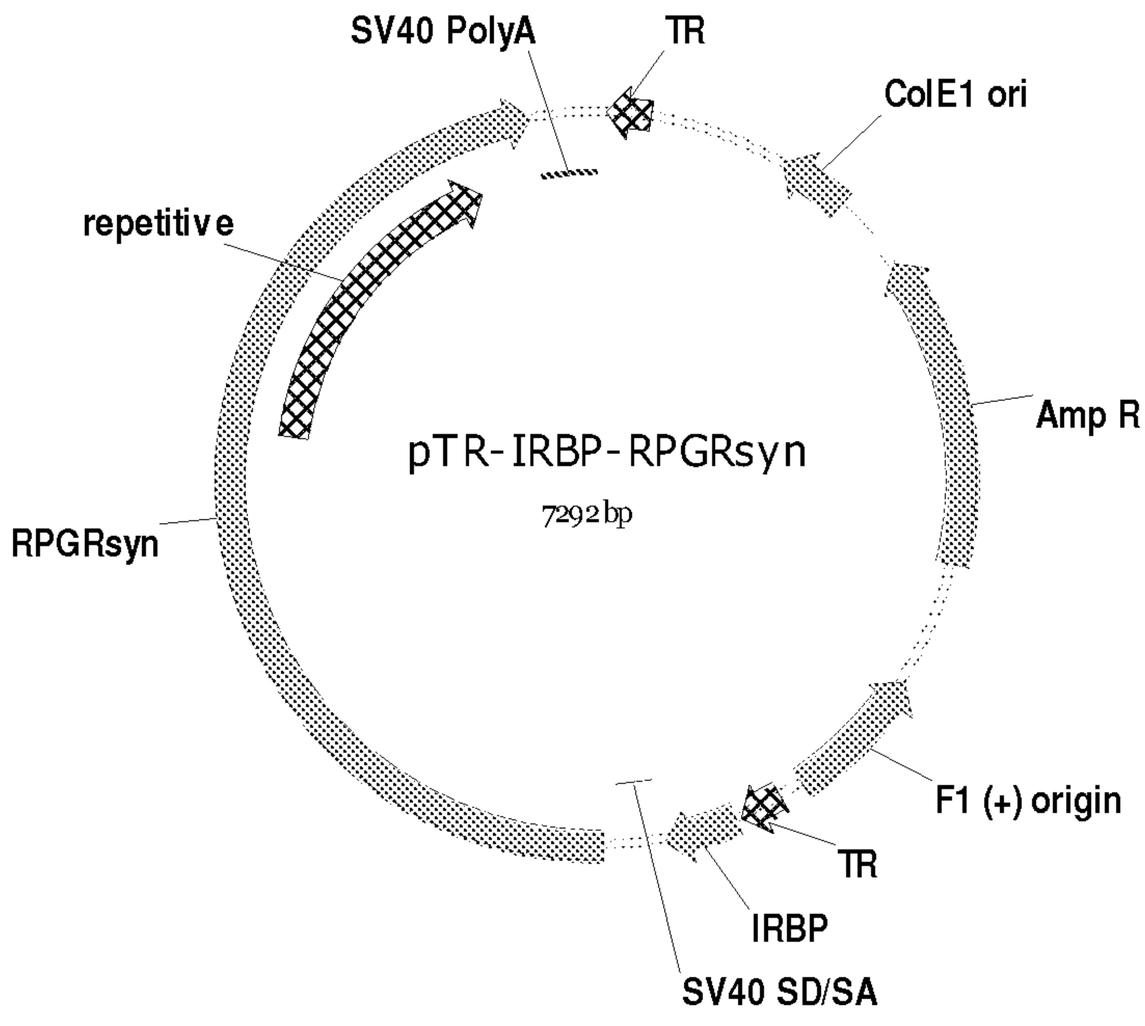
SV40 PolyA
TR
ColE1 ori
repetitive
Amp R
pTR-IRBP-RPGRsyn
7292 bp
RPGRsyn
F1 (+) origin
TR
IRBP
SV40 SD/SA

FIG. 5

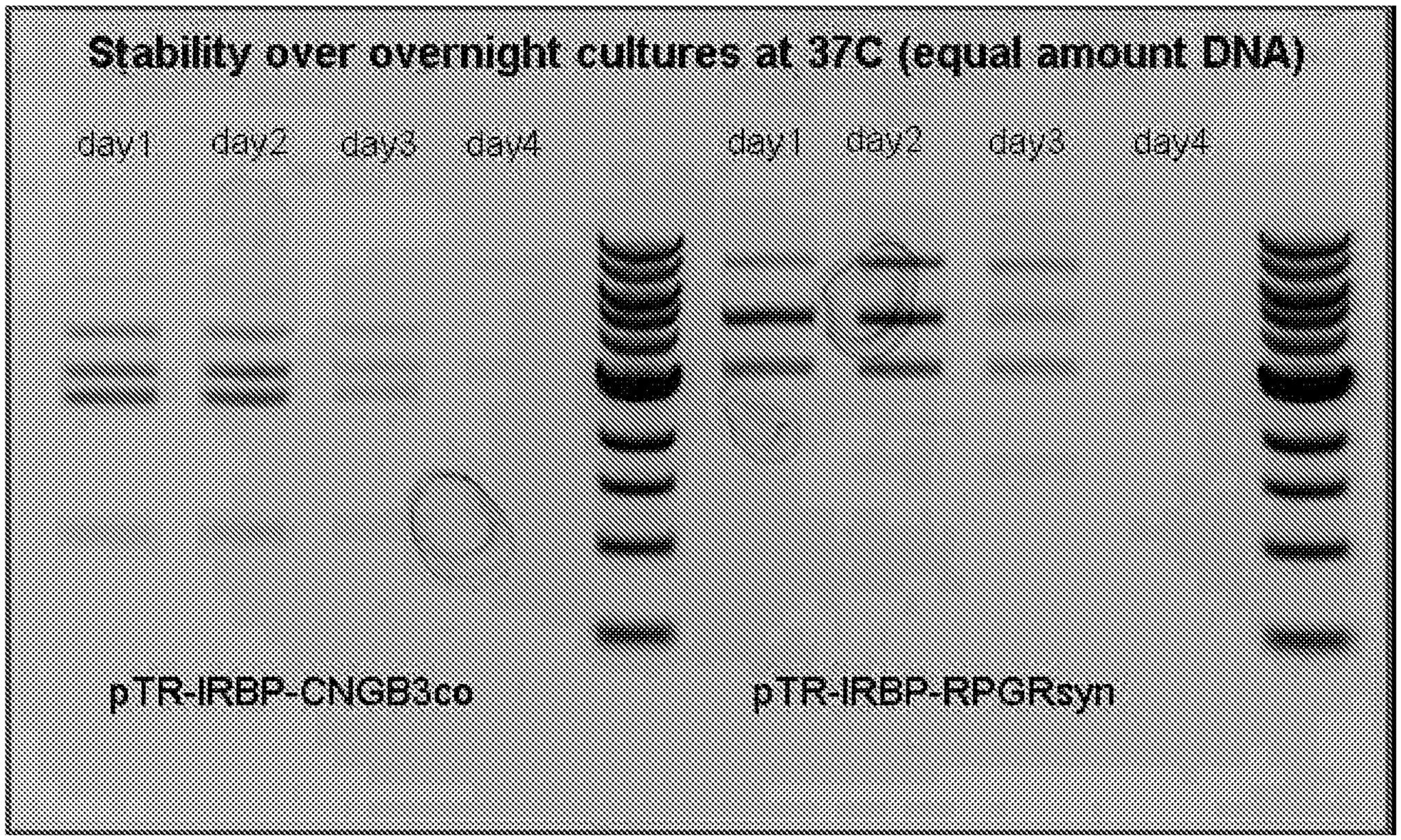

FIG. 6

```
Contig sequence      (1) ----------AGATCTGAATTCAGCACAGTGTCTG
pTR-IRBP-RPGRsyn  (3151) GCAGAGAGGGAGTGGCCAACCTCCTAGATCTGAATTCAGCACAGTGTCTG
                         3201                                3250
 Contig sequence    (26) GCATGTAGCAGGAACTAAAATAATGGCAGTGATTAATGTTATGATATGCA
pTR-IRBP-RPGRsyn  (3201) GCATGTAGCAGGAACTAAAATAATGGCAGTGATTAATGTTATGATATGCA
                         3251                                3300
 Contig sequence    (76) GACACAACACAGCAAGATAAGATGCAATGTACCTTCTGGGTCAAACCACC
pTR-IRBP-RPGRsyn  (3251) GACACAACACAGCAAGATAAGATGCAATGTACCTTCTGGGTCAAACCACC
                         3301                                3350
 Contig sequence   (126) CTGGCCACTCCTCCCCGATACCCAGGGTTGATGTGCTTGAATTAGACAGG
pTR-IRBP-RPGRsyn  (3301) CTGGCCACTCCTCCCCGATACCCAGGGTTGATGTGCTTGAATTAGACAGG
                         3351                                3400
 Contig sequence   (176) ATTAAAGGCTTACTGGAGCTGGAAGCCTTGCCCCAACTCAGGAGTTTAGC
pTR-IRBP-RPGRsyn  (3351) ATTAAAGGCTTACTGGAGCTGGAAGCCTTGCCCCAACTCAGGAGTTTAGC
                         3401                                3450
 Contig sequence   (226) CCCAGACCTTCTGTCCACCAGCTCTAGACTCGAGGAACTGAAAAACCAGA
pTR-IRBP-RPGRsyn  (3401) CCCAGACCTTCTGTCCACCAGCTCTAGACTCGAGGAACTGAAAAACCAGA
                         3451                                3500
 Contig sequence   (276) AAGTTAACTGGTAAGTTTAGTCTTTTTGTCTTTTATTTCAGGTCCCGGAT
pTR-IRBP-RPGRsyn  (3451) AAGTTAACTGGTAAGTTTAGTCTTTTTGTCTTTTATTTCAGGTCCCGGAT
                         3501                                3550
 Contig sequence   (326) CCGGTGGTGGTGCAAATCAAAGAACTGCTCCTCAGTGGATGTTGCCTTTA
pTR-IRBP-RPGRsyn  (3501) CCGGTGGTGGTGCAAATCAAAGAACTGCTCCTCAGTGGATGTTGCCTTTA
                         3551                                3600
 Contig sequence   (376) CTTCTAGGCCTGTACGGAAGTGTTACTTCTGCTCTAAAAGCTGCGGAATT
pTR-IRBP-RPGRsyn  (3551) CTTCTAGGCCTGTACGGAAGTGTTACTTCTGCTCTAAAAGCTGCGGAATT
                         3601                                3650
 Contig sequence   (426) GTACCCGCGGCCGCGCCACCATGAGAGAGCCAGAGGAGCTGATGCCAGAT
pTR-IRBP-RPGRsyn  (3601) GTACCCGCGGCCGCGCCACCATGAGAGAGCCAGAGGAGCTGATGCCAGAT
                         3651                                3700
```

```
 Contig sequence   (476) AGCGGAGCAGTGTTTACCTTCGGAAAGTCCAAGTTCGCAGAGAATAACCC
pTR-IRBP-RPGRsyn  (3651) AGCGGAGCAGTGTTTACCTTCGGAAAGTCCAAGTTCGCAGAGAATAACCC
                         3701                                          3750
 Contig sequence   (526) AGGAAAGTTCTGGTTTAAAAACGACGTGCCCGTCCACCTGTCTTGTGGCG
pTR-IRBP-RPGRsyn  (3701) AGGAAAGTTCTGGTTTAAAAACGACGTGCCCGTCCACCTGTCTTGTGGCG
                         3751                                          3800
 Contig sequence   (576) ATGAGCATAGTGCCGTGGTCACTGGGAACAATAAGCTGTATATGTTCGGG
pTR-IRBP-RPGRsyn  (3751) ATGAGCATAGTGCCGTGGTCACTGGGAACAATAAGCTGTATATGTTCGGG
                         3801                                          3850
 Contig sequence   (626) TCCAACAATTGGGGACAGCTGGGGCTGGGATCCAAATCTGCTATCTCTAA
pTR-IRBP-RPGRsyn  (3801) TCCAACAATTGGGGACAGCTGGGGCTGGGATCCAAATCTGCTATCTCTAA
                         3851                                          3900
 Contig sequence   (676) GCCAACCTGCGTGAAGGCACTGAAACCCGAGAAGGTCAAACTGGCCGCTT
pTR-IRBP-RPGRsyn  (3851) GCCAACCTGCGTGAAGGCACTGAAACCCGAGAAGGTCAAACTGGCCGCTT
                         3901                                          3950
 Contig sequence   (726) GTGGCAGAAACCACACTCTGGTGAGCACCGAGGGCGGGAATGTCTATGCC
pTR-IRBP-RPGRsyn  (3901) GTGGCAGAAACCACACTCTGGTGAGCACCGAGGGCGGGAATGTCTATGCC
                         3951                                          4000
 Contig sequence   (776) ACCGGAGGCAACAATGAGGGACAGCTGGGACTGGGGGACACTGAGGAAAG
pTR-IRBP-RPGRsyn  (3951) ACCGGAGGCAACAATGAGGGACAGCTGGGACTGGGGGACACTGAGGAAAG
                         4001                                          4050
 Contig sequence   (826) GAATACCTTTCACGTGATCTCCTTCTTTACATCTGAGCATAAGATCAAGC
pTR-IRBP-RPGRsyn  (4001) GAATACCTTTCACGTGATCTCCTTCTTTACATCTGAGCATAAGATCAAGC
                         4051                                          4100
 Contig sequence   (876) AGCTGAGCGCCGGCTCCAACACATCTGCAGCCCTGACTGAGGACGGGCGC
pTR-IRBP-RPGRsyn  (4051) AGCTGAGCGCCGGCTCCAACACATCTGCAGCCCTGACTGAGGACGGGCGC
                         4101                                          4150
 Contig sequence   (926) CTGTTCATGTGGGGAGATAATTCAGAGGGCCAGATTGGGCTGAAAAACGT
pTR-IRBP-RPGRsyn  (4101) CTGTTCATGTGGGGAGATAATTCAGAGGGCCAGATTGGGCTGAAAAACGT
                         4151                                          4200
 Contig sequence   (976) GAGCAACGTGTGCGTGCCTCAGCAGGTGACCATCGGAAAGCCAGTCAGTT
pTR-IRBP-RPGRsyn  (4151) GAGCAACGTGTGCGTGCCTCAGCAGGTGACCATCGGAAAGCCAGTCAGTT
                         4201                                          4250
 Contig sequence  (1026) GGATTTCATGTGGCTACTATCATAGCGCCTTCGTGACCACAGATGGCGAG
pTR-IRBP-RPGRsyn  (4201) GGATTTCATGTGGCTACTATCATAGCGCCTTCGTGACCACAGATGGCGAG
                         4251                                          4300
 Contig sequence  (1076) CTGTACGTCTTTGGGGAGCCCGAAAACGGAAAACTGGGCCTGCCTAACCA
pTR-IRBP-RPGRsyn  (4251) CTGTACGTCTTTGGGGAGCCCGAAAACGGAAAACTGGGCCTGCCTAACCA
                         4301                                          4350
 Contig sequence  (1126) GCTGCTGGGCAATCACCGGACACCCCAGCTGGTGTCCGAGATCCCTGAAA
pTR-IRBP-RPGRsyn  (4301) GCTGCTGGGCAATCACCGGACACCCCAGCTGGTGTCCGAGATCCCTGAAA
                         4351                                          4400
 Contig sequence  (1176) AAGTGATCCAGGTCGCCTGCGGGGGAGAGCATACAGTGGTCCTGACTGAG
pTR-IRBP-RPGRsyn  (4351) AAGTGATCCAGGTCGCCTGCGGGGGAGAGCATACAGTGGTCCTGACTGAG
                         4401                                          4450
 Contig sequence  (1226) AATGCCGTGTACACCTTCGGACTGGGCCAGTTTGGCCAGCTGGGGCTGGG
pTR-IRBP-RPGRsyn  (4401) AATGCCGTGTACACCTTCGGACTGGGCCAGTTTGGCCAGCTGGGGCTGGG
                         4451                                          4500
 Contig sequence  (1276) AACCTTCCTGTTTGAGACATCCGAACCAAAAGTGATCGAGAACATTCGCG
pTR-IRBP-RPGRsyn  (4451) AACCTTCCTGTTTGAGACATCCGAACCAAAAGTGATCGAGAACATTCGCG
                         4501                                          4550
 Contig sequence  (1326) ACCAGACTATCAGCTACATTTCCTGCGGAGAGAATCACACCGCACTGATC
pTR-IRBP-RPGRsyn  (4501) ACCAGACTATCAGCTACATTTCCTGCGGAGAGAATCACACCGCACTGATC
                         4551                                          4600
 Contig sequence  (1376) ACAGACATTGGCCTGATGTATACCTTTGGCGATGGGCGGCACGGGAAGCT
pTR-IRBP-RPGRsyn  (4551) ACAGACATTGGCCTGATGTATACCTTTGGCGATGGGCGGCACGGGAAGCT
                         4601                                          4650
 Contig sequence  (1426) GGGACTGGGCCTGGAGAACTTCACTAATCACTTCATCCCCACCCTGTGCT
pTR-IRBP-RPGRsyn  (4601) GGGACTGGGCCTGGAGAACTTCACTAATCACTTCATCCCCACCCTGTGCT
                         4651                                          4700
 Contig sequence  (1476) CTAACTTCCTGCGGTTCATCGTGAAACTGGTCGCTTGCGGCGGGTGTCAC
```

FIG. 6 (Continued)

```
pTR-IRBP-RPGRsyn  (4651) CTAACTTCCTGCGGTTCATCGTGAAACTGGTCGCTTGCGGCGGGTGTCAC
                         4701                                        4750
 Contig sequence  (1526) ATGGTGGTCTTCGCTGCACCTCATAGGGGCGTGGCTAAGGAGATCGAATT
pTR-IRBP-RPGRsyn  (4701) ATGGTGGTCTTCGCTGCACCTCATAGGGGCGTGGCTAAGGAGATCGAATT
                         4751                                        4800
 Contig sequence  (1576) TGACGAGATTAACGATACATGCCTGAGCGTGGCAACTTTCCTGCCATACA
pTR-IRBP-RPGRsyn  (4751) TGACGAGATTAACGATACATGCCTGAGCGTGGCAACTTTCCTGCCATACA
                         4801                                        4850
 Contig sequence  (1626) GCTCCCTGACTTCTGGCAATGTGCTGCAGAGAACCCTGAGTGCAAGGATG
pTR-IRBP-RPGRsyn  (4801) GCTCCCTGACTTCTGGCAATGTGCTGCAGAGAACCCTGAGTGCAAGGATG
                         4851                                        4900
 Contig sequence  (1676) CGGAGAAGGGAGAGGGAACGCTCTCCTGACAGTTTCTCAATGCGACGAAC
pTR-IRBP-RPGRsyn  (4851) CGGAGAAGGGAGAGGGAACGCTCTCCTGACAGTTTCTCAATGCGACGAAC
                         4901                                        4950
 Contig sequence  (1726) CCTGCCACCTATCGAGGGGACACTGGGACTGAGTGCCTGCTTCCTGCCTA
pTR-IRBP-RPGRsyn  (4901) CCTGCCACCTATCGAGGGGACACTGGGACTGAGTGCCTGCTTCCTGCCTA
                         4951                                        5000
 Contig sequence  (1776) ACTCAGTGTTTCCACGATGTAGCGAGCGGAATCTGCAGGAGTCTGTCCTG
pTR-IRBP-RPGRsyn  (4951) ACTCAGTGTTTCCACGATGTAGCGAGCGGAATCTGCAGGAGTCTGTCCTG
                         5001                                        5050
 Contig sequence  (1826) AGTGAGCAGGATCTGATGCAGCCAGAGGAACCCGACTACCTGCTGGATGA
pTR-IRBP-RPGRsyn  (5001) AGTGAGCAGGATCTGATGCAGCCAGAGGAACCCGACTACCTGCTGGATGA
                         5051                                        5100
 Contig sequence  (1876) GATGACCAAGGAGGCCGAAATCGACAACTCTAGTACAGTGGAGTCCCTGG
pTR-IRBP-RPGRsyn  (5051) GATGACCAAGGAGGCCGAAATCGACAACTCTAGTACAGTGGAGTCCCTGG
                         5101                                        5150
 Contig sequence  (1926) GCGAGACTACCGATATCCTGAATATGACACACATTATGTCACTGAACAGC
pTR-IRBP-RPGRsyn  (5101) GCGAGACTACCGATATCCTGAATATGACACACATTATGTCACTGAACAGC
                         5151                                        5200
 Contig sequence  (1976) AATGAGAAGAGTCTGAAACTGTCACCAGTGCAGAAGCAGAAGAAACAGCA
pTR-IRBP-RPGRsyn  (5151) AATGAGAAGAGTCTGAAACTGTCACCAGTGCAGAAGCAGAAGAAACAGCA
                         5201                                        5250
 Contig sequence  (2026) GACTATTGGCGAGCTGACTCAGGACACCGCCCTGACAGAGAACGACGATA
pTR-IRBP-RPGRsyn  (5201) GACTATTGGCGAGCTGACTCAGGACACCGCCCTGACAGAGAACGACGATA
                         5251                                        5300
 Contig sequence  (2076) GCGATGAGTATGAGGAAATGTCCGAGATGAAGGAAGGCAAAGCTTGTAAG
pTR-IRBP-RPGRsyn  (5251) GCGATGAGTATGAGGAAATGTCCGAGATGAAGGAAGGCAAAGCTTGTAAG
                         5301                                        5350
 Contig sequence  (2126) CAGCATGTGAGTCAGGGGATCTTCATGACACAGCCAGCCACAACTATTGA
pTR-IRBP-RPGRsyn  (5301) CAGCATGTGAGTCAGGGGATCTTCATGACACAGCCAGCCACAACTATTGA
                         5351                                        5400
 Contig sequence  (2176) GGCTTTTTCAGACGAGGAAGTGGAGATCCCCGAGGAAAAAGAGGGCGCAG
pTR-IRBP-RPGRsyn  (5351) GGCTTTTTCAGACGAGGAAGTGGAGATCCCCGAGGAAAAAGAGGGCGCAG
                         5401                                        5450
 Contig sequence  (2226) AAGATTCCAAGGGGAATGGAATTGAGGAACAGGAGGTGGAAGCCAACGAG
pTR-IRBP-RPGRsyn  (5401) AAGATTCCAAGGGGAATGGAATTGAGGAACAGGAGGTGGAAGCCAACGAG
                         5451                                        5500
 Contig sequence  (2276) GAAAATGTGAAAGTCCACGGAGGCAGGAAGGAGAAAACAGAAATCCTGTC
pTR-IRBP-RPGRsyn  (5451) GAAAATGTGAAAGTCCACGGAGGCAGGAAGGAGAAAACAGAAATCCTGTC
                         5501                                        5550
 Contig sequence  (2326) TGACGATCTGACTGACAAGGCCGAGGTGTCCGAAGGCAAGGCAAAATCTG
pTR-IRBP-RPGRsyn  (5501) TGACGATCTGACTGACAAGGCCGAGGTGTCCGAAGGCAAGGCAAAATCTG
                         5551                                        5600
 Contig sequence  (2376) TCGGAGAGGCAGAAGACGGACCAGAGGGACGAGGGGATGGAACCTGCGAG
pTR-IRBP-RPGRsyn  (5551) TCGGAGAGGCAGAAGACGGACCAGAGGGACGAGGGGATGGAACCTGCGAG
                         5601                                        5650
 Contig sequence  (2426) GAAGGCTCAAGCGGGGCTGAGCATTGGCAGGACGAGGAACGAGAGAAGGG
pTR-IRBP-RPGRsyn  (5601) GAAGGCTCAAGCGGGGCTGAGCATTGGCAGGACGAGGAACGAGAGAAGGG
                         5651                                        5700
 Contig sequence  (2476) CGAAAAGGATAAAGGCCGCGGGGAGATGGAACGACCTGGAGAGGGCGAAA
pTR-IRBP-RPGRsyn  (5651) CGAAAAGGATAAAGGCCGCGGGGAGATGGAACGACCTGGAGAGGGCGAAA
```

FIG. 6 (Continued)

```
                        5701                                                5750
 Contig sequence  (2526) AAGAGCTGGCAGAGAAGGAGGAATGGAAGAAAAGGGACGGCGAGGAACAG
pTR-IRBP-RPGRsyn  (5701) AAGAGCTGGCAGAGAAGGAGGAATGGAAGAAAAGGGACGGCGAGGAACAG
                        5751                                                5800
 Contig sequence  (2576) GAGCAGAAAGAAAGGGAGCAGGGCCACCAGAAGGAGCGCAACCAGGAGAT
pTR-IRBP-RPGRsyn  (5751) GAGCAGAAAGAAAGGGAGCAGGGCCACCAGAAGGAGCGCAACCAGGAGAT
                        5801                                                5850
 Contig sequence  (2626) GGAAGAGGGCGGCGAGGAAGAGCATGGCGAGGGAGAAGAGGAAGAGGGCG
pTR-IRBP-RPGRsyn  (5801) GGAAGAGGGCGGCGAGGAAGAGCATGGCGAGGGAGAAGAGGAAGAGGGCG
                        5851                                                5900
 Contig sequence  (2676) ATAGAGAAGAGGAAGAGGAAAAAGAAGGCGAAGGGAAGGAGGAAGGAGAG
pTR-IRBP-RPGRsyn  (5851) ATAGAGAAGAGGAAGAGGAAAAAGAAGGCGAAGGGAAGGAGGAAGGAGAG
                        5901                                                5950
 Contig sequence  (2726) GGCGAGGAAGTGGAAGGCGAGAGGGAAAAGGAGGAAGGAGAACGGAAGAA
pTR-IRBP-RPGRsyn  (5901) GGCGAGGAAGTGGAAGGCGAGAGGGAAAAGGAGGAAGGAGAACGGAAGAA
                        5951                                                6000
 Contig sequence  (2776) AGAGGAAAGAGCCGGCAAAGAGGAAAAGGGCGAGGAAGAGGGCGATCAGG
pTR-IRBP-RPGRsyn  (5951) AGAGGAAAGAGCCGGCAAAGAGGAAAAGGGCGAGGAAGAGGGCGATCAGG
                        6001                                                6050
 Contig sequence  (2826) GCGAAGGCGAGGAGGAAGAGACCGAGGGCCGCGGGGAAGAGAAAGAGGAG
pTR-IRBP-RPGRsyn  (6001) GCGAAGGCGAGGAGGAAGAGACCGAGGGCCGCGGGGAAGAGAAAGAGGAG
                        6051                                                6100
 Contig sequence  (2876) GGAGGAGAGGTGGAGGGCGGAGAGGTCGAAGAGGGAAAGGGCGAGCGCGA
pTR-IRBP-RPGRsyn  (6051) GGAGGAGAGGTGGAGGGCGGAGAGGTCGAAGAGGGAAAGGGCGAGCGCGA
                        6101                                                6150
 Contig sequence  (2926) AGAGGAAGAGGAAGAGGGCGAGGGCGAGGAAGAAGAGGGCGAGGGGGAAG
pTR-IRBP-RPGRsyn  (6101) AGAGGAAGAGGAAGAGGGCGAGGGCGAGGAAGAAGAGGGCGAGGGGGAAG
                        6151                                                6200
 Contig sequence  (2976) AAGAGGAGGGAGAGGGCGAAGAGGAAGAGGGGGAGGGAAAGGGCGAAGAG
pTR-IRBP-RPGRsyn  (6151) AAGAGGAGGGAGAGGGCGAAGAGGAAGAGGGGGAGGGAAAGGGCGAAGAG
                        6201                                                6250
 Contig sequence  (3026) GAAGGAGAGGAAGGGGAGGGAGAGGAAGAGGGGGAGGAGGGCGAGGGGGA
pTR-IRBP-RPGRsyn  (6201) GAAGGAGAGGAAGGGGAGGGAGAGGAAGAGGGGGAGGAGGGCGAGGGGGA
                        6251                                                6300
 Contig sequence  (3076) AGGCGAGGAGGAAGAAGGAGAGGGGGAAGGCGAAGAGGAAGGCGAGGGGG
pTR-IRBP-RPGRsyn  (6251) AGGCGAGGAGGAAGAAGGAGAGGGGGAAGGCGAAGAGGAAGGCGAGGGGG
                        6301                                                6350
 Contig sequence  (3126) AAGGAGAGGAGGAAGAAGGGGAAGGCGAAGGCGAAGAGGAGGGAGAAGGA
pTR-IRBP-RPGRsyn  (6301) AAGGAGAGGAGGAAGAAGGGGAAGGCGAAGGCGAAGAGGAGGGAGAAGGA
                        6351                                                6400
 Contig sequence  (3176) GAGGGGAGGAAGAGGAAGGAGAAGGGAAGGGCGAGGAGGAAGGCGAAGA
pTR-IRBP-RPGRsyn  (6351) GAGGGGAGGAAGAGGAAGGAGAAGGGAAGGGCGAGGAGGAAGGCGAAGA
                        6401                                                6450
 Contig sequence  (3226) GGGAGAGGGGGAAGGCGAGGAAGAGGAAGGCGAGGGCGAAGGAGAGGACG
pTR-IRBP-RPGRsyn  (6401) GGGAGAGGGGGAAGGCGAGGAAGAGGAAGGCGAGGGCGAAGGAGAGGACG
                        6451                                                6500
 Contig sequence  (3276) GCGAGGGCGAGGGAGAAGAGGAGGAAGGGGAATGGGAAGGCGAAGAAGAG
pTR-IRBP-RPGRsyn  (6451) GCGAGGGCGAGGGAGAAGAGGAGGAAGGGGAATGGGAAGGCGAAGAAGAG
                        6501                                                6550
 Contig sequence  (3326) GAAGGCGAAGGCGAAGGCGAAGAAGAGGGCGAAGGGGAGGGCGAGGAGGG
pTR-IRBP-RPGRsyn  (6501) GAAGGCGAAGGCGAAGGCGAAGAAGAGGGCGAAGGGGAGGGCGAGGAGGG
                        6551                                                6600
 Contig sequence  (3376) CGAAGGCGAAGGGGAGGAAGAGGAAGGCGAAGGAGAAGGCGAGGAAGAAG
pTR-IRBP-RPGRsyn  (6551) CGAAGGCGAAGGGGAGGAAGAGGAAGGCGAAGGAGAAGGCGAGGAAGAAG
                        6601                                                6650
 Contig sequence  (3426) AGGGAGAGGAGGAAGGCGAGGAGGAAGGAGAGGGGGAGGAGGAGGGAGAA
pTR-IRBP-RPGRsyn  (6601) AGGGAGAGGAGGAAGGCGAGGAGGAAGGAGAGGGGGAGGAGGAGGGAGAA
                        6651                                                6700
 Contig sequence  (3476) GGCGAGGGCGAAGAAGAAGAAGAGGGAGAAGTGGAGGGCGAAGTCGAGGG
pTR-IRBP-RPGRsyn  (6651) GGCGAGGGCGAAGAAGAAGAAGAGGGAGAAGTGGAGGGCGAAGTCGAGGG
                        6701                                                6750
```

FIG. 6 (Continued)

```
 Contig sequence   (3526) GGAGGAGGGAGAAGGGGAAGGGGAGGAAGAAGAGGGCGAAGAAGAAGGCG
pTR-IRBP-RPGRsyn   (6701) GGAGGAGGGAGAAGGGGAAGGGGAGGAAGAAGAGGGCGAAGAAGAAGGCG
                          6751                                          6800
 Contig sequence   (3576) AGGAAAGAGAAAAAGAGGGAGAAGGCGAGGAAAACCGGAGAAATAGGGAA
pTR-IRBP-RPGRsyn   (6751) AGGAAAGAGAAAAAGAGGGAGAAGGCGAGGAAAACCGGAGAAATAGGGAA
                          6801                                          6850
 Contig sequence   (3626) GAGGAGGAAGAGGAAGAGGGAAAGTACCAGGAGACAGGCGAAGAGGAAAA
pTR-IRBP-RPGRsyn   (6801) GAGGAGGAAGAGGAAGAGGGAAAGTACCAGGAGACAGGCGAAGAGGAAAA
                          6851                                          6900
 Contig sequence   (3676) CGAGCGGCAGGATGGCGAGGAATATAAGAAAGTGAGCAAGATCAAAGGAT
pTR-IRBP-RPGRsyn   (6851) CGAGCGGCAGGATGGCGAGGAATATAAGAAAGTGAGCAAGATCAAAGGAT
                          6901                                          6950
 Contig sequence   (3726) CCGTCAAGTACGGCAAGCACAAAACCTATCAGAAGAAAAGCGTGACCAAC
pTR-IRBP-RPGRsyn   (6901) CCGTCAAGTACGGCAAGCACAAAACCTATCAGAAGAAAAGCGTGACCAAC
                          6951                                          7000
 Contig sequence   (3776) ACACAGGGGAATGGAAAAGAGCAGCGAAGTAAAATGCCTGTGCAGTCAAA
pTR-IRBP-RPGRsyn   (6951) ACACAGGGGAATGGAAAAGAGCAGCGAAGTAAAATGCCTGTGCAGTCAAA
                          7001                                          7050
 Contig sequence   (3826) ACGGCTGCTGAAGAATGGCCCAAGCGGGTCTAAAAAATTCTGGAACAATG
pTR-IRBP-RPGRsyn   (7001) ACGGCTGCTGAAGAATGGCCCAAGCGGGTCTAAAAAATTCTGGAACAATG
                          7051                                          7100
 Contig sequence   (3876) TCCTGCCACACTATCTGGAACTGAAGTAAGCGGCCGCGCGGATCCAGACA
pTR-IRBP-RPGRsyn   (7051) TCCTGCCACACTATCTGGAACTGAAGTAAGCGGCCGCGCGGATCCAGACA
                          7101                                          7150
 Contig sequence   (3926) TGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGA
pTR-IRBP-RPGRsyn   (7101) TGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGA
                          7151                                          7200
 Contig sequence   (3976) AAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAAC
pTR-IRBP-RPGRsyn   (7151) AAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAAC
                          7201                                          7250
 Contig sequence   (4026) CATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTA
pTR-IRBP-RPGRsyn   (7201) CATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTA
                          7251                                  7292
 Contig sequence   (4076) TGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTT----------
pTR-IRBP-RPGRsyn   (7251) TGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAGCATG
```

FIG. 6 sequence alignment of the assembled contig obtained from sequencing of pTR-IRBP-RPGRsyn plasmid DNA to the reference pTR-IRBP-RPGRsyn sequence. The start and stop codons of rht RPGRsyn cDNA were bold and underlined.

FIG. 6 (Continued)

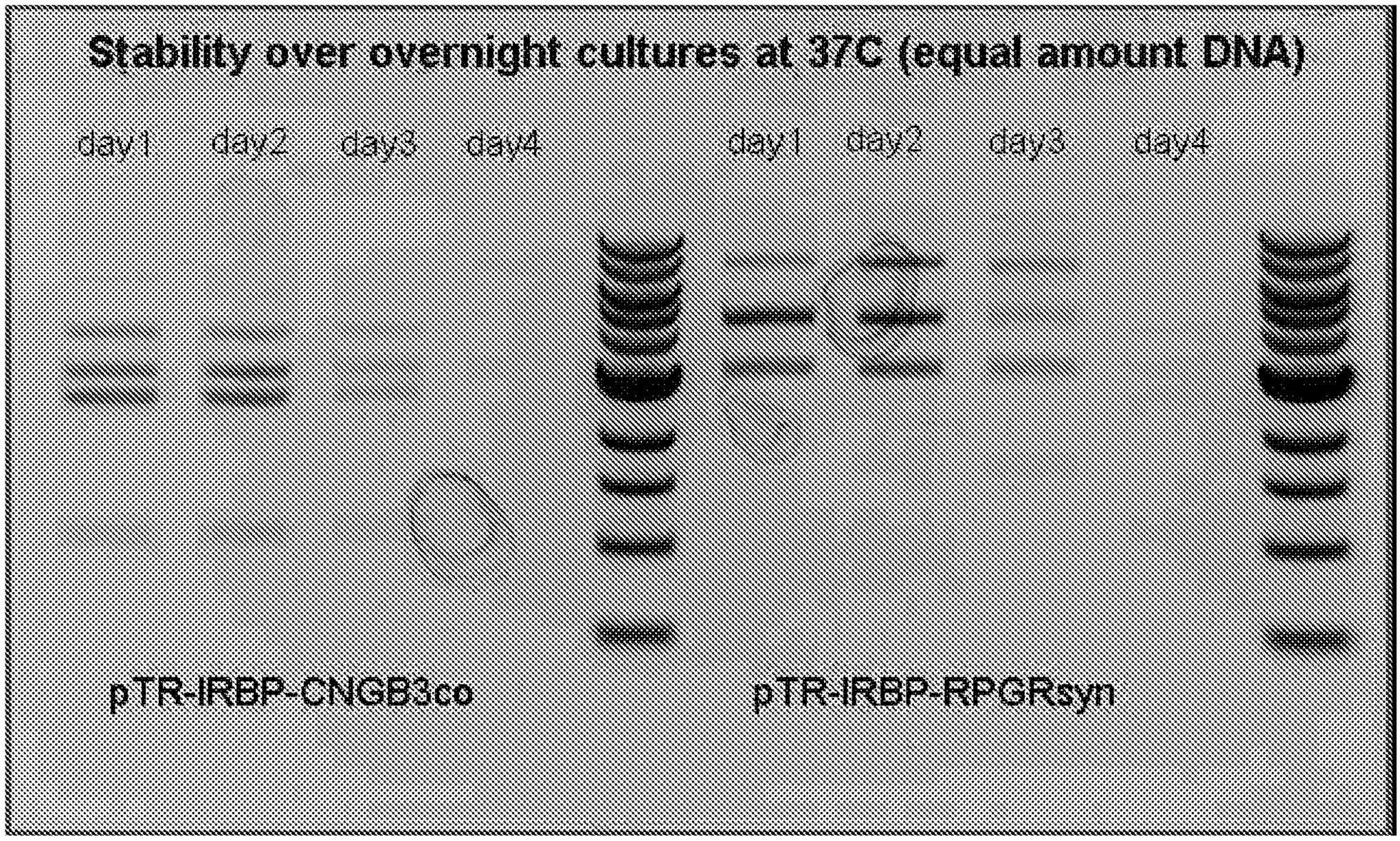
Stability over overnight cultures at 37C (equal amount DNA)
day1
day2
day3
day4
day1
day2
day3
day4
pTR-IRBP-CNGB3co
pTR-IRBP-RPGRsyn

FIG. 7