US 20250257365A1

(19) United States
(12) Patent Application Publication
Demirer et al.
(10) Pub. No.: US 2025/0257365 A1
(43) Pub. Date: Aug. 14, 2025

(54) **TARGETED DNA INTEGRATION IN PLANTS BY CRISPR-ASSOCIATED TRANSPOSASES (CASTS)**

(71) Applicant: **California Institute of Technology**, Pasadena, CA (US)

(72) Inventors: **Gozde Demirer**, Altadena, CA (US); **Yunqing Wang**, Pasadena, CA (US); **Kimberley T. Muchenje**, Pasadena, CA (US)

(21) Appl. No.: **19/048,891**

(22) Filed: **Feb. 8, 2025**

Related U.S. Application Data

(60) Provisional application No. 63/551,713, filed on Feb. 9, 2024.

Publication Classification

(51) **Int. Cl.**

C12N 15/82	(2006.01)
C12N 9/12	(2006.01)
C12N 9/22	(2006.01)
C12N 15/10	(2006.01)
C12N 15/11	(2006.01)
C12Q 1/6869	(2018.01)
C12Q 1/6895	(2018.01)

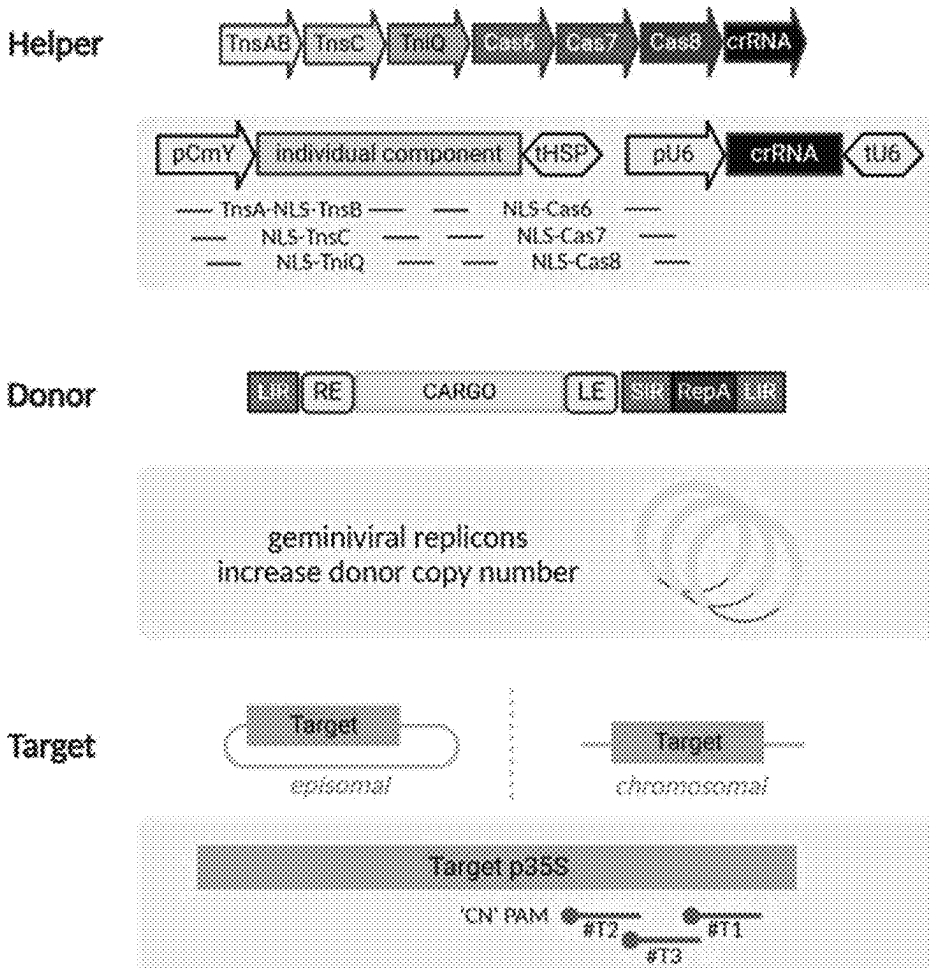
(52) **U.S. Cl.**

CPC ***C12N 15/8213*** (2013.01); ***C12N 9/1241*** (2013.01); ***C12N 9/22*** (2013.01); ***C12N 15/1086*** (2013.01); ***C12N 15/11*** (2013.01); ***C12N 15/8205*** (2013.01); ***C12N 15/8242*** (2013.01); ***C12N 15/8261*** (2013.01); ***C12Q 1/6869*** (2013.01); ***C12Q 1/6895*** (2013.01); *C07K 2319/09* (2013.01); *C07K 2319/40* (2013.01); *C12N 2310/20* (2017.05); *C12Q 2600/13* (2013.01)

(57) **ABSTRACT**

Disclosed herein include methods, compositions, and kits suitable for use in integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the methods, systems and nucleic acid compositions comprise polynucleotides encoding and RNA-guided DNA binding complex, a transposition complex, for insertion of cargo sequence using transposition.

Specification includes a Sequence Listing.



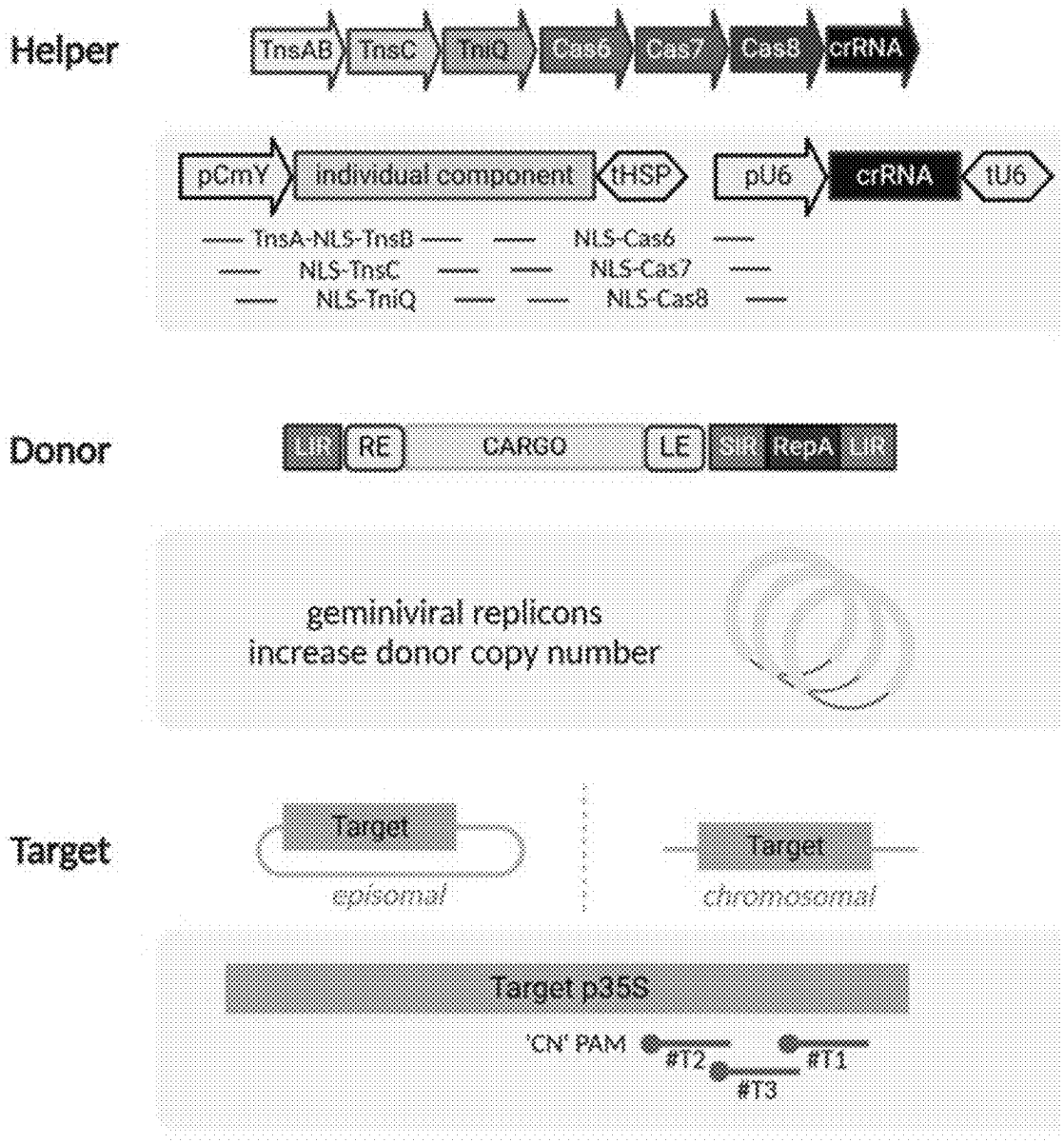

Helper
TnsAB
TnsC
TniQ
Cas6
Cas7
Cas8
crRNA
pCmY
individual component
tHSP
pU6
crRNA
tU6
TnsA-NLS-TnsB
NLS-Cas6
NLS-TnsC
NLS-Cas7
NLS-TniQ
NLS-Cas8
Donor
LIR
RE
CARGO
LE
SIR
RepA
LIR
geminiviral replicons
increase donor copy number
Target
Target
episomal
Target
chromosomal
Target p35S
'CN' PAM
#T2
#T1
#T3

FIG. 1

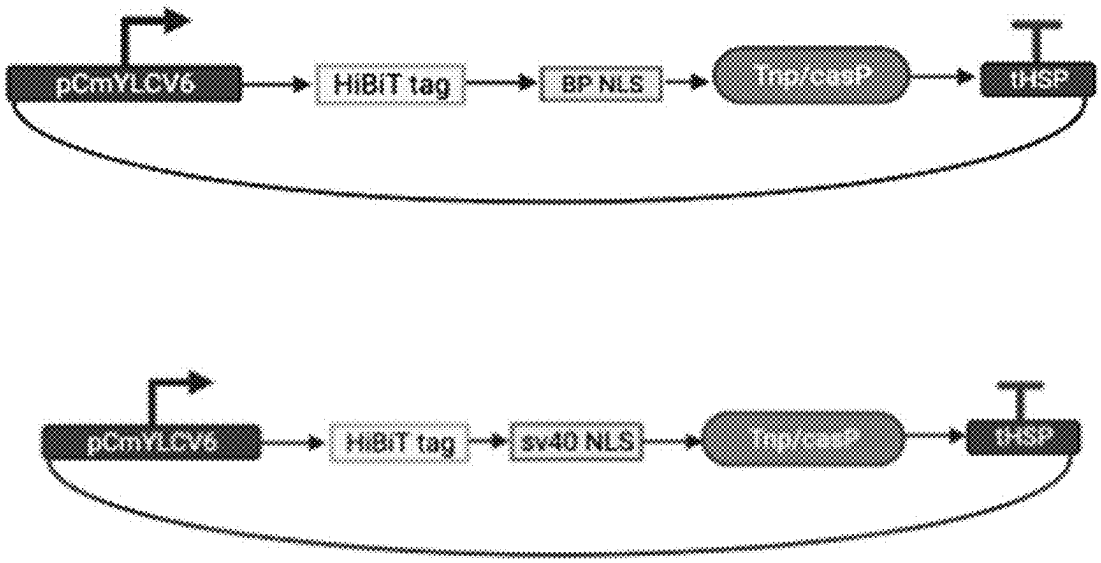
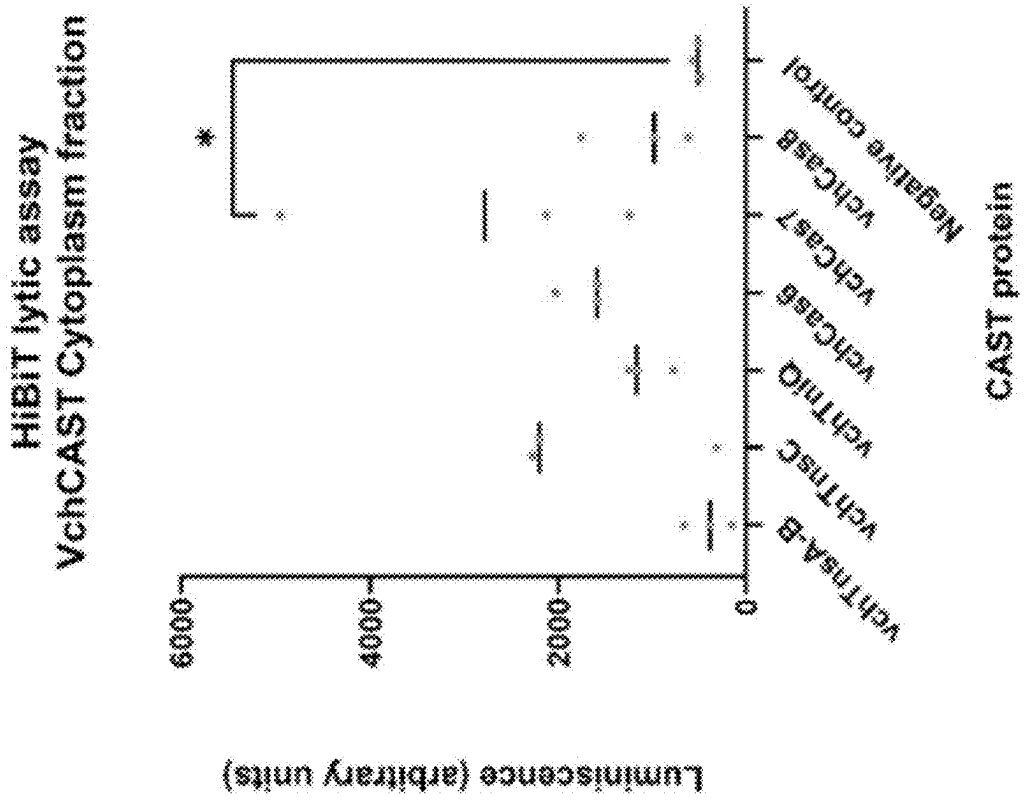

pCmYLCV6
HiBiT tag
BP NLS
Inp/cadP
tHSP
pCmYLCV6
HiBiT tag
sv40 NLS
Inp/cadP
tHSP

FIG. 2A


HiBiT lytic assay
VchCAST Cytoplasm fraction
Luminiscence (arbitrary units)
6000
4000
2000
0
*
VchTnsA-B
VchTnsC
VchTnsD
VchCasB
VchCas7
VchCas8
Negative control
CAST protein

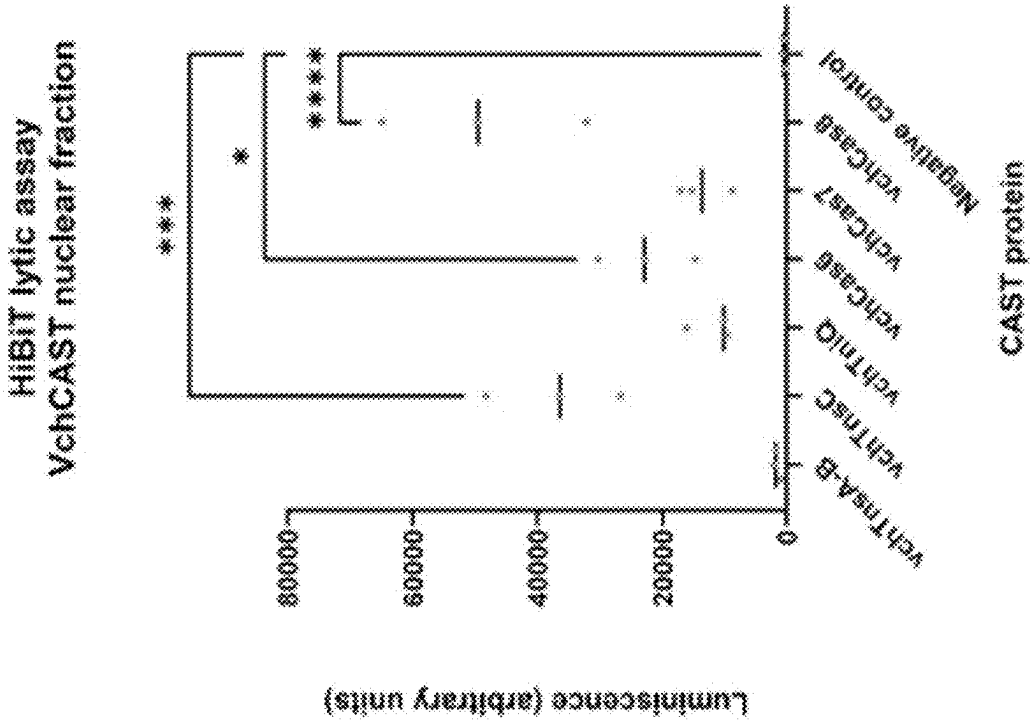
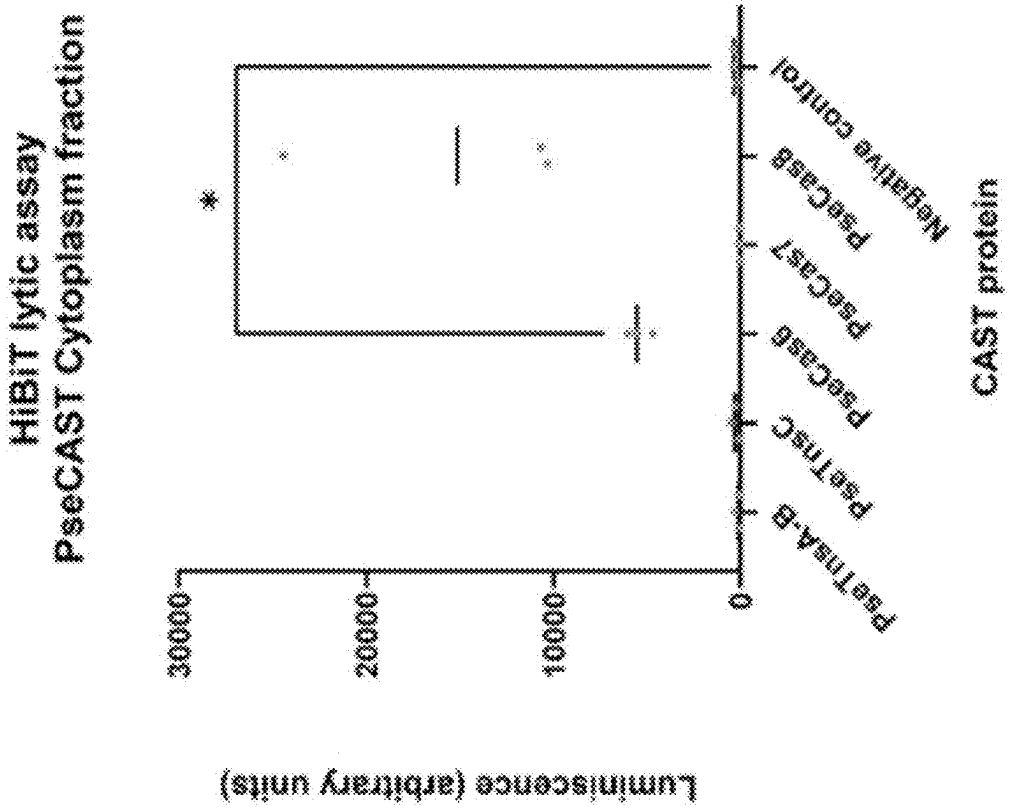

HiBiT lytic assay
VchCAST nuclear fraction
Luminiscence (arbitrary units)
80000
60000
40000
20000
0
***
*
****
VchTnsA-B
VchTnsC
VchTnsD
VchCasB
VchCas7
VchCas8
Negative control
CAST protein

FIG. 2B


HiBiT lytic assay
PseCAST Cytoplasm fraction
Luminiscence (arbitrary units)
30000
20000
10000
0
*
PseTnsA-B
PseTnsC
PseCas9
PseCas7
PseCas6
Negative control
CAST protein

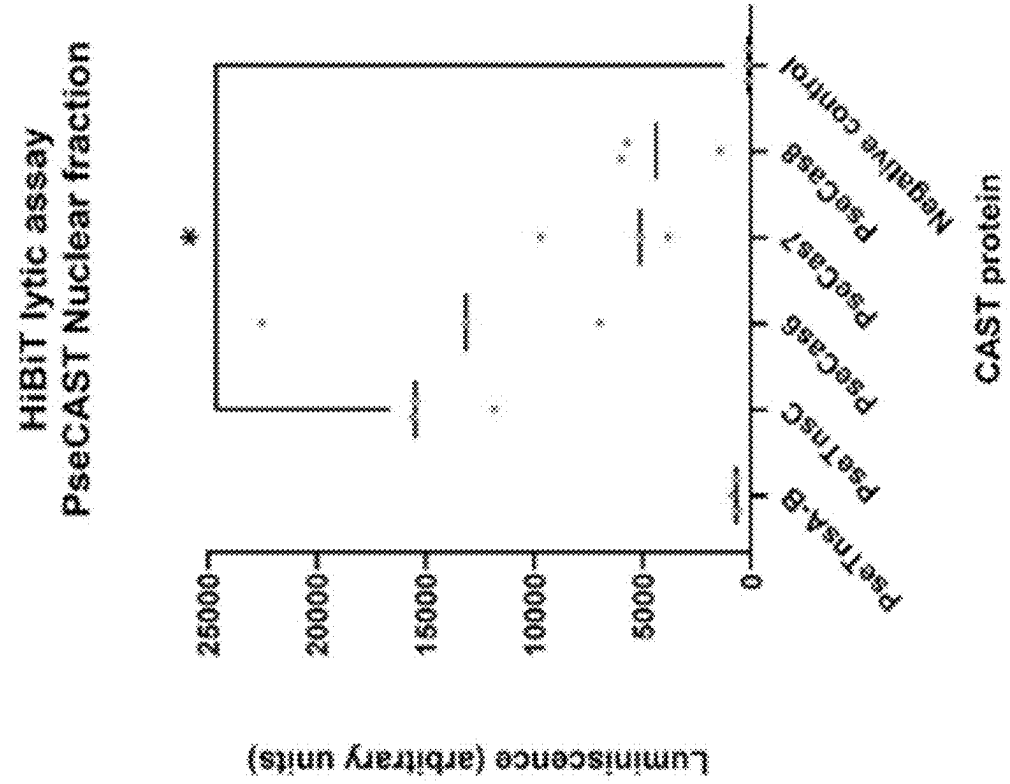

HiBiT lytic assay
PseCAST Nuclear fraction
Luminiscence (arbitrary units)
25000
20000
15000
10000
5000
0
*
PseTnsA-B
PseTnsC
PseCas9
PseCas7
PseCas6
Negative control
CAST protein

FIG. 2C

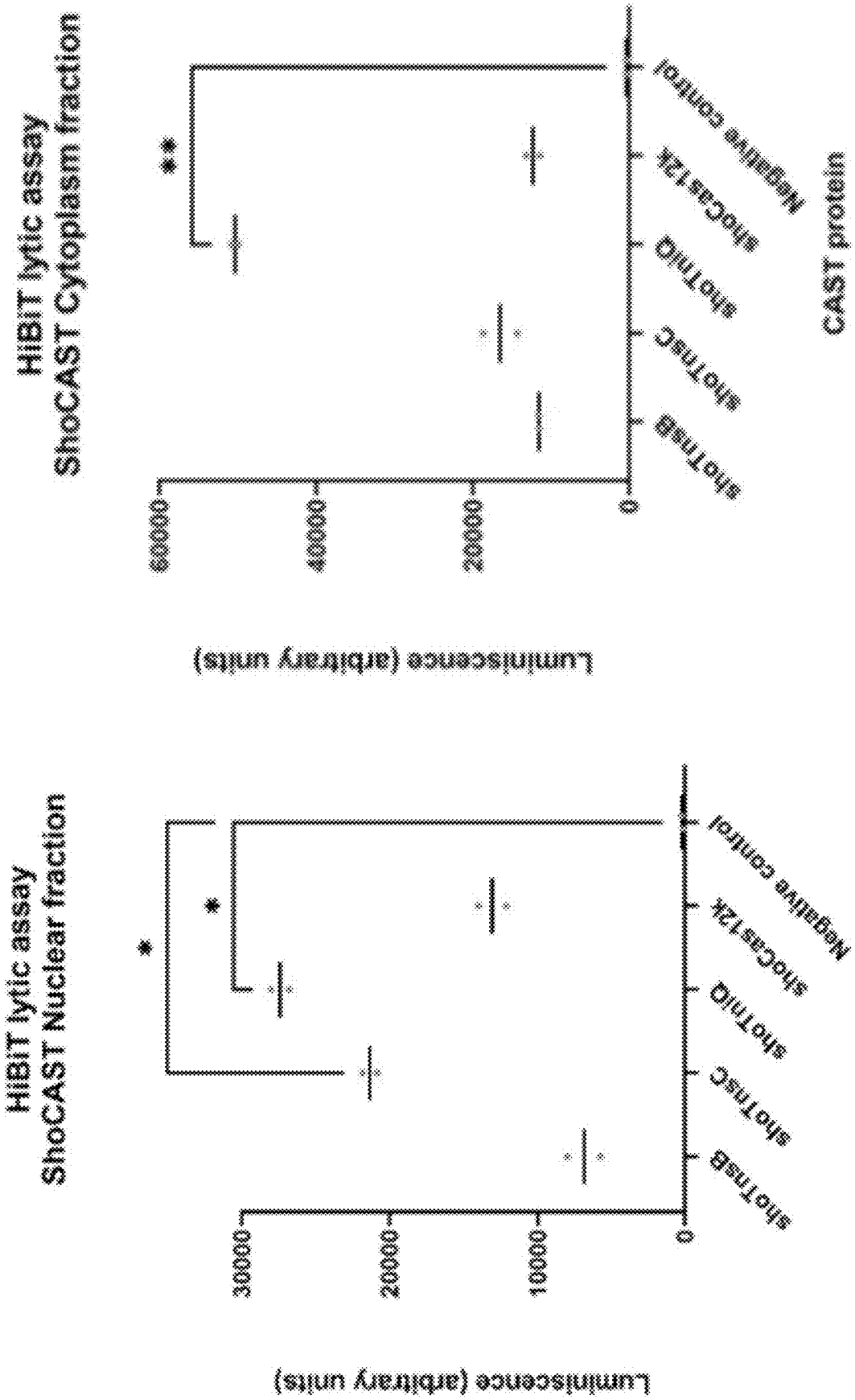

HiBiT lytic assay
ShoCAST Cytoplasm fraction
Luminiscence (arbitrary units)
60000
40000
20000
0
**
shoTnsB
shoTnsC
shoTnsQ
shoCas12k
Negative control
CAST protein
HiBiT lytic assay
ShoCAST Nuclear fraction
Luminiscence (arbitrary units)
30000
20000
10000
0
*
*
shoTnsB
shoTnsC
shoTnsQ
shoCas12k
Negative control
CAST protein

FIG. 2D

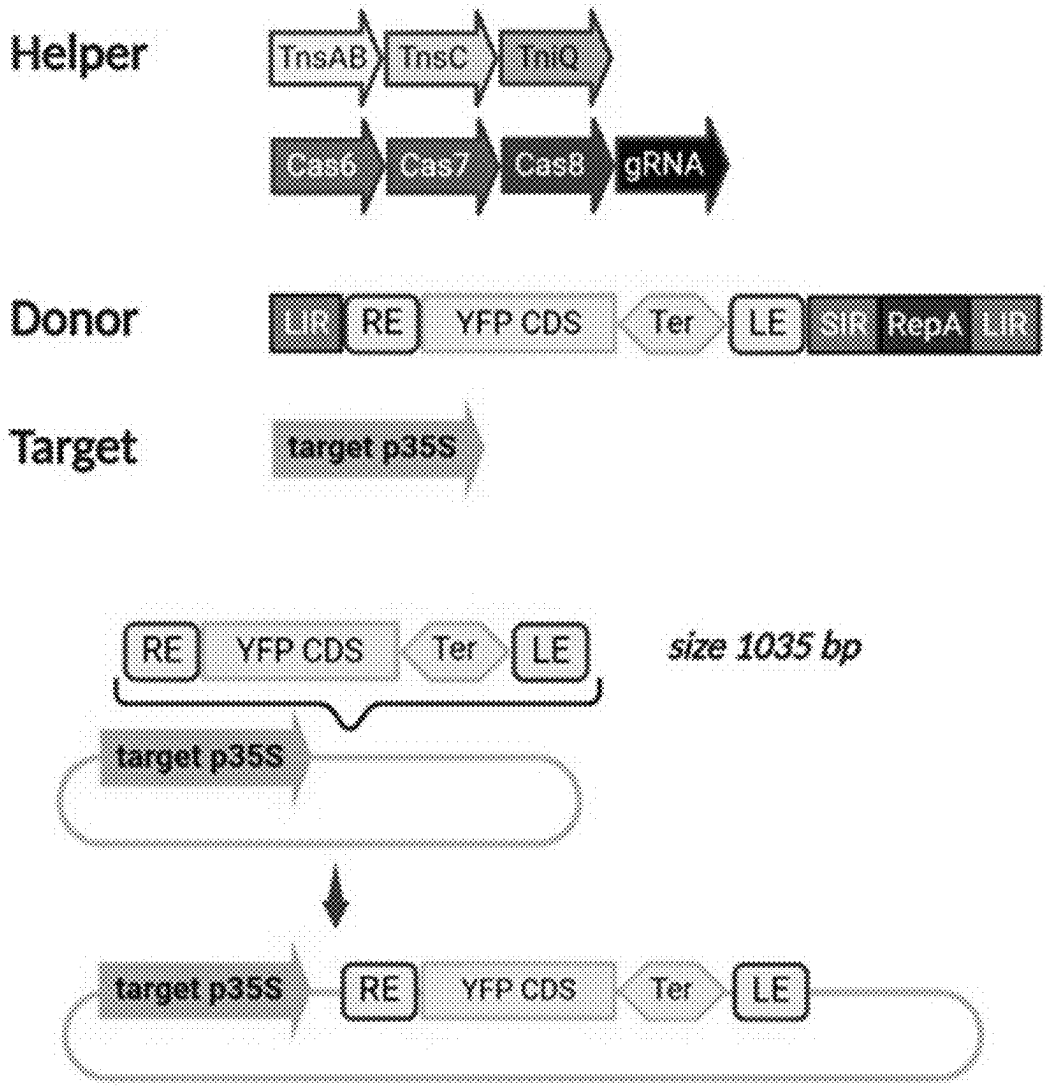

Helper
TnsAB
TnsC
TniQ
Cas6
Cas7
Cas8
gRNA
Donor
LIR
RE
YFP CDS
Ter
LE
SIR
RepA
LIR
Target
target p35S
RE
YFP CDS
Ter
LE
size 1035 bp
target p35S
target p35S
RE
YFP CDS
Ter
LE

FIG. 3A

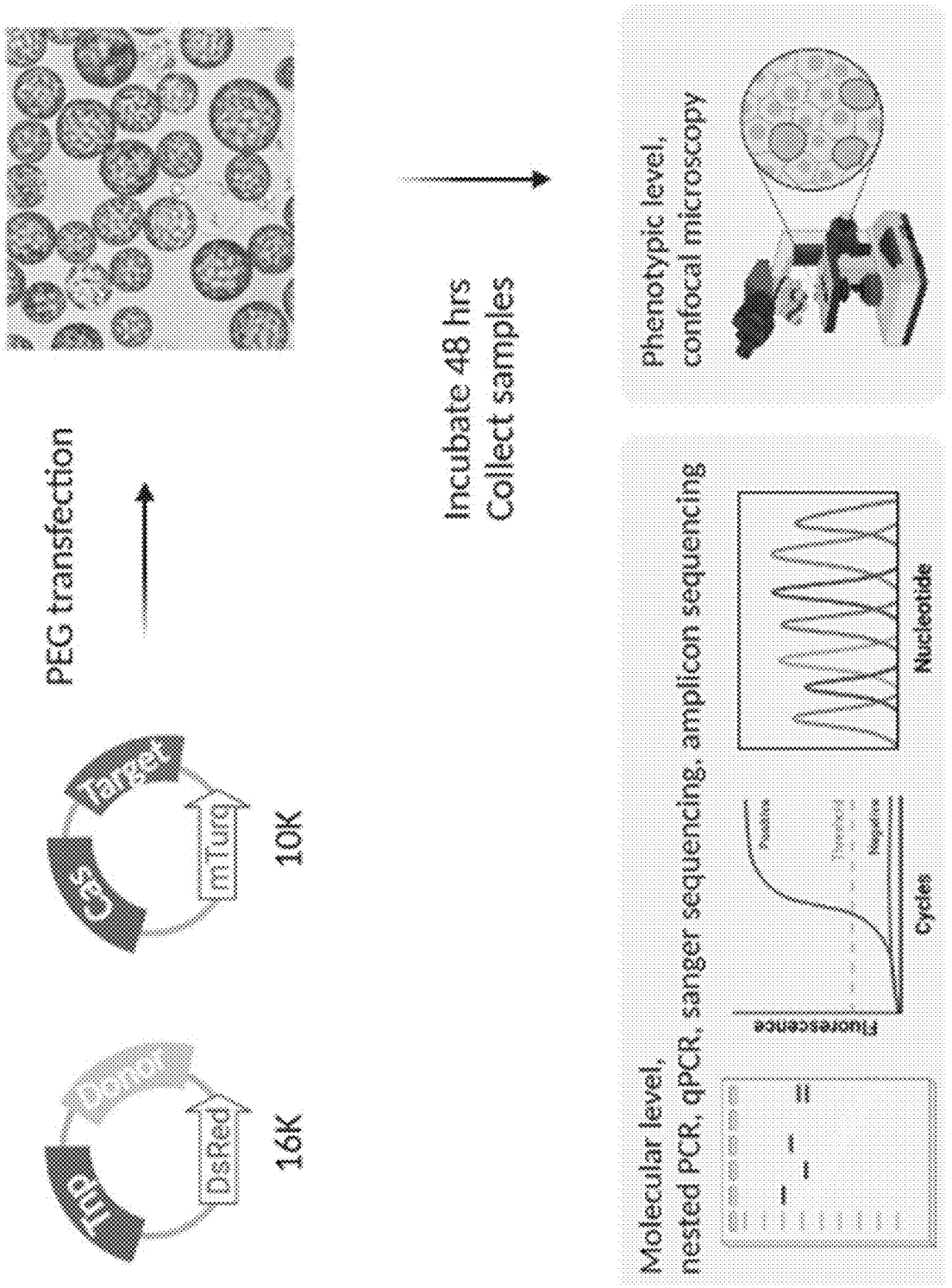

Target
DsRed
16K
Cas
Target
mTurq
10K
PEG transfection
Incubate 48 hrs
Collect samples
Phenotypic level,
confocal microscopy
Molecular level,
nested PCR, qPCR, sanger sequencing, amplicon sequencing
Fluorescence
Cycles
Nucleotide

FIG. 3B

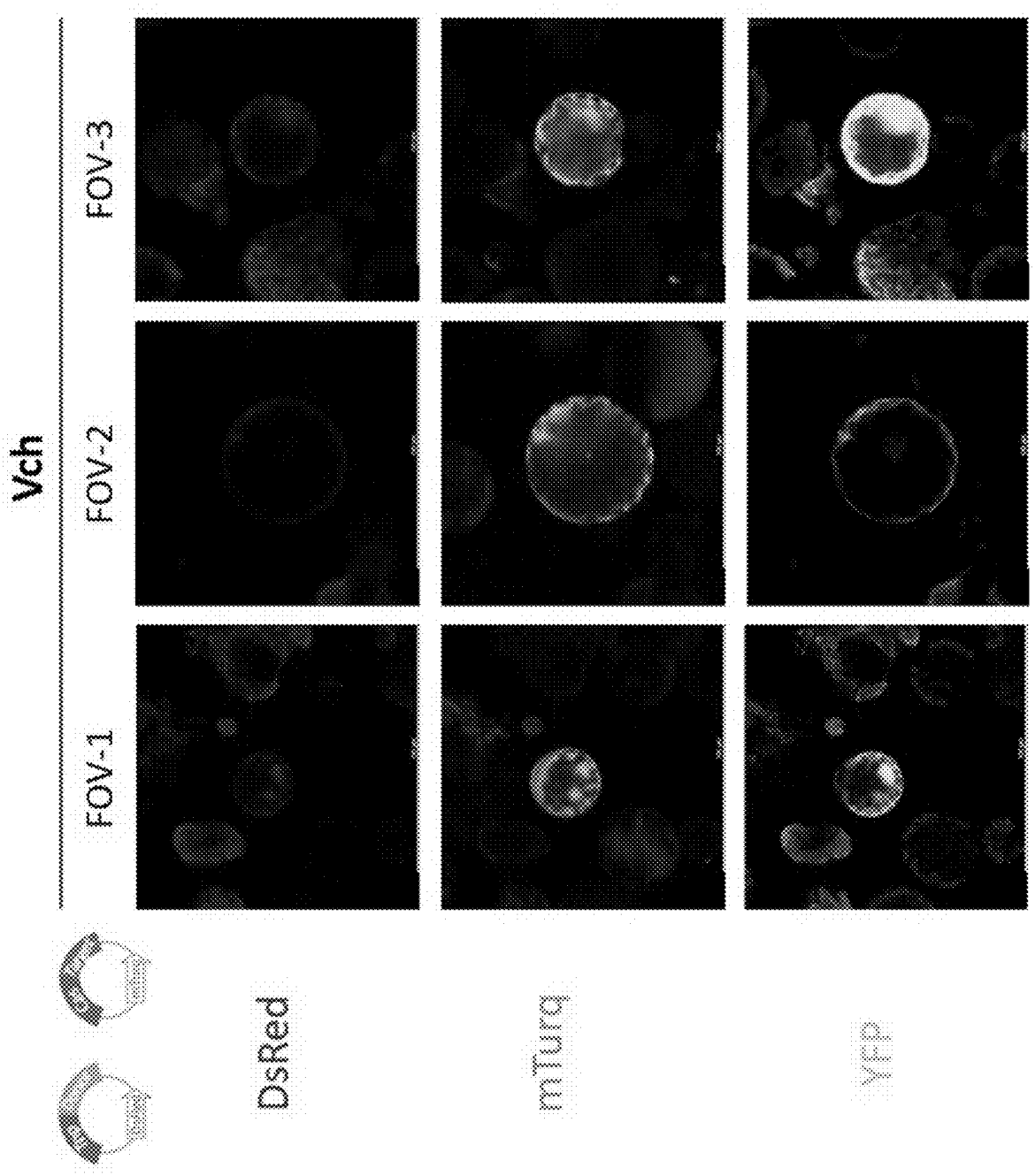


FIG. 4

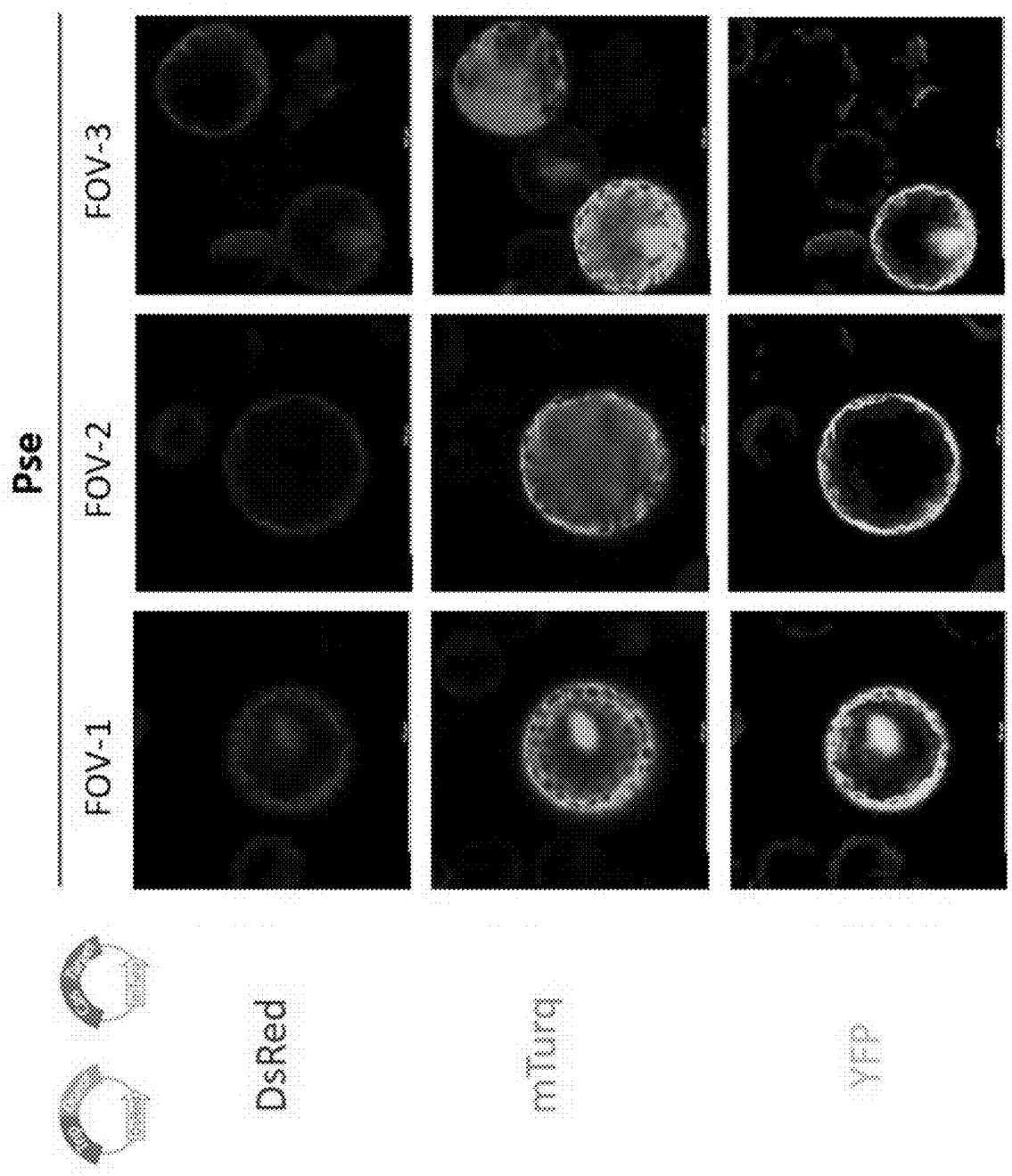


FIG. 4 (Cont'd)

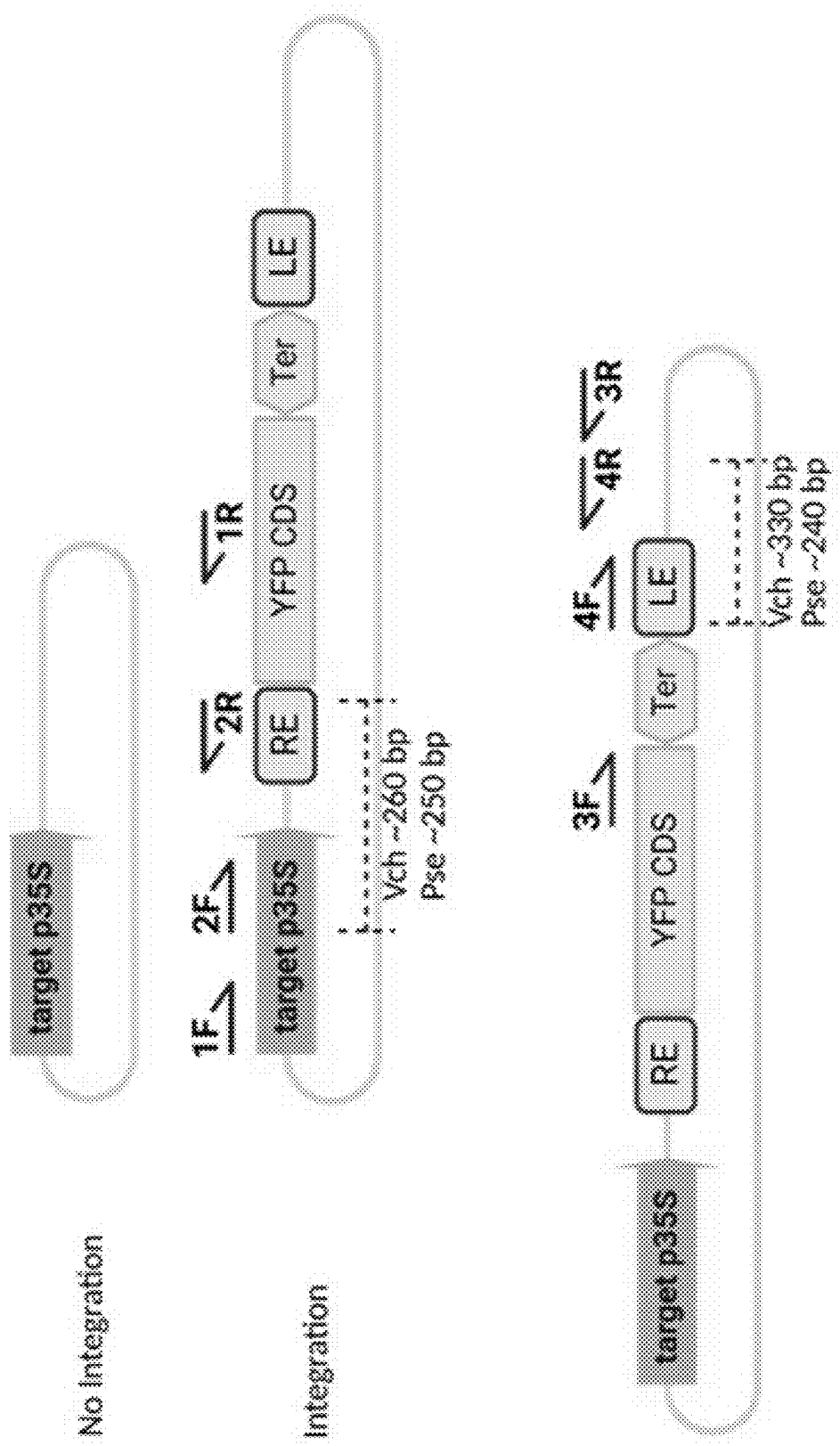

No Integration
target p35S
Integration
1F
2F
target p35S
2R
RE
1R
YFP CDS
Ter
LE
Vch ~260 bp
Pse ~250 bp
3F
4F
3R
4R
target p35S
RE
YFP CDS
Ter
LE
Vch ~330 bp
Pse ~240 bp

FIG. 5A

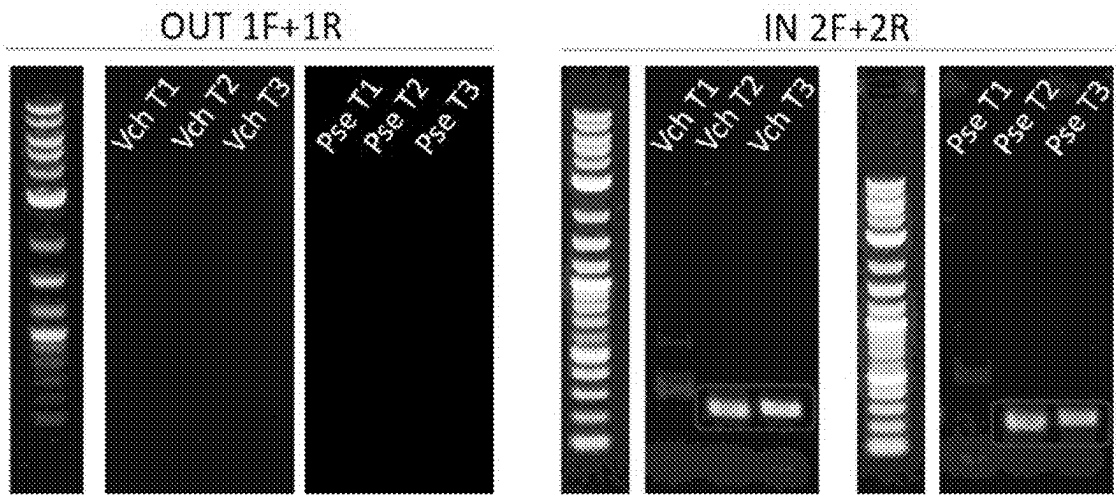

OUT 1F+1R
Vch T1
Vch T2
Vch T3
Pse T1
Pse T2
Pse T3
IN 2F+2R
Vch T1
Vch T2
Vch T3
Pse T1
Pse T2
Pse T3

FIG. 5B

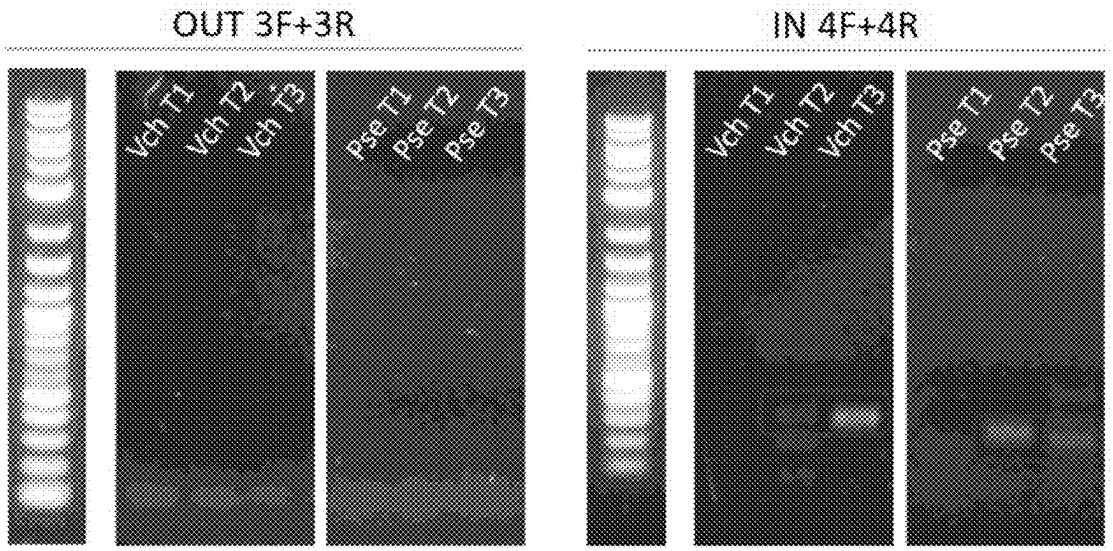

OUT 3F+3R
Vch T1
Vch T2
Vch T3
Pse T1
Pse T2
Pse T3
IN 4F+4R
Vch T1
Vch T2
Vch T3
Pse T1
Pse T2
Pse T3

FIG. 5C

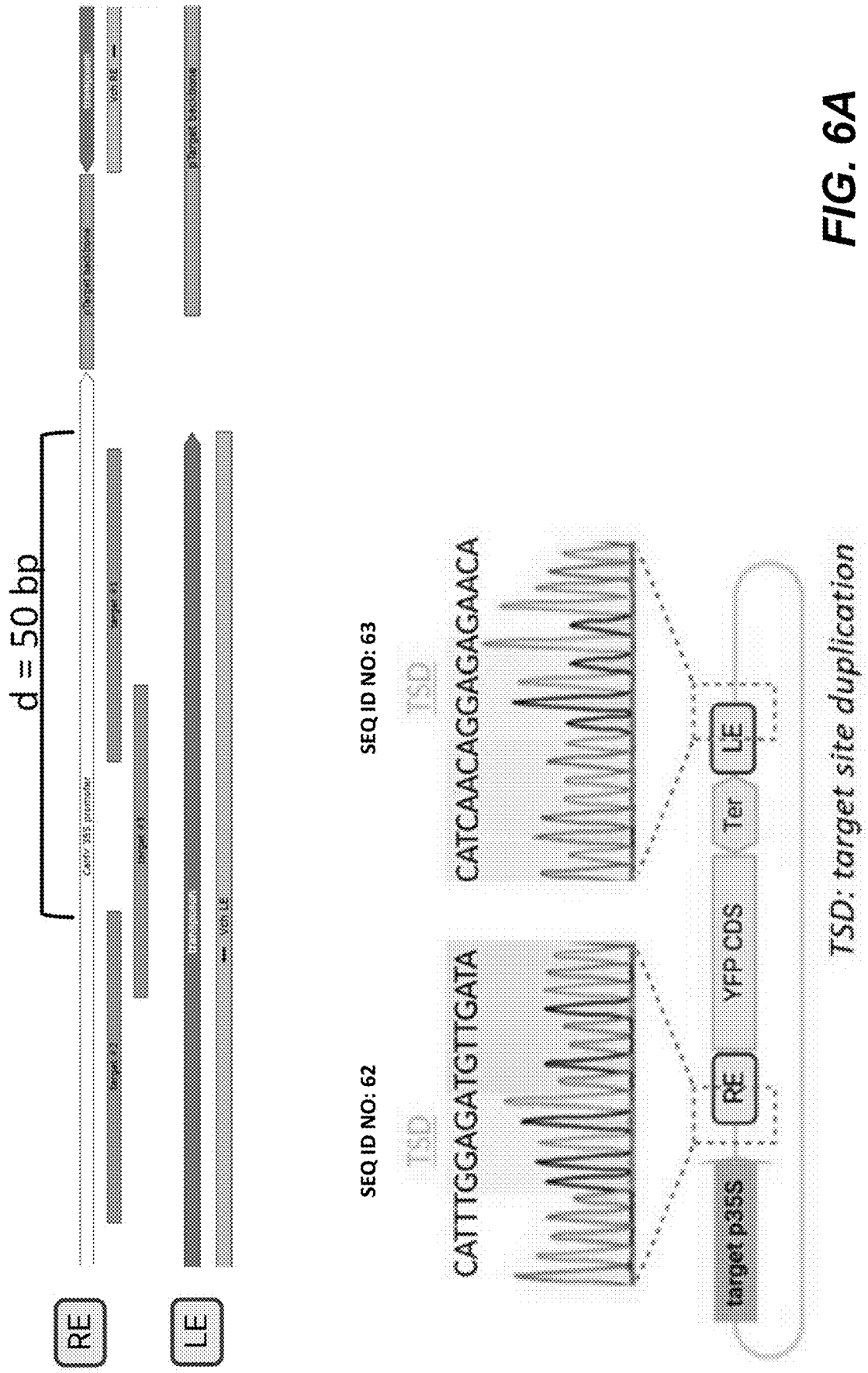
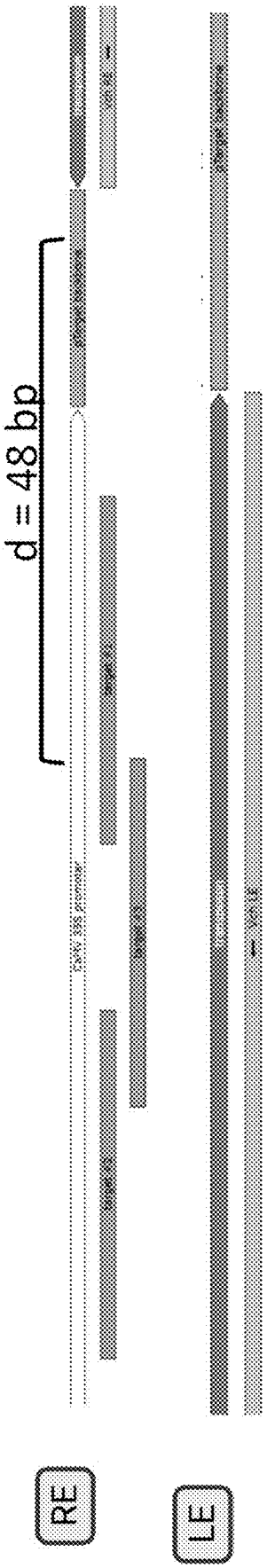

d = 50 bp
RE
LE
SEQ ID NO: 62
TSD
CATTTGGAGATGTTGATA
SEQ ID NO: 63
TSD
CATCAACAGGAGAGAACA
target p35S
RE
YFP CDS
Ter
LE
TSD: target site duplication

FIG. 6A


d = 48 bp
RE
LE

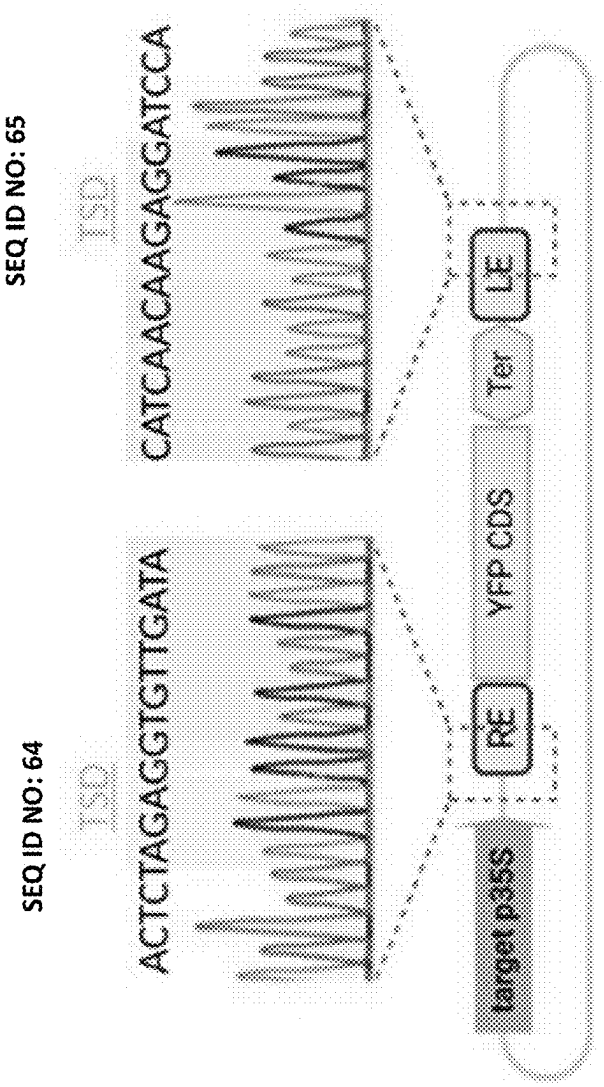

SEQ ID NO: 64
TSD
ACTCTAGAGGTGTTGATA
SEQ ID NO: 65
TSD
CATCAACAAGAGGATCCA
target p35S
RE
YFP CDS
Ter
LE

FIG. 6B

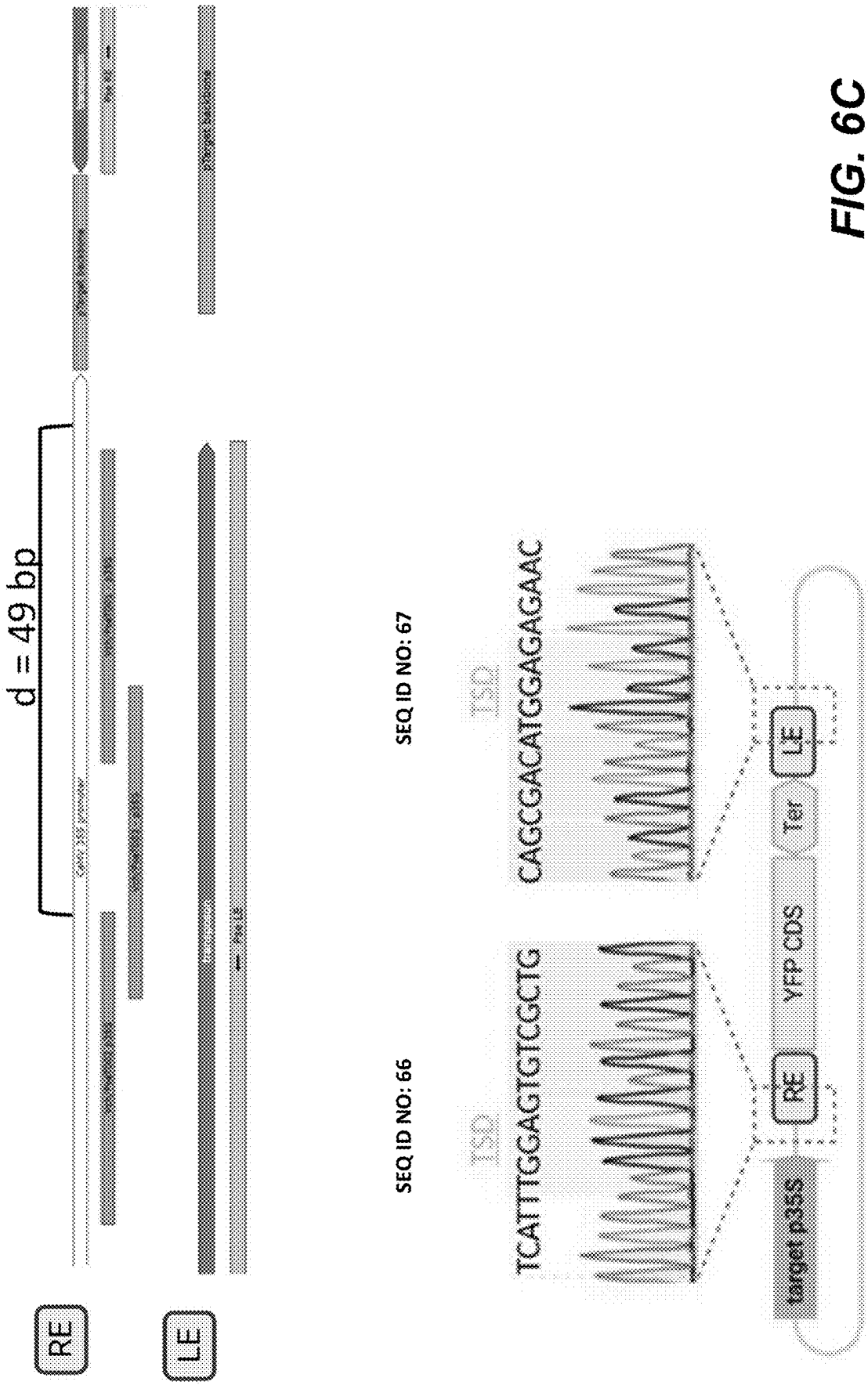

d = 49 bp
RE
LE
SEQ ID NO: 66
TSD
TCATTTGGAGTGTCGCTG
SEQ ID NO: 67
TSD
CAGCGACATGGAGAGAAC
target p35S
RE
YFP CDS
Ter
LE

FIG. 6C

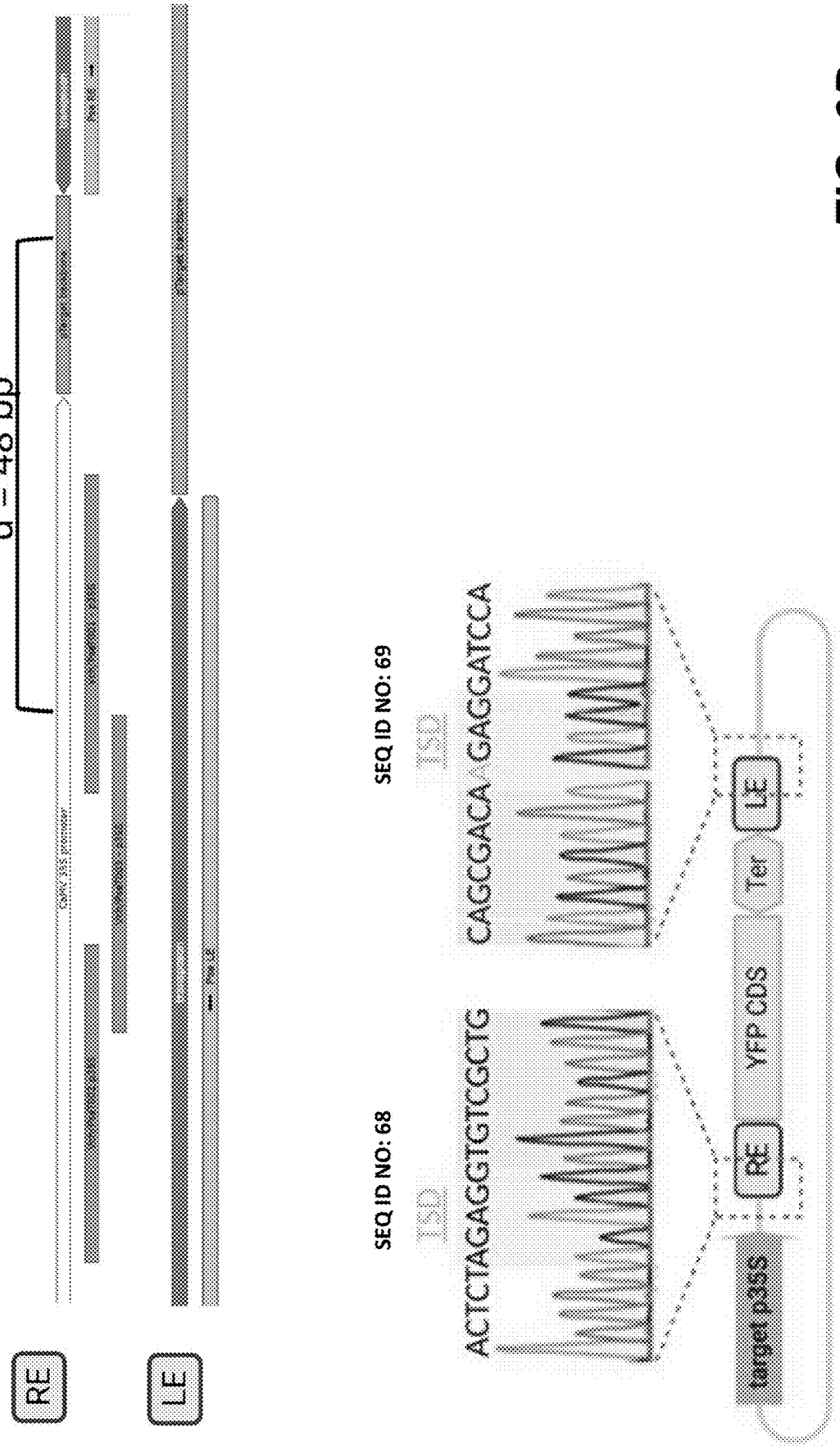

d = 48 bp
RE
LE
SEQ ID NO: 68
TSD
ACTCTAGAGGTGTCGCTG
SEQ ID NO: 69
TSD
CAGCGACA GAGGATCCA
target p35S
RE
YFP CDS
Ter
LE

FIG. 6D

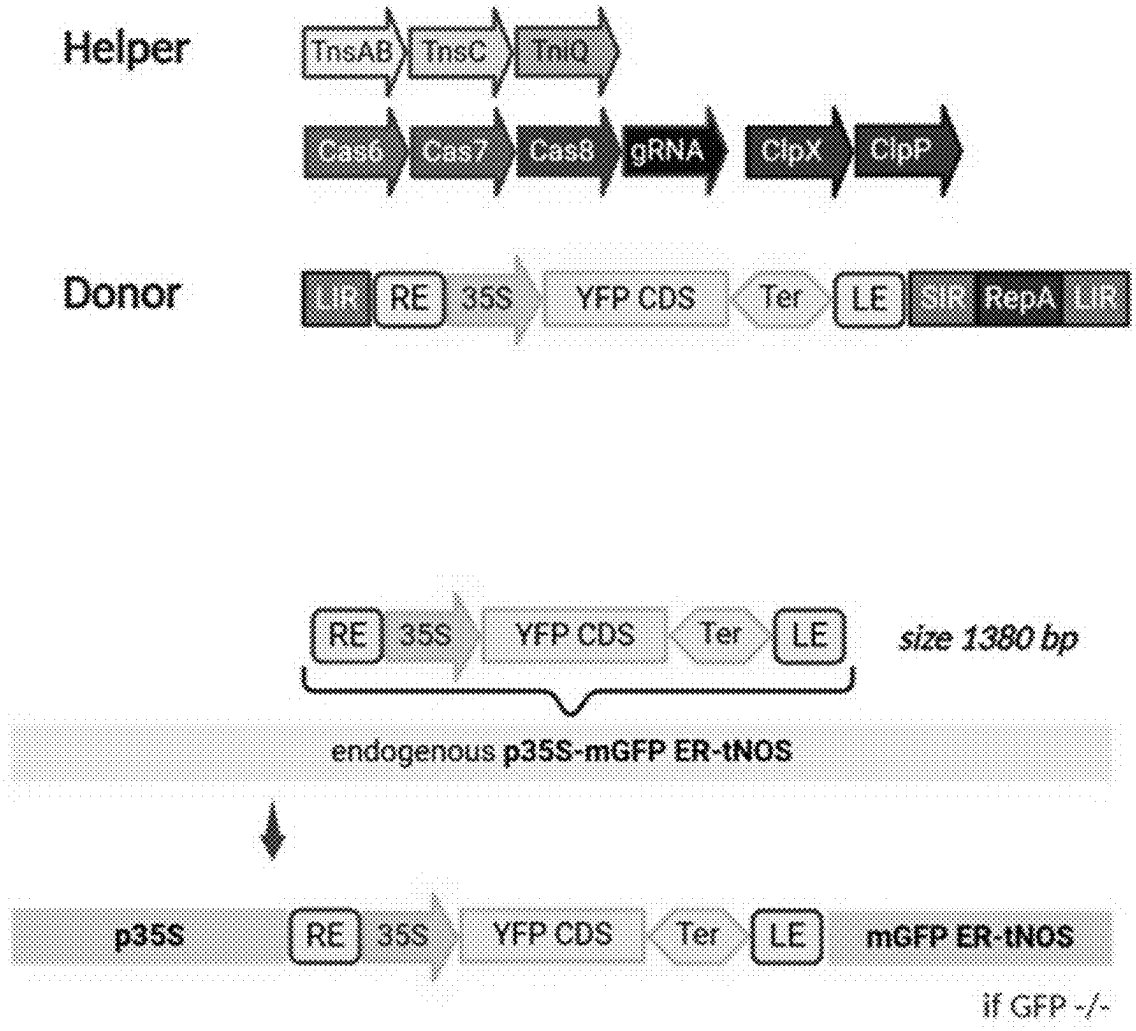

Helper
TnsAB
TnsC
TniQ
Cas6
Cas7
Cas8
gRNA
ClpX
ClpP
Donor
LIR
RE
35S
YFP CDS
Ter
LE
SIR
RepA
LIR
RE
35S
YFP CDS
Ter
LE
size 1380 bp
endogenous p35S-mGFP ER-tNOS
p35S
RE
35S
YFP CDS
Ter
LE
mGFP ER-tNOS
If GFP -/-

FIG. 7A

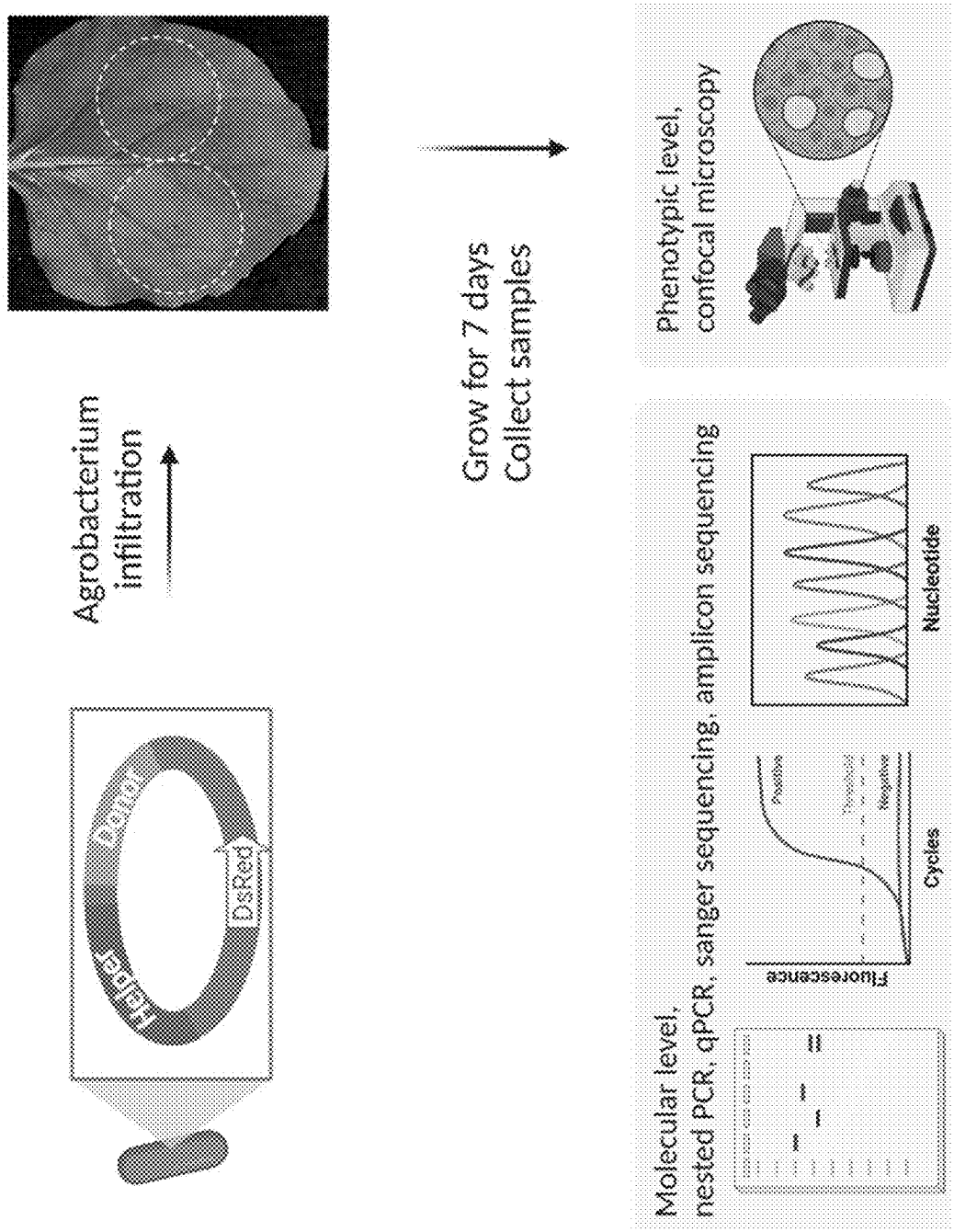

Agrobacterium
infiltration
Grow for 7 days
Collect samples
Phenotypic level,
confocal microscopy
Molecular level,
nested PCR, qPCR, sanger sequencing, amplicon sequencing
Fluorescence
Cycles
Nucleotide

FIG. 7B

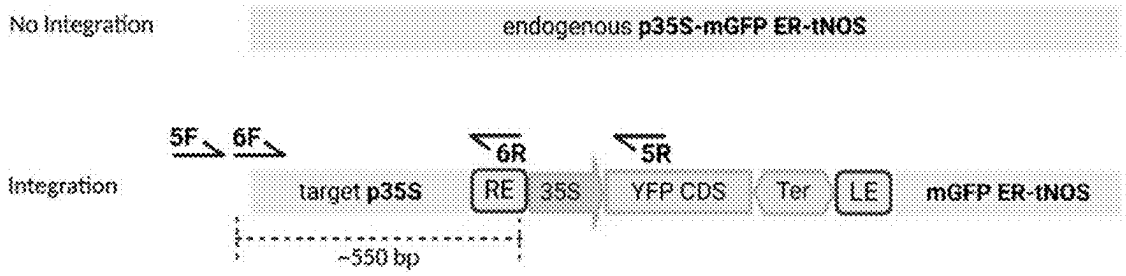

No Integration
endogenous p35S-mGFP ER-tNOS
5F
6F
6R
5R
Integration
target p35S
RE
35S
YFP CDS
Ter
LE
mGFP ER-tNOS
~550 bp

FIG. 8A

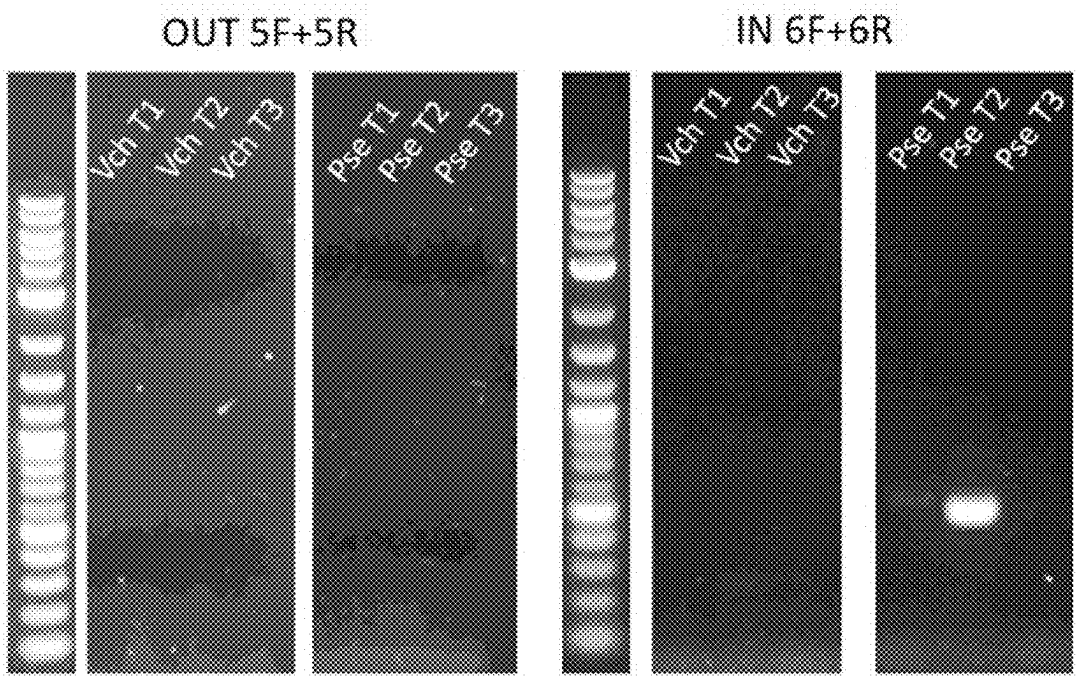

OUT 5F+5R
IN 6F+6R
Vch T1
Vch T2
Vch T3
Pse T1
Pse T2
Pse T3
Vch T1
Vch T2
Vch T3
Pse T1
Pse T2
Pse T3

FIG. 8B

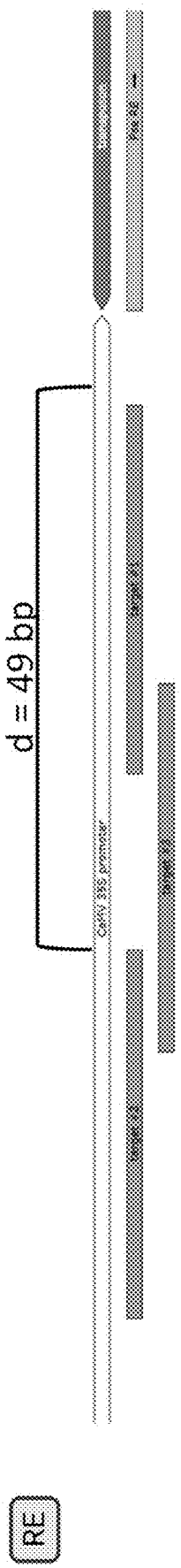


FIG. 8C

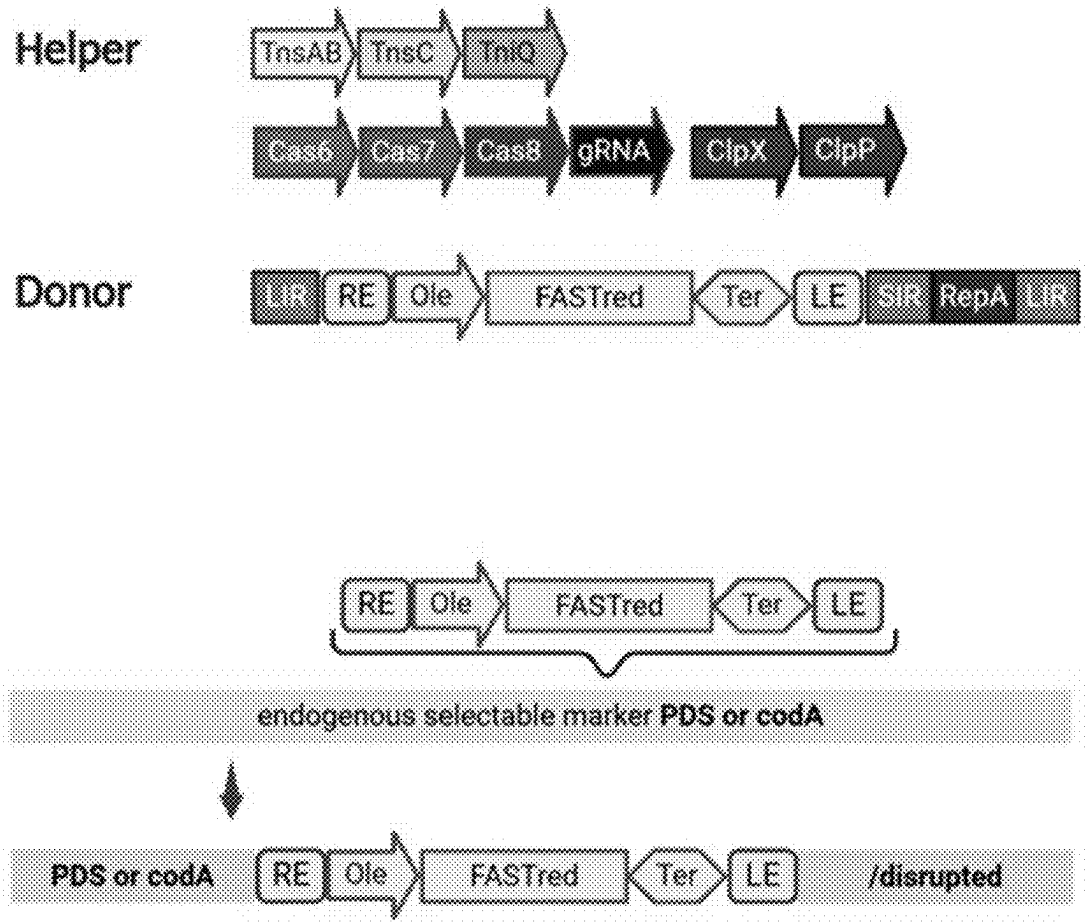

Helper
TnsAB
TnsC
TniQ
Cas6
Cas7
Cas8
gRNA
ClpX
ClpP
Donor
LIR
RE
Ole
FASTred
Ter
LE
SIR
RepA
LIR
RE
Ole
FASTred
Ter
LE
endogenous selectable marker PDS or codA
PDS or codA
RE
Ole
FASTred
Ter
LE
/disrupted

FIG. 9A

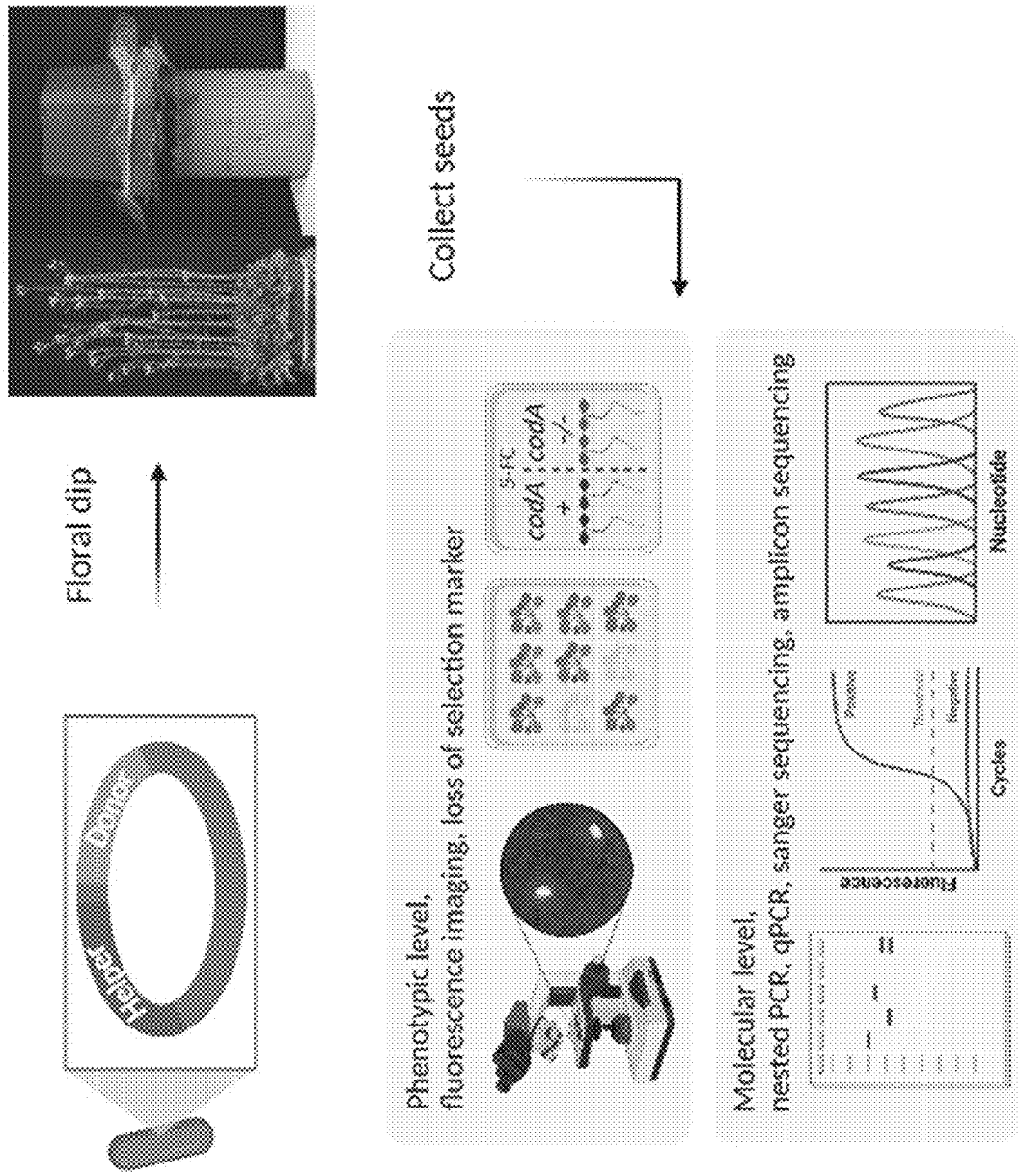

Floral dip
codA
Collect seeds
Phenotypic level,
fluorescence imaging, loss of selection marker
5-FC
codA
codA
+
-/-
Molecular level,
nested PCR, qPCR, sanger sequencing, amplicon sequencing
Fluorescence
Cycles
Nucleotide

FIG. 9B

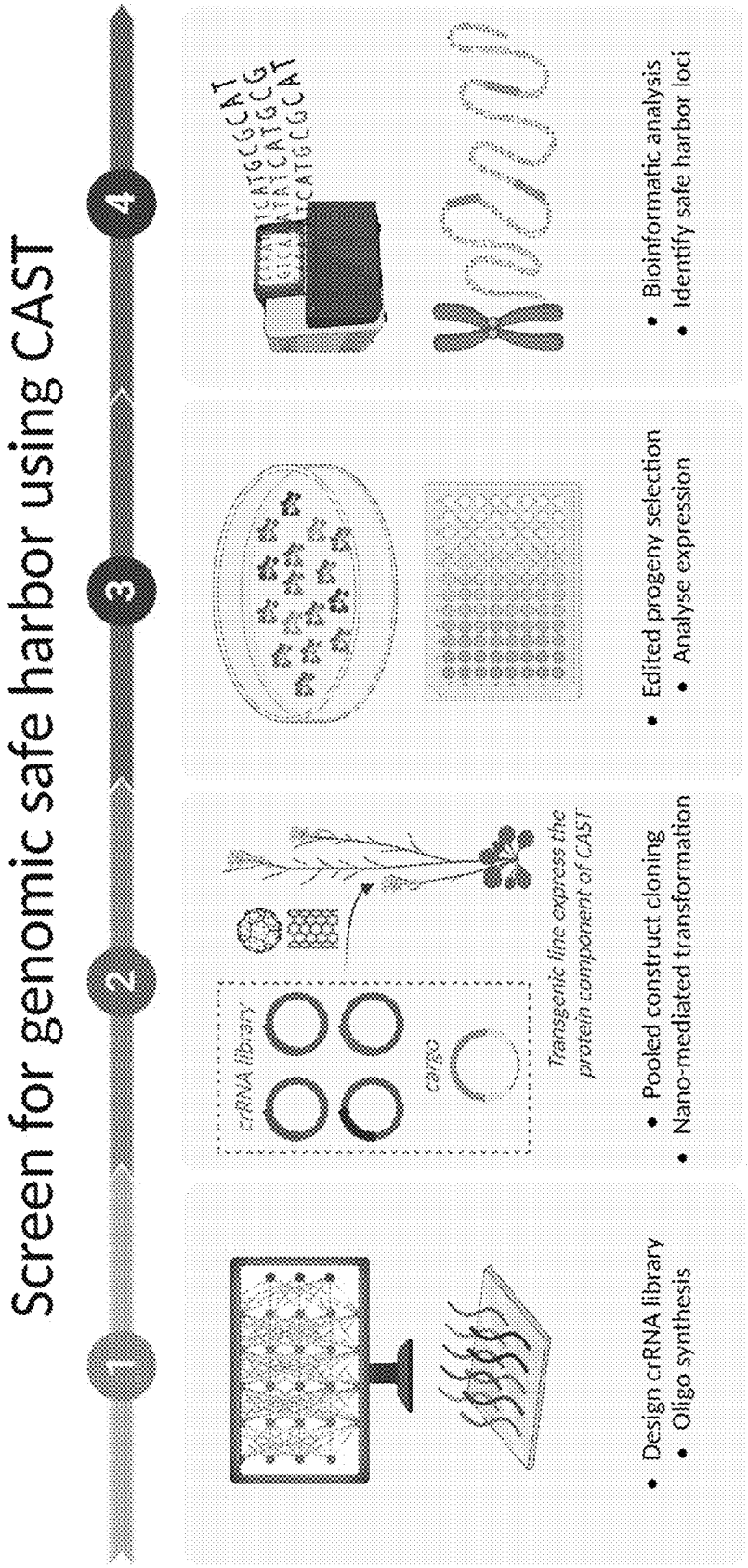

Screen for genomic safe harbor using CAST
1
2
3
4
crRNA library
cargo
Transgenic line express the protein component of CAST
TCATGCGCAT
ATATCATGCG
CATGCGCAT
Design crRNA library
Oligo synthesis
Pooled construct cloning
Nano-mediated transformation
Edited progeny selection
Analyse expression
Bioinformatic analysis
Identify safe harbor loci

FIG. 10

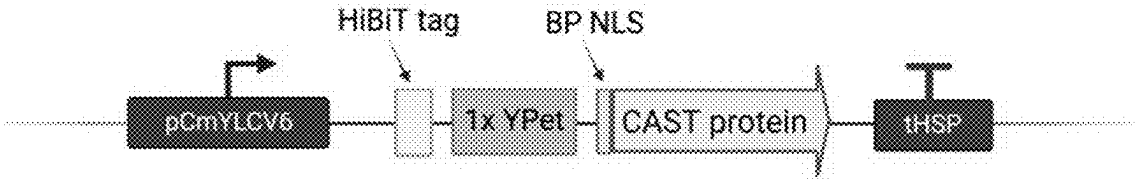


FIG. 11A

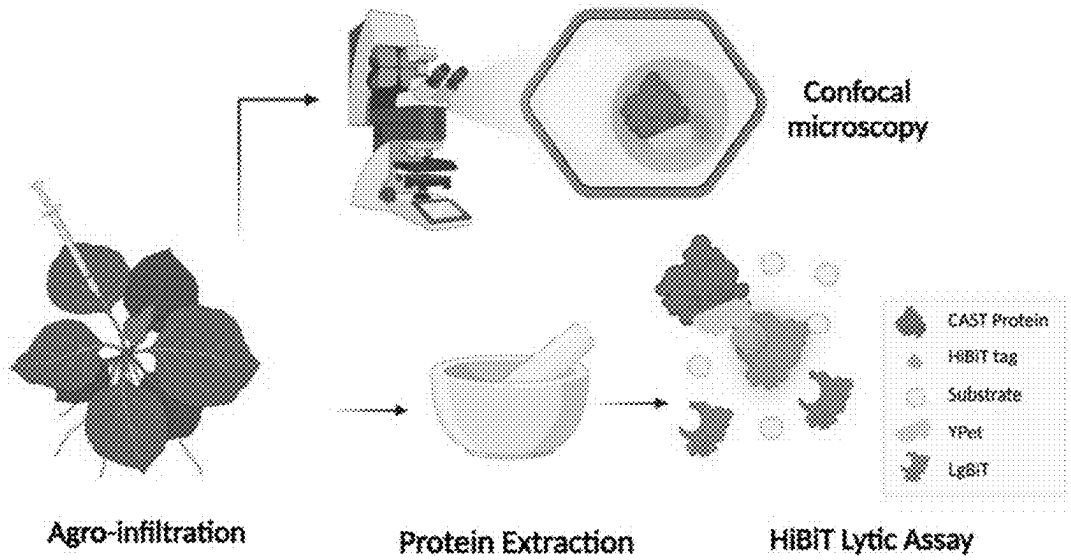


FIG. 11B

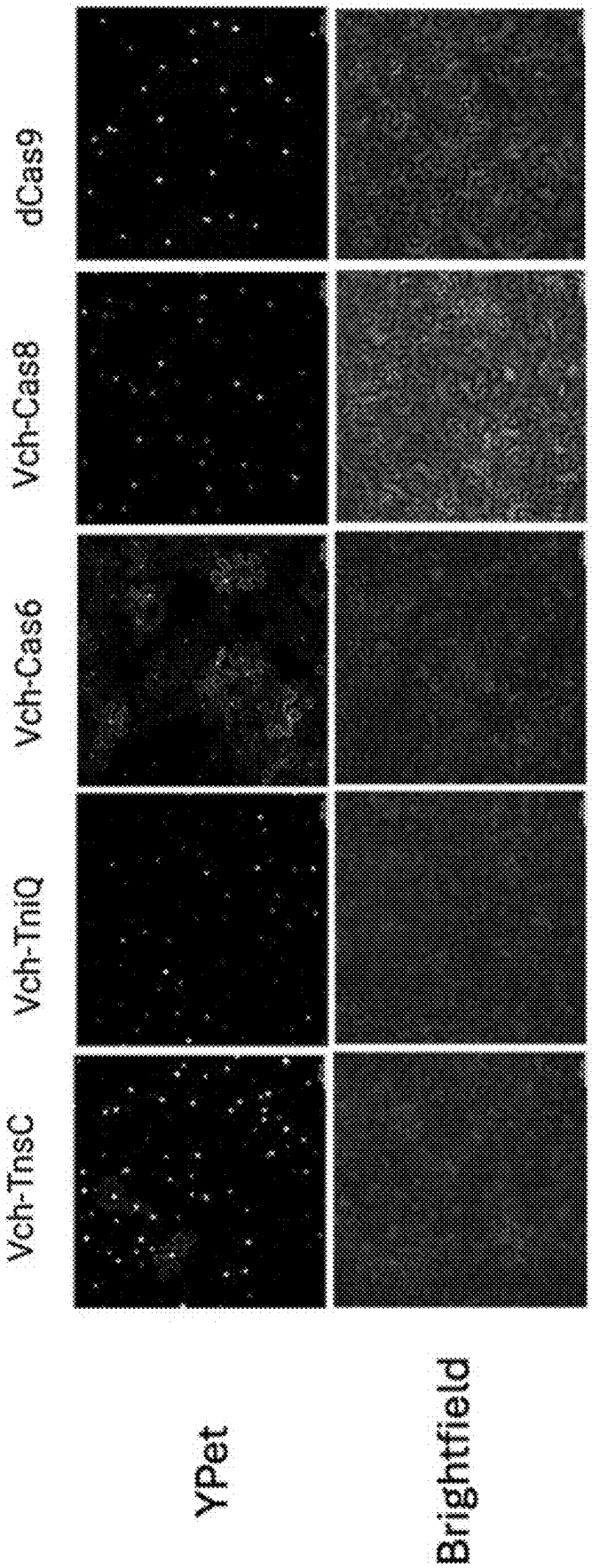


FIG. 11C

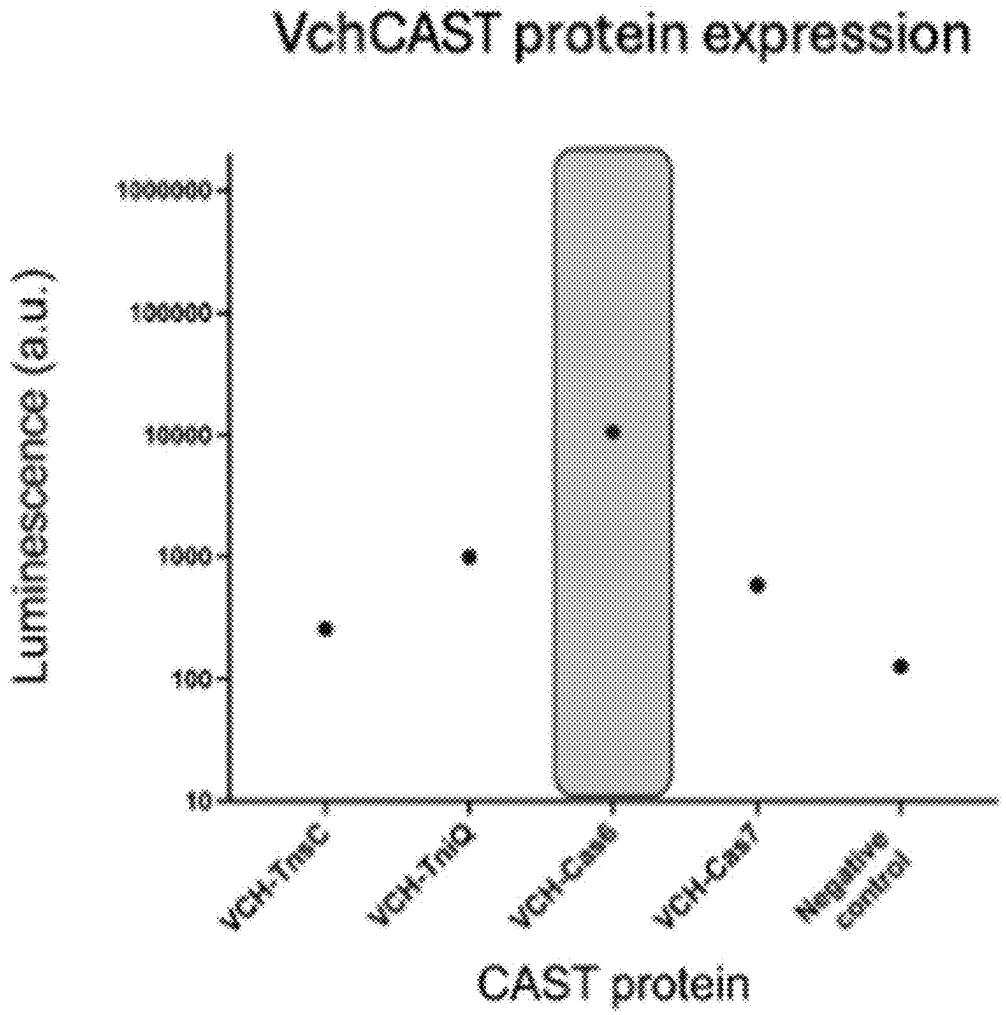

VchCAST protein expression
Luminescence (a.u.)
1000000
100000
10000
1000
100
10
VCH-TnsC
VCH-TnsB
VCH-Cas8
VCH-Cas7
Negative control
CAST protein

FIG. 11D

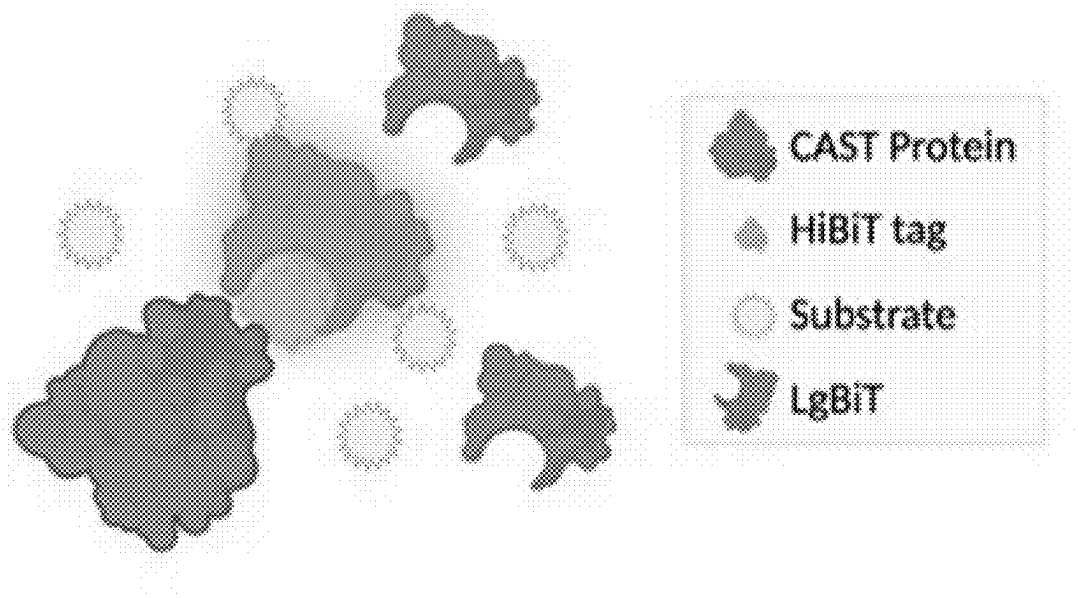


FIG. 12A

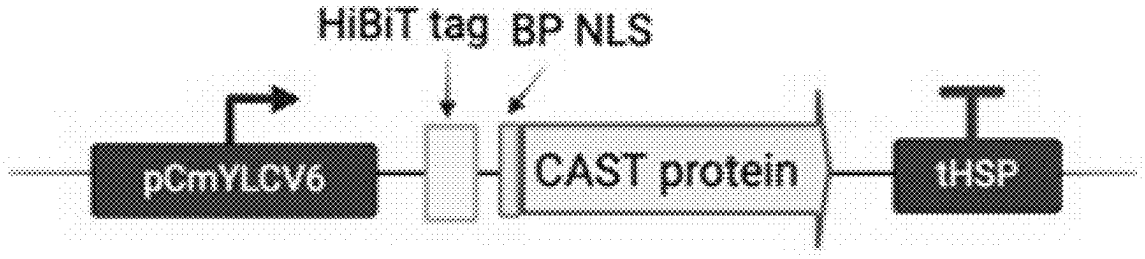


FIG. 12B

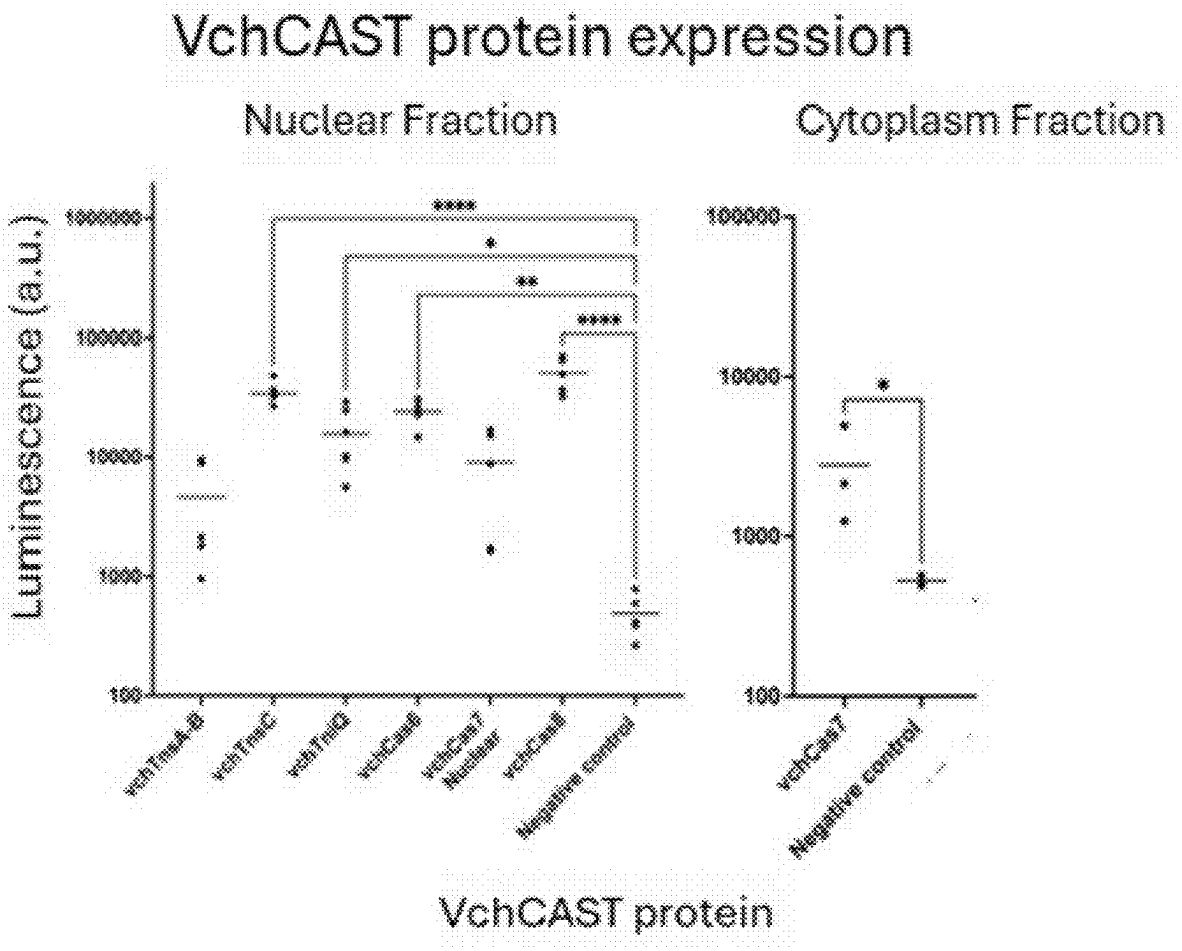

VchCAST protein expression
Nuclear Fraction
Cytoplasm Fraction
Luminescence (a.u.)
1000000
100000
10000
1000
100
****
*
**
****
vchTnsA-B
vchTnsC
vchTnsD
vchCas6
vchCas7
Buffer
vchCas8
Negative control
VchCAST protein

FIG. 12C

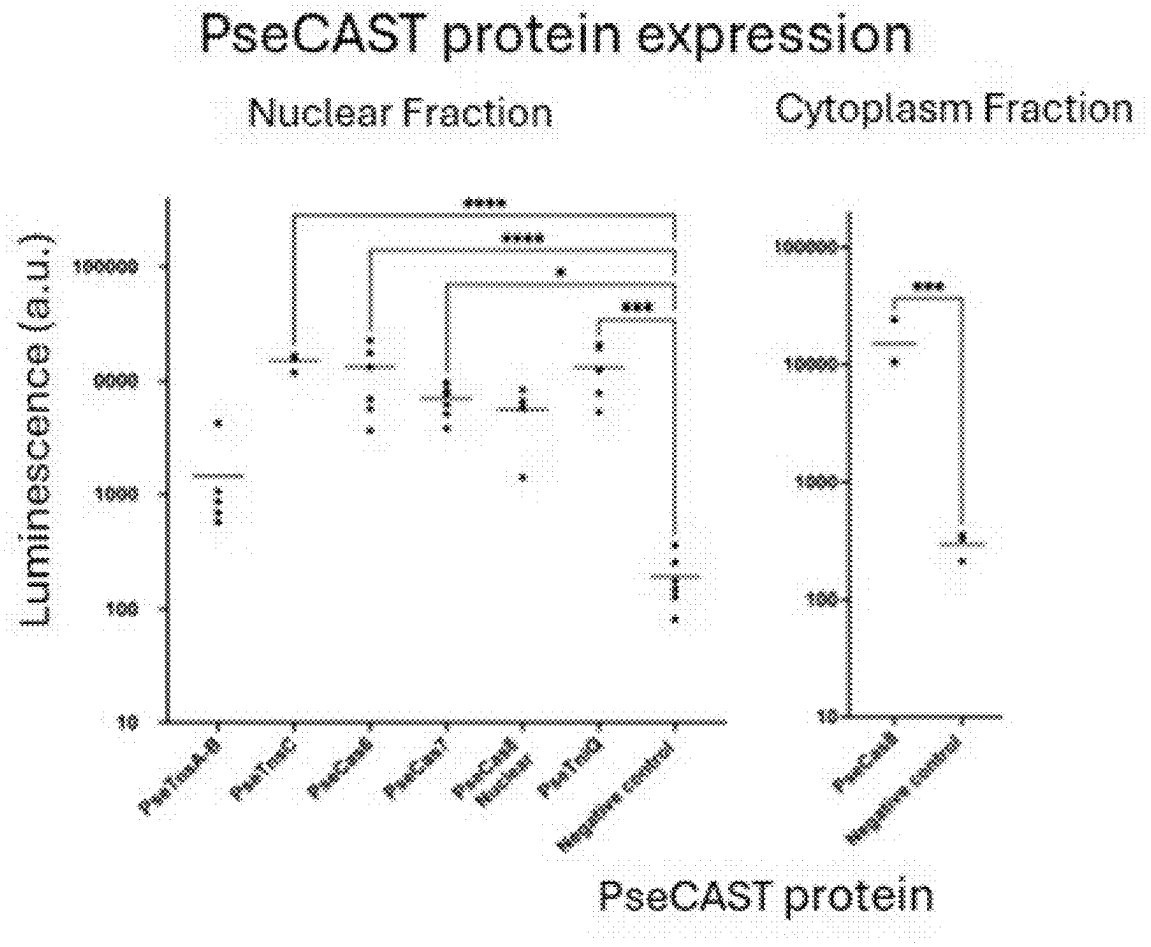
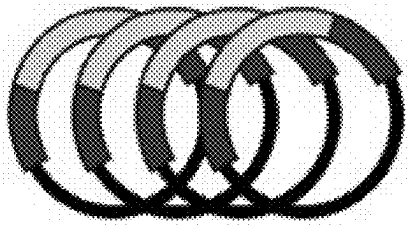

PseCAST protein expression
Nuclear Fraction
Cytoplasm Fraction
Luminescence (a.u.)
PseCAST protein

FIG. 12D



Geminiviral replicons (GV)initiate a rolling cycle replication, independent of cellular replication – **Increased copy number**

Construct map

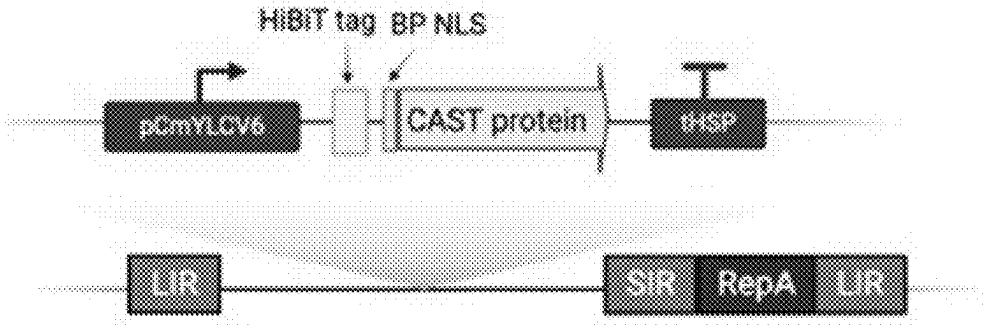


FIG. 12E

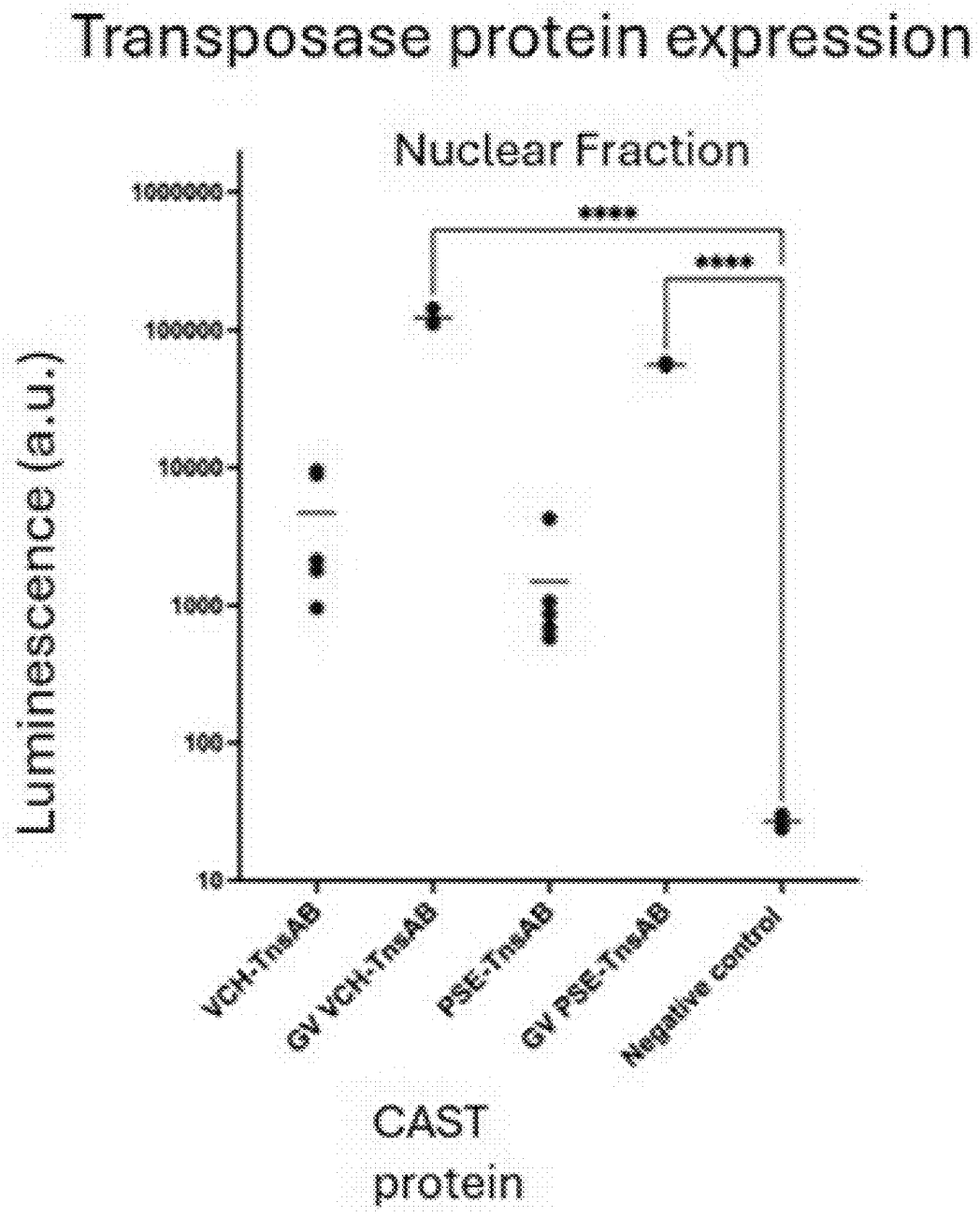

Transposase protein expression
Nuclear Fraction
Luminescence (a.u.)
1000000
100000
10000
1000
100
10
****
****
VCH-TnsAB
GV VCH-TnsAB
PSE-TnsAB
GV PSE-TnsAB
Negative control
CAST protein

FIG. 12F

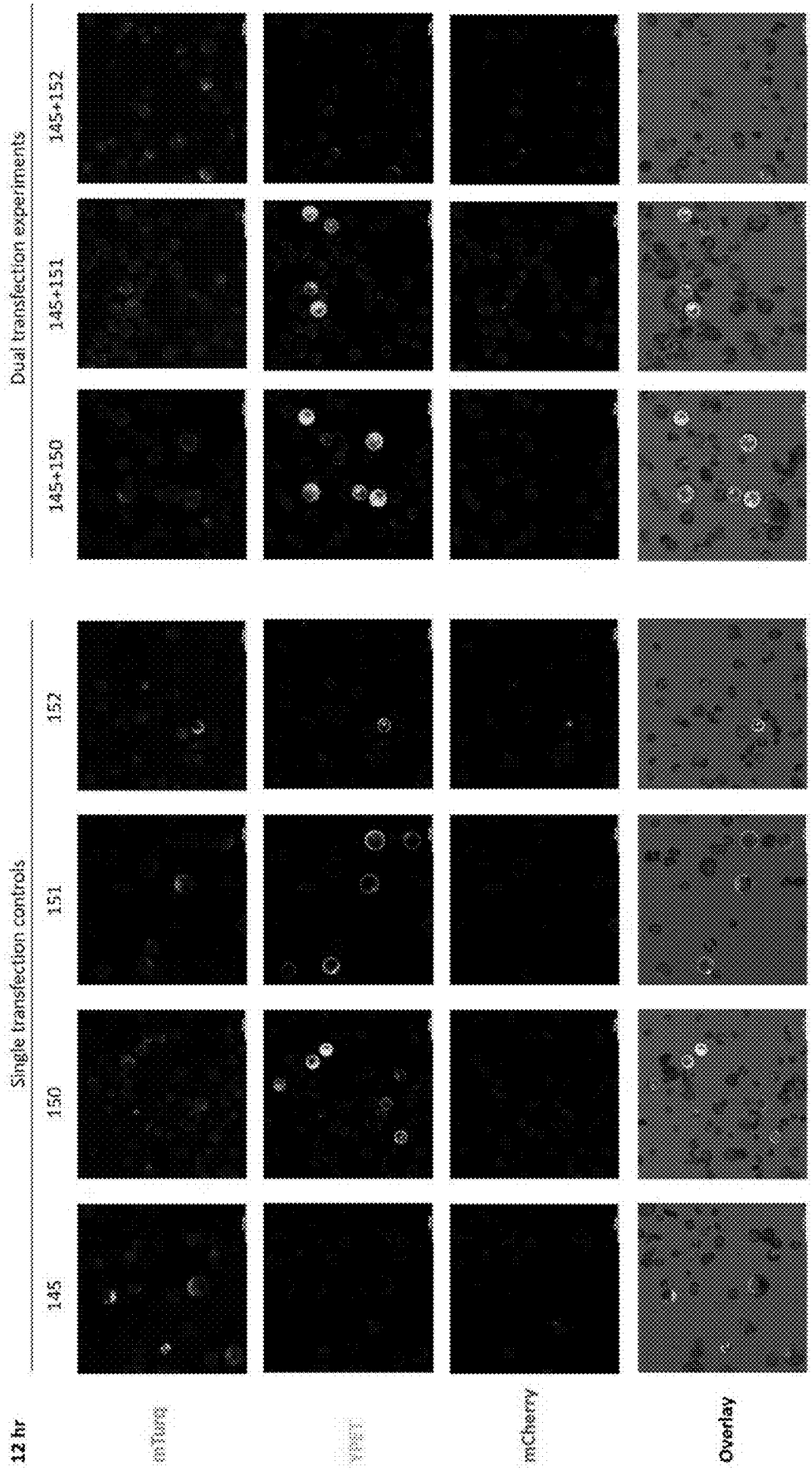

12 hr
Single transfection controls
145
150
151
152
Dual transfection experiments
145+150
145+151
145+152
mTurq
YFP
mCherry
Overlay

FIG. 13A

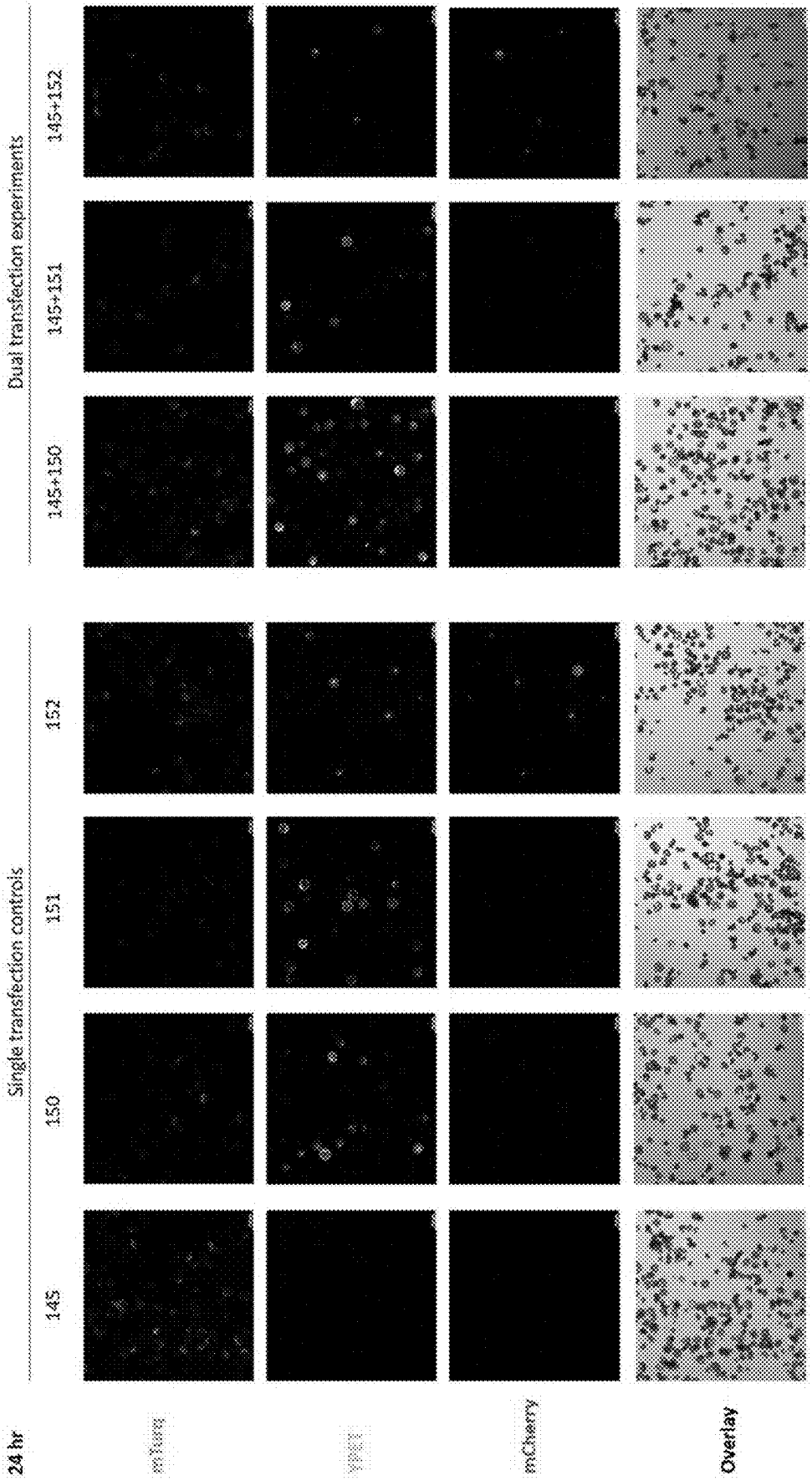

24 hr
Single transfection controls
145
150
151
152
Dual transfection experiments
145+150
145+151
145+152
mCherry
Overlay

FIG. 13B

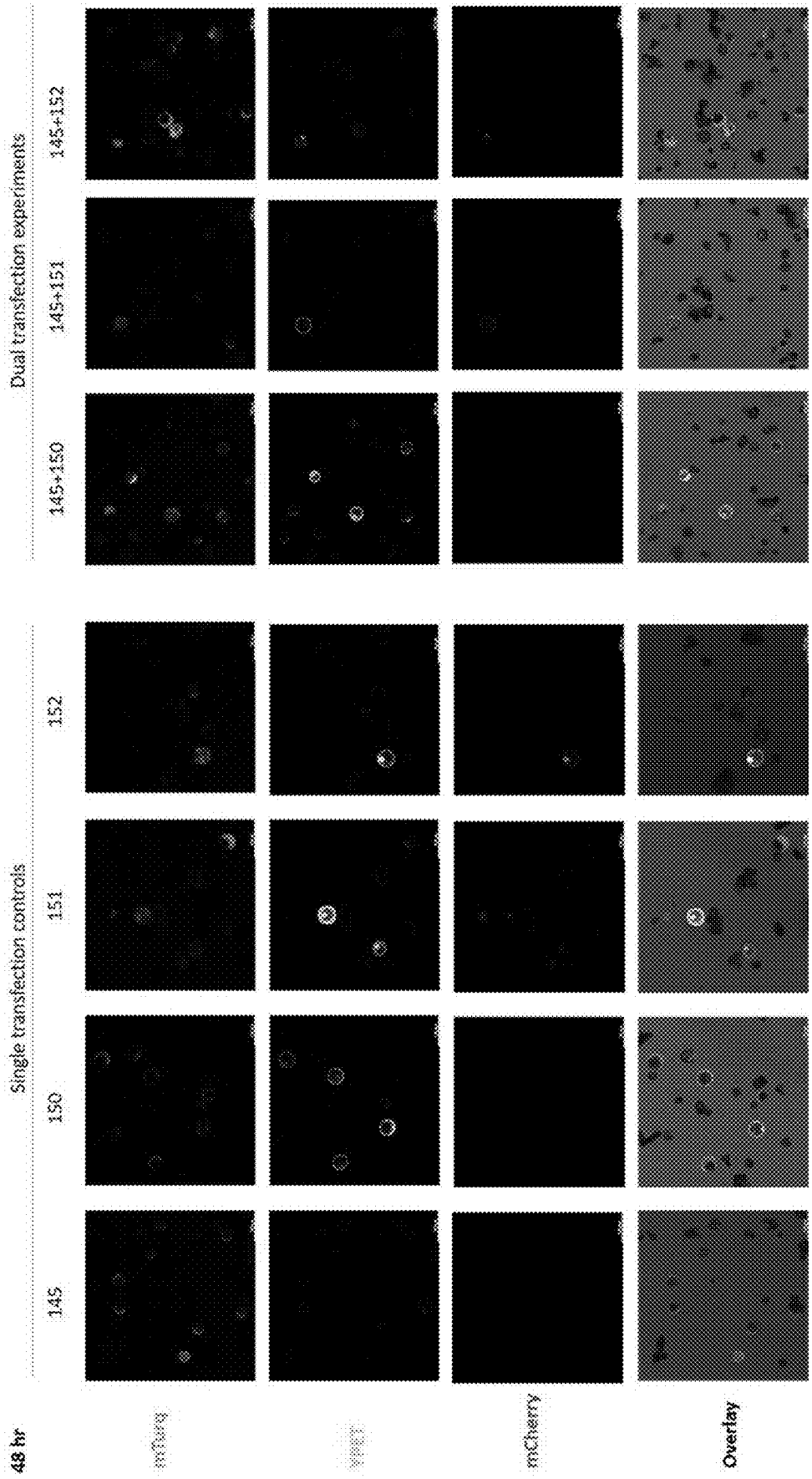

48 hr
Single transfection controls
145
150
151
152
Dual transfection experiments
145+150
145+151
145+152
mCherry
Overlay

FIG. 13C

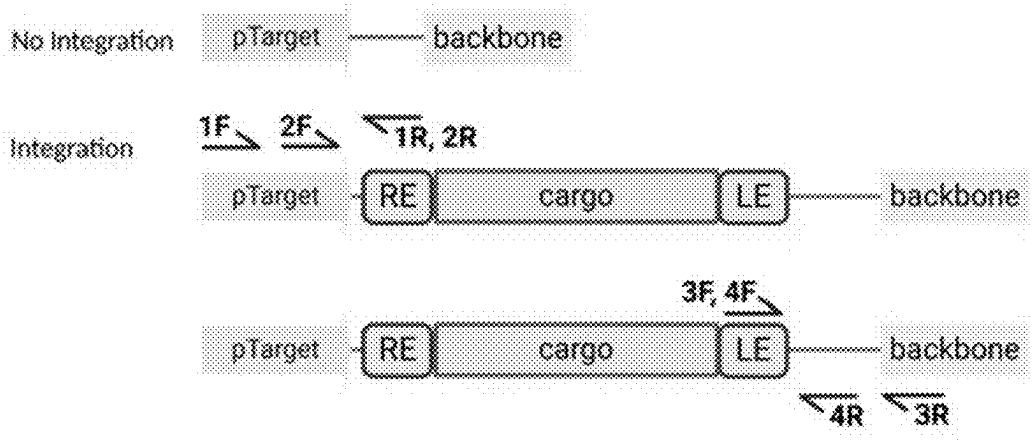

No Integration
pTarget
backbone
Integration
1F
2F
1R, 2R
pTarget
RE
cargo
LE
backbone
3F, 4F
pTarget
RE
cargo
LE
backbone
4R
3R

FIG. 14A

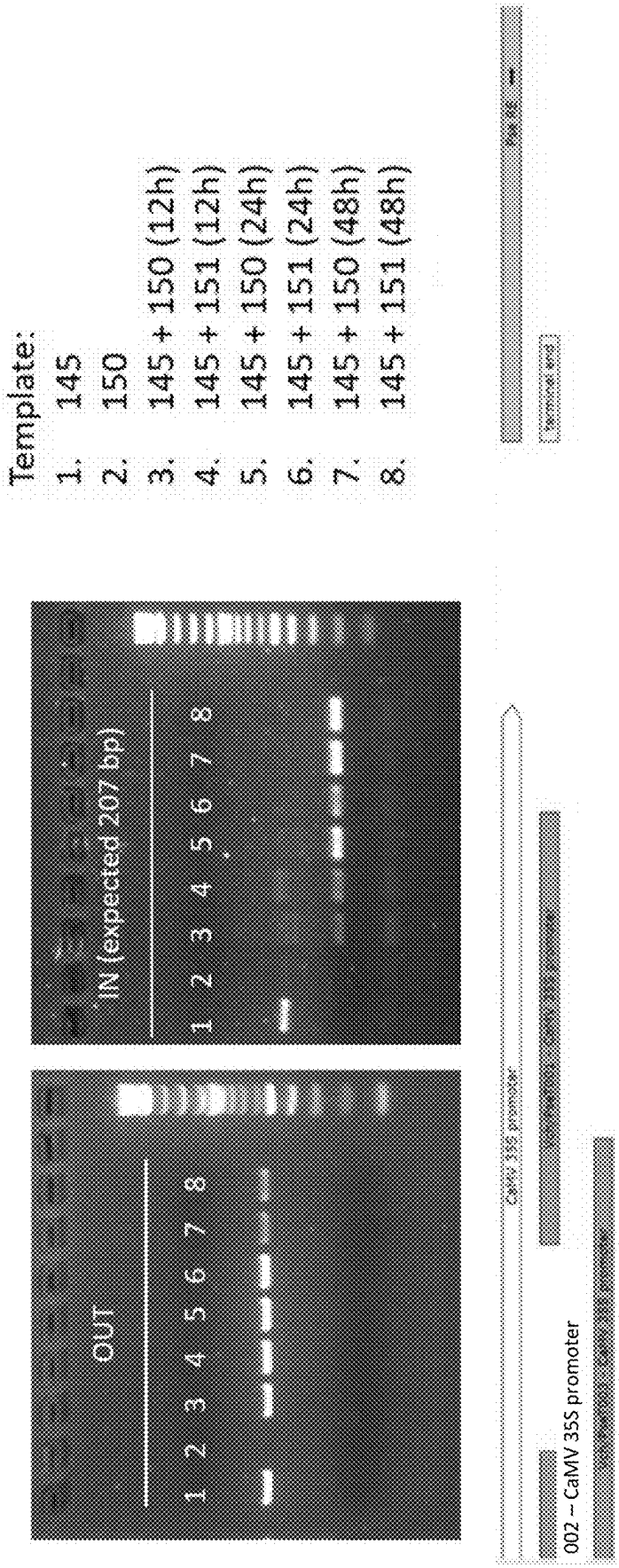

Template:
1. 145
2. 150
3. 145 + 150 (12h)
4. 145 + 151 (12h)
5. 145 + 150 (24h)
6. 145 + 151 (24h)
7. 145 + 150 (48h)
8. 145 + 151 (48h)
OUT
1 2 3 4 5 6 7 8
IN (expected 207 bp)
1 2 3 4 5 6 7 8
CaMV 35S promoter
002 – CaMV 35S promoter

FIG. 14B

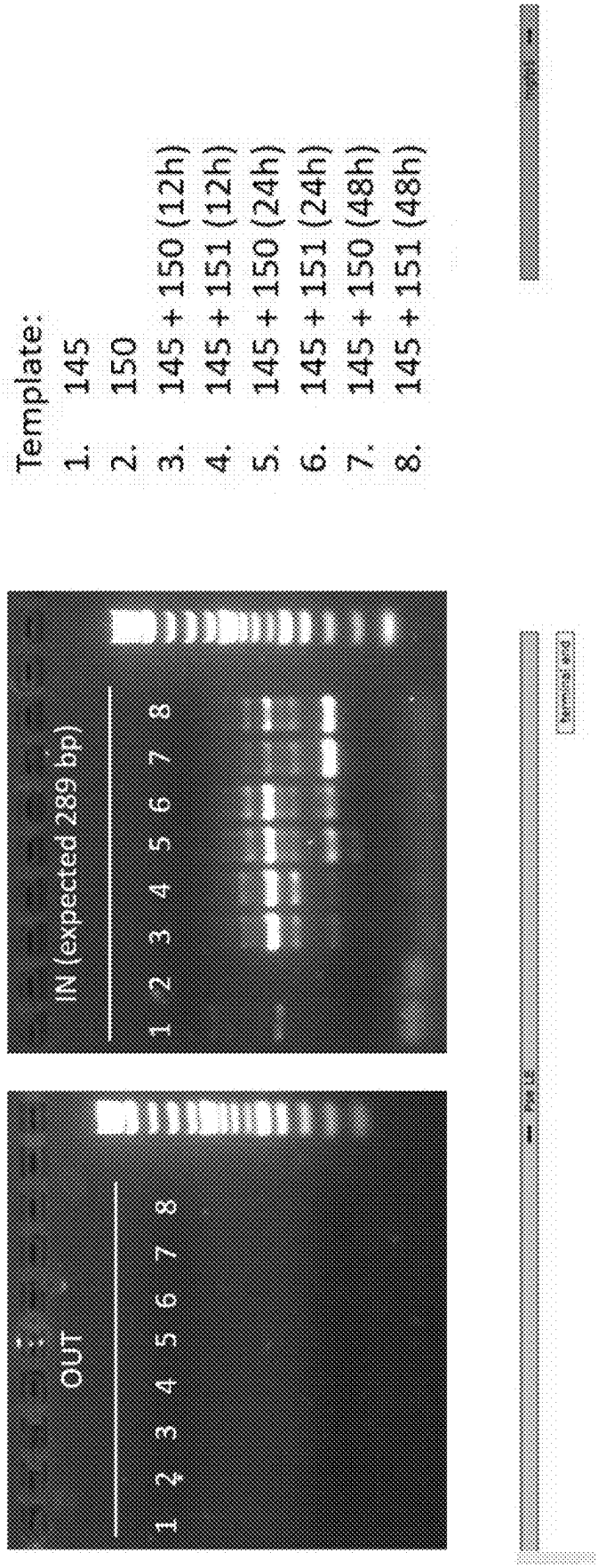

Template:
1. 145
2. 150
3. 145 + 150 (12h)
4. 145 + 151 (12h)
5. 145 + 150 (24h)
6. 145 + 151 (24h)
7. 145 + 150 (48h)
8. 145 + 151 (48h)
IN (expected 289 bp)
1 2 3 4 5 6 7 8
OUT
1 2 3 4 5 6 7 8

FIG. 14C

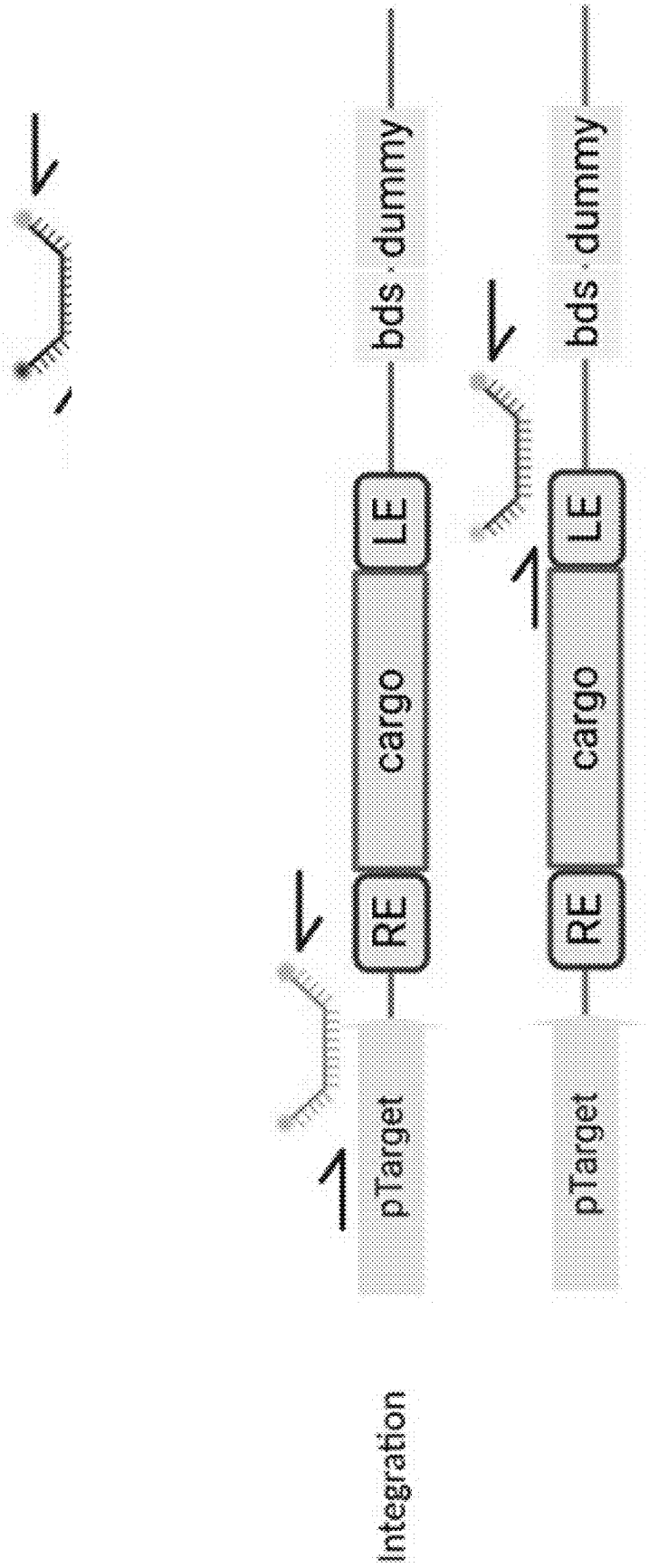

Integration
pTarget
RE
cargo
LE
bds · dummy
pTarget
RE
cargo
LE
bds · dummy

FIG. 15A

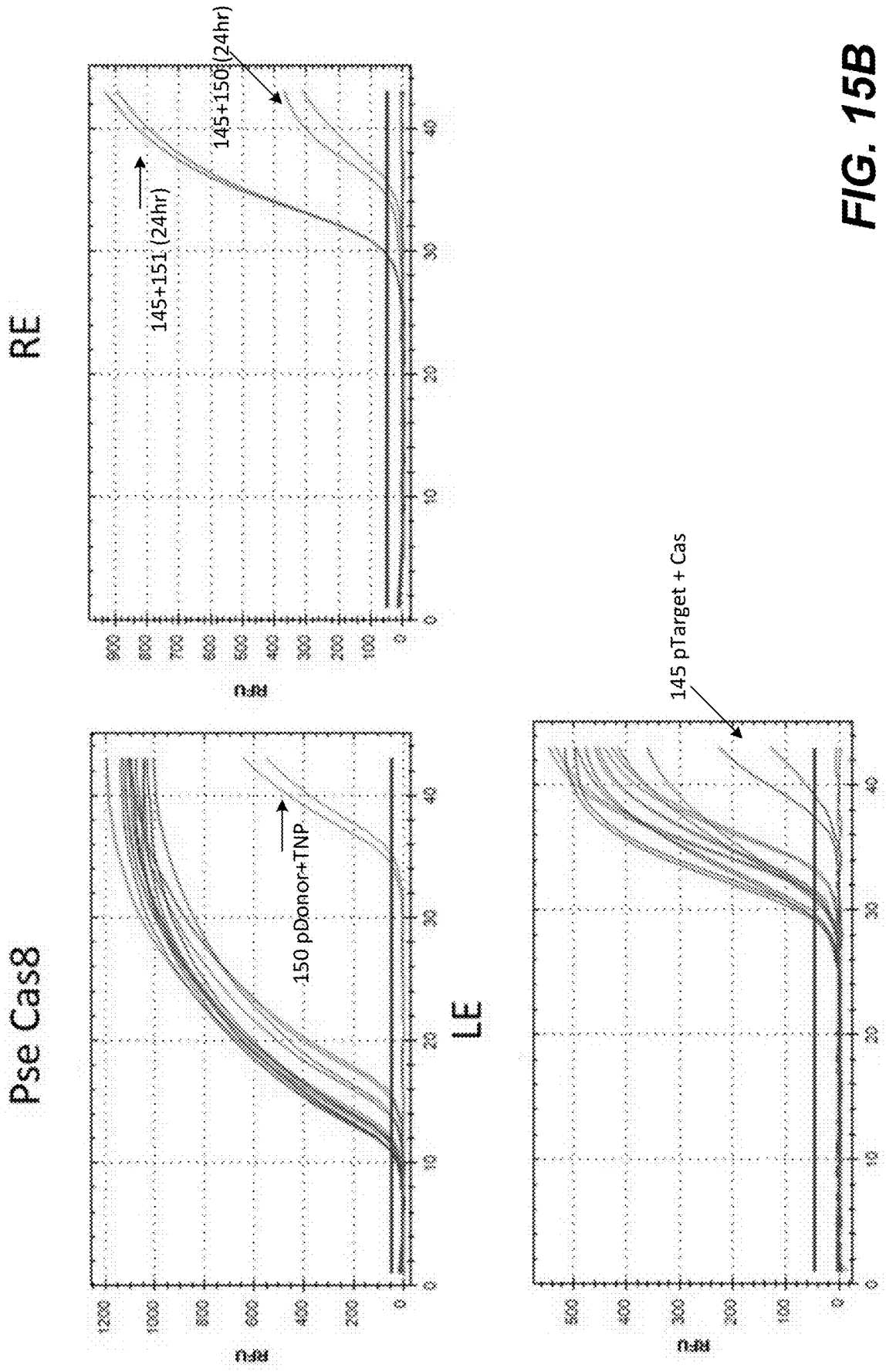
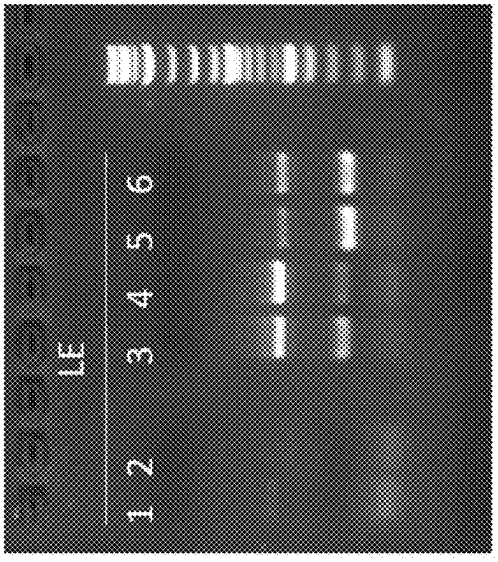
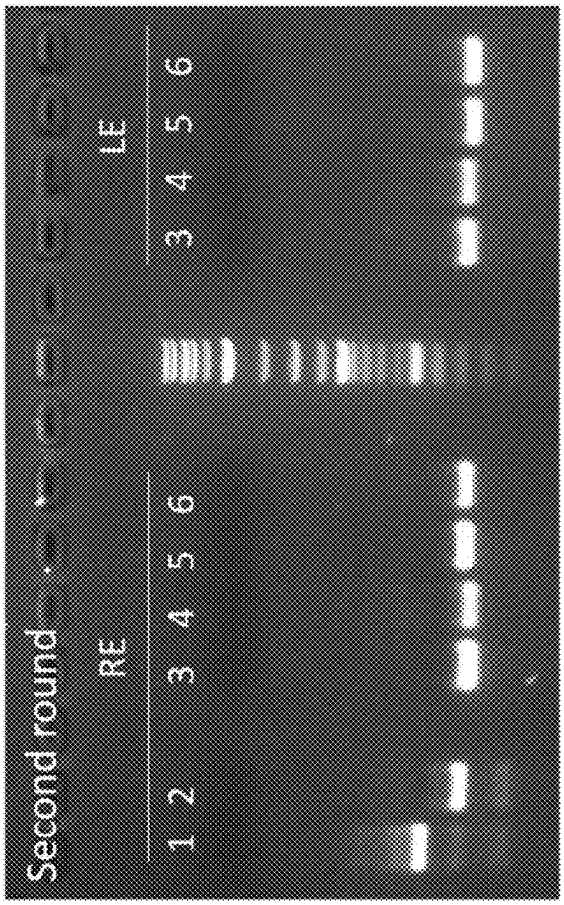

Pse Cas8
RE
LE
RFU
145+151 (24hr)
145+150 (24hr)
150 pDonor+TNP
145 pTarget + Cas

FIG. 15B





Template

1. 145 pTarget + CAS
2. 150 pDonor + TNP
3. #147 Random 200 – T2
4. #148 proless mCherry– T3
5. #150 Random 200 – T2
6. #151 proless mCherry– T3

FIG. 16A

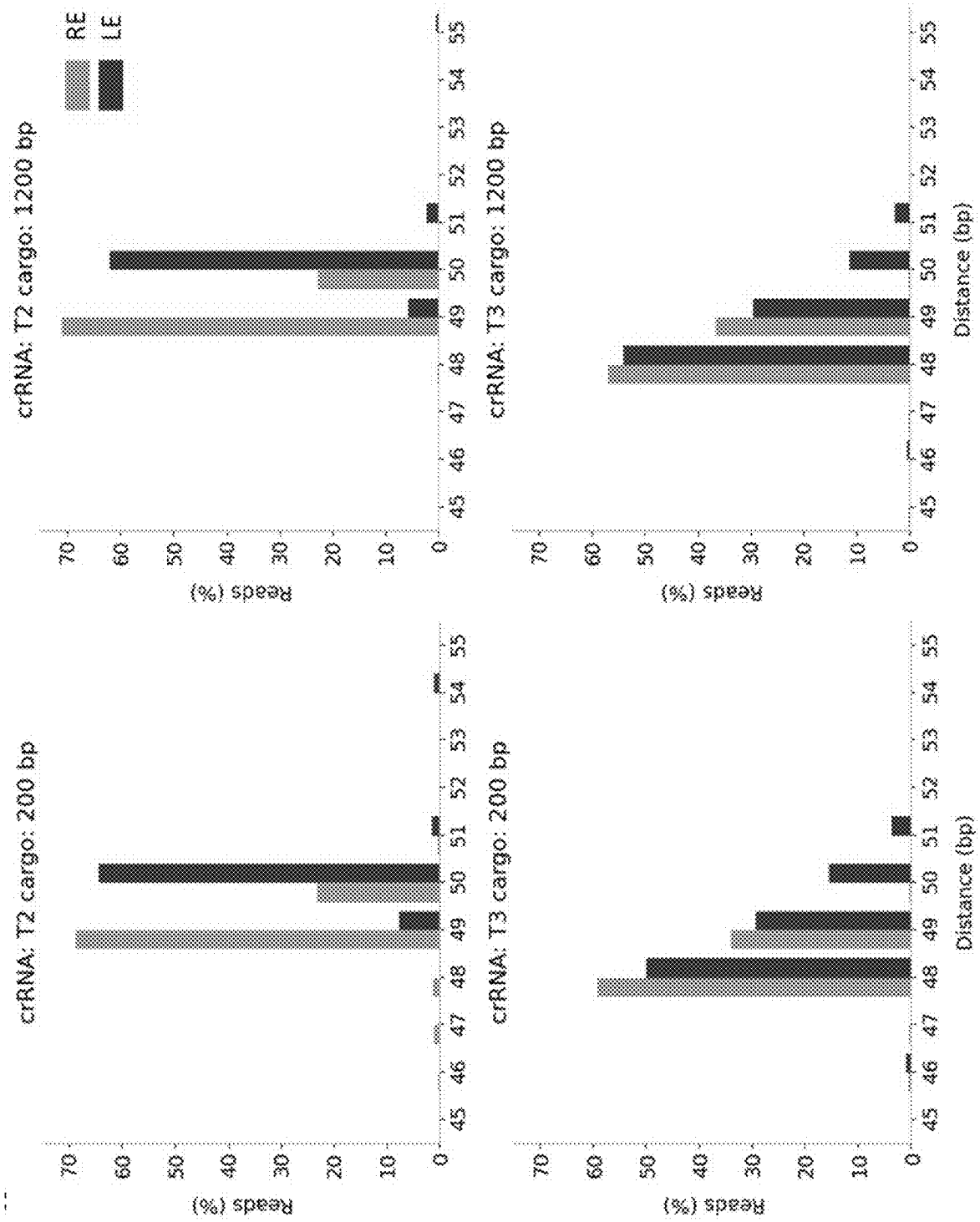

crRNA: T2 cargo: 200 bp
crRNA: T3 cargo: 200 bp
crRNA: T2 cargo: 1200 bp
crRNA: T3 cargo: 1200 bp
RE
LE
Reads (%)
Distance (bp)
0
10
20
30
40
50
60
70
45
46
47
48
49
50
51
52
53
54
55

FIG. 16B

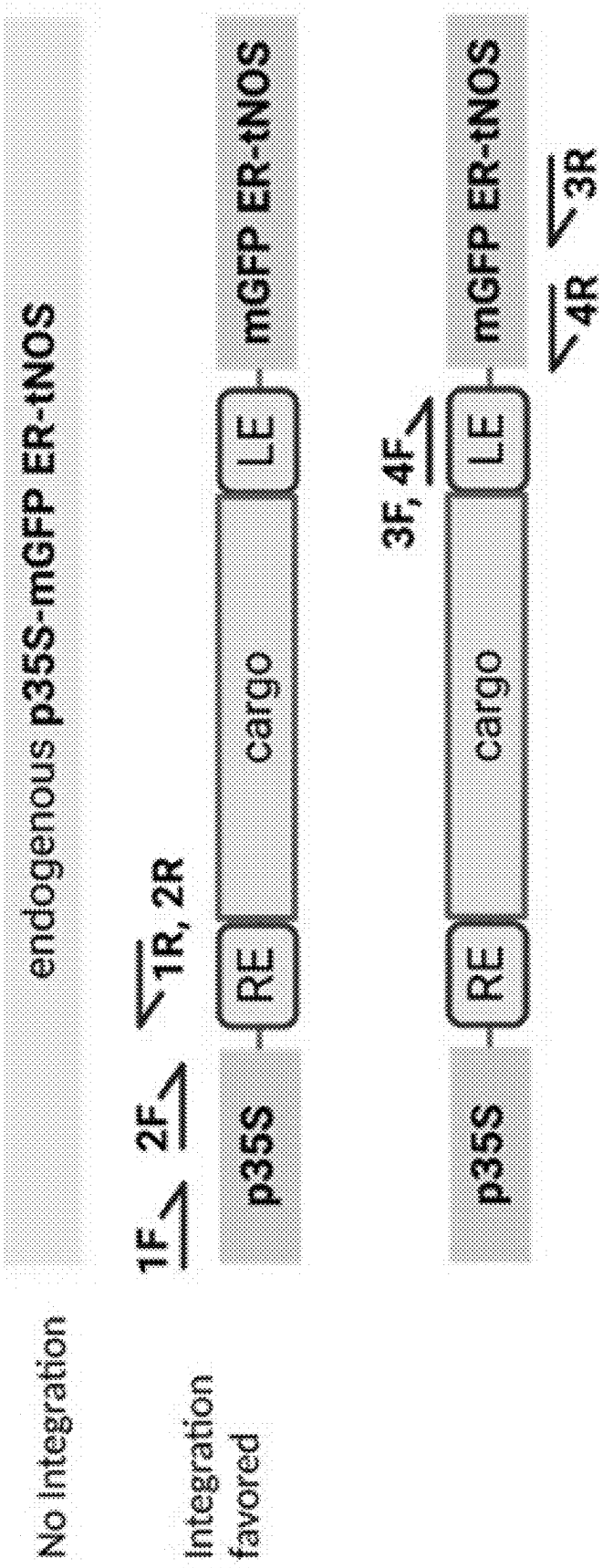
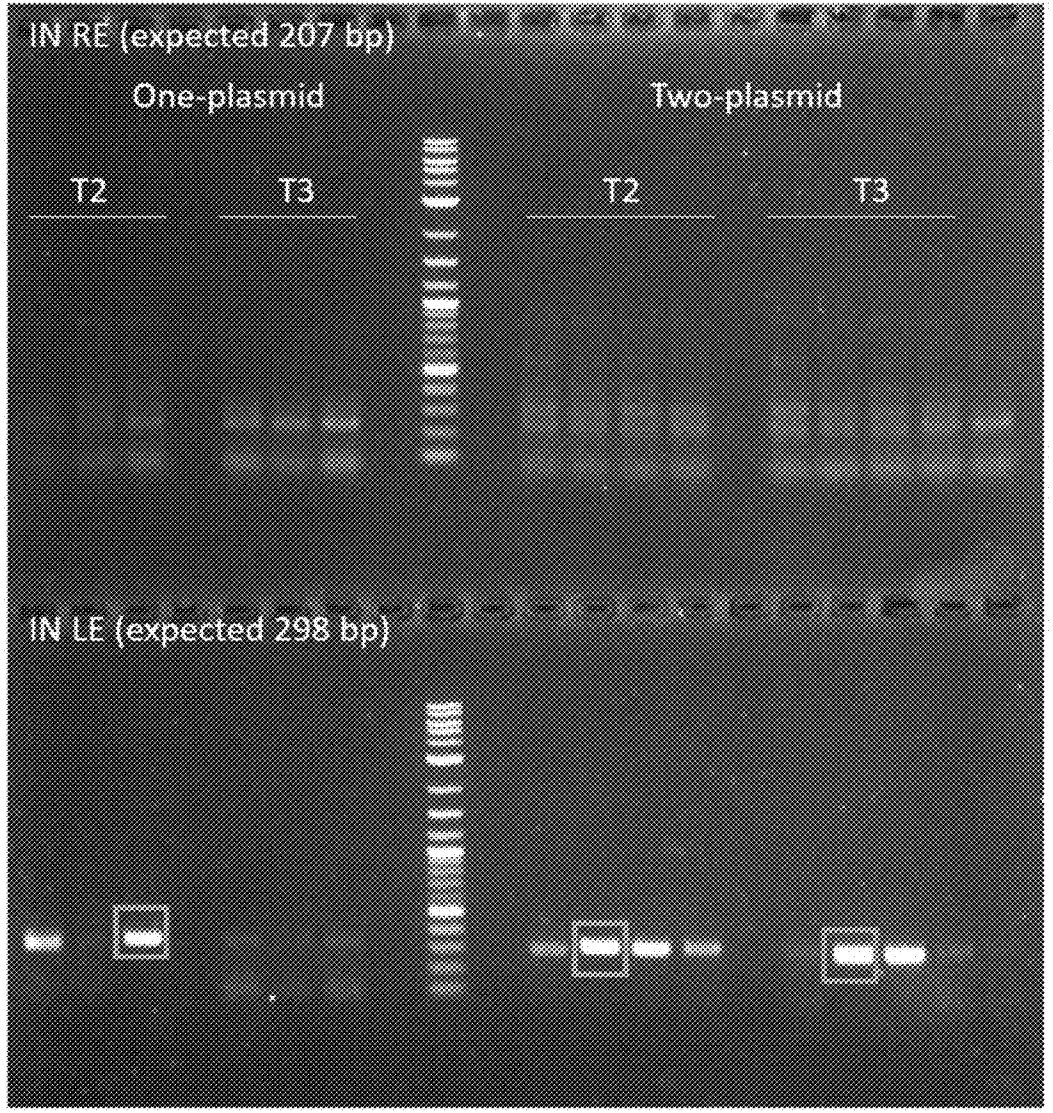

No Integration
endogenous p35S-mGFP ER-tNOS
1F
2F
1R, 2R
p35S
RE
cargo
LE
mGFP ER-tNOS
Integration favored
3F, 4F
4R
3R

FIG. 17A



Left to ladder
1. 120T2 #1
2. 120T2 #2
3. 120T2 #3
4. 120T3 #1
5. 120T3 #2
6. 120T3 #3

Right to ladder
1. 158+427T2 #1
2. 158+427T2 #2
3. 158+427T2 #3
4. 158+427T2 #4
5. 158+427T3 #1
6. 158+427T3 #2
7. 158+427T3 #3
8. 158+427T3 #4

FIG. 17B

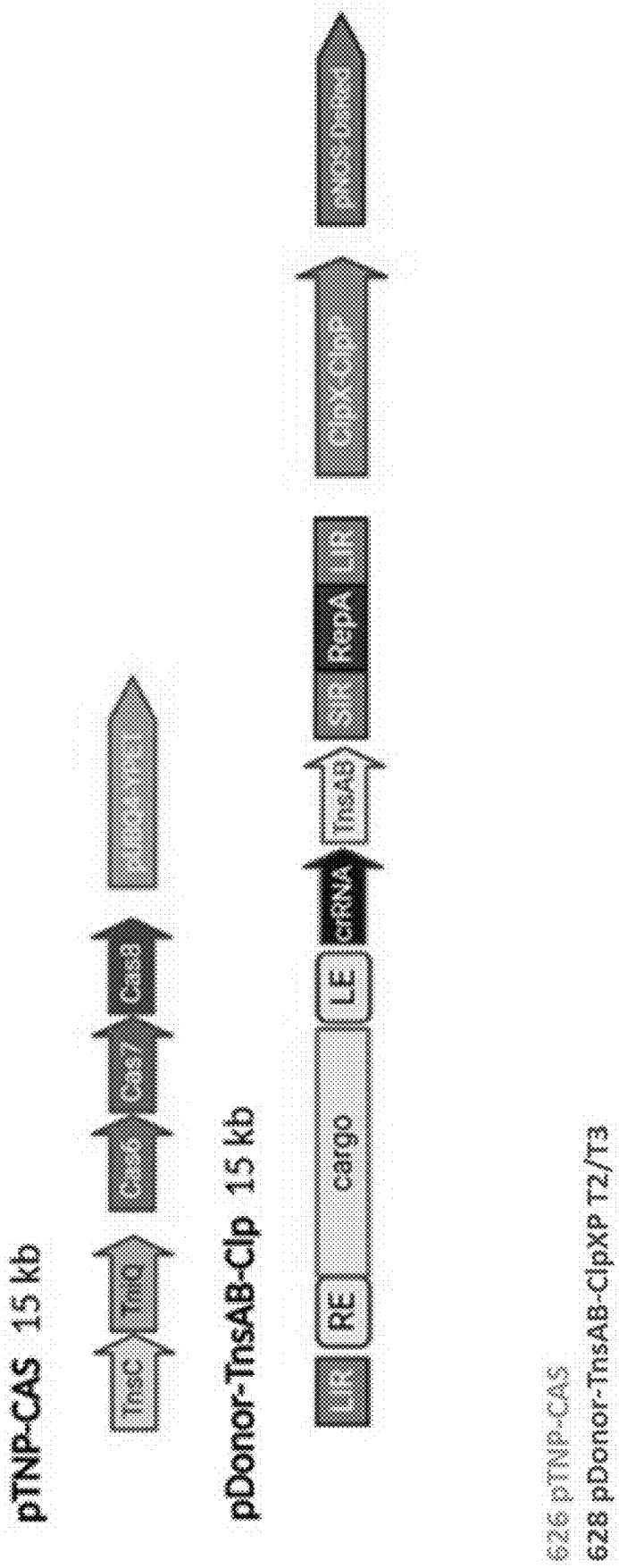

pTNP-CAS 15 kb
TnsC
TnsD
pDonor-TnsAB-Clp 15 kb
LIR
RE
cargo
LE
crRNA
TnsAB
SIR
RepA
LIR
626 pTNP-CAS
628 pDonor-TnsAB-ClpXP T2/T3

FIG. 18A

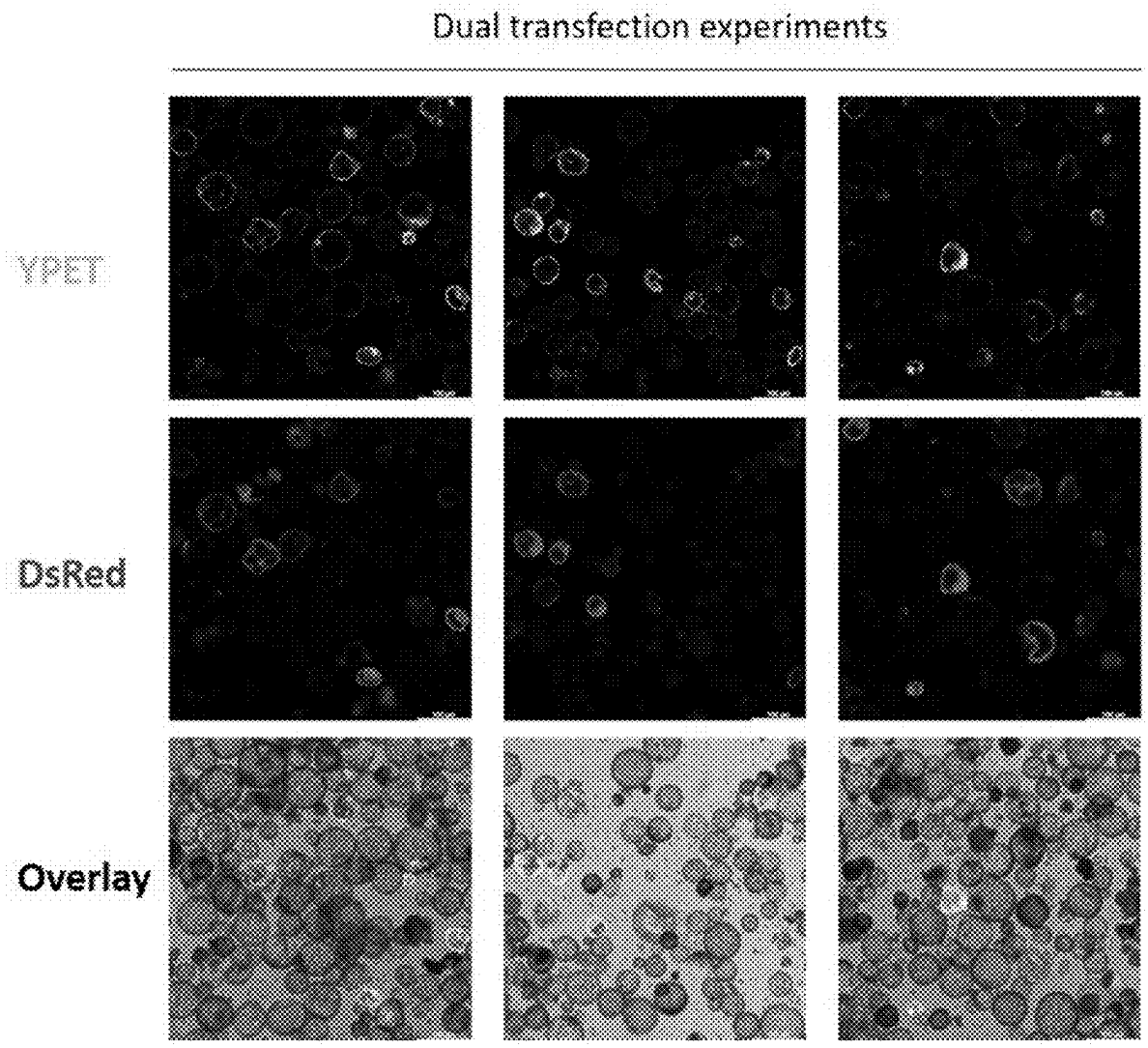


FIG. 18B

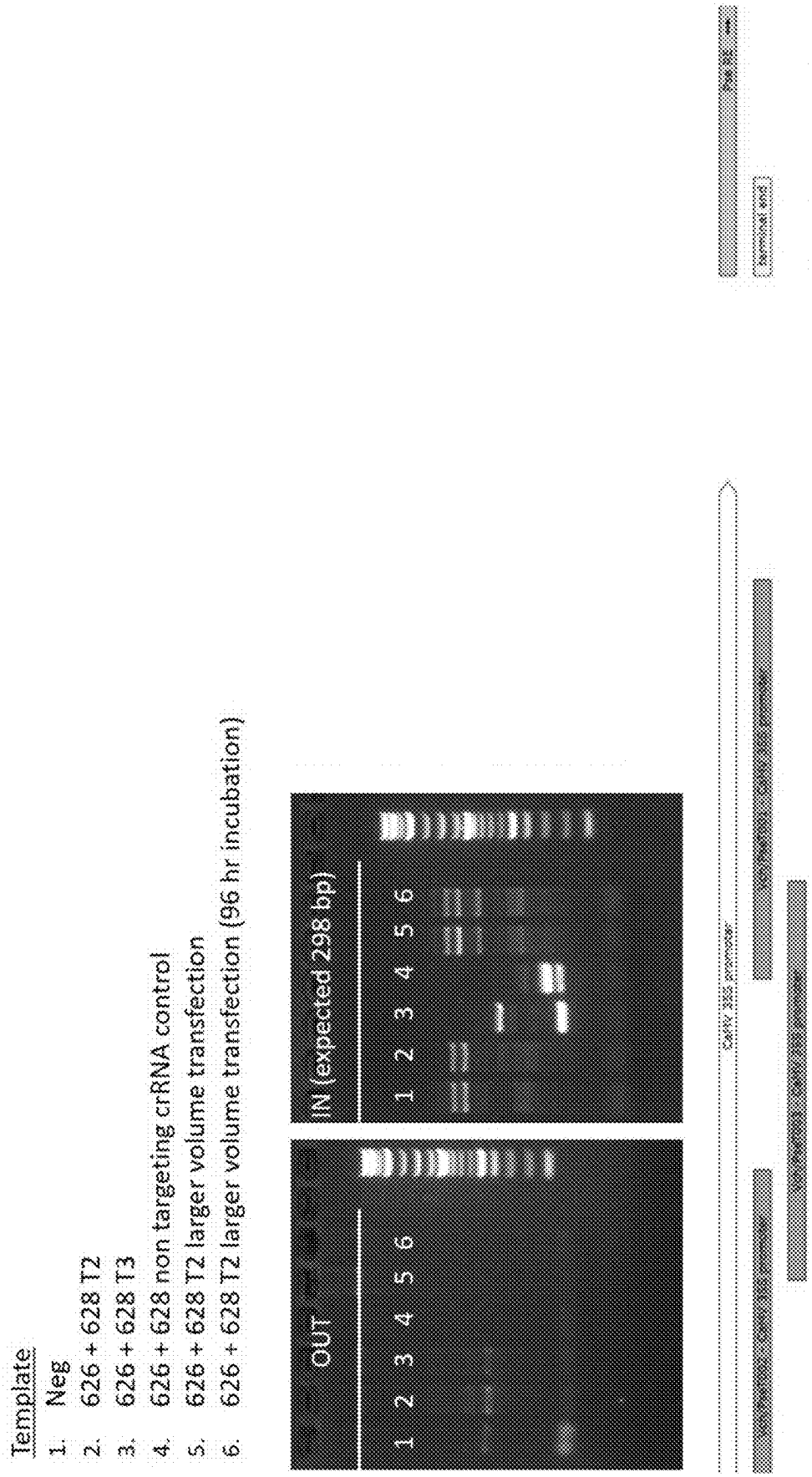

Template
1. Neg
2. 626 + 628 T2
3. 626 + 628 T3
4. 626 + 628 non targeting crRNA control
5. 626 + 628 T2 larger volume transfection
6. 626 + 628 T2 larger volume transfection (96 hr incubation)
OUT
1 2 3 4 5 6
IN (expected 298 bp)
1 2 3 4 5 6
Cas9 355 promoter
Spectinomycin

FIG. 19

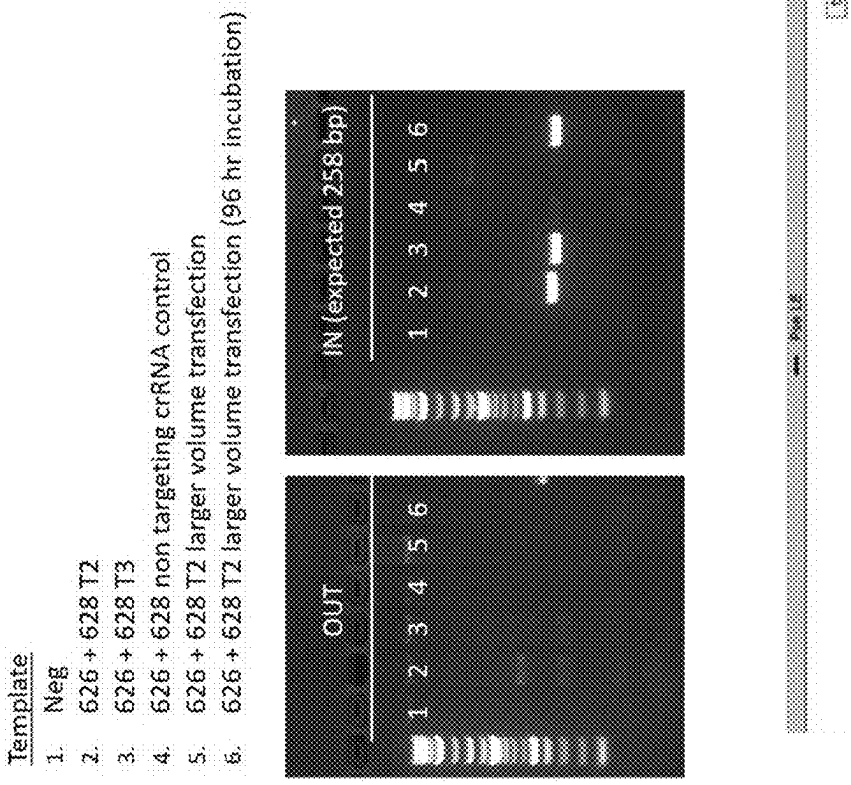

Template
1. Neg
2. 626 + 628 T2
3. 626 + 628 T3
4. 626 + 628 non targeting crRNA control
5. 626 + 628 T2 larger volume transfection
6. 626 + 628 T2 larger volume transfection (96 hr incubation)
OUT
1 2 3 4 5 6
IN (expected 258 bp)
1 2 3 4 5 6

FIG. 19

TARGETED DNA INTEGRATION IN PLANTS BY CRISPR-ASSOCIATED TRANSPOSASES (CASTS)

RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Application No. 63/551,713, filed Feb. 9, 2024. The entire contents of this application is hereby expressly incorporated by reference in their entireties.

REFERENCE TO SEQUENCE LISTING

[0002] The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is provided as a file entitled 30KJ-810004-US_Sequence-Listing, created Feb. 6, 2025, which is 332,753 bytes in size. The information in the electronic format of the Sequence Listing is incorporated herein by reference in its entirety.

BACKGROUND

Field

[0003] The present disclosure relates generally to the field of genetic engineering.

Description of the Related Art

[0004] Climate change poses unprecedented challenges to food safety. The world's population is projected to reach 9.6 billion by 2050, resulting in a 60% surge in demand for crops, while detrimental environmental conditions continue to limit plant production. Given the projected reductions in crop yields, it becomes imperative to explore novel strategies to address the food security issue.

[0005] Plant genetic engineering has emerged as a promising avenue for imparting new traits, including enhanced resilience to environmental conditions, heightened crop yields, and improved quality. Nevertheless, engineering of plants faces multiple hurdles, ranging from the need for better understanding of plant genomes and expression profiles to the development of advanced DNA editing tools. Several challenges persist in the current genome editing toolbox that rely on targeted nucleases such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and CRISPR/Cas system. One critical consideration is that double-strand breaks (DSBs) can give rise to undesirable byproducts, including small insertions and deletions (indels), large-scale genomic deletions, and chromosomal translocations. Additionally, homology-directed repair (HDR)-mediated DNA insertion in plants is challenging because HDR is only initiated in the S and G2 phase of the cell cycle and necessitates the supply of enough repair template. There is a need for improved gene editing techniques in plants.

SUMMARY

[0006] Disclosed herein include systems for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the system comprises: i) an RNA-guided DNA binding complex or one or more first helper polynucleotides each comprising a sequence encoding a component of the RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the system comprises one or more helper accessory proteins or one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP; ii) a transposition complex or one or more second helper polynucleotides each comprising a sequence encoding a component of the transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

[0007] Disclosed herein include nucleic acid compositions for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the nucleic acid composition comprises: i) one or more first helper polynucleotides each comprising a sequence encoding a component of an RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the nucleic acid composition comprises one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP; ii) one or more second helper polynucleotides each comprising a sequence encoding a component of a transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

[0008] In some embodiments, at least one of the one or more first helper polynucleotides and/or at least one of the one or more second helper polynucleotides are comprised within a second autonomous replicon. In some embodiments, the first autonomous replicon, the second autonomous replicon, or both is derived from a geminivirus. In some embodiments, the geminivirus comprises cabbage leaf curl virus, tomato golden mosaic virus, bean yellow dwarf virus, African cassava mosaic virus, wheat dwarf virus, *miscanthus* streak mastrevirus, tobacco yellow dwarf virus, tomato yellow leaf curl virus, bean golden mosaic virus, beet curly top virus, maize streak virus, or tomato pseudo-curly top virus. In some embodiments, the donor polynucleotide comprised within the first autonomous replicon comprises, from 5' to 3': a first long intergenic region (LIR), the RE, the cargo sequence, the LE, a short intergenic region (SIR), a sequence encoding RepA, and a second LTR. In some embodiments, the at least one of the one or more first helper polynucleotides and/or the at least one of the one or more second helper polynucleotides comprised within the second autonomous replicon comprises, from 5' to 3': a first long intergenic region (LIR), the first helper polynucleotide or the second helper polynucleotide, a short intergenic region

(SIR), a sequence encoding RepA, and a second LIR. In some embodiments, the first and/or second LIR comprise the sequence of SEQ ID NO: 1 or a sequence having one, two, or three mismatches relative to the sequence of SEQ ID NO: 1; the SIR comprises the sequence of SEQ ID NO: 2 or a sequence having one, two, or three mismatches relative to the sequence of SEQ ID NO: 2; and the sequence encoding RepA comprises the sequence of SEQ ID NO: 3 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 3. In some embodiments, the amount of the donor polynucleotide in the plant cell is capable of increasing by at least 2-fold following the onset of autonomous replication. In some embodiments, the amount of the donor polynucleotide in the plant cell is capable of increasing by at least 10-fold following the onset of autonomous replication. In some embodiments, the amount of a gene product encoded by the donor polynucleotide in the plant cell is capable of increasing by at least 2-fold following the onset of autonomous replication. In some embodiments, the amount of the gene product encoded by the donor polynucleotide in the plant cell is capable of increasing by at least 10-fold following the onset of autonomous replication.

[0009] In some embodiments, the LE comprises the sequence of any one of SEQ ID NOs: 4-5 or a sequence having one, two or three mismatches relative to the sequence of any one of SEQ ID NOs: 4-5 and wherein the RE comprises the sequence of any one of SEQ ID NOs: 6-7 or a sequence having one, two or three mismatches relative to the sequence of any one of SEQ ID NOs: 6-7. In some embodiments, the cargo sequence is 0.2 to 1000 kb in length. In some embodiments, the cargo sequence is 200 to 1200 bp in length.

[0010] In some embodiments, the cargo sequence comprises one or more exogenous sequences each selected from the group comprising: a nitrogen fixation gene, a plant stress-induced gene, a nutrient utilization gene, a gene that affects plant pigmentation, a gene that encodes an antisense or ribozyme molecule, a gene encoding an antigen capable of being secreted, a toxin gene, a receptor gene, a ligand gene, a seed storage gene, a hormone gene, an enzyme gene, an interleukin gene, a cytokine gene, a growth factor gene, a transcription factor gene, a transcriptional repressor gene, a DNA-binding protein gene, a recombination gene, a DNA replication gene, a programmed cell death gene, a kinase gene, a phosphatase gene, a G protein gene, a cyclin gene, a cell cycle control gene, a gene involved in transcription, a gene involved in translation, a gene involved in RNA processing, a gene involved in RNAi, an organellar gene, a intracellular trafficking gene, an integral membrane protein gene, a transporter gene, a membrane channel protein gene, a cell wall gene, a gene involved in protein processing, a gene involved in protein modification, a gene involved in protein degradation, a gene involved in metabolism, a gene involved in biosynthesis, a gene involved in assimilation of nitrogen or other elements or nutrients, a gene involved in controlling carbon flux, gene involved in respiration, a gene involved in photosynthesis, a gene involved in light sensing, a gene involved in organogenesis, a gene involved in embryogenesis, a gene involved in differentiation, a gene involved in meiotic drive, a gene involved in self incompatibility, a gene involved in development, a gene involved in nutrient, metabolite or mineral transport, a gene involved in nutrient, metabolite or mineral storage, a calcium-binding protein gene, and a lipid-binding protein gene. In some embodiments, the cargo sequence comprises one or more exogenous sequences each selected from the group comprising: a gene encoding an enzyme involved in metabolizing biochemical wastes for use in bioremediation, a gene that encodes an enzyme for modifying pathways that produce secondary plant metabolites, a gene that encodes an enzyme that produces a pharmaceutical, a gene that encodes an enzyme that improves or changes the nutritional content of a plant, a gene that encodes an enzyme involved in vitamin synthesis, a gene that encodes an enzyme involved in carbohydrate, polysaccharide or starch synthesis, a gene that encodes an enzyme involved in mineral accumulation or availability, a gene that encodes a phytase, a gene that encodes an enzyme involved in fatty acid, fat or oil synthesis, a gene that encodes an enzyme involved in synthesis of chemicals or plastics, a gene that encodes an enzyme involved in synthesis of a fuel, a gene that encodes an enzyme involved in synthesis of a fragrance, a gene that encodes an enzyme involved in synthesis of a flavor, a gene that encodes an enzyme involved in synthesis of a pigment or dye, a gene that encodes an enzyme involved in synthesis of a hydrocarbon, a gene that encodes an enzyme involved in synthesis of a structural or fibrous compound, a gene that encodes an enzyme involved in synthesis of a food additive, a gene that encodes an enzyme involved in synthesis of a chemical insecticide, a gene that encodes an enzyme involved in synthesis of an insect repellent, and a gene controlling carbon flux in a plant. In some embodiments, the cargo sequence comprises an exogenous sequence encoding a fluorescent protein. In some embodiments, the fluorescent protein comprises mPlum, mCherry, DsRed, FASTred, mOrange, EYFP, VENUS, YPet, GFP, EGFP, EmGFP, mCFP, Cerulean, CyPet, Kaede, or any combination thereof.

[0011] In some embodiments, the one or more Cas proteins comprise a Cas6 protein, a Cas7 protein, and a Cas8 protein. In some embodiments, the Cas6 protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 8-9; the Cas7 protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 10-11; and/or the Cas8 protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 12-13. In some embodiments, the transposase of the RNA-guided DNA binding complex comprises a TniQ protein. In some embodiments, the TniQ protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 14-15. In some embodiments, the one or more transposases of the transposition complex comprise a TnsA transposase, a TnsB transposase, and a TnsC protein. In some embodiments, the one or more transposases of the transposition complex comprise a TnsAB fusion protein and a TnsC protein. In some embodiments, the TnsAB fusion protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 16-17; and/or the TnsC protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 18-19. In some embodiments, the system, the RNA-guided DNA binding complex and/or the transposition complex is derived from a Type I-B, Type I-D, TypeI-F, or Type V-K Crispr-associated transposase system of a bacte-

ria. In some embodiments, the bacteria comprise *Vibrio cholera* (Vch), *Pseudoalteromonas* (Pse), or Scytonema hofmanni (Sho). In some embodiments, the ClpX comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 21; and/or the ClpP comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 20. In some embodiments, at least one of the one or more Cas proteins and/or the transposase of the RNA-guided DNA binding complex comprise a nuclear localization signal (NLS). In some embodiments, the NLS is an N-terminal NLS or a C-terminal NLS; at least one of the one or more transposases of the transposition complex comprises an NLS. In some embodiments, the NLS is an N-terminal NLS or a C-terminal NLS; and/or at least one of the one or more helper accessory proteins comprises an NLS. In some embodiments, the NLS is an N-terminal NLS or a C-terminal NLS. In some embodiments, the TnsAB fusion protein comprises an NLS between the TnsA amino acid sequence and the TnsB amino acid sequence. In some embodiments, the TnsAB fusion protein comprises, from N-terminus to C-terminus: TnsA, the NLS, and TnsB. In some embodiments, the NLS comprises an amino acid sequence encoded by a nucleotide sequence of any one of SEQ ID NOs: 22-23 or a sequence have one, two, or three mismatches relative to any one of SEQ ID NOs: 22-23.

[0012] In some embodiments, the crRNA comprises a spacer that is complementary to a search target sequence on a first strand of the double stranded target sequence. In some embodiments, the crRNA comprises a [repeat scaffold]-[spacer]-[repeat scaffold] structure. In some embodiments, the first strand of the double stranded target sequence is the sense strand. In some embodiments, the cargo sequence is capable of being integrated at an integration site following binding of the RNA-guided DNA binding complex to the search target sequence, wherein the integration site is about 48 to 52 base pairs downstream of the double stranded target sequence. In some embodiments, the double stranded target sequence is situated within a selectable marker gene of the genome of the plant cell. In some embodiments, the selectable marker gene comprises a fluorescent protein coding gene, a phytoene desaturase (PDS) gene, a codA gene, a diphtheria toxin a subunits (DT-A) gene, an exotoxin A gene, a ricin toxin A gene, a cytochrome P-450 gene, an RNase T1 gene, or a barnase gene. In some embodiments, the double stranded target sequence is situated within a safe harbor locus of the genome of the plant cell.

[0013] In some embodiments, each of the one or more first helper polynucleotides comprises a first promoter operably linked to the sequence encoding the component of the RNA-guided DNA binding complex; each of the one or more second helper polynucleotides comprises a second promoter operably linked to the sequence encoding the component of the transposition complex; and/or each of the one or more helper accessory polynucleotides comprises a third promoter operably linked to the sequence encoding at least one of the one or more helper accessory proteins. In some embodiments, the first, second, and/or third promoters are the same or different. In some embodiments, the first, second, and/or third promoters comprise a ubiquitous promoter, a constitutive promoter, a cell-type specific promoter, a tissue-specific promoter, an inducible promoter, or any combination thereof. In some embodiments, the constitutive promoter is selected from the group comprising: pCmYLCV911 (pCmY), pU6, pU3, pU6, pAct2, pAct-1, pUBQ10, pUBQ4, pUbi1, and PUbi2; the tissue-specific promoter is selected from the group comprising: pSIREO, pNAC10, pPAT21, phspr, pPFn2, pPEPC, PLhcb, pTA29, pLat52, pZml3, pOleosin, pGlutenin, pD-hordein, and pE8; the inducible promoter is selected from the group comprising: pAdh-1, pwunl, pGBSS, pHSP18.2, pRd29, pSR2, pCCA1, pUGT71C5, pGSE, pwin3.12, pR2329, pBs3, pCaPrx, p4xM1.1, p4xM2.3, pIFS2, pSAG12, pSEOF1, pEm, pRd29, pSAUR15A, and pChn48. In some embodiments, each of the one or more first helper polynucleotides comprises a first transcription terminator operably linked to the sequence encoding the component of the RNA-guided DNA binding complex; each of the one or more second helper polynucleotides comprises a second transcription terminator operably linked to the sequence encoding the component of the transposition complex; and/or each of the one or more helper accessory polynucleotides comprises a third transcription terminator operably linked to the sequence encoding at least one of the one or more helper accessory proteins. In some embodiments, the first, second, and/or third transcription terminators are the same or different. In some embodiments, the first, second, and/or third transcription terminators comprise AtHSP18.2 (tHSP), tU6, tACT3, tACT3-tRb7MAR, tACT3-tTM6MAR, tEU, tEU-tTM6MAR, tEU (intronless), tEU (intronless)-tACT3-tRB7MAR, tHSP18-tEU-tRb7MAR, tHSP18-tACT3, tHSP18-tACT3-tRb7, tHSP18-tPINII-tRb7MAR, tHSP18-tPINII-tTM6MAR, tHSP18-tRb7MAR, tProteinase inhibitor II (tPINII), trbcS, or any combination thereof.

[0014] The system or the nucleic acid composition can comprise: at least three first helper polynucleotides each comprising a sequence encoding a Cas protein, wherein the sequence FIG. **3** helper polynucleotide comprising a sequence encoding a transposase protein, wherein the sequence encoding the transposase protein is operably linked to a pCmY promoter and a tHSP terminator; a first helper polynucleotide comprising a sequence encoding a crRNA, wherein the sequence encoding the crRNA is operably linked to a pU6 promoter and a tU6 terminator; and at least two second helper polynucleotides each comprising a sequence encoding a transposase, wherein the sequence encoding a transposase is operably linked to a pCmY promoter and a tHSP terminator. The system or the nucleic acid composition can comprise at least two helper accessory polynucleotides, wherein the sequence encoding at least one of the one or more helper accessory proteins is operably linked to a pCmY promoter and a tHSP terminator.

[0015] In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex, the sequence encoding the component of the transposition complex, the sequence encoding at least one of the one or more helper accessory proteins, or any combination thereof, is codon optimized for expression in the plant cell. In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas6 protein. In some embodiments, the sequence encoding the Cas6 protein comprises the nucleotide sequence of any one of SEQ ID NOs: 24-25 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 24-25. In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas7 protein. In some embodiments, the sequence encoding the Cas7 protein comprises the nucleo-

tide sequence of any one of SEQ ID NOs: 26-27 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 26-27. In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas8 protein. In some embodiments, the sequence encoding the Cas8 protein comprises the nucleotide sequence of any one of SEQ ID NOs: 28-29 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 28-29. In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex encodes a TniQ protein. In some embodiments, the sequence encoding the TniQ protein comprises the nucleotide sequence of any one of SEQ ID NOs: 30-31 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 30-31. In some embodiments, the sequence encoding the component of the transposition complex encodes a TnsAB fusion protein. In some embodiments, the sequence encoding the TnsAB fusion protein comprises the nucleotide sequence of any one of SEQ ID NOs: 32-33 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 32-33. In some embodiments, the sequence encoding the component of the transposition complex encodes a TnsC protein. In some embodiments, the sequence encoding the TnsC protein comprises the nucleotide sequence of any one of SEQ ID NOs: 34-35 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 34-35. In some embodiments, the sequence encoding ClpX comprises the nucleotide sequence of SEQ ID NO: 37 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 37 and the sequence encoding ClpP comprises the nucleotide sequence of SEQ ID NO: 36 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 36.

[0016] In some embodiments, the component of the RNA-guided DNA binding complex, the component of the transposition complex, or both, comprises an N-terminal or a C-terminal tag. In some embodiments, the tag is an epitope tag. In some embodiments, the epitope tag comprises a myc tag, a FLAG tag, a polyHistidine tag, a HiBiT tag, HA tag, S-peptide tag, calmodulin-binding peptide (CBP), glutathione S-transferase (GST), maltose binding protein (MBP), or any combination thereof. In some embodiments, the tag comprises a fluorescent protein. In some embodiments, the fluorescent protein comprises mPlum, mCherry, DsRed, FASTred, mOrange, EYFP, VENUS, YPet, GFP, EGFP, EmGFP, mCFP, Cerulean, CyPet, Kaede, or any combination thereof.

[0017] In some embodiments, the one or more first helper polynucleotides the one or more second helper polynucleotides, and/or the donor polynucleotide are situated on the same nucleic acid or different nucleic acids. In some embodiments, the one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide are comprised within one or more vectors. In some embodiments, the one or more vectors comprise an RNA viral vector, a DNA viral vector, a plasmid vector, an artificial chromosome, or any combination thereof. In some embodiments, the one or more vectors comprise an *Agrobacterium tumefaciens* Ti vector. In some embodiments, the one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide are comprised within a T-DNA region of the *Agrobacterium tumefaciens* Ti vector. In some embodiments, the T-DNA region comprising the one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide comprises the sequence of any one of SEQ ID NOs: 38-46.

[0018] Disclosed herein include methods for integration of a nucleic acid sequence into double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the method comprises: contacting the plant cell with a system or the nucleic acid composition of the disclosure, wherein the cargo sequence is integrated at an integration site in the genome of the plant cell or at a target site of the target plasmid upon expression of the RNA-guided DNA binding complex and the transposition complex in the plant cell.

[0019] In some embodiments, the integration site is about 48 to 52 base pairs downstream of the double stranded target sequence. In some embodiments, the one or more Cas proteins and the transposase of the RNA-guided DNA binding complex are pre-complexed with the crRNA prior to the contacting. In some embodiments, the plant cell is comprised within a plant. In some embodiments, the plant cell is comprised within a flower, a leaf, a stem, a root, terminal bud, a seed, or any other tissue of the plant. In some embodiments, the plant cell is a monocot plant cell or a eudicot plant cell. In some embodiments, the integration of the cargo sequence confers i) a change in one or more of the following traits to the plant: grain number, grain size, grain weight, panicle size, tiller number, fragrance, nutritional value, shelf life, lycopene content, starch content and/or ii) lower gluten content, reduced levels of a toxin, reduced levels of steroidal glycoalkaloids, a substitution of mitosis for meiosis, asexual propagation, improved haploid breeding, and/or shortened growth time. In some embodiments, the integration of the cargo sequence confers one or more of the following traits to the plant cell and/or the plant: herbicide tolerance, drought tolerance, male sterility, insect resistance, abiotic stress tolerance, modified fatty acid metabolism, modified carbohydrate metabolism, modified seed yield, modified oil percent, modified protein percent, resistance to bacterial disease, resistance to fungal disease, and resistance to viral disease. In some embodiments, the system or the nucleic acid composition is introduced into the plant cell by a technique comprising: pollen tube pathway, polyethylene glycol (PEG)-mediated gene transfer, electroporation, microinjection, microparticle bombardment, nanomaterial-mediated delivery, *Agrobacterium tumefaciens*-mediated transformation, or any combination thereof. In some embodiments, the nanomaterial-mediated delivery comprises: clay nanosheets, carbon nanotubes, carbon nanodots, self-assembled protein nanoparticles, peptides, DNA nanostructures, quantum dots, or any combination thereof. In some embodiments, the one or more vectors are introduced into the plant cell via *Agrobacterium tumefaciens*-mediated transformation of the plant cell.

[0020] Disclosed herein include methods for screening for safe harbor loci in plants. In some embodiments, the method comprises: (a) generating a genome-wide crRNA library; (b) contacting a plant cell comprised within a plant with a system or the nucleic acid composition of disclosure, wherein: the system comprises pooled single or combinatorial crRNAs generated in step (a); or the one or more first helper polynucleotides comprise pooled single or combina-

torial crRNAs generated in step a), wherein the cargo sequence is integrated into one or more double-stranded targets sites in the genome of the plant cell upon expression of the RNA-guided DNA binding complex and the transposition complex in the plant cell; (c) identifying integrants by expression of a gene product encoded by the cargo sequence; (d) subjecting the integrants to next-generation sequencing; and (e) performing bioinformatics analysis, a high-throughput phenotypic assay, or both to identify a safe harbor locus.

[0021] In some embodiments, the plant is a monocot plant or a eudicot plant. In some embodiments, integration of the cargo sequence at the identified safe harbor locus does not affect the growth, lifespan, health, gene expression profile, or any combination thereof, of the plant. In some embodiments, the system or the nucleic acid composition is introduced into the plant cell by a technique comprising: pollen tube pathway, polyethylene glycol (PEG)-mediated gene transfer, electroporation, microinjection, microparticle bombardment, nanomaterial-mediated delivery, *Agrobacterium tumefaciens*-mediated transformation, or any combination thereof. In some embodiments, the nanomaterial-mediated delivery comprises: clay nanosheets, carbon nanotubes, carbon nanodots, self-assembled protein nanoparticles, peptides, DNA nanostructures, quantum dots, or any combination thereof. In some embodiments, the one or more vectors are introduced into the plant cell via *Agrobacterium tumefaciens*-mediated transformation of the plant cell. In some embodiments, the T-DNA region of the Ti vector comprises a bi-directional selection marker comprising a positive selection marker and a negative selection marker, wherein the identifying of step c) comprises: (i) generating a first filial generation (F1) plant comprising the T-DNA of the Ti vector by positive selection; and (ii) generating a second filial generation (F2) plant that does not comprise the T-DNA of the Ti vector from the first filial generation plant comprising the T-DNA of the Ti vector, by negative selection.

[0022] Disclosed herein include kits comprising a system or nucleic acid composition described herein, and a set of instructions for use.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0024] FIG. **1** displays a construct design for helper, donor, and target plasmids.

[0025] FIG. **2**A-FIG. **2**D display exemplary data related to quantification of luminescence signal from HiBiT-tagged CAST proteins. FIG. **2**A displays a schematic of construct design. The upper panel was used to express the VchCAST and PseCAST proteins with the exception of the TnsA-B construct in which the Bipartite (BP) NLS was inserted in between the two proteins. The lower panel was used to express the ShoCAST proteins. FIG. **2**B-FIG. **2**D display exemplary expression data. Data is presented as luminescence values of three replicates normalized to total protein concentration per sample for the VchCAST, PseCAST and ShoCAST nuclear and cytoplasmic fractions respectively. One way ANOVA test ($p<0.05$).

[0026] FIG. **3**A-FIG. **3**B display schematics of CAST-mediated episomal DNA integration in protoplast cells. FIG. **3**A shows an exemplary construct design: Upper panel illustrates the generic Helper, Donor, and Target plasmid designs. Lower panel shows the target loci before and after integration. FIG. **3**B displays a schematic of an experimental procedure of CAST-mediated episomal DNA integration: PEG transfection of genetic materials into *A. thaliana* protoplasts and downstream phenotypic and molecular analysis.

[0027] FIG. **4** displays representative images showing DsRed- and mTurq-positive protoplasts confirmed the co-transfection of plasmids and YFP-positive protoplasts and the successful integration events.

[0028] FIG. **5**A-FIG. **5**C display exemplary data related to the successful integration junctions detected via nested PCR. FIG. **5**A shows primer design for OUT and IN nested PCR reactions for detecting the junctions in favored orientation of Target-RE-Cargo-LE. Nested PCR results of the (FIG. **5**B) RE and (FIG. **5**C) LE junctions of Vch and Pse-mediated integration in protoplasts. Squares indicate the expected bands with the correct integration product size.

[0029] FIG. **6**A-FIG. **6**D display data and schematics showing successful integration junctions detected via Sanger sequencing. Integration junctions from (FIG. **6**A) Vch T2 (FIG. **6**B) Vch T3 (FIG. **6**C) Pse T2 and (FIG. **6**D) Pse T3 RE and LE amplification products are shown.

[0030] FIG. **7**A-FIG. **7**B display representative schematics of CAST-mediated DNA integration into the plant genome. FIG. **7**A displays a representative construct design. Upper panel illustrates the generic Helper and Donor plasmids design. Lower panel illustrates the genomic target loci before and after integration. FIG. **7**B displays a schematic of an experimental procedure of CAST-mediated chromosomal DNA integration, including *Agrobacterium*-mediated transformation of genetic materials into *N. benthamiana* leaves, and downstream phenotypic and molecular analysis.

[0031] FIG. **8**A-FIG. **8**C display exemplary data related to the successful integration junctions detected via nested PCR and Sanger sequencing. FIG. **8**A shows primer design for OUT and IN nested PCR reactions for detecting the RE junction in favored orientation Target-RE-Cargo-LE. FIG. **8**B shows exemplary nested PCR results of the RE junctions of Vch and Pse mediated integration in protoplasts. Square indicates the expected band with the correct integration product size. Shown in FIG. **8**C displays a diagram of the integration site from Pse T2 RE.

[0032] FIG. **9**A-FIG. **9**B display schematics of engineering a stable transgenic *Arabidopsis* with CAST. FIG. **9**A shows the construct design. Upper panel illustrates the generic Helper and Donor plasmids design. Lower panel illustrates the genomic target loci PDS or codA before and after integration. Shown in FIG. **9**B is an experimental procedure of CAST mediated stable chromosomal DNA integration, including *Agrobacterium*-mediated transformation of genetic materials into *A. thaliana* flower, and downstream phenotypic and molecular analysis.

[0033] FIG. **10** displays a schematic of an exemplary protocol for application of CAST to screen for genomic safe harbor loci.

[0034] FIG. **11**A-FIG. **11**D display exemplary data related to confocal imaging and HiBiT lytic detection of CAST protein expression. FIG. **11**A displays schematics of construct design. FIG. **11**B displays a schematic of protein expression and detection workflow. FIG. **11**C displays exemplary images of confocal microscopy detection of HiBiT tagged YPet-VchCAST fusion proteins with an N-ter-

minus BP NLS. FIG. **11**D displays quantification of HiBiT tagged YPet-VchCAST fusion proteins detected as luminescence signal. n=1 biological replicate.

[0035] FIG. **12**A-FIG. **12**F display exemplary data related to quantification of luminescence signal from HiBiT-tagged CAST proteins. FIG. **12**A displays a schematic of the HiBiT Lytic Assay. FIG. **12**B displays a schematic of the construct design used to express the VchCAST and PseCAST proteins tagged with an N-terminal HiBiT peptide tag. Shown in FIG. **12**C-FIG. **12**D are luminescence values for the VchCAST and PseCAST proteins in nuclear and cytoplasmic fractions, respectively. Data are normalized to total protein concentration per sample and represent a minimum of three replicates. Ordinary one-way ANOVA test, *$p<0.05$, **$p<0.01$, ***$p<0.001$, and ****$p<0.0001$. FIG. **12**E displays a schematic of the construct design for CAST transposases (TnsA-B fusion) flanked by a Geminiviral replicon. FIG. **12**F shows a graph of luminescence values reflecting transposase protein expression for the VchCAST and PseCAST proteins. Data are normalized to total protein concentration per sample and represent a minimum of three replicates. Ordinary one-way ANOVA test, *$p<0.05$, **$p<0.01$, ***$p<0.001$, and ****$p<0.0001$. For all graphs, data are shown as aligned dot plots with horizontal lines indicating the mean.

[0036] FIG. **13**A-FIG. **13**C display exemplary data showing mTurq- and YPET-positive protoplasts confirmed the co-transfection at different timepoints. The time points shown are 12 hr (FIG. **13**A), 24 hr (FIG. **13**B), and 48 hr (FIG. **13**C). mTurq- and YPET-positive protoplasts confirmed the co-transfection at different timepoints. mCherry signal was detected only in experimental groups at 12 and 24 hours, but detected in both negative and experiment groups at 48 hours.

[0037] FIG. **14**A-FIG. **14**C display exemplary data showing successful integration junctions detected via nested PCR. Primer design for OUT and IN nested PCR reactions for detecting the junctions in favored orientation of Target-RE-Cargo-LE is shown in FIG. **14**A. Also shown are nested PCR results of the RE (FIG. **14**B) and LE (FIG. **14**C) junctions of PseCAST-mediated integration in protoplasts.

[0038] FIG. **15**A-FIG. **15**B display data showing quantification of integration events via nested TaqMan probe-based qPCR. TaqMan probes and primers targeting Cas8 gene, RE junction, and LE junction are shown in FIG. **15**A. FIG. **15**B shows qPCR amplification curve and Ct values.

[0039] FIG. **16**A-FIG. **16**B show exemplary amplicon sequencing to determine the integration site distribution. FIG. **16**A displays gel images of the first and second round of PCR, showing the expected bands. Shown in FIG. **16**B are graphs of integration distance profiles of different samples.

[0040] FIG. **17**A-FIG. **17**B display molecular analysis of CAST function in leaf genome. Primer design for OUT and IN nested PCR reactions for detecting the RE and LE junctions in favored orientation Target-RE-Cargo-LE is shown in FIG. **17**B. Shown in FIG. **17**B are nested PCR results of the RE and LE junctions from PseCAST-mediated integration in 16c *N. benthamiana* leaves. Square indicates the bands with the expected integrated product size.

[0041] FIG. **18**A-FIG. **18**B display construct design and confocal images of CAST integration in protoplast genome. FIG. **18**A displays a diagram of the constructs used for CAST integration: pTNP-CAS and pDonor-TnsAB-Clp. FIG. **18**B shows exemplary confocal images of YPET- and DsRed-positive protoplasts, confirming the successful co-transfection of both plasmids.

[0042] FIG. **19** displays exemplary data related to the successful integration junctions detected via nested PCR. Nested PCR results of the RE and LE junctions from Pse mediated integration in 16c *N. benthamiana* protoplasts. Also shown is a diagram of a simulated integrated sequence.

DETAILED DESCRIPTION

[0043] In the following detailed description, reference is made to the accompanying drawings, which form a part hereof. In the drawings, similar symbols typically identify similar components, unless context dictates otherwise. The illustrative embodiments described in the detailed description, drawings, and claims are not meant to be limiting. Other embodiments may be utilized, and other changes may be made, without departing from the spirit or scope of the subject matter presented herein. It will be readily understood that the aspects of the present disclosure, as generally described herein, and illustrated in the Figures, can be arranged, substituted, combined, separated, and designed in a wide variety of different configurations, all of which are explicitly contemplated herein and made part of the disclosure herein.

[0044] All patents, published patent applications, other publications, and sequences from GenBank, and other databases referred to herein are incorporated by reference in their entirety with respect to the related technology.

[0045] Disclosed herein include systems for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the system comprises: i) an RNA-guided DNA binding complex or one or more first helper polynucleotides each comprising a sequence encoding a component of the RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the system comprises one or more helper accessory proteins or one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP; ii) a transposition complex or one or more second helper polynucleotides each comprising a sequence encoding a component of the transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

[0046] Disclosed herein include nucleic acid compositions for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the nucleic acid composition comprises: i) one or more first helper polynucleotides each comprising a sequence encoding a component of an RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the nucleic acid composi-

tion comprises one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP; ii) one or more second helper polynucleotides each comprising a sequence encoding a component of a transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

[0047] Disclosed herein include methods for integration of a nucleic acid sequence into double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the method comprises: contacting the plant cell with a system or the nucleic acid composition of the disclosure, wherein the cargo sequence is integrated at an integration site in the genome of the plant cell or at a target site of the target plasmid upon expression of the RNA-guided DNA binding complex and the transposition complex in the plant cell.

[0048] Disclosed herein include methods for screening for safe harbor loci in plants. In some embodiments, the method comprises: (a) generating a genome-wide crRNA library; (b) contacting a plant cell comprised within a plant with a system or the nucleic acid composition of disclosure, wherein: the system comprises pooled single or combinatorial crRNAs generated in step (a); or the one or more first helper polynucleotides comprise pooled single or combinatorial crRNAs generated in step a), wherein the cargo sequence is integrated into one or more double-stranded targets sites in the genome of the plant cell upon expression of the RNA-guided DNA binding complex and the transposition complex in the plant cell; (c) identifying integrants by expression of a gene product encoded by the cargo sequence; (d) subjecting the integrants to next-generation sequencing; and (e) performing bioinformatics analysis, a high-throughput phenotypic assay, or both to identify a safe harbor locus.

[0049] Disclosed herein include kits comprising a system or nucleic acid composition described herein, and a set of instructions for use.

Definitions

[0050] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the present disclosure belongs. See, e.g. Singleton et al., Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, NY 1994); Sambrook et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, NY 1989). For purposes of the present disclosure, the following terms are defined below.

[0051] As used herein, the term "about" means plus or minus 5% of the provided value.

[0052] As used herein, the term "double-stranded target DNA" refers to a DNA that includes a "target site" or "target sequence." The term "target sequence" is used herein to refer to a nucleic acid sequence present in a double-stranded target DNA to which a DNA-targeting sequence or segment (also referred to herein as a "spacer") of a crRNA can hybridize, provided sufficient conditions for hybridization exist. For example, the target sequence 5'-GAGCATATC-3' within a target DNA is targeted by (or is capable of hybridizing with, or is complementary to) the RNA sequence 5'-GAUAUGCUC-3'. Hybridization between the DNA-targeting sequence or segment of a crRNA and the target sequence can, for example, be based on Watson-Crick base pairing rules, which enables programmability in the DNA-targeting sequence or segment. The DNA-targeting sequence or segment of a crRNA can be designed, for instance, to hybridize with any target sequence.

[0053] The terms "polynucleotide" and "nucleic acid" are used interchangeably herein and refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. A polynucleotide can be single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids/triple helices, or a polymer including purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. In some embodiments, a polynucleotide comprises a nucleotide sequence encoding a gene product operably linked to one or more expression control elements (e.g., a promoter), as an expression cassette.

[0054] As used herein, the term "binding" refers to a non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid). While in a state of non-covalent interaction, the macromolecules are said to be "associated" or "interacting" or "binding" (e g., when a molecule X is said to interact with a molecule Y, it means that the molecule X binds to molecule Y in a non-covalent manner). Binding interactions can be characterized by a dissociation constant (Kd), for example a Kd of, or a Kd less than, 10'6 M, 10" 7 M, 10'8 M, 10'9M, 10'10 M, 10'11 M, 10'12M, 10'13 M, IO'14 M, 10'15M, or a number or a range between any two of these values. Kd can be dependent on environmental conditions, e.g., pH and temperature. "Affinity" refers to the strength of binding, and increased binding affinity is correlated with a lower Kd.

[0055] The terms "complementarity" and "complementary" mean that a nucleic acid can form hydrogen bond(s) with another nucleic acid based on traditional Watson-Crick base paring rule, that is, adenine (A) pairs with thymine (U) and guanine (G) pairs with cytosine (C). Complementarity can be perfect (e.g. complete complementarity) or imperfect (e.g. partial complementarity). Perfect or complete complementarity indicates that each and every nucleic acid base of one strand is capable of forming hydrogen bonds according to Watson-Crick canonical base pairing with a corresponding base in another, antiparallel nucleic acid sequence. Partial complementarity indicates that only a percentage of the contiguous residues of a nucleic acid sequence can form Watson-Crick base pairing with the same number of contiguous residues in another, antiparallel nucleic acid sequence. In some embodiments, the complementarity can be at least 70%, 80%, 90%, 100% or a number or a range between any two of these values. In some embodiments, the complementarity is perfect, i.e. 100%. For example, the complementary candidate sequence segment is perfectly complementary to the candidate sequence segment, whose sequence can be deducted from the candidate sequence segment using the Watson-Crick base pairing rules.

[0056] The term "vector" as used herein, can refer to a vehicle for carrying or transferring a nucleic acid. Non-

limiting examples of vectors include plasmids, bacteria, and viruses (for example, *Agrobacterium tumefaciens* Ti vectors).

[0057] The term "construct," as used herein, can refer to a recombinant nucleic acid that has been generated for the purpose of the expression of a specific nucleotide sequence (s), or that is to be used in the construction of other recombinant nucleotide sequences. As used herein, the term "plasmid" can refer to a nucleic acid that can be used to replicate recombinant DNA sequences within a host organism. The sequence can be a double stranded DNA.

[0058] As used herein, the term "promoter" is a nucleotide sequence that permits binding of RNA polymerase and directs the transcription of a gene. Typically, a promoter is located in the 5' non-coding region of a gene, proximal to the transcriptional start site of the gene. Sequence elements within promoters that function in the initiation of transcription are often characterized by consensus nucleotide sequences. Examples of promoters include, but are not limited to, promoters from bacteria, yeast, plants, viruses, and mammals (including humans). A promoter can be inducible, repressible, and/or constitutive. Inducible promoters initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, such as a change in temperature.

[0059] As used herein, the term "operably linked" is used to describe the connection between regulatory elements and a gene or its coding region. Typically, gene expression is placed under the control of one or more regulatory elements, for example, without limitation, constitutive or inducible promoters, tissue-specific regulatory elements, and enhancers. A gene or coding region is said to be "operably linked to" or "operatively linked to" or "operably associated with" the regulatory elements, meaning that the gene or coding region is controlled or influenced by the regulatory element. For instance, a promoter is operably linked to a coding sequence if the promoter effects transcription or expression of the coding sequence.

[0060] As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to the nucleotide bases or amino acid residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity or similarity is used in reference to proteins, it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted with a functionally equivalent residue of the amino acid residues with similar physiochemical properties and therefore do not change the functional properties of the molecule.

[0061] Described herein are systems, compositions, and methods, for insertion of a cargo sequence into a target double-stranded DNA in a plant cell. Described herein is a CRISPR-associated transposases (CAST) system in plants for programmable and high efficiency DNA integration, merging CRISPR RNA-guided targeting with high insertion efficiency of transposases. CAST systems were first engineered into a powerful RNA-guided DNA-insertion tool in *E. coli* with nearly 100% efficiency upon selection, obviating the need for DSB in the target DNA or homology arms in the donor DNA. In 2023, two CASTs derived from *Vibrio cholerae* and *Pseudoalteromonas*, designated as Vch CAST and Pse CAST, have demonstrated ability in catalyzing the insertion of large DNA sequences in a targeted manner without inducing DSBs in mammalian cells.

[0062] Such biotechnology for plants will enable basic discoveries in plant genomics, such as the identification of essential genes and screening of ideal locus for exogenous gene insertion and expression. It will also allow improved capabilities, such as building developmental or metabolic pathways to provide biotic and abiotic stress tolerance, battle new plant epidemics and adverse effects of climate change, and enable scalable and affordable biosynthesis of valuable products in plants.

[0063] Disclosed herein are novel methodologies for targeted DNA integration in plants through the utilization of an engineered CAST system. This work creates an advanced genetic engineering toolbox, presenting opportunities for researchers engaged in the exploration of fundamental plant biology, as well as in the engineering of plants to gain desired traits and to facilitate molecular farming.

[0064] The disclosure covers the establishment of a CAST-mediated DNA integration technique in plants, along with the validation of its functionality through the integration of fluorescent cargo in *Arabidopsis thaliana* protoplasts and *Nicotiana benthamiana* leaves. Additionally, the disclosure delineates the proposed procedure to engineer a stable plant and employ it for safe harbor loci screening.

[0065] CRISPR-associated transposons or CASTs are mobile genetic elements (MGEs) that have evolved to make use of minimal CRISPR systems for RNA-guided transposition of their DNA. Unlike traditional CRISPR systems that contain interference mechanisms to degrade targeted DNA, CASTs lack proteins and/or protein domains responsible for DNA cleavage. Specialized transposon machinery, similar to that of Tn7 transposon, complexes with the CRISPR RNA (crRNA) and associated Cas proteins for transposition. CAST systems have been characterized in a wide range of bacteria and make use of variable CRISPR configurations including Type I-F, Type I-B, Type I-C, Type I-D, Type I-E, Type IV, and Type V-K.

[0066] Many CRISPR-associated transposons are similar to the Tn7 transposon which functions with a cut and paste mechanism. It contains a heteromeric transposase consisting of TnsA and TnsB proteins, and a regulator protein TnsC. Structural analysis has shown binding of the TnsB protein and sequence specific motifs on the ends of the transposon which allows for excision and mobility. Targeting for integration is done by the TnsD or TnsE proteins which preferentially target safe sites within the host chromosome or mobile elements (plasmids or bacteriophages), respectively. TnsE is not found in CASTs but a TnsD homolog, TniQ, is present and functions to bridge the gap between the transposase and CRISPR-Cas. Multiple CRISPR types have been found to associate with transposons with two of the most studied being Type I-F, which makes use of a multi-subunit effector, and Type V-K, which makes use of a single Cas12k effector. In both cases, Tn7 transposons have evolved to make use of these effectors to create R loops for site-specific integration. While TnsA is present in Type I-F systems, it is notably absent in Type V-K systems which showed higher off-target integrations during initial characterization.

[0067] A Type IF-3 CAST (Tn6677) was initially identified in *Vibrio Cholerae* and has been extensively studied. This system contains proteins TnsA, TnsB, and TnsC that complex with Cas6, Cas7, and a Cas5-Cas8 fusion through interactions with TniQ. Initial integration steps include TniQ

complexed with Cas proteins, which binds at the target site, and TnsA and TnsB excision of the transposon, which is followed by TnsC binding to TniQ and transposase binding to TnsC. There can be off-targeting prior to this final step, but TnsB and TnsC binding leads to a final proofreading step to maintain a high on-target percentage. Tn6677 integration has been validated at near 100% on-target efficiency at site specific locations in multiple points in the host genome. Other systems have also been characterized and validated in this class with varying ranges of efficiency, and include orthogonal systems for multiplexed insertions up to 10 kb.

[0068] A Type V-K system was originally characterized from a cyanobacteria, *Scytonema hofmanni*, and contains a single Cas effector, Cas12k, that functions with a tracrRNA. This system functions similarly to Tn7 but does not have a TnsA protein which can result in off-targeting and chimera formation during over-expression. The Cas12k and tracrRNA complex bind to the target site and TnsC is polymerized directly adjacent prior to TniQ attachment and TnsB recognition and integration. While these systems use traditional tracrRNA characteristic of Type II CRISPR systems, they can also target with short crRNA located adjacent to the transposon end. Type V-K spacers preferentially target locations near tRNA genes, but other sites have been observed in these short crRNA guides.

[0069] Disclosed herein include systems for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the system comprises: i) an RNA-guided DNA binding complex or one or more first helper polynucleotides each comprising a sequence encoding a component of the RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the system comprises one or more helper accessory proteins or one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP; ii) a transposition complex or one or more second helper polynucleotides each comprising a sequence encoding a component of the transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

[0070] Disclosed herein include nucleic acid compositions for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the nucleic acid composition comprises: i) one or more first helper polynucleotides each comprising a sequence encoding a component of an RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the nucleic acid composition comprises one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP; ii) one or more second helper polynucleotides each comprising a sequence encoding a component of a transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

Autonomous Replicons

[0071] In some embodiments of the systems and nucleic acid composition of the disclosure: a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon. At least one of the one or more first helper polynucleotides and/or at least one of the one or more second helper polynucleotides can be comprised within a second autonomous replicon. The first autonomous replicon, the second autonomous replicon, or both can be derived from a geminivirus. The geminivirus can comprise cabbage leaf curl virus, tomato golden mosaic virus, bean yellow dwarf virus, African cassava mosaic virus, wheat dwarf virus, miscanthus streak mastrevirus, tobacco yellow dwarf virus, tomato yellow leaf curl virus, bean golden mosaic virus, beet curly top virus, maize streak virus, or tomato pseudo-curly top virus.

[0072] Geminiviruses replicate through a rolling circle replication (RCR) cycle, and consequently, viral replicons can achieve high copy number, increasing the transient expression of, e.g., a donor polynucleotide and/or one or more first and/or second helper polynucleotides. For example, a deconstructed version of bean yellow dwarf virus (BeYDV), was used to deliver ZFNs and a repair template to tobacco cells to achieve gene targeting at an integrated reporter gene (Baltes et al., Plant Cell 2014, 26:151-163); BeYDV also have been used for targeted knock-in of a strong promoter upstream of a tomato gene that regulates anthocyanin synthesis (Cermik et al., Genome Biol 16:232, 2015). WDV is a ssDNA virus (Mastrevirus) that infects a variety of grasses, including most cereals. WDV-derived replicons can be used to express foreign proteins in cells from plants such as wheat and maize cells (Ugaki et al., supra; Matzeit et al., Plant Cell 1991, 3:247-258; and Suarez-Lopez and Gutierrez, Virology 1997, 227:389-399). Tomato leaf curl virus (ToLCV) also is a ssDNA virus (Begomovirus), and although its natural hosts are normally Solanaceous species, ToLCV-derived replicons can efficiently replicate and express GFP in rice (Pandey et al., Virol J 2009, 6:152).

[0073] Gemini virus-based replicons can be particularly useful. Geminiviruses are a large family of plant viruses that contain circular, single-stranded DNA genomes. Examples of geminiviruses include the cabbage leaf curl virus, tomato golden mosaic virus, bean yellow dwarf virus (BeYDV; also referred to as chickpea chlorotic dwarf virus), African cassava mosaic virus, wheat dwarf virus (WDV), miscanthus streak mastrevirus, tobacco yellow dwarf virus, tomato yellow leaf curl virus, bean golden mosaic virus, beet curly top virus, maize streak virus, and tomato pseudo-curly top virus.

[0074] The engineered replicon can be generated by, for example, replacing non-essential geminivirus nucleotide sequence (e.g., CP sequence) with a desired cargo sequence. Other methods for adding sequence to viral vectors include, without limitation, those discussed in Peretz et al. (*Plant Physiol.,* 145:1251-1263, 2007).

[0075] In one example of the autonomous replicons of the disclosure, the LIR (long intergenic region) region initiates transcription of the cargo sequence, while the SIR (short intergenic region) terminates transcription. Geminivirus-derived vectors can be sent to cells in two different pathways: the cis or autonomous and trans or tethered route wherein cis employs the Rep in its native position to the LIR, driven by a C-sense promoter, and can give rise to thousands of copies of the replicons and trans employs persistent production of Rep protein through stable integration to drive production of the replicon.

[0076] In some embodiments, the donor polynucleotide comprised within the first autonomous replicon comprises, from 5' to 3': a first long intergenic region (LIR), the RE, the cargo sequence, the LE, a short intergenic region (SIR), a sequence encoding RepA, and a second LTR. In some embodiments, the at least one of the one or more first helper polynucleotides and/or the at least one of the one or more second helper polynucleotides comprised within the second autonomous replicon comprises, from 5' to 3': a first long intergenic region (LIR), the first helper polynucleotide or the second helper polynucleotide, a short intergenic region (SIR), a sequence encoding RepA, and a second LIR. The first and/or second LIR can comprise or consist of the sequence of SEQ ID NO: 1 or a sequence having one, two, or three mismatches relative to the sequence of SEQ ID NO: 1. The SIR can comprise or consist of the sequence of SEQ ID NO: 2 or a sequence having one, two, or three mismatches relative to the sequence of SEQ ID NO: 2. The sequence encoding RepA can comprise or consist of the sequence of SEQ ID NO: 3 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to SEQ ID NO: 3.

[0077] The amount of the donor polynucleotide in the plant cell can be capable of increasing by at least 2-fold (e.g., 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold, or a number or a range between any of these values) following the onset of autonomous replication. The amount of the donor polynucleotide in the plant cell can be capable of increasing by at least 10-fold following the onset of autonomous replication.

[0078] The amount of a gene product encoded by the donor polynucleotide in the plant cell can be capable of increasing by at least 2-fold (e.g., 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold, or a number or a range between any of these values) following the onset of autonomous replication. The amount of the gene product encoded by the donor polynucleotide in the plant cell can be capable of increasing by at least 10-fold following the onset of autonomous replication.

Donor Polynucleotide

[0079] A donor polynucleotide can comprise a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence. Cargo sequence flanked by an RE and LE is thus a transposable element (e.g., transposon) capable of being inserted into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell using the systems, nucleic compositions, and methods disclosed herein.

[0080] In some embodiments, the LE comprises the sequence of any one of SEQ ID NOs: 4-5 or a sequence having one, two or three mismatches relative to the sequence of any one of SEQ ID NOs: 4-5 and wherein the RE comprises the sequence of any one of SEQ ID NOs: 6-7 or a sequence having one, two or three mismatches relative to the sequence of any one of SEQ ID NOs: 6-7. In some embodiments, the LE comprises or consists of the sequence of any one of SEQ ID NOs: 4-5 or a sequence that is at least 80% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to the sequence of any one of SEQ ID NOs: 4-5. In some embodiments, the RE comprises or consists of the sequence of any one of SEQ ID NOs: 6-7 or a that is at least 80% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to the sequence of any one of SEQ ID NOs: 6-7.

Cargo Sequence

[0081] The length of the cargo sequence can vary. The cargo sequence can be 0.2 to 1000 kilobase pairs (kb) (e.g., 0.2 kb, 0.5 kb, 0.75 kb, 1.0 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 300 kb, 400 kb, 500 kb, 600 kb, 700 kb, 800 kb, 900 kb, 1000 kb, or a number or a range between any two of these values) in length. The cargo sequence can be 200 to 1200 base pairs (bp) (e.g., 200 bp, 210 bp, 220 bp, 230 bp, 240 bp, 250 bp, 260 bp, 270 bp, 280 bp, 290 bp, 300 bp, 310 bp, 320 bp, 330 bp, 340 bp, 350 bp, 360 bp, 370 bp, 380 bp, 390 bp, 400 bp, 410 bp, 420 bp, 430 bp, 440 bp, 450 bp, 460 bp, 470 bp, 480 bp, 490 bp, 500 bp, 510 bp, 520 bp, 530 bp, 540 bp, 550 bp, 560 bp, 570 bp, 580 bp, 590 bp, 600 bp, 610 bp, 620 bp, 630 bp, 640 bp, 650 bp, 660 bp, 670 bp, 680 bp, 690 bp, 700 bp, 710 bp, 720 bp, 730 bp, 740 bp, 750 bp, 760 bp, 770 bp, 780 bp, 790 bp, 800 bp, 810 bp, 820 bp, 830 bp, 840 bp, 850 bp, 860 bp, 870 bp, 880 bp, 890 bp, 900 bp, 910 bp, 920 bp, 930 bp, 940 bp, 950 bp, 960 bp, 970 bp, 980 bp, 990 bp, 1000 bp, 1025 bp, 1050 bp, 1075 bp, 1100 bp, 1125 bp, 1150 bp, 1175 bp, 1200 bp or a number or a range between any two of these values) in length.

[0082] Any cargo sequence contemplated by the skilled artisan can be used according to the systems, compositions, and methods of the disclosure. In some embodiments, the cargo sequence comprises one or more exogenous sequences which when introduced into plants will alter the phenotype of the plant, a plant organ, plant tissue, or portion of the plant. Exemplary exogenous sequences encode polypeptides involved in one or more important biological properties in

plants. Other exemplary exogenous sequences can alter expression of exogenous or endogenous genes, either increasing or decreasing expression, optionally in response to a specific signal or stimulus.

[0083] As used herein, the term "trait" can refer either to the altered phenotype of interest or the nucleic acid which causes the altered phenotype of interest. One of the major purposes of transformation of crop plants is to add some commercially desirable, agronomically important traits to the plant. Such traits include, but are not limited to, herbicide resistance or tolerance; insect (pest) resistance or tolerance; disease resistance or tolerance (viral, bacterial, fungal, nematode or other pathogens); stress tolerance and/or resistance, as exemplified by resistance or tolerance to drought, heat, chilling, freezing, excessive moisture, salt stress, mechanical stress, extreme acidity, alkalinity, toxins, UV light, ionizing radiation or oxidative stress; increased yields, whether in quantity or quality; enhanced or altered nutrient acquisition and enhanced or altered metabolic efficiency; enhanced or altered nutritional content and makeup of plant tissues used for food, feed, fiber or processing; physical appearance; male sterility; drydown; standability; prolificacy; starch quantity and quality; oil quantity and quality; protein quality and quantity; amino acid composition; modified chemical production; altered pharmaceutical or nutraceutical properties; altered bioremediation properties; increased biomass; altered growth rate; altered fitness; altered biodegradability; altered CO_2 fixation; presence of bioindicator activity; altered digestibility by humans or animals; altered allergenicity; altered mating characteristics; altered pollen dispersal; improved environmental impact; altered nitrogen fixation capability; the production of a pharmaceutically active protein; the production of a small molecule with medicinal properties; the production of a chemical including those with industrial utility; the production of nutraceuticals, food additives, carbohydrates, RNAs, lipids, fuels, dyes, pigments, vitamins, scents, flavors, vaccines, antibodies, hormones, and the like; and alterations in plant architecture or development, including changes in developmental timing, photosynthesis, signal transduction, cell growth, reproduction, or differentiation. Additionally one could create a library of an entire genome from any organism or organelle including mammals, plants, microbes, fungi, or bacteria, represented on one or more donor polynucleotides.

[0084] In some embodiments, the modified plant may exhibit increased or decreased expression or accumulation of a product of the plant, which may be a natural product of the plant or a new or altered product of the plant. Exemplary products include an enzyme, an RNA molecule, a nutritional protein, a structural protein, an amino acid, a lipid, a fatty acid, a polysaccharide, a sugar, an alcohol, an alkaloid, a carotenoid, a propanoid, a phenylpropanoid, or terpenoid, a steroid, a flavonoid, a phenolic compound, an anthocyanin, a pigment, a vitamin or a plant hormone. In some embodiments, the modified plant has enhanced or diminished requirements for light, water, nitrogen, or trace elements. In some embodiments, the modified plant has an enhance ability to capture or fix nitrogen from its environment. In some embodiments, the modified plant is enriched for an essential amino acid as a proportion of a protein fraction of the plant. The protein fraction may be, for example, total seed protein, soluble protein, insoluble protein, water-extractable protein, and lipid-associated protein. The modification can include, e.g., overexpression, underexpression, antisense modulation, sense suppression, inducible expression, inducible repression, or inducible modulation of a gene.

[0085] A brief summary of exemplary improved properties and polypeptides of interest for either increased or decreased expression is provided below.

Herbicide Resistance

[0086] A herbicide resistance (or tolerance) trait is a characteristic of a modified plant that is resistant to dosages of an herbicide that is typically lethal to a non-modified plant. Exemplary herbicides for which resistance is useful in a plant include glyphosate herbicides, phosphinothricin herbicides, oxynil herbicides, imidazolinone herbicides, dinitroaniline herbicides, pyridine herbicides, sulfonylurea herbicides, bialaphos herbicides, sulfonamide herbicides and glufosinate herbicides. Other herbicides would be useful as would combinations of herbicide genes.

[0087] The genes encoding phosphinothricin acetyltransferase (bar), glyphosate tolerant EPSP synthase genes, glyphosate acetyltransferase, the glyphosate degradative enzyme gene gox encoding glyphosate oxidoreductase, deh (encoding a dehalogenase enzyme that inactivates dalapon), herbicide resistant (e.g., sulfonylurea and imidazolinone) acetolactate synthase, and bxn genes (encoding a nitrilase enzyme that degrades bromoxynil) are good examples of herbicide resistant genes for use in transformation. The bar gene codes for an enzyme, phosphinothricin acetyltransferase (PAT), which inactivates the herbicide phosphinothricin and prevents this compound from inhibiting glutamine synthetase enzymes. The enzyme 5 enolpyruvylshikimate 3 phosphate synthase (EPSP Synthase), is normally inhibited by the herbicide N (phosphonomethyl)glycine (glyphosate). However, genes are known that encode glyphosate resistant EPSP synthase enzymes. These genes are particularly contemplated for use in plant transformation. The deh gene encodes the enzyme dalapon dehalogenase and confers resistance to the herbicide dalapon. The bxn gene codes for a specific nitrilase enzyme that converts bromoxynil to a non herbicidal degradation product. The glyphosate acetyl transferase gene inactivates the herbicide glyphosate and prevents this compound from inhibiting EPSP synthase. Polypeptides that may produce plants having tolerance to plant herbicides include polypeptides involved in the shikimate pathway, which are of interest for providing glyphosate tolerant plants. Such polypeptides include polypeptides involved in biosynthesis of chorismate, phenylalanine, tyrosine and tryptophan.

Insect Resistance

[0088] Potential insect resistance (or tolerance) genes that can be introduced include *Bacillus thuringiensis* toxin genes or Bt genes (Watrud et al., In: Engineered Organisms and the Environment, 1985). Bt genes may provide resistance to lepidopteran or coleopteran pests such as European Corn Borer (ECB). Preferred Bt toxin genes for use in such embodiments include the CryIA(b) and CryIA(c) genes. Endotoxin genes from other species of *B. thuringiensis* which affect insect growth or development also may be employed in this regard. It is contemplated that preferred Bt genes for use according to the present disclosure will be those in which the coding sequence has been modified to

effect increased expression in plants, and for example, in monocot plants. Means for preparing synthetic genes are well known in the art and are disclosed in, for example, U.S. Pat. Nos. 5,500,365 and 5,689,052, each of the disclosures of which are specifically incorporated herein by reference in their entirety. Examples of such modified Bt toxin genes include a synthetic Bt CryIA(b) gene (Perlak et al., Proc. Natl. Acad. Sci. USA, 88:3324-3328, 1991), and the synthetic CryIA(c) gene termed 1800b (PCT Application WO 95/06128).

[0089] Protease inhibitors also may provide insect resistance (Johnson et al., Proc Natl Acad Sci USA. 1989 December; 86(24): 9871-9875.), and will thus have utility in plant transformation. The use of a protease inhibitor II gene, pinII, from tomato or potato is envisioned to be particularly useful. Even more advantageous is the use of a pinII gene in combination with a Bt toxin gene, the combined effect of which has been discovered to produce synergistic insecticidal activity. Other genes which encode inhibitors of the insect's digestive system, or those that encode enzymes or co factors that facilitate the production of inhibitors, also may be useful. This group may be exemplified by oryzacystatin and amylase inhibitors such as those from wheat and barley.

[0090] Amylase inhibitors are found in various plant species and are used to ward off insect predation via inhibition of the digestive amylases of attacking insects. Several amylase inhibitor genes have been isolated from plants and some have been introduced as exogenous nucleic acids, conferring an insect resistant phenotype that is potentially useful ("Plants, Genes, and Crop Biotechnology" by Maarten J. Chrispeels and David E. Sadava (2003) Jones and Bartlett Press).

[0091] Genes encoding lectins may confer additional or alternative insecticide properties. Lectins are multivalent carbohydrate binding proteins which have the ability to agglutinate red blood cells from a range of species. Lectins have been identified recently as insecticidal agents with activity against weevils, ECB and rootworm (Murdock et al., Phytochemistry, 29:85-89, 1990, Czapla & Lang, J. Econ. Entomol., 83:2480-2485, 1990). Lectin genes contemplated to be useful include, for example, barley and wheat germ agglutinin (WGA) and rice lectins (Gatehouse et al., J. Sci. Food. Agric, 35:373-380, 1984), with WGA being preferred. Genes controlling the production of large or small polypeptides active against insects when introduced into the insect pests, such as, e.g., lytic peptides, peptide hormones and toxins and venoms, form another aspect of the disclosure. For example, it is contemplated that the expression of juvenile hormone esterase, directed towards specific insect pests, also may result in insecticidal activity, or perhaps cause cessation of metamorphosis (Hammock et al., Nature, 344:458-461, 1990).

[0092] Genes which encode enzymes that affect the integrity of the insect cuticle form yet another aspect of the disclosure. Such genes include those encoding, e.g., chitinase, proteases, lipases and also genes for the production of nikkomycin, a compound that inhibits chitin synthesis, the introduction of any of which is contemplated to produce insect resistant plants. Genes that code for activities that affect insect molting, such as those affecting the production of ecdysteroid UDP glucosyl transferase, also fall within the scope of the useful exogenous nucleic acids of the present disclosure. Genes that code for enzymes that facilitate the production of compounds that reduce the nutritional quality of the host plant to insect pests also are encompassed by the present disclosure. In some embodiments, insecticidal activity can be conferred on a plant by altering its sterol composition. Sterols are obtained by insects from their diet and are used for hormone synthesis and membrane stability. Therefore alterations in plant sterol composition by expression of novel genes, e.g., those that directly promote the production of undesirable sterols or those that convert desirable sterols into undesirable forms, can, in some embodiments, have a negative effect on insect growth and/or development and hence endow the plant with insecticidal activity. Lipoxygenases are naturally occurring plant enzymes that have been shown to exhibit anti nutritional effects on insects and to reduce the nutritional quality of their diet. Therefore, further embodiments of the disclosure concern modified plants with enhanced lipoxygenase activity which may be resistant to insect feeding.

[0093] *Tripsacum dactyloides* is a species of grass that is resistant to certain insects, including corn root worm. It is anticipated that genes encoding proteins that are toxic to insects or are involved in the biosynthesis of compounds toxic to insects will be isolated from *Tripsacum* and that these novel genes will be useful in conferring resistance to insects. It is known that the basis of insect resistance in *Tripsacum* is genetic, because said resistance has been transferred to *Zea mays* via sexual crosses (Branson and Guss, Proceedings North Central Branch Entomological Society of America, 27:91-95, 1972). It is further anticipated that other cereal, monocot or dicot plant species may have genes encoding proteins that are toxic to insects which would be useful for producing insect resistant plants.

[0094] Further genes encoding proteins characterized as having potential insecticidal activity also may be used as exogenous nucleic acids in accordance herewith. Such genes include, for example, the cowpea trypsin inhibitor (CpTI; Hilder et al., Nature, 330:160-163, 1987) which may be used as a rootworm deterrent; genes encoding avermectin (Avermectin and Abamectin., Campbell, W. C., Ed., 1989; Ikeda et al., J. Bacteriol., 169:5615-5621, 1987) which may prove particularly useful as a corn rootworm deterrent; ribosome inactivating protein genes; and even genes that regulate plant structures. Modified plants including anti insect antibody genes and genes that code for enzymes that can convert a non toxic insecticide (pro insecticide) applied to the outside of the plant into an insecticide inside the plant also are contemplated.

[0095] Polypeptides that may improve plant tolerance to the effects of plant pests or pathogens include proteases, polypeptides involved in anthocyanin biosynthesis, polypeptides involved in cell wall metabolism, including cellulases, glucosidases, pectin methylesterase, pectinase, polygalacturonase, chitinase, chitosanase, and cellulose synthase, and polypeptides involved in biosynthesis of terpenoids or indole for production of bioactive metabolites to provide defense against herbivorous insects. Vegetative Insecticidal Proteins (VIP) are a class of proteins originally found to be produced in the vegetative growth phase of the bacterium, *Bacillus cereus*, but do have a spectrum of insect lethality similar to the insecticidal genes found in strains of *Bacillus* thuriengensis. Both the vipla and vip3 A genes have been isolated and have demonstrated insect toxicity. It is anticipated that such genes may be used in modified plants to confer insect resistance ("Plants, Genes, and Crop Bio-

technology" by Maarten J. Chrispeels and David E. Sadava (2003) Jones and Bartlett Press).

Environment or Stress Resistance

[0096] Improvement of a plant's ability to tolerate various environmental stresses such as, but not limited to, drought, excess moisture, chilling, freezing, high temperature, salt, and oxidative stress, also can be effected through expression of novel genes. Benefits may be realized in terms of increased resistance to freezing temperatures through the introduction of an "antifreeze" protein such as that of the Winter Flounder (Cutler et al, J. Plant Physiol., 135:351-354, 1989) or synthetic gene derivatives thereof. Improved chilling tolerance also may be conferred through increased expression of glycerol 3 phosphate acetyltransferase in chloroplasts (Wolter et al., The EMBO J., 4685-4692, 1992). Resistance to oxidative stress (often exacerbated by conditions such as chilling temperatures in combination with high light intensities) can be conferred by expression of superoxide dismutase (Gupta et al., 1993), and may be improved by glutathione reductase (Bowler et al., Ann Rev. Plant Physiol., 43:83-116, 1992). Such strategies can allow for tolerance to freezing in newly emerged fields as well as extending later maturity higher yielding varieties to earlier relative maturity zones.

[0097] It is contemplated that the expression of novel genes that favorably affect plant water content, total water potential, osmotic potential, or turgor will enhance the ability of the plant to tolerate drought. As used herein, the terms "drought resistance" and "drought tolerance" are used to refer to a plant's increased resistance or tolerance to stress induced by a reduction in water availability, as compared to normal circumstances, and the ability of the plant to function and survive in lower water environments. Expression of genes encoding for the biosynthesis of osmotically active solutes, such as polyol compounds, may impart protection against drought. Within this class are genes encoding for mannitol L phosphate dehydrogenase (Lee and Saier, 1982) and trehalose 6 phosphate synthase (Kaasen et al., J. Bacteriology, 174:889-898, 1992). Through the subsequent action of native phosphatases in the cell or by the introduction and coexpression of a specific phosphatase, these introduced genes will result in the accumulation of either mannitol or trehalose, respectively, both of which have been well documented as protective compounds able to mitigate the effects of stress. Mannitol accumulation in transgenic tobacco has been verified and preliminary results indicate that plants expressing high levels of this metabolite are able to tolerate an applied osmotic stress (Tarczynski et al., Science, 259:508-510, 1993, Tarczynski et al Proc. Natl. Acad. Sci. USA, 89:1-5, 1993).

[0098] Similarly, the efficacy of other metabolites in protecting either enzyme function (e.g., alanopine or propionic acid) or membrane integrity (e.g., alanopine) has been documented (Loomis et al., J. Expt. Zoology, 252:9-15, 1989), and therefore expression of genes encoding for the biosynthesis of these compounds might confer drought resistance in a manner similar to or complimentary to mannitol. Other examples of naturally occurring metabolites that are osmotically active and/or provide some direct protective effect during drought and/or desiccation include fructose, erythritol (Coxson et al., Biotropica, 24:121-133, 1992), sorbitol, dulcitol (Karsten et al., Botanica Marina, 35:11-19, 1992), glucosylglycerol (Reed et al., J. Gen. Microbiology, 130:1-4, 1984; Erdmann et al., J. Gen. Microbiology, 138:363-368, 1992), sucrose, stachyose (Koster and Leopold, Plant Physiol., 88:829-832, 1988; Blackman et al., Plant Physiol., 100:225-230, 1992), raffmose (Bernal Lugo and Leopold, Plant Physiol., 98:1207-1210, 1992), proline (Rensburg et al., J. Plant Physiol., 141:188-194, 1993), glycine betaine, ononitol and pinitol (Vernon and Bohnert, The EMBO J., 11:2077-2085, 1992). Continued canopy growth and increased reproductive fitness during times of stress will be augmented by introduction and expression of genes such as those controlling the osmotically active compounds discussed above and other such compounds. Currently preferred genes which promote the synthesis of an osmotically active polyol compound are genes which encode the enzymes mannitol 1 phosphate dehydrogenase, trehalose 6 phosphate synthase and myo-inositol O-methyltransferase.

[0099] It is contemplated that the expression of specific proteins also may increase drought tolerance. Three classes of Late Embryo genie Abundant (LEA) Proteins have been assigned based on structural similarities (see Dure et al., Plant Molecular Biology, 12:475-486, 1989). All three classes of LEAs have been demonstrated in maturing (e.g. desiccating) seeds. Within these 3 types of LEA proteins, the Type II (dehydrin type) have generally been implicated in drought and/or desiccation tolerance in vegetative plant parts (e.g. Mundy and Chua, The EMBO J., 7:2279-2286, 1988; Piatkowski et al., Plant Physiol., 94:1682-1688, 1990; Yamaguchi Shinozaki et al., Plant Cell Physiol., 33:217-224, 1992). Expression of a Type III LEA (HVA 1) in tobacco was found to influence plant height, maturity and drought tolerance (Fitzpatrick, Gen. Engineering News, 22:7, 1993). In rice, expression of the HVA 1 gene influenced tolerance to water deficit and salinity (Xu et al., Plant Physiol., 110:249-257, 1996). Expression of structural genes from any of the three LEA groups may therefore confer drought tolerance. Other types of proteins induced during water stress include thiol proteases, aldolases or transmembrane transporters (Guerrero et al., Plant Molecular Biology, 15:11-26, 1990), which may confer various protective and/or repair type functions during drought stress. It also is contemplated that genes that effect lipid biosynthesis and hence membrane composition might also be useful in conferring drought resistance on the plant. Many of these genes for improving drought resistance have complementary modes of action. Thus, it is envisaged that combinations of these genes might have additive and/or synergistic effects in improving drought resistance in plants. Many of these genes also improve freezing tolerance (or resistance); the physical stresses incurred during freezing and drought are similar in nature and may be mitigated in similar fashion. Benefit may be conferred via constitutive expression of these genes, but the preferred means of expressing these novel genes may be through the use of a turgor induced promoter (such as the promoters for the turgor induced genes described in Guerrero et al., Plant Molecular Biology, 15:11-26, 1990 and Shagan et al., Plant Physiol., 101:1397-1398, 1993 which are incorporated herein by reference). Spatial and temporal expression patterns of these genes may enable plants to better withstand stress.

[0100] Expression of genes that are involved with specific morphological traits that allow for increased water extractions from drying soil can be of benefit. For example, introduction and expression of genes that alter root characteristics may enhance water uptake. It also is contemplated

that expression of genes that enhance reproductive fitness during times of stress would be of significant value. For example, expression of genes that improve the synchrony of pollen shed and receptiveness of the female flower parts, e.g., silks, would be of benefit. In addition, expression of genes that minimize kernel abortion during times of stress may increase the amount of grain to be harvested and hence be of value. Given the overall role of water in determining yield, it is contemplated that enabling plants to utilize water more efficiently, through the introduction and expression of novel genes, will improve overall performance even when soil water availability is not limiting. By introducing genes that improve the ability of plants to maximize water usage across a full range of stresses relating to water availability, yield stability or consistency of yield performance may be realized.

[0101] Polypeptides that may improve stress tolerance under a variety of stress conditions include polypeptides involved in gene regulation, such as serine/threonine-protein kinases, MAP kinases, MAP kinase kinases, and MAP kinase kinase kinases; polypeptides that act as receptors for signal transduction and regulation, such as receptor protein kinases; intracellular signaling proteins, such as protein phosphatases, GTP binding proteins, and phospholipid signaling proteins; polypeptides involved in arginine biosynthesis; polypeptides involved in ATP metabolism, including for example ATPase, adenylate transporters, and polypeptides involved in ATP synthesis and transport; polypeptides involved in glycine betaine, jasmonic acid, flavonoid or steroid biosynthesis; and hemoglobin. Enhanced or reduced activity of such polypeptides in modified plants will provide changes in the ability of a plant to respond to a variety of environmental stresses, such as chemical stress, drought stress and pest stress. Other polypeptides that can improve plant tolerance to cold or freezing temperatures include polypeptides involved in biosynthesis of trehalose or raffinose, polypeptides encoded by cold induced genes, fatty acyl desaturases and other polypeptides involved in glycerolipid or membrane lipid biosynthesis, which find use in modification of membrane fatty acid composition, alternative oxidase, calcium-dependent protein kinases, LEA proteins or uncoupling protein.

[0102] Other polypeptides for improvement of plant tolerance to heat include polypeptides involved in biosynthesis of trehalose, polypeptides involved in glycerolipid biosynthesis or membrane lipid metabolism (for altering membrane fatty acid composition), heat shock proteins or mitochondrial NDK. Other polypeptides that may improve tolerance to extreme osmotic conditions include polypeptides involved in proline biosynthesis. Other polypeptides for improvement of plant tolerance to drought conditions include aquaporins, polypeptides involved in biosynthesis of trehalose or wax, LEA proteins or invertase.

Disease Resistance

[0103] Increased resistance (or tolerance) to diseases can, in some embodiments, be realized through introduction of genes into plants, for example, into monocotyledonous plants such as maize. It is possible to produce resistance to diseases caused by viruses, viroids, bacteria, fungi and nematodes. It also is contemplated that control of mycotoxin producing organisms may be realized through expression of introduced genes. Resistance can be affected through suppression of endogenous factors that encourage disease-causing interactions, expression of exogenous factors that are toxic to or otherwise provide protection from pathogens, or expression of factors that enhance the plant's own defense responses.

[0104] Resistance to viruses can be produced through expression of novel genes. For example, it has been demonstrated that expression of a viral coat protein in a modified plant can impart resistance to infection of the plant by that virus and perhaps other closely related viruses (Cuozzo et al, Bio/Technology, 6:549-553, 1988, Hemenway et al., The EMBO J., 7:1273-1280, 1988, Abel et al., Science, 232:738-743, 1986). It is contemplated that expression of antisense genes targeted at essential viral functions can also impart resistance to viruses. For example, an antisense gene targeted at the gene responsible for replication of viral nucleic acid may inhibit replication and lead to resistance to the virus. Interference with other viral functions through the use of antisense genes also may increase resistance to viruses.

[0105] Increased resistance to diseases caused by bacteria and fungi may be realized through introduction of novel genes. It is contemplated that genes encoding so called "peptide antibiotics," pathogenesis related (PR) proteins, toxin resistance, or proteins affecting host pathogen interactions such as morphological characteristics will be useful. Peptide antibiotics are polypeptide sequences which are inhibitory to growth of bacteria and other microorganisms. For example, the classes of peptides referred to as cecropins and magainins inhibit growth of many species of bacteria and fungi. It is proposed that expression of PR proteins in plants, for example, monocots such as maize, may be useful in conferring resistance to bacterial disease. These genes are induced following pathogen attack on a host plant and have been divided into at least five classes of proteins (Bol, Linthorst, and Cornelissen, 1990). Included amongst the PR proteins are beta 1, 3 glucanases, chitinases, and osmotin and other proteins that are believed to function in plant resistance to disease organisms. Other genes have been identified that have antifungal properties, e.g., UDA (stinging nettle lectin), or herein (Broakaert et al., 1989; Barkai Golan et al., 1978). It is known that certain plant diseases are caused by the production of phytotoxins. Resistance to these diseases can, in some embodiments, be achieved through expression of a novel gene that encodes an enzyme capable of degrading or otherwise inactivating the phytotoxin. It also is contemplated that expression of novel genes that alter the interactions between the host plant and pathogen may be useful in reducing the ability of the disease organism to invade the tissues of the host plant,—e.g., an increase in the waxiness of the leaf cuticle or other morphological characteristics.

[0106] Polypeptides useful for imparting improved disease responses to plants include polypeptides encoded by cercosporin induced genes, antifungal proteins and proteins encoded by R-genes or SAR genes. Agronomically important diseases caused by fungal phytopathogens include: glume or leaf blotch, late blight, stalk/head rot, rice blast, leaf blight and spot, corn smut, wilt, sheath blight, stem canker, root rot, blackleg or kernel rot.

[0107] Exemplary plant viruses include tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, maize dwarf mosaic virus, etc. Specific fungal, bacterial and viral pathogens of major crops include, but are not limited to:

[0108] RICE: rice brown spot fungus (*Cochliobolus miyabeanus*), rice blast fungus-*Magnaporthe grisea* (*Pyricularia*

grisea), *Magnaporthe salvinii* (*Sclerotium oryzae*), *Xanthomonas oryzae* pv. *oryzae, Xanthomonas oryzae* pv. oryzicola, *Rhizoctonia* spp. (including but not limited to *Rhizoctonia solani, Rhizoctonia oryzae* and *Rhizoctonia oryzae-sativae*), *Pseudomonas* spp. (including but not limited to *Pseudomonas plantarii, Pseudomonas avenae, Pseudomonas glumae, Pseudomonas fuscovaginae, Pseudomonas alboprecipitans, Pseudomonas syringae* pv. *panici, Pseudomonas syringae* pv. *syringae, Pseudomonas syringae* pv. *oryzae* and *Pseudomonas syringae* pv. *aptata*), *Erwinia* spp. (including but not limited to *Erwinia herbicola, Erwinia amylovaora, Erwinia chrysanthemi* and *Erwinia carotovora*), *Achyla* spp. (including but not limited to *Achyla conspicua* and *Achyia klebsiana*), *Pythium* spp. (including but not limited to *Pythium dissotocum, Pythium irregulare, Pythium arrhenomanes, Pythium myriotylum, Pythium catenulatum, Pythium graminicola* and *Pythium spinosum*), *Saprolegnia* spp., *Dictyuchus* spp., *Pythiogeton* spp., *Phytophthora* spp., *Alternaria padwickii, Cochliobolus miyabeanus, Curvularia* spp. (including but not limited to *Curvularia lunata, Curvularia affmis, Curvularia clavata, Curvularia eragrostidis, Curvularia fallax, Curvularia geniculata, Curvularia inaequalis, Curvularia intermedia, Curvularia oryzae, Curvularia oryzae-sativae, Curvularia pallescens, Curvularia senegalensis, Curvularia tuberculata, Curvularia uncinata* and *Curvularia verruculosa*), *Sarocladium oryzae, Gerlachia oryzae, Fusarium* spp. (including but not limited *Fusarium graminearum, Fusarium nivale* and to different pathovars of *Fusarium monoliforme*, including pvs. *fujikuroi* and *zeae*), *Sclerotium rolfsii, Phoma exigua, Mucor fragilis, Trichoderma viride, Rhizopus* spp., *Cercospora oryzae, Entyloma oryzae, Dreschlera gigantean, Scierophthora macrospora, Mycovellosiella oryzae, Phomopsis oryzae-sativae, Puccinia graminis, Uromyces coronatus, Cylindrocladium scoparium, Sarocladium oryzae, Gaeumannomyces graminis* pv. *graminis, Myrothecium verrucaria, Pyrenochaeta oryzae, Ustilaginoidea virens, Neovossia* spp. (including but not limited to *Neovossia horrida*), *Tilletia* spp., *Balansia oryzae-sativae, Phoma* spp. (including but not limited to *Phoma sorghina, Phoma insidiosa, Phoma glumarum, Phoma glumicola* and *Phoma oryzina*), *Nigrospora* spp. (including but not limited to *Nigrospora oryzae, Nigrospora* sphaerica, *Nigrospora panici* and *Nigrospora padwickii*), *Epiococcum nigrum, Phyllostica* spp., *Wolkia decolorans, Monascus purpureus, Aspergillus* spp., *Penicillium* spp., *Absidia* spp., *Mucor* spp., *Chaetomium* spp., *Dematium* spp., *Monilia* spp., *Streptomyces* spp., *Syncephalastrum* spp., *Verticillium* spp., *Nematospora coryli, Nakataea sigmoidea, Cladosporium* spp., *Bipolaris* spp., *Coniothyrium* spp., *Diplodia oryzae, Exserophilum rostratum, Helococera oryzae, Melanomma glumarum, Metashaeria* spp., *Mycosphaerella* spp., *Oidium* spp., *Pestalotia* spp., *Phaeoseptoria* spp., *Sphaeropsis* spp., *Trematosphaerella* spp., rice black-streaked dwarf virus, rice dwarf virus, rice gall dwarf virus, barley yellow dwarf virus, rice grassy stunt virus, rice hoja blanca virus, rice necrosis mosaic virus, rice ragged stunt virus, rice stripe virus, rice stripe necrosis virus, rice transitory yellowing virus, rice tungro bacilliform virus, rice tungro spherical virus, rice yellow mottle virus, rice tarsonemid mite virus, *Echinochloa* hoja blanca virus, *Echinochloa* ragged stunt virus, orange leaf mycoplasma-like organism, yellow dwarf mycoplasma-like organism, *Aphelenchoides besseyi, Ditylenchus angustus, Hirschmanniella* spp., *Criconemella* spp., *Meloidogyne* spp., *Heterodera* spp., *Pratylenchus* spp., Hoplolaimus *indicus*. SOYBEANS: *Phytophthora sojae, Fusarium solani* f. sp. Glycines, *Macrophomina phaseolina, Fusarium, Pythium, Rhizoctonia, Phialophora gregata, Sclerotinia sclerotiorum, Diaporthe phaseolorum* var. *sojae, Colletotrichum truncatum, Phomopsis longicolla, Cercospora kikuchii, Diaporthe phaseolonum* var. *meridionalis* (and var. caulivora), *Phakopsora pachyrhyzi, Fusarium solani, Microsphaera diffusa, Septoria glycines, Cercospora kikuchii, Macrophomina phaseolina, Sclerotinia sclerotiorum, Corynespora cassiicola, Rhizoctonia solani, Cercospora sojina, Phytophthora megasperma* fsp. *glycinea, Macrophomina phaseolina, Fusarium oxysporum, Diapothe phaseolorum* var. *sojae* (*Phomopsis sojae*), *Diaporthe phaseolorum* var. caulivora, *Sclerotium rolfsii, Cercospora kikuchii, Cercospora sojina, Peronospora manshurica, Colletotrichum dematium* (*Colletotichum truncatum*), *Corynespora cassiicola, Phyllosticta sojicola, Alternaria alternata, Pseudomonas syringae* p.v. *glycinea, Xanthomonas campestris* p.v. *phaseoli, Microspaera diffusa, Fusarium semitectum, Phialophora gregata*, Soybean mosaic virus, *Glomerella* glycines, Tobacco Ring spot virus, Tobacco Streak virus, *Phakopsora pachyrhizi, Pythium aphanidermatum, Pythium ultimum, Pythium dearyanum*, Tomato spotted wilted virus, *Heterodera glycines, Fusarium solani*, Soybean cyst and root knot nematodes.

[0109] CORN: *Fusarium moniliforme* var. *subglutinans, Erwinia stewartii, Fusarium moniliforme, Gibberella zeae* (*Fusarium Graminearum*), *Stenocarpella maydi* (*Diplodia maydis*), *Pythium irregulare, Pythium debaryanum, Pythium graminicola, Pythium splendens, Pythium ultimum, Pythium aphanidermatum, Aspergillus flavus, Bipolaris maydis* O, T (*cochliobolus heterostrophus*), *Helminthosporium carbonum* I, II, and III (*Cochliobolus carborium*), *Exserohilum turcicum* I, II and III, *Helminthosporium pedicellatum, Physoderma maydis, Phyllosticta maydis, Kabatie-maydis, Cercospora sorghi, Ustilago maydis, Puccinia sorghi, Puccinia polysora, Macrophomina phaseolina, Penicillium oxalicum, Nigrospora oryzae, Cladosporium herbarum, Curvularia lunata, Curvularia inaequalis, Curvularia pallescens, Clavibacter michiganese* subsp. *Nebraskense, Trichoderma viride*, Maize dwarf Mosaic Virus A and B, Wheat Streak Mosaic Virus, Maize Chlorotic Dwarf Virus, *Claviceps sorghi, Pseudonomas avenae, Erwinia chrysantemi* p.v. *Zea, Erwinia corotovora, Cornstun spiroplasma, Diplodia macrospora, Sclerophthora macrospora, Peronosclerospora sorghi, Peronoscherospora philippinesis, Peronosclerospora maydis, Peronosclerospora sacchari, Spacelotheca reiliana, Physopella zea, Cephalosporium maydis, Caphalosporium acremonium*, Maize Chlorotic Mottle Virus, High Plains Virus, Maize Mosaic Virus, Maize Rayado Fino Virus, Maize Streak Virus, Maize Stripe Virus, Maize Rought Dwarf Virus:

[0110] WHEAT: *Pseudomonas syringae* p.v. *atrofaciens, Urocystis agropyri, Xanthomonas campestris* p.v. *translucens, Pseudomonas syringae* p.v. *syringae, Alternaria alternata, Cladosporium herbarum, Fusarium graminearum, Fusarium avenaceum, Fusarium cuhnorum, Ustilago tritici, Ascochyta tritici, Cephalosporium gramineum, Collotetrichum graminicola, Erysiphe graminis* f. sp. *Tritici, Puccinia graminis* f. sp. *Tritici, Puccinia recondite* f. sp. *tritici, puccinia striiformis, Pyrenophora triticirepentis, Septoria nodorum, Septoria tritici, Spetoria avenae, Pseudocercosporella herpotrichoides, Rhizoctonia solani, Rhizoctonia*

cerealis, Gaeumannomyces graminis var. *tritici, Pythium aphanidermatum, Pythium arrhenomanes, Pythium ultimum, Bipolaris sorokiniana*, Barley Yellow Dwarf Virus, Brome Mosaic Virus, Soil Borne Wheat Mosaic Virus, Wheat Streak Virus, Wheat Spindle Streak Virus, American Wheat Striate Virus, *Claviceps purpurea, Tilletia tritici, Tilletia laeyis, Pstilago tritici, Tilletia indica, Rhizoctonia solani, Pythium arrhenomannes, Pythium gramicola, Pythium aphanidermatum*, High Plains Virus, European Wheat Striate Virus:

[0111] CANOLA: *Albugo Candida, Alternaria brassicae, Leptosharia maculans, Rhizoctonia solani, Sclerotinia sclerotiorum, Mycospaerella brassiccola, Pythium ultimum, Peronospora parasitica, Fusarium roseum, Fusarium oxysporum, Tilletia* foetida, *Tilletia caries, Alternaria alternata*: SUNFLOWER: *Plasmophora halstedii, Scherotinia sclerotiorum, Aster* Yellows, *Septoria helianthi, Phomopsis helianthi, Altemaria helianthi, Alternaria zinniae, Botrytis cinera, Phoma macdonaldii, Macrophomina phaseolina, Erysiphe cichoracearum, Phizopus oryzae, Rhizopus arrhizus, Rhizopus stolonifer, Puccinia helianthi, Verticillium Dahliae, Erwinia carotovoram* p.v. *carotovora, Cephalosporium acremonium, Phytophthora cryptogea, Albugo tragopogonis.*

[0112] SORGHUM: *Exserohilum turcicum, Colletotrichum graminicola* (*Glomerella graminicola*), *Cercospora sorghi, Gloeocercospora sorghi, Ascochyta sorghi, Pseudomonas syringae* p.v. *syringae, Xanthomonas campestris* p.v. *holcicola, Pseudomonas andropogonis, Puccinia purpurea, Macrophomina phaseolina, Periconia circinata, Fusarium moniliforme, Alternaria alternate, Bipolaris sorghicola, Helminthosporium sorghicola, Curvularia lunata, Phoma insidiosa, Pseudomonas avenae* (*Pseudomonas alboprecipitans*), *Ramulispora sorghi, Ramulispora sorghicola, Phyllachara sacchari Sporisorium relianum* (*Sphacelotheca reliana*), *Sphacelotheca cruenta, Sporisorium sorghi*, Sugarcane mosaic H, Maize Dwarf Mosaic Virus A & B, *Claviceps sorghi, Rhizoctonia solani, Acremonium* strictum, *Sclerophthona macrospora, Peronosclerospora sorghi, Peronosclerospora philippinensis, Sclerospora graminicola, Fusarium graminearum, Fusarium Oxysporum, Pythium arrhenomanes, Pythium graminicola*. ALFALFA: *Clavibater michiganensis* subsp. *Insidiosum, Pythium ultimum, Pythium irregulare, Pythium splendens, Pythium debaryanum, Pythium aphanidermatum, Phytophthora megasperma, Peronospora trifoliorum, Phoma medicaginis* var. *medicaginis, Cercospora medicaginis, Pseudopeziza medicaginis, Leptotrochila medicaginis, Fusarium oxysporum, Rhizoctonia solani, Uromyces striatus, Colletotrichum trifolii* race 1 and race 2, *Leptosphaerulina briosiana, Stemphylium botryosum, Stagonospora meliloti, Sclerotinia trifoliorum, Alfalfa Mosaic Virus, Verticillium alboatrum, Xanthomonas campestris* p.v. *alfalfae, Aphanomyces euteiches, Stemphylium herbarum, Stemphylium alfalfae.*

Plant Agronomic Characteristics

[0113] Two of the factors determining where crop plants can be grown are the average daily temperature during the growing season and the length of time between frosts. Within the areas where it is possible to grow a particular crop, there are varying limitations on the maximal time it is allowed to grow to maturity and be harvested. For example, a variety to be grown in a particular area is selected for its ability to mature and dry down to harvestable moisture content within the required period of time with maximum possible yield. Therefore, crops of varying maturities are developed for different growing locations. Apart from the need to dry down sufficiently to permit harvest, it is desirable to have maximal drying take place in the field to minimize the amount of energy required for additional drying post harvest. Also, the more readily a product such as grain can dry down, the more time there is available for growth and kernel fill. Genes that influence maturity and/or dry down can be identified and introduced into plant lines using transformation techniques to create new varieties adapted to different growing locations or the same growing location, but having improved yield to moisture ratio at harvest. Expression of genes that are involved in regulation of plant development can be useful. It is contemplated that genes can be introduced into plants that would improve standability and other plant growth characteristics. Expression of novel genes in plants which confer stronger stalks, improved root systems, or prevent or reduce ear droppage or shattering would be of great value to the farmer. Introduction and expression of genes that increase the total amount of photoassimilate available by, for example, increasing light distribution and/or interception would be advantageous. In addition, the expression of genes that increase the efficiency of photosynthesis and/or the leaf canopy would further increase gains in productivity. It is contemplated that expression of a phytochrome gene in crop plants can be advantageous. Expression of such a gene may reduce apical dominance, confer semidwarfism on a plant, or increase shade tolerance (U.S. Pat. No. 5,268,526). Such approaches would allow for increased plant populations in the field.

Nutrient Utilization

[0114] The ability to utilize available nutrients may be a limiting factor in growth of crop plants. In some embodiments, one can alter nutrient uptake, tolerate pH extremes, mobilization through the plant, storage pools, and availability for metabolic activities by the introduction of novel genes. These modifications would allow a plant, for example, maize to more efficiently utilize available nutrients. It is contemplated that an increase in the activity of, for example, an enzyme that is normally present in the plant and involved in nutrient utilization would increase the availability of a nutrient or decrease the availability of an anti-nutritive factor. An example of such an enzyme would be phytase. It is further contemplated that enhanced nitrogen utilization by a plant is desirable. Expression of a glutamate dehydrogenase gene in plants, e.g., *E. coli* gdhA genes, may lead to increased fixation of nitrogen in organic compounds. Furthermore, expression of gdhA in plants may lead to enhanced resistance to the herbicide glufosinate by incorporation of excess ammonia into glutamate, thereby detoxifying the ammonia. It also is contemplated that expression of a novel gene may make a nutrient source available that was previously not accessible, e.g., an enzyme that releases a component of nutrient value from a more complex molecule, perhaps a macromolecule.

[0115] Polypeptides useful for improving nitrogen flow, sensing, uptake, storage and/or transport include those involved in aspartate, glutamine or glutamate biosynthesis, polypeptides involved in aspartate, glutamine or glutamate transport, polypeptides associated with the TOR (Target of Rapamycin) pathway, nitrate transporters, nitrate reductases, amino transferases, ammonium transporters, chlorate trans-

porters or polypeptides involved in tetrapyrrole biosynthesis. Polypeptides useful for increasing the rate of photosynthesis include phytochrome, ribulose bisphosphate carboxylase-oxygenase, Rubisco activase, photosystem I and II proteins, electron carriers, ATP synthase, NADH dehydrogenase or cytochrome oxidase.

[0116] Polypeptides useful for increasing phosphorus uptake, transport or utilization include phosphatases or phosphate transporters.

Male Sterility

[0117] Male sterility is useful in the production of hybrid seed. Male sterility may be produced through expression of novel genes. For example, it has been shown that expression of genes that encode proteins, RNAs, or peptides that interfere with development of the male inflorescence and/or gametophyte result in male sterility. Chimeric ribonuclease genes that express in the anthers of transgenic tobacco and oilseed rape have been demonstrated to lead to male sterility (Mariani et al, Nature, 347:737-741, 1990).

[0118] A number of mutations were discovered in maize that confer cytoplasmic male sterility. One mutation in particular, referred to as T cytoplasm, also correlates with sensitivity to Southern corn leaf blight. A DNA sequence, designated TURF 13 (Levings, Science, 250:942-947, 1990), was identified that correlates with T cytoplasm. Therefore, in some embodiments, it would be possible through the introduction of TURF 13 via transformation, to separate male sterility from disease sensitivity. As it is necessary to be able to restore male fertility for breeding purposes and for grain production, genes encoding restoration of male fertility also may be introduced.

Altered Nutritional Content

[0119] Genes may be introduced into plants to improve or alter the nutrient quality or content of a particular crop. Introduction of genes that alter the nutrient composition of a crop may greatly enhance the feed or food value. For example, the protein of many grains is suboptimal for feed and food purposes, especially when fed to pigs, poultry, and humans. The protein is deficient in several amino acids that are essential in the diet of these species, requiring the addition of supplements to the grain. Limiting essential amino acids may include lysine, methionine, tryptophan, threonine, valine, arginine, and histidine. Some amino acids become limiting only after corn is supplemented with other inputs for feed formulations. The levels of these essential amino acids in seeds and grain may be elevated by mechanisms which include, but are not limited to, the introduction of genes to increase the biosynthesis of the amino acids, decrease the degradation of the amino acids, increase the storage of the amino acids in proteins, or increase transport of the amino acids to the seeds or grain.

[0120] Polypeptides useful for providing increased seed protein quantity and/or quality include polypeptides involved in the metabolism of amino acids in plants, particularly polypeptides involved in biosynthesis of methionine/cysteine and lysine, amino acid transporters, amino acid efflux carriers, seed storage proteins, proteases, or polypeptides involved in phytic acid metabolism. The protein composition of a crop may be altered to improve the balance of amino acids in a variety of ways including elevating expression of native proteins, decreasing expression of those with poor composition, changing the composition of native proteins, or introducing genes encoding entirely new proteins.

[0121] The introduction of genes that alter the oil content of a crop plant may also be of value. Increases in oil content may result in increases in metabolizable-energy-content and density of the seeds for use in feed and food. The introduced genes may encode enzymes that remove or reduce rate-limitations or regulated steps in fatty acid or lipid biosynthesis. Such genes may include, but are not limited to, those that encode acetyl-CoA carboxylase, ACP-acyltransferase, alpha-ketoacyl-ACP synthase, or other well known fatty acid biosynthetic activities. Other possibilities are genes that encode proteins that do not possess enzymatic activity such as acyl carrier protein. Genes may be introduced that alter the balance of fatty acids present in the 5 oil providing a more healthful or nutritive feedstuff. The introduced DNA also may encode sequences that block expression of enzymes involved in fatty acid biosynthesis, altering the proportions of fatty acids present in crops.

[0122] Genes may be introduced that enhance the nutritive value of crops, or of foods derived from crops by increasing the level of naturally occurring phytosterols, or by encoding for proteins to enable the synthesis of phytosterols in crops. The phytosterols from these crops can be processed directly into foods, or extracted and used to manufacture food products.

[0123] Genes may be introduced that enhance the nutritive value of the starch component of crops, for example by increasing the degree of branching, resulting in 5 improved utilization of the starch in livestock by delaying its metabolism. Additionally, other major constituents of a crop may be altered, including genes that affect a variety of other nutritive, processing, or other quality aspects. For example, pigmentation may be increased or decreased.

[0124] Carbohydrate metabolism may be altered, for example by increased sucrose production and/or transport. Polypeptides useful for affecting on carbohydrate metabolism include polypeptides involved in sucrose or starch metabolism, carbon assimilation or carbohydrate transport, including, for example sucrose transporters or glucose/hexose transporters, enzymes involved in glycolysis/gluconeogenesis, the pentose phosphate cycle, or raffinose biosynthesis, or polypeptides involved in glucose signaling, such as SNFI complex proteins.

[0125] Feed or food crops may also possess sub-optimal quantities of vitamins, antioxidants or other nutraceuticals, requiring supplementation to provide adequate nutritive value and ideal health value. Introduction of genes that enhance vitamin biosynthesis may be envisioned including, for example, vitamins A, E, B 12, choline, or the like. Mineral content may also be sub-optimal. Thus genes that affect the accumulation or availability of compounds containing phosphorus, sulfur, calcium, manganese, zinc, or iron among others would be valuable.

[0126] Numerous other examples of improvements of crops can be used according to the systems, nucleic acid compositions, and methods of the disclosure. The improvements may not necessarily involve grain, but may, for example, improve the value of a crop for silage. Introduction of DNA to accomplish this might include sequences that alter lignin production such as those that result in the "brown midrib" phenotype associated with superior feed value for cattle. Other genes may encode for enzymes that alter the

structure of extracellular carbohydrates in the stover, or that facilitate the degradation of the carbohydrates in the non-grain portion of the crop so that it can be efficiently fermented into ethanol or other useful carbohydrates. It may be desirable to modify the nutritional content of plants by reducing undesirable components such as fats, starches, etc. This may be done, for example, by the use of exogenous nucleic acids that encode enzymes which increase plant use or metabolism of such components so that they are present at lower quantities. Alternatively, it may be done by use of exogenous nucleic acids that reduce expression levels or activity of native plant enzymes that synthesize such components.

[0127] Likewise the elimination of certain undesirable traits may improve the food or feed value of the crop. Many undesirable traits must currently be eliminated by special post-harvest processing steps and the degree to which these can be engineered into the plant prior to harvest and processing would provide significant value. Examples of such traits are the elimination of anti-nutritionals such as phytates and phenolic compounds which are commonly found in many crop species. Also, the reduction of fats, carbohydrates and certain phytohormones may be valuable for the food and feed industries as they may allow a more efficient mechanism to meet specific dietary requirements.

[0128] In addition to direct improvements in feed or food value, genes also may be introduced which improve the processing of crops and improve the value of the products resulting from the processing. One use of crops is via wetmilling. Thus novel genes that increase the efficiency and reduce the cost of such processing, for example by decreasing steeping time, may also find use. Improving the value of wetmilling products may include altering the quantity or quality of starch, oil, corn gluten meal, or the components of gluten feed. Elevation of starch may be achieved through the identification and elimination of rate limiting steps in starch biosynthesis by expressing increased amounts of enzymes involved in biosynthesis or by decreasing levels of the other components of crops resulting in proportional increases in starch. Oil is another product of wetmilling, the value of which may be improved by introduction and expression of genes. Oil properties may be altered to improve its performance in the production and use of cooking oil, shortenings, lubricants or other oil-derived products or improvement of its health attributes when used in the food-related applications. Novel fatty acids also may be synthesized which upon extraction can serve as starting materials for chemical syntheses. The changes in oil properties may be achieved by altering the type, level, or lipid arrangement of the fatty acids present in the oil. This in turn may be accomplished by the addition of genes that encode enzymes that catalyze the synthesis of novel fatty acids (e.g. fatty acid elongases, desaturases) and the lipids possessing them or by increasing levels of native fatty acids while, in some embodiments, reducing levels of precursors or breakdown products. Alternatively, DNA sequences may be introduced which slow or block steps in fatty acid biosynthesis resulting in the increase in precursor fatty acid intermediates. Genes that might be added include desaturases, epoxidases, hydratases, dehydratases, or other enzymes that catalyze reactions involving fatty acid intermediates. Representative examples of catalytic steps that might be blocked include the desaturations from stearic to oleic acid or oleic to linolenic acid resulting in the respective accumulations of stearic and oleic acids. Another example is the blockage of elongation steps resulting in the accumulation of C8 to C12 saturated fatty acids.

[0129] Polypeptides useful for providing increased seed oil quantity and/or quality include polypeptides involved in fatty acid and glycerolipid biosynthesis, beta-oxidation enzymes, enzymes involved in biosynthesis of nutritional compounds, such as carotenoids and tocopherols, or polypeptides that increase embryo size or number or thickness of aleurone.

[0130] Polypeptides involved in production of galactomannans or arabinogalactans are of interest for providing plants having increased and/or modified reserve polysaccharides for use in food, pharmaceutical, cosmetic, paper and paint industries.

[0131] Polypeptides involved in modification of flavonoid/isoflavonoid metabolism in plants include cinnamate-4-hydroxylase, chalcone synthase or flavones synthase. Enhanced or reduced activity of such polypeptides in modified plants will provide changes in the quantity and/or speed of flavonoid metabolism in plants and may improve disease resistance by enhancing synthesis of protective secondary metabolites or improving signaling pathways governing disease resistance.

[0132] Polypeptides involved in lignin biosynthesis are of interest for increasing plants' resistance to lodging and for increasing the usefulness of plant materials as biofuels.

Production or Assimilation of Chemicals or Biologicals

[0133] It may further be considered that a modified plant prepared in accordance with the disclosure may be used for the production or manufacturing of useful biological compounds that were either not produced at all, or not produced at the same level, in the plant previously. Alternatively, plants produced in accordance with the disclosure may be made to metabolize or absorb and concentrate certain compounds, such as hazardous wastes, thereby allowing bioremediation of these compounds.

[0134] The novel plants producing these compounds are made possible by the introduction and expression of one or potentially many genes with the constructs provided by the disclosure. The vast array of possibilities include but are not limited to any biological compound which is presently produced by any organism such as proteins, nucleic acids, primary and intermediary metabolites, carbohydrate polymers, enzymes for uses in bioremediation, enzymes for modifying pathways that produce secondary plant metabolites such as falconoid or vitamins, enzymes that can produce pharmaceuticals, and for introducing enzymes that may produce compounds of interest to the manufacturing industry such as specialty chemicals and plastics. The compounds may be produced by the plant, extracted upon harvest and/or processing, and used for any presently recognized useful purpose such as pharmaceuticals, fragrances, and industrial enzymes to name a few.

Other Characteristics

[0135] Cell cycle modification: Polypeptides encoding cell cycle enzymes and regulators of the cell cycle pathway are useful for manipulating growth rate in plants to provide early vigor and accelerated maturation. Improvements in quality traits, such as seed oil content, may also be obtained by expression of cell cycle enzymes and cell cycle regula-

tors. Polypeptides of interest for modification of cell cycle pathway include cycling and EIF5a pathway proteins, polypeptides involved in polyamine metabolism, polypeptides which act as regulators of the cell cycle pathway, including cyclin-dependent kinases (CDKs), CDK-activating kinases, cell cycle-dependent phosphatases, CDK-inhibitors, Rb and Rb-binding proteins, or transcription factors that activate genes involved in cell proliferation and division, such as the E2F family of transcription factors, proteins involved in degradation of cyclins, such as cullins, and plant homologs of tumor suppressor polypeptides.

[0136] Plant growth regulators: Polypeptides involved in production of substances that regulate the growth of various plant tissues are of interest in the present disclosure and may be used to provide modified plants having altered morphologies and improved plant growth and development profiles leading to improvements in yield and stress response. Of particular interest are polypeptides involved in the biosynthesis, or degradation of plant growth hormones, such as gibberellins, brassinosteroids, cytokinins, auxins, ethylene or abscisic acid, and other proteins involved in the activity, uptake and/or transport of such polypeptides, including for example, cytokinin oxidase, cytokinin/purine permeases, F-box proteins, G-proteins or phytosulfokines.

[0137] Transcription factors in plants: Transcription factors play a key role in plant growth and development by controlling the expression of one or more genes in temporal, spatial and physiological specific patterns. Enhanced or reduced activity of such polypeptides in modified plants will provide significant changes in gene transcription patterns and provide a variety of beneficial effects in plant growth, development and response to environmental conditions. Transcription factors of interest include, but are not limited to myb transcription factors, including helix-turn-helix proteins, homeodomain transcription factors, leucine zipper transcription factors, MADS transcription factors, transcription factors having AP2 domains, zinc finger transcription factors, CCAAT binding transcription factors, ethylene responsive transcription factors, transcription initiation factors or UV damaged DNA binding proteins.

[0138] Homologous recombination: Increasing the rate of homologous recombination in plants is useful for accelerating the introgression of trans genes into breeding varieties by backcrossing, and to enhance the conventional breeding process by allowing rare recombinants between closely linked genes in phase repulsion to be identified more easily. Polypeptides useful for expression in plants to provide increased homologous recombination include polypeptides involved in mitosis and/or meiosis, DNA replication, nucleic acid metabolism, DNA repair pathways or homologous recombination pathways including for example, recombinases, nucleases, proteins binding to DNA double-strand breaks, single-strand DNA binding proteins, strand-exchange proteins, resolvases, ligases, helicases and polypeptide members of the RAD52 epistasis group.

[0139] Non-Protein-Expressing Exogenous Nucleic Acids Plants with decreased expression of a gene of interest can also be achieved, for example, by expression of antisense nucleic acids, dsRNA or RNAi, catalytic RNA such as ribozymes, sense expression constructs that exhibit cosuppression effects, aptamers or zinc finger proteins.

[0140] Antisense RNA reduces production of the polypeptide product of the target messenger RNA, for example by blocking translation through formation of RNA:RNA duplexes or by inducing degradation of the target mRNA. Antisense approaches are a way of preventing or reducing gene function by targeting the genetic material as disclosed in U.S. Pat. Nos. 4,801,540; 5,107,065; 5,759,829; 5,910,444; 6,184,439; and 6,198,026, all of which are incorporated herein by reference. In one approach, an antisense gene sequence is introduced that is transcribed into antisense RNA that is complementary to the target mRNA. For example, part or all of the normal gene sequences are placed under a promoter in inverted orientation so that the 'wrong' or complementary strand is transcribed into a non-protein expressing antisense RNA. The promoter used for the antisense gene may influence the level, timing, tissue, specificity, or inducibility of the antisense inhibition.

[0141] RNAi gene suppression in plants by transcription of a dsRNA is described in U.S. Pat. No. 6,506,559, U.S. patent application Publication No. 2002/0168707, WO 98/53083, WO 99/53050 and WO 99/61631, all of which are incorporated herein by reference. The double-stranded RNA or RNAi constructs can trigger the sequence-specific degradation of the target messenger RNA. Suppression of a gene by RNAi can be achieved using a recombinant DNA construct having a promoter operably linked to a DNA element comprising a sense and anti-sense element of a segment of genomic DNA of the gene, e.g., a segment of at least about 23 nucleotides, more preferably about 50 to 200 nucleotides where the sense and anti-sense DNA components can be directly linked or joined by an intron or artificial DNA segment that can form a loop when the transcribed RNA hybridizes to form a hairpin structure. Catalytic RNA molecules or ribozymes can also be used to inhibit expression of the target gene or genes or facilitate molecular reactions. Ribozymes are targeted to a given sequence by hybridization of sequences within the ribozyme to the target mRNA. Two stretches of homology are required for this targeting, and these stretches of homologous sequences flank the catalytic ribozyme structure. It is possible to design ribozymes that specifically pair with virtually any target mRNA and cleave the target mRNA at a specific location, thereby inactivating it. A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs that are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include Tobacco Ringspot Virus (Prody et ah, Science, 231:1577-1580, 1986), Avocado Sunblotch Viroid (Palukaitis et ah, Virology, 99:145-151, 1979; Symons, Nuch Acids Res., 9:6527-6537, 1981), and Lucerne Transient Streak Virus (Forster and Symons, Cell, 49:211-220, 1987), and the satellite RNAs from velvet tobacco mottle virus, *Solanum nodiflorum* mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff, et al., Nature 334:585-591 (1988). Several different ribozyme motifs have been described with RNA cleavage activity (Symons, Annu. Rev. Biochem., 61:641-671, 1992). Other suitable ribozymes include sequences from RNase P with RNA cleavage activity (Yuan et ah, Proc. Natl. Acad. Sd. USA, 89:8006-8010, 1992; Yuan and Altman, Science, 263:1269-1273, 1994; U.S. Pat. Nos. 5,168,053 and 5,624,824), hairpin ribozyme structures (Berzal-Herranz et ah, Genes and Deveh, 6:129-134, 1992; Chowrira et ah, J. Biol. Chem., 269:25856-25864, 1994) and Hepatitis Delta virus based ribozymes (U.S. Pat. No. 5,625,047). The general design and

optimization of ribozyme directed RNA cleavage activity has been discussed in detail (Haseloff and Gerlach, 1988, Nature. 1988 Aug. 18; 334(6183):585-91, Chowrira et al., J. Biol. Chem., 269:25856-25864, 1994).

[0142] Another method of reducing protein expression utilizes the phenomenon of cosuppression or gene silencing (for example, U.S. Pat. Nos. 6,063,947; 5,686,649; or 5,283,184; each of which is incorporated herein by reference). Cosuppression of an endogenous gene using a full-length cDNA sequence as well as a partial cDNA sequence are known (for example, Napoli et al., Plant Cell 2:279-289 [1990]; van der Krol et al., Plant Cell 2:291-299 [1990]; Smith et al., MoI. Gen. Genetics 224:477-481 [1990]). The phenomenon of cosuppression has also been used to inhibit plant target genes in a tissue-specific manner. In some embodiments, nucleic acids from one species of plant are expressed in another species of plant to effect cosuppression of a homologous gene. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed, for example, about 65%, 80%, 85%, 90%, or preferably 95% or greater identical. Higher identity may result in a more effective repression of expression of the endogenous sequence. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Generally, where inhibition of expression is desired, some transcription of the introduced sequence occurs. The effect may occur where the introduced sequence contains no coding sequence per se, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence.

[0143] Yet another method of reducing protein activity is by expressing nucleic acid ligands, so-called aptamers, which specifically bind to the protein. Aptamers may be obtained by the SELEX (Systematic Evolution of Ligands by Exponential Enrichment) method. See U.S. Pat. No. 5,270,163, incorporated herein by reference. In the SELEX method, a candidate mixture of single stranded nucleic acids having regions of randomized sequence is contacted with the protein and those nucleic acids having an increased affinity to the target are selected and amplified. After several iterations a nucleic acid with optimal affinity to the polypeptide is obtained and is used for expression in modified plants.

[0144] A zinc finger protein that binds a polypeptide-encoding sequence or its regulatory region is also used to alter expression of the nucleotide sequence. Transcription of the nucleotide sequence may be reduced or increased. Zinc finger proteins are, for example, described in Beerli et al. (1998) PNAS 95:14628-14633., or in WO 95/19431, WO 98/54311, or WO 96/06166, all incorporated herein by reference.

[0145] Other examples of non-protein expressing sequences specifically envisioned for use with the disclosure include tRNA sequences, for example, to alter codon usage, and rRNA variants, for example, which may confer resistance to various agents such as antibiotics. It is contemplated that unexpressed DNA sequences, including novel synthetic sequences, may be introduced into cells as proprietary "labels" of those cells and plants and seeds thereof. It would not be necessary for a label DNA element to disrupt the function of a gene endogenous to the host organism, as the sole function of this DNA would be to identify the origin of the organism. For example, one can introduce a unique DNA sequence into a plant and this DNA element would identify all cells, plants, and progeny of these cells as having arisen from that labeled source. Inclusion of label DNAs would enable one to distinguish proprietary germplasm or germplasm derived from such, from unlabelled germplasm.

[0146] In some embodiments, the cargo sequence comprises a detectable protein, for example a protein tag or a fluorescent protein. The tag can be an epitope tag. The epitope tag can comprise a myc tag, a FLAG tag, a polyHistidine tag, a HiBiT tag, HA tag, S-peptide tag, calmodulin-binding peptide (CBP), glutathione S-transferase (GST), maltose binding protein (MBP), or any combination thereof. The cargo sequence can comprise an exogenous sequence encoding a fluorescent protein. The fluorescent protein can comprise mPlum, mCherry, DsRed, FASTred, mOrange, EYFP, VENUS, YPet, GFP, EGFP, EmGFP, mCFP, Cerulean, CyPet, Kaede, or any combination thereof.

Rna-Guided DNA Binding Complex and Transposition Complex

[0147] As provided above, CRISPR-associated transposons or CASTs are mobile genetic elements (MGEs) that have evolved to make use of minimal CRISPR systems for RNA-guided transposition of their DNA. Unlike traditional CRISPR systems that contain interference mechanisms to degrade targeted DNA, CASTs lack proteins and/or protein domains responsible for DNA cleavage. Specialized transposon machinery, similar to that of Tn7 transposon, complexes with the CRISPR RNA (crRNA) and associated Cas proteins for transposition. CAST systems have been characterized in a wide range of bacteria and make use of variable CRISPR configurations including Type I-F, Type I-B, Type I-C, Type I-D, Type I-E, Type IV, and Type V-K.

[0148] CRISPR-associated transposons are, in some instances, similar to the Tn7 transposon which functions with a cut and paste mechanism. It contains a heteromeric transposase consisting of TnsA and TnsB proteins, and a regulator protein TnsC. Structural analysis has shown binding of the TnsB protein and sequence specific motifs on the ends of the transposon which allows for excision and mobility. Targeting for integration is done by the TnsD or TnsE proteins which preferentially target safe sites within the host chromosome or mobile elements (plasmids or bacteriophages), respectively. TnsE is not found in CASTs but a TnsD homolog, TniQ, is present and functions to bridge the gap between the transposase and CRISPR-Cas. Multiple CRISPR types have been found to associate with transposons with two of the most studied being Type I-F, which makes use of a multi-subunit effector, and Type V-K, which makes use of a single Cas12k effector. In both cases, Tn7 transposons have evolved to make use of these effectors to create R loops for site-specific integration. While TnsA is present in Type I-F systems, it is notably absent in Type V-K systems which showed higher off-target integrations during initial characterization.

[0149] Provided herein are systems comprising an RNA-guided DNA binding complex or one or more first helper polynucleotides each comprising a sequence encoding a component of the RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof. In some embodiments, there is provided nucleic acid compositions comprising one or more

first helper polynucleotides each comprising a sequence encoding a component of an RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof. The system, the RNA-guided DNA binding complex and/or the transposition complex can be derived from a Type I-B, Type I-D, TypeI-F, or Type V-K Crispr-associated transposase system of a bacteria. The bacteria can comprise *Vibrio cholera* (Vch), *Pseudoalteromonas* (Pse), or *Scytonema hofmanni* (Sho).

[0150] The one or more Cas proteins can comprise a Cas6 protein. The one or more Cas proteins can comprise a Cas7 protein. The one or more Cas proteins can comprise a Cas8 protein. The one or more Cas proteins can comprise a Cas6 protein, a Cas7 protein, and a Cas8 protein.

[0151] The Cas6 protein can comprise or consist of an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 8-9. The Cas7 protein can comprise or consist of an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 10-11. The Cas8 protein can comprise or consist of an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 12-13. The transposase of the RNA-guided DNA binding complex can comprise a TniQ protein. The TniQ protein can comprise an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 14-15.

[0152] The one or more transposases of the transposition complex can comprise a TnsA transposase. The one or more transposases of the transposition complex can comprise a TnsB transposase. The one or more transposases of the transposition complex can comprise a TnsC protein. The one or more transposases of the transposition complex can comprise a TnsA transposase, a TnsB transposase, and a TnsC protein. The one or more transposases of the transposition complex can comprise a TnsAB fusion protein and a TnsC protein. The TnsAB fusion protein can comprise an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 16-17. The TnsC protein can comprise an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 18-19.

[0153] In some embodiments, the system comprises one or more helper accessory proteins or one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and/or ClpP. ClpX (e.g., ClpX AAA+ATPase) and ClpP (e.g., ClpP peptidase) are bacterial proteins that can promote disassembly of post-transposition complex (PTC) in bacterial Tn7 and Mu transposon systems. In some embodiments, the activity of ClpX and/or ClpP can enhance the transposition efficiency, compared to methods and systems that lack said proteins.

[0154] The ClpX can comprise or consist of an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to SEQ ID NO: 21. The ClpP can comprise or consist of an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to SEQ ID NO: 20.

crRNAs

[0155] The RNA-guided DNA-binding complex can comprise a crRNA. The terms "gRNA," "guide RNA", "CRISPR guide sequence" or "crRNA" may be used interchangeably throughout and refer to a nucleic acid comprising a sequence that determines the binding specificity of the RNA-guided DNA-binding complex. A crRNA hybridizes to (complementary to, partially or completely) a double-stranded target sequence (e.g., in the genome) in a cell, e.g., a plant cell. The crRNA can comprise a spacer that hybridizes to a search target sequence on a first strand of the double stranded target sequence (a target site). The spacer may be between 15-35 nucleotides, 18-33 nucleotides, or 19-35 nucleotides in length. In some embodiments, the crRNA (e.g., gRNA) sequence that hybridizes to the target site is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 26, 27, 28, 29, 30, 31, 32, or 33 nucleotides in length. In some embodiments, the crRNA sequence that hybridizes to the target site is between 10-30, or between 15-25, or 25-35 nucleotides in length.

[0156] To facilitate crRNA design, many computational tools have been developed (See Prykhozhij et al (PLoS ONE, 10(3): (2015)); Zhu et al. (PLoS ONE, 9(9) (2014)); Xiao et al. (Bioinformatics. January 21 (2014)); Heigwer et al. (Nat Methods, 11(2); 122-123 (2014)). Methods and tools for guide RNA design are discussed by Zhu (Frontiers in Biology, 10 (4) pp 289-296 (2015)), which is incorporated by reference herein. Additionally, there are many publicly available software tools that can be used to facilitate the design of crRNA(s); including but not limited to, Genscript Interactive CRISPR gRNA Design Tool, WU-CRISPR, and Broad Institute GPP sgRNA Designer. There are also publicly available pre-designed gRNA sequences to target many genes and locations within the genomes of many species (human, mouse, rat, zebrafish, *C. elegans*), including but not limited to, IDT DNA Predesigned Alt-R CRISPR-Cas9 guide RNAs, Addgene Validated gRNA Target Sequences, and GenScript Genome-wide gRNA databases.

[0157] In addition to a sequence that binds to a target nucleic acid (e.g., a spacer), in some embodiments, the gRNA may also comprise a scaffold sequence. In some

embodiments, such a gRNA may be referred to as a single guide RNA (sgRNA) or a crRNA. Exemplary scaffold sequences will be evident to one of skill in the art and can be found, for example, in Jinek, et al. Science (2012) 337(6096):816-821, and Ran, et al. Nature Protocols (2013) 8:2281-2308. crRNA(s) used in the present disclosure can be between about 5 and 100 nucleotides long, or longer (e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59 60, 61, 62, 63, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91 92, 93, 94, 95, 96, 97, 98, 99, or 100 nucleotides in length, or longer).

[0158] In some embodiments, the gRNA sequence does not comprise a scaffold sequence and a scaffold sequence is expressed as a separate transcript. In such embodiments, the gRNA sequence further comprises an additional sequence that is complementary to a portion of the scaffold sequence and functions to bind (hybridize) the scaffold sequence. The crRNA can comprise a [repeat scaffold]-[spacer]-[repeat scaffold] structure. The first strand of the double stranded target sequence can be the sense strand.

[0159] In some embodiments, the spacer sequence of the crRNA is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or at least 100% complementary to a search target sequence on a first strand of the double stranded target sequence. In some embodiments, the gRNA sequence is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or at least 100% complementary to the 3' end of the search target sequence on a first strand of the double stranded target sequence (e.g., the last 5, 6, 7, 8, 9, or 10 nucleotides of the 3' end of the target sequence). The crRNA can comprise a spacer that is complementary to a search target sequence on a first strand of the double stranded target sequence. The first strand of the double stranded target sequence can be the sense strand.

[0160] The target sequence may be flanked by a protospacer adjacent motif (PAM). A PAM site is a nucleotide sequence in proximity to a target sequence. For example, a PAM may be a DNA sequence immediately following the DNA sequence targeted by the RNA-guided DNA binding complex.

[0161] The target sequence may or may not be flanked by a protospacer adjacent motif (PAM) sequence. In certain embodiments, a nucleic acid-guided nuclease can only bind a target sequence if an appropriate PAM is present, see, for example Doudna et al., Science, 2014, 346(6213): 1258096, incorporated herein by reference. A PAM can be 5' or 3' of a target sequence. A PAM can be upstream or downstream of a target sequence. In some embodiments, the target sequence is immediately flanked on the 3' end by a PAM sequence. A PAM can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides in length. In certain embodiments, a PAM is between 2-6 nucleotides in length. The target sequence may or may not be located adjacent to a PAM sequence (e.g., PAM sequence located immediately 3' of the target sequence) (e.g., for Type I CRISPR/Cas systems and Type II CRISPR/Cas systems). In some embodiments, e.g., Type I systems, the PAM is on the alternate side of the protospacer (the 5' end). Makarova et al. describes the nomenclature for all the classes, types and subtypes of CRISPR systems (Nature Reviews Microbiology 13:722-736 (2015)). Guide structures and PAMs are described in by R. Barrangou (Genome Biol. 16:247 (201 5)).

[0162] Non-limiting examples of the PAM sequences include: CC, CA, AG, GT, TA, AC, CA, GC, CG, GG, CT, TG, GA, AGG, TGG, T-rich PAMs (such as ITT, TTG, TTC, TTTT, NGG, NGA, NAG, NGGNG and NNAGAAW (W=A or T), NNNNGATT, NAAR (R=A or G), NNGRR (R=A or G), NNAGAA and NAAAAC, where "N" is any nucleotide. In some embodiments, e.g., for Type I-F CAST systems and the systems, nucleic acid compositions and methods described herein, the PAM sequence can be 5'-CN-3', where "N" is any nucleotide.

[0163] "Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule, which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization. There may be mismatches distal from the PAM.

[0164] The cargo sequence can be capable of being integrated at an integration site following binding of the RNA-guided DNA binding complex to the search target sequence. The integration site can be about 48 to 52 base pairs (e.g., 48, 49, 50, 51, or 52 base pairs) downstream of the double stranded target sequence. The double stranded target sequence can be situated within a selectable marker gene of the genome of the plant cell. The selectable marker gene can comprise a fluorescent protein coding gene, a phytoene desaturase (PDS) gene, a codA gene, a diphtheria toxin a subunits (DT-A) gene, an exotoxin A gene, a ricin toxin A gene, a cytochrome P-450 gene, an RNase T1 gene, or a barnase gene. The double stranded target sequence can be situated within a safe harbor locus of the genome of the plant cell.

Polynucleotides and Nucleic Acid Compositions

[0165] Disclosed herein include systems for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the system comprises: i) an RNA-guided DNA binding complex or one or more first helper polynucleotides each comprising a sequence encoding a component of the RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the system comprises one or more helper accessory proteins or one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP; ii) a transposition complex or one or more second helper polynucleotides each comprising a sequence encoding a component of the transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

[0166] Disclosed herein include nucleic acid compositions for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the nucleic acid composition comprises: i) one or more first helper polynucleotides each comprising a sequence encoding a component of an RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the nucleic acid composition comprises one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP; ii) one or more second helper polynucleotides each comprising a sequence encoding a component of a transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

[0167] Each of the one or more first and/or second helper polynucleotides, the one or more helper accessory polynucleotides or both, can be operably linked to one omore expression control elements (e.g., a promoter and/or a transcriptional terminator). Each of the one or more first helper polynucleotides can comprise a first promoter operably linked to the sequence encoding the component of the RNA-guided DNA binding complex. Each of the one or more second helper polynucleotides can comprise a second promoter operably linked to the sequence encoding the component of the transposition complex. Each of the one or more helper accessory polynucleotides can comprise a third promoter operably linked to the sequence encoding at least one of the one or more helper accessory proteins.

[0168] The first, second, and/or third promoters can be the same or different. The first, second, and/or third promoters can comprise a ubiquitous promoter, a constitutive promoter, a cell-type specific promoter, a tissue-specific promoter, an inducible promoter, or any combination thereof. In some embodiments, the constitutive promoter is selected from the group comprising: pCmYLCV911 (pCmY), pU6, pU3, pU6, pAct2, pAct-1, pUBQ10, pUBQ4, pUbi1, and PUbi2; the tissue-specific promoter is selected from the group comprising: pSIREO, pNAC10, pPAT21, phspr, pPFn2, pPEPC, PLhcb, pTA29, pLat52, pZml3, pOleosin, pGlutenin, pDhordein, and pE8; the inducible promoter is selected from the group comprising: pAdh-1, pwunl, pGBSS, pHSP18.2, pRd29, pSR2, pCCA1, pUGT71C5, pGSE, pwin3.12, pR2329, pBs3, pCaPrx, p4xM1.1, p4xM2.3, pIFS2, pSAG12, pSEOF1, pEm, pRd29, pSAUR15A, and pChn48.

[0169] Each of the one or more first helper polynucleotides can comprise a first transcription terminator operably linked to the sequence encoding the component of the RNA-guided DNA binding complex. Each of the one or more second helper polynucleotides can comprise a second transcription terminator operably linked to the sequence encoding the component of the transposition complex; and/or Each of the one or more helper accessory polynucleotides can comprise a third transcription terminator operably linked to the sequence encoding at least one of the one or more helper accessory proteins.

[0170] The first, second, and/or third transcription terminators can be the same or different. The first, second, and/or third transcription terminators can comprise AtHSP18.2 (tHSP), tU6, tACT3, tACT3-tRb7MAR, tACT3-tTM6MAR, tEU, tEU-tTM6MAR, tEU (intronless), tEU (intronless)-tACT3-tRB7MAR, tHSP18-tEU-tRb7MAR, tHSP18-tACT3, tHSP18-tACT3-tRb7, tHSP18-tPINII-tRb7MAR, tHSP18-tPINII-tTM6MAR, tHSP18-tRb7MAR, tProteinase inhibitor II (tPINII), trbcS, or any combination thereof.

[0171] The system or the nucleic acid composition can comprise: at least three first helper polynucleotides each comprising a sequence encoding a Cas protein, wherein the sequence encoding the Cas protein is operably linked to a pCmY promoter and a tHSP terminator; a first helper polynucleotide comprising a sequence encoding a transposase protein, wherein the sequence encoding the transposase protein is operably linked to a pCmY promoter and a tHSP terminator; a first helper polynucleotide comprising a sequence encoding a crRNA, wherein the sequence encoding the crRNA is operably linked to a pU6 promoter and a tU6 terminator; and at least two second helper polynucleotides each comprising a sequence encoding a transposase, wherein the sequence encoding a transposase is operably linked to a pCmY promoter and a tHSP terminator. The system or the nucleic acid composition can comprise at least two helper accessory polynucleotides, wherein the sequence encoding at least one of the one or more helper accessory proteins is operably linked to a pCmY promoter and a tHSP terminator.

[0172] Further examples of promoters and terminators that can be used systems, compositions, and methods of the disclosure are describe below.

[0173] Exemplary classes of plant promoters are described below. Constitutive Expression promoters: Exemplary constitutive expression promoters include the ubiquitin promoter (e.g., sunflower-Binet et al. Plant Science 79: 87-94 (1991); maize-Christensen et al. Plant Molec. Biol. 12: 619-632 (1989); and *Arabidopsis*-Callis et al., J. Biol. Chem. 265: 12486-12493 (1990) and Norris et al., Plant Mol. Biol. 21: 895-906 (1993)); the CaMV 35S promoter (U.S. Pat. Nos. 5,858,742 and 5,322,938); or the actin promoter (e.g., rice—U.S. Pat. No. 5,641,876; McElroy et al. Plant Cell 2: 163-171 (1990), McElroy et al. Mol. Gen. Genet. 231: 150-160 (1991), and Chibbar et al. Plant Cell Rep. 12: 506-509 (1993)). Inducible Expression promoters: Exemplary inducible expression promoters include the chemically regulatable tobacco PR-I promoter (e.g., tobacco—U.S. Pat. No. 5,614,395; *Arabidopsis*-Lebel et al., Plant J. 16: 223-233 (1998); maize—U.S. Pat. No. 6,429,362). Various chemical regulators may be employed to induce expression, including the benzothiadiazole, isonicotinic acid, and salicylic acid compounds disclosed in U.S. Pat. Nos. 5,523,311 and 5,614,395. Other promoters inducible by certain alcohols or ketones, such as ethanol, include, for example, the alcA gene promoter from *Aspergillus nidulans* (Caddick et al. (1998) Nat. Biotechnol 16:177-180). A glucocorticoid-mediated induction system is described in Aoyama and Chua (1997) The Plant Journal 11: 605-612 wherein gene expression is induced by application of a glucocorticoid, for example a dexamethasone. Another

class of useful promoters are water-deficit-inducible promoters, e.g. promoters which are derived from the 5' regulatory region of genes identified as a heat shock protein 17.5 gene (HSP 17.5), an HVA22 gene (HVA22), and a cinnamic acid 4-hydroxylase (CA4H) gene of *Zea mays*. Another water-deficit-inducible promoter is derived from the rab-17 promoter as disclosed by Vilardell et al., Plant Molecular Biology, 17(5):985-993, 1990. See also U.S. Pat. No. 6,084,089 which discloses cold inducible promoters, U.S. Pat. No. 6,294,714 which discloses light inducible promoters, U.S. Pat. No. 6,140,078 which discloses salt inducible promoters, U.S. Pat. No. 6,252,138 which discloses pathogen inducible promoters, and U.S. Pat. No. 6,175,060 which discloses phosphorus deficiency inducible promoters.

[0174] As another example, numerous wound-inducible promoters have been described (e.g. Xu et al. Plant Molec. Biol. 22: 573-588 (1993), Logemann et al. Plant Cell 1: 151-158 (1989), Rohrmeier & Lehle, Plant Molec. Biol. 22: 783-792 (1993), Firek et al. Plant Molec. Biol. 22: 129-142 (1993), Warner et al. Plant J. 3: 191-201 (1993)). Logemann describe 5' upstream sequences of the potato wunl gene. Xu et al. show that a wound-inducible promoter from the dicotyledon potato (pin2) is active in the monocotyledon rice. Rohrmeier & Lehle describe maize Wipl cDNA which is wound induced and which can be used to isolate the cognate promoter. Firek et al. and Warner et al. have described a wound-induced gene from the monocotyledon *Asparagus officinalis*, which is expressed at local wound and pathogen invasion sites.

[0175] Tissue-Specific Promoters: Exemplary promoters that express genes only in certain tissues are useful according to the present disclosure. For example, root specific expression can be attained using the promoter of the maize metallothionein-like (MTL) gene described by de Framond (FEBS 290: 103-106 (1991)) and also in U.S. Pat. No. 5,466,785, incorporated herein by reference. U.S. Pat. No. 5,837,848 discloses a root specific promoter. Another exemplary promoter confers pith-preferred expression (see PCT Pub. No. WO 93/07278, herein incorporated by reference, which describes the maize trpA gene and promoter that is preferentially expressed in pith cells). Leaf-specific expression may be attained, for example, by using the promoter for a maize gene encoding phosphoenol carboxylase (PEPC) (see Hudspeth & Grula, Plant Molec Biol 12: 579-589 (1989)). Pollen-specific expression may be conferred by the promoter for the maize calcium-dependent protein kinase (CDPK) gene which is expressed in pollen cells (WO 93/07278). U.S. Pat. Appl. Pub. No. 20040016025 describes tissue-specific promoters. Pollen-specific expression may be conferred by the tomato LAT52 pollen-specific promoter (Bate et. al., Plant Mol Biol. 1998 July; 37(5):859-69). See also U.S. Pat. No. 6,437,217 which discloses a root-specific maize RS81 promoter, U.S. Pat. No. 6,426,446 which discloses a root specific maize RS324 promoter, U.S. Pat. No. 6,232,526 which discloses a constitutive maize A3 promoter, U.S. Pat. No. 6,177,611 which discloses constitutive maize promoters, U.S. Pat. No. 6,433,252 which discloses a maize L3 oleosin promoter that are aleurone and seed coat-specific promoters, U.S. Pat. No. 6,429,357 which discloses a constitutive rice actin 2 promoter and intron, U.S. patent application Pub. No. 20040216189 which discloses an inducible constitutive leaf specific maize chloroplast aldolase promoter.

[0176] Optionally a plant transcriptional terminator can be used in place of the plant-expressed gene native transcriptional terminator. Exemplary transcriptional terminators are those that are known to function in plants and include the CaMV 35S terminator, the tml terminator, the nopaline synthase terminator and the pea rbcS E9 terminator. These can be used. in both monocotyledons and dicotyledons.

[0177] Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize Adhl gene have been found to significantly enhance expression. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis et al., Genes Develop. 1: 1183-1200 (1987)). The intron from the maize bronzel gene also enhances expression. Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader. U.S. Patent Application Publication 2002/0192813 discloses 5', 3' and intron elements useful in the design of effective plant expression vectors.

[0178] A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically, leader sequences from Tobacco Mosaic Virus (TMV, the "omega-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (e.g. Gallie et al. Nucl. Acids Res. 15: 8693-8711 (1987); Skuzeski et al. Plant Molec. Biol. 15: 65-79 (1990)). Other leader sequences known in the art include but are not limited to: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein, O., Fuerst, T. R., and Moss, B. PNAS USA 86:6126-6130 (1989)); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Allison et al., 1986); MDMV leader (Maize Dwarf Mosaic Virus); Virology 154:9-20); human immunoglobulin heavy-chain binding protein (BiP) leader, (Macejak, D. G., and Sarnow, P., Nature 353: 90-94 (1991); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), (Jobling, S. A., and Gehrke, L., Nature 325:622-625 (1987); tobacco mosaic virus leader (TMV), (Gallie et al., Molecular Biology of RNA, pages 237-256 (1989); or Maize Chlorotic Mottle Virus leader (MCMV) (Lommel et al., Virology 81:382-385 (1991). See also, Della-Cioppa et al., Plant Physiology 84:965-968 (1987). A minimal promoter may also be incorporated. Such a promoter has low background activity in plants when there is no transactivator present or when enhancer or response element binding sites are absent. One exemplary minimal promoter is the BzI minimal promoter, which is obtained from the bronzel gene of maize. Roth et al., Plant Cell 3: 317 (1991). A minimal promoter may also be created by use of a synthetic TATA element. The TATA element allows recognition of the promoter by RNA. polymerase factors and confers a basal level of gene expression in the absence of activation (see generally, Mukumoto (1993) Plant Mol Biol 23: 995-1003; Green (2000) Trends Biochem Sci 25: 59-63).

[0179] Sequences controlling the targeting of gene products also may be included. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the amino terminal end of various proteins which is cleaved during chloroplast import to yield the mature protein (e.g. Comai et al. J. Biol. Chem. 263: 15104-15109 (1988)). These signal sequences can be fused

to heterologous gene products to effect the import of heterologous products into the chloroplast (van den Broeck, et al. Nature 313: 358-363 (1985)). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the RUBISCO protein, the CAB protein, the EPSP synthase enzyme, the GS2 protein or many other proteins which are known to be chloroplast localized. Other gene products are localized to other organelles such as the mitochondrion and the peroxisome (e.g. Unger et al. Plant Molec. Biol. 13: 411-418 (1989)). Examples of sequences that target to such organelles are the nuclear-encoded ATPases or specific aspartate amino transferase isoforms for mitochondria. Targeting cellular protein bodies has been described by Rogers et al. (Proc. Natl. Acad. Sci. USA 82: 6512-6516 (1985)). In addition, amino terminal and carboxy-terminal sequences are responsible for targeting to the ER, the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, Plant Cell 2: 769-783 (1990)). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for vacuolar targeting of gene products (Shinshi et al. Plant Molec. Biol. 14: 357-368 (1990)).

[0180] Another element which may be introduced is a matrix attachment region element (MAR), such as the chicken lysozyme A element (Stief, 1989), which can be positioned around an expressible gene of interest to effect an increase in overall expression of the gene and diminish position dependent effects upon incorporation into the plant genome (Stief et al., Nature, 341:343, 1989; Phi-Van et al., MoI. Cell. Biol., 10:2302-2307.1990).

[0181] Use of non-plant promoter regions isolated from *Drosophila melanogaster* and *Saccharomyces cerevisiae* can be used to express genes in plants. The promoter can be derived from plant or non-plant species. In some embodiments, the nucleotide sequence of the promoter is derived from non-plant species for the expression of genes in plant cells, including but not limited to dicotyledon plant cells such as tobacco, tomato, potato, soybean, canola, sunflower, alfalfa, cotton and *Arabidopsis*, or monocotyledonous plant cell, such as wheat, maize, rye, rice, turf grass, oat, barley, sorghum, millet, and sugarcane. In some embodiments, the non-plant promoters are constitutive or inducible promoters derived from insect, e.g., *Drosophila melanogaster* or yeast, e.g., *Saccharomyces cerevisiae*. Promoters derived from any animal, protist, or fungi are also contemplated. These non-plant promoters can be operably linked to nucleic acid sequences encoding polypeptides or non-protein-expressing sequences including, but not limited to, antisense RNA and ribozymes, to form nucleic acid constructs, vectors, and host cells (prokaryotic or eukaryotic), comprising the promoters.

[0182] The sequence encoding the component of the RNA-guided DNA binding complex, the sequence encoding the component of the transposition complex, the sequence encoding at least one of the one or more helper accessory proteins, or any combination thereof, can be codon optimized for expression in the plant cell.

[0183] In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas6 protein. The sequence encoding the Cas6 protein can comprise or consist of the nucleotide sequence of any one of SEQ ID NOs: 24-25 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 24-25. In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas7 protein. The sequence encoding the Cas7 protein can comprise or consist of the nucleotide sequence of any one of SEQ ID NOs: 26-27 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 26-27. In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas8 protein. The sequence encoding the Cas8 protein can comprise or consist of the nucleotide sequence of any one of SEQ ID NOs: 28-29 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 28-29. In some embodiments, the sequence encoding the component of the RNA-guided DNA binding complex encodes a TniQ protein. The sequence encoding the TniQ protein can comprise or consist of the nucleotide sequence of any one of SEQ ID NOs: 30-31 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 30-31.

[0184] In some embodiments, the sequence encoding the component of the transposition complex encodes a TnsAB fusion protein. The sequence encoding the TnsAB fusion protein can comprise or consist of the nucleotide sequence of any one of SEQ ID NOs: 32-33 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 32-33. In some embodiments, the sequence encoding the component of the transposition complex encodes a TnsC protein. The sequence encoding the TnsC protein can comprise or consist of the nucleotide sequence of any one of SEQ ID NOs: 34-35 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to any one of SEQ ID NOs: 34-35.

[0185] The sequence encoding ClpX can comprise or consist of the nucleotide sequence of SEQ ID NO: 37 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to SEQ ID NO: 37 and the sequence encoding ClpP can comprise or consist of the nucleotide sequence of SEQ ID NO: 36 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) identical to SEQ ID NO: 36.

[0186] The component of the RNA-guided DNA binding complex, the component of the transposition complex, or

both, can comprise an N-terminal or a C-terminal tag. The tag can be an epitope tag. The epitope tag can comprise a myc tag, a FLAG tag, a polyHistidine tag, a HiBiT tag, HA tag, S-peptide tag, calmodulin-binding peptide (CBP), glutathione S-transferase (GST), maltose binding protein (MBP), or any combination thereof. The tag can comprise a fluorescent protein. The fluorescent protein can comprise mPlum, mCherry, DsRed, FASTred, mOrange, EYFP, VENUS, YPet, GFP, EGFP, EmGFP, mCFP, Cerulean, CyPet, Kaede, or any combination thereof.

[0187] The one or more first helper polynucleotides the one or more second helper polynucleotides, and/or the donor polynucleotide can be situated on the same nucleic acid or different nucleic acids. The one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide can be comprised within one or more vectors. The one or more vectors can comprise an RNA viral vector, a DNA viral vector, a plasmid vector, an artificial chromosome, or any combination thereof. The one or more vectors can comprise an *Agrobacterium tumefaciens* Ti vector. The one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide can be comprised within a T-DNA region of the *Agrobacterium tumefaciens* Ti vector.

[0188] The T-DNA region comprising the one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide can comprise the sequence of any one of SEQ ID NOs: 38-46. The T-DNA region comprising the one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide can comprise the sequence of any one of SEQ ID NOs: 38-46 or a sequence that is at least 80% identical (e.g., 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values) to any one of SEQ ID NOs: 38-46.

Plant Cell Transformation and Delivery

[0189] Any method known in the art can be used to deliver a system or nucleic acid composition of the disclosure to a plant cell. *Agrobacterium* (e.g., *Agrobacterium tumefaciens*)-mediated transformation is one method for introducing a desired genetic element into a plant. Several *Agrobacterium* species mediate the transfer of a specific DNA known as "T-DNA" that can be genetically engineered to carry a desired piece of DNA into many plant species. Plasmids used for delivery contain the T-DNA flanking the nucleic acid to be inserted into the plant. The major events marking the process of T-DNA mediated pathogenesis are induction of virulence genes, processing and transfer of T-DNA.

[0190] There are several methods known in the art to transform plant cells with *Agrobacterium*. One method is co-cultivation of *Agrobacterium* with cultured isolated protoplasts. This method requires an established culture system that allows culturing protoplasts and plant regeneration from cultured protoplasts. A second method is transformation of cells or tissues with *Agrobacterium*. This method requires (a) that the plant cells or tissues can be modified by *Agrobacterium* and (b) that the modified cells or tissues can be induced to regenerate into whole plants. A third method is transformation of seeds, apices or meristems with *Agrobacterium*. This method requires exposure of the meristematic cells of these tissues to *Agrobacterium* and micropropagation of the shoots or plan organs arising from these meristematic cells. Those of skill in the art are familiar with procedures for growth and suitable culture conditions for *Agrobacterium* as well as subsequent inoculation procedures. Liquid or semi-solid culture media can be used. The density of the *Agrobacterium* culture used for inoculation and the ratio of *Agrobacterium* cells to explant can vary from one system to the next, as can media, growth procedures, timing and lighting conditions.

[0191] Transformation of dicotyledons using *Agrobacterium* has long been known in the art, and transformation of monocotyledons using *Agrobacterium* has also been described. See, WO 94/00977 and U.S. Pat. No. 5,591,616, both of which are incorporated herein by reference. See also, Negrotto et al., Plant Cell Reports 19: 798-803 (2000), incorporated herein by reference.

[0192] A number of wild-type and disarmed strains of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes* harboring Ti or Ri plasmids can be used for gene transfer into plants. Preferably, the *Agrobacterium* hosts contain disarmed Ti and Ri plasmids that do not contain the oncogenes that cause tumorigenesis or rhizogenesis. Exemplary strains include *Agrobacterium tumefaciens* strain C58, a nopaline-type strain that is used to mediate the transfer of DNA into a plant cell, octopine-type strains such as LBA4404 or succinamopine-type strains, e.g., EHAlOl or EHAl 05. The use of these strains for plant transformation has been reported and the methods are familiar to those of skill in the art.

[0193] U.S. Application No. 20040244075 published Dec. 2, 2004 describes improved methods of *Agrobacterium*-mediated transformation. The efficiency of transformation by *Agrobacterium* may be enhanced by using a number of methods known in the art. For example, the inclusion of a natural wound response molecule such as acetosyringone (AS) to the *Agrobacterium* culture has been shown to enhance transformation efficiency with *Agrobacterium tumefaciens* (Shahla et al., (1987) Plant Molec. Biol. 8:291-298). Alternatively, transformation efficiency may be enhanced by wounding the target tissue to be modified or transformed. Wounding of plant tissue may be achieved, for example, by punching, maceration, bombardment with microprojectiles, etc. (See e.g., Bidney et al., (1992) Plant Molec. Biol. 18:301-313).

[0194] In addition, a method described by Broothaerts, et. al. (Nature 433: 629-633, 2005) expands the bacterial genera that can be used to transfer genes into plants. This work involved the transfer of a disarmed Ti plasmid without T-DNA and another vector with T-DNA containing the marker enzyme beta-glucuronidase, into three different bacteria. Gene transfer was successful and this method significantly expands the tools available for gene delivery into plants.

[0195] Another widely used technique to genetically transform plants involves the use of microprojectile bombardment, in this process, a nucleic acid containing the desired genetic elements to be introduced into the plant is deposited on or in small dense particles, e.g., tungsten, platinum, or preferably 1 micron gold particles, which are then delivered at a high velocity into the plant tissue or plant cells using a specialized biolistics device. Many such devices have been designed and constructed; one in particular, the PDS 1000/He sold by BioRad, is the instrument most commonly used for biolistics of plant cells. The advantage

of this method is that no specialized sequences need to be present on the nucleic acid molecule to be delivered into plant cells; delivery of any nucleic acid sequence is possible.

[0196] For the bombardment, cells in suspension are concentrated on filters or solid culture medium. Alternatively, immature embryos, seedling explants, or any plant tissue or target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate.

[0197] Various biolistics protocols have been described that differ in the type of particle or the manner in which DNA is coated onto the particle. Any technique for coating microprojectiles that allows for delivery of transforming DNA to the target cells may be used. For example, particles may be prepared by functionalizing the surface of a gold oxide particle by providing free amine groups. DNA, having a strong negative charge, will then bind to the functionalized particles. Parameters such as the concentration of DNA used to coat microprojectiles may influence the recovery of transformants containing a single copy of the transgene. For example, a lower concentration of DNA may not necessarily change, the efficiency of the transformation but may instead increase the proportion of single copy insertion events. In this regard, ranges of approximately 1 ng to approximately 10 g (10,000 ng), approximately 5 ng to 8 g or approximately 20 ng, 50 ng, 100 ng, 200 ng, 500 ng, 1 g, 2 g, 5 g, or 7 g of transforming DNA may be used per each 1.0-2.0 mg of starting 1.0 micron gold particles.

[0198] Other physical and biological parameters may be varied, such as manipulation of the DNA/microprojectile precipitate, factors that affect the flight and velocity of the projectiles, manipulation of the cells before and immediately after bombardment (including osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells), the orientation of an immature embryo or other target tissue relative to the particle trajectory, and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. One may particularly wish to adjust physical parameters such as DNA concentration, gap distance, flight distance, tissue distance, and helium pressure. The particles delivered via biolistics can be "dry" or "wet." In the "dry" method, the coated particles such as gold are applied onto a macrocarrier (such as a metal plate, or a carrier sheet made of a fragile material such as mylar) and dried. The gas discharge then accelerates the macrocarrier into a stopping screen, which halts the macrocarrier but allows the particles to pass through; the particles then continue their trajectory until they impact the tissue being bombarded. For the "wet" method, the droplet containing the coated particles is applied to the bottom part of a filter holder, which is attached to a base which is itself attached to a rupture disk holder used to hold the rupture disk to the helium egress tube for bombardment. The gas discharge directly displaces the DNA/gold droplet from the filter holder and accelerates the particles and their DNA cargo into the tissue being bombarded. The wet biolistics method has been described in detail elsewhere but has not previously been applied in the context of plants (Mialhe et al., MoI Mar Biol Biotechnol. 4(4):275-831995). The concentrations of the various components for coating particles and the physical parameters for delivery can be optimized using procedures known in the art.

[0199] More recently, nanomaterial mediated delivery can be used to transform and/or deliver biomolecules (e.g., nucleic acids and/or proteins) or particles comprising said biomolecules (e.g., an LNP), to a plant cell. In some embodiments, the one or more Cas proteins and the transposase of the RNA-guided DNA binding complex can be pre-complexed with the crRNA prior to the contacting (e.g., as a ribonucleoprotein particle or RNP). Any component of the systems or nucleic acid compositions described herein, or, e.g., an RNP can be formulated into a nanoparticle for delivery. For instance, the nanomaterial-mediated delivery can comprise: clay nanosheets, carbon nanotubes, carbon nanodots, self-assembled protein nanoparticles, peptides, DNA nanostructures, quantum dots, or any combination thereof. Methods for nanomaterial-mediated transformation of plant cells can also be found in U.S. Pat. No. 11,661,606, the content of which is hereby incorporated by reference in its entirety.

[0200] A variety of plant cells/tissues are suitable for transformation, including immature embryos, scutellar tissue, suspension cell cultures, immature inflorescence, shoot meristem, epithelial peels, nodal explants, callus tissue, hypocotyl tissue, cotyledons, roots, and leaves, meristem cells, and gametic cells such as microspores, pollen, sperm and egg cells. It is contemplated that any cell from which a fertile plant may be regenerated is useful as a recipient cell. Callus may be initiated from tissue sources including, but not limited to, immature embryos, seedling apical meristems, microspore-derived embryos, roots, hypocotyls, cotyledons and the like. Those cells which are capable of proliferating as callus also are recipient cells for genetic transformation.

[0201] Any suitable plant culture medium can be used. Examples of suitable media would include but are not limited to MS-based media (Murashige and Skoog, Physiol. Plant, 15:473-497, 1962) or N6-based media (Chu et al., Scientia *Sinica* 18:659, 1975) supplemented with additional plant growth regulators including but not limited to auxins such as picloram (4-amino-3,5,6-trichloropicolinic acid), 2,4-D (2,4-dichlorophenoxyacetic acid), naphalene-acetic acid (NAA) and dicamba (3,6-dichloroanisic acid), cytokinins such as BAP (6-benzylaminopurine) and kinetin, and gibberellins. Other media additives can include but are not limited to amino acids, macroelements, iron, microelements, vitamins and organics, carbohydrates, undefined media components such as casein hydrolysates, an appropriate gelling agent such as a form of agar, a low melting point agarose or Gelrite if desired. Those of skill in the art are familiar with the variety of tissue culture media, which when supplemented appropriately, support plant tissue growth and development and are suitable for plant transformation and regeneration. These tissue culture media can either be purchased as a commercial preparation, or custom prepared and modified. Examples of such media would include but are not limited to Murashige and Skoog (Mursahige and Skoog, Physiol. Plant, 15:473-497, 1962), N6 (Chu et al., Scientia Sinica 18:659, 1975), Linsmaier and Skoog (Linsmaier and Skoog, Physio. Plant., 18:100, 1965), Uchimiya and Murashige (Uchimiya and Murashige, Plant Physiol. 15:473, 1962), Gamborg's B5 media (Gamborg et al., Exp. Cell Res., 50:151, 1968), D medium (Duncan et al., Planta, 165:322-332, 1985), Mc-Cown's Woody plant media (McCown and Lloyd,—HortScience 6:453, 1981), Nitsch and Nitsch (Nitsch and Nitsch, Science 163:85-87, 1969), and Schenk and Hildebrandt (Schenk and Hildebrandt, Can. J. Bot. 50:199-204, 1972) or derivations of these media

supplemented accordingly. Those of skill in the art are aware that media and media supplements such as nutrients and growth regulators for use in transformation and regeneration and other culture conditions such as light intensity during incubation, pH, and incubation temperatures can be varied.

[0202] Those of skill in the art are aware of the numerous modifications in selective regimes, media, and growth conditions that can be varied depending on the plant system and the selective agent. Typical selective agents include but are not limited to antibiotics such as geneticin (G418), kanamycin, paromomycin or other chemicals such as glyphosate or other herbicides. Consequently, such media and culture conditions disclosed in the present disclosure can be modified or substituted with nutritionally equivalent components, or similar processes for selection and recovery of transgenic events, and fall within the scope of the present disclosure.

Methods

[0203] Disclosed herein include methods for integration of a nucleic acid sequence into double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell. In some embodiments, the method comprises: contacting the plant cell with a system or the nucleic acid composition of the disclosure, wherein the cargo sequence is integrated at an integration site in the genome of the plant cell or at a target site of the target plasmid upon expression of the RNA-guided DNA binding complex and the transposition complex in the plant cell.

[0204] In some embodiments, the integration site is about 48 to 52 base pairs (e.g., 48, 49, 50, 51 or 52 base pairs) downstream of the double stranded target sequence. In some embodiments, the one or more Cas proteins and the transposase of the RNA-guided DNA binding complex can be pre-complexed with the crRNA prior to the contacting.

[0205] The plant cell can be comprised within a plant. The plant cell can be comprised within a flower, a leaf, a stem, a root, terminal bud, a seed, or any other tissue of the plant. The plant cell can be a monocot plant cell or a eudicot plant cell.

[0206] In some embodiments, the integration of the cargo sequence confers i) a change in one or more of the following traits to the plant: grain number, grain size, grain weight, panicle size, tiller number, fragrance, nutritional value, shelf life, lycopene content, starch content and/or ii) lower gluten content, reduced levels of a toxin, reduced levels of steroidal glycoalkaloids, a substitution of mitosis for meiosis, asexual propagation, improved haploid breeding, and/or shortened growth time. In some embodiments, the integration of the cargo sequence confers one or more of the following traits to the plant cell and/or the plant: herbicide tolerance, drought tolerance, male sterility, insect resistance, abiotic stress tolerance, modified fatty acid metabolism, modified carbohydrate metabolism, modified seed yield, modified oil percent, modified protein percent, resistance to bacterial disease, resistance to fungal disease, and resistance to viral disease.

[0207] The system or the nucleic acid composition can be introduced into the plant cell by a technique comprising: pollen tube pathway, polyethylene glycol (PEG)-mediated gene transfer, electroporation, microinjection, microparticle bombardment, nanomaterial-mediated delivery, *Agrobacterium tumefaciens*-mediated transformation, or any combination thereof. The nanomaterial-mediated delivery can comprise: clay nanosheets, carbon nanotubes, carbon nanodots, self-assembled protein nanoparticles, peptides, DNA nanostructures, quantum dots, or any combination thereof. The one or more vectors can be introduced into the plant cell via *Agrobacterium tumefaciens*-mediated transformation of the plant cell.

[0208] Exemplary monocotyledonous plants (e.g., monocots) include, without limitation, wheat, maize, rice, orchids, onion, aloe, true lilies, grasses (e.g., *Setaria*), woody shrubs and trees (e.g., palms and bamboo), and food plants such as pineapple and sugar cane. Exemplary dicotyledonous plants (e.g., eudicots or dicots) include, without limitation, tomato, cassava, soybean, tobacco, potato, *Arabidopsis, N. benthamiana*, rose, pansy, sunflower, grape, strawberry, squash, bean, pea, and peanut.

[0209] In some embodiments, the methods described herein can include screening the plant, plant tissue, or plant cell to determine if an insertion has occurred at or near the sequence targeted by the crRNA. Any method known in the art can be used to detect insertions, such as PCR (endpoint PCR or quantitative PCR) and next generation sequences methods.

[0210] In addition, in some embodiments in which a plant part or plant cell is used, the methods provided herein can include regenerating a plant from the plant part or plant cell. The methods also can include breeding the plant (e.g., the plant into which the nucleic acids were introduced, or the plant obtained after regeneration of the plant part or plant cell used as a starting material) to obtain a genetically desired plant lineage. Methods for regenerating and breeding plants are well established in the art.

[0211] Disclosed herein include methods for screening for safe harbor loci in plants. In some embodiments, the method comprises: (a) generating a genome-wide crRNA library; (b) contacting a plant cell comprised within a plant with a system or the nucleic acid composition of disclosure, wherein: the system comprises pooled single or combinatorial crRNAs generated in step (a); or the one or more first helper polynucleotides comprise pooled single or combinatorial crRNAs generated in step a), wherein the cargo sequence is integrated into one or more double-stranded targets sites in the genome of the plant cell upon expression of the RNA-guided DNA binding complex and the transposition complex in the plant cell; (c) identifying integrants by expression of a gene product encoded by the cargo sequence; (d) subjecting the integrants to next-generation sequencing; and (e) performing bioinformatics analysis, a high-throughput phenotypic assay, or both to identify a safe harbor locus.

[0212] The plant can be a monocot plant or a eudicot plant. In some embodiments, integration of the cargo sequence at the identified safe harbor locus does not affect the growth, lifespan, health, gene expression profile, or any combination thereof, of the plant. The system or the nucleic acid composition can be introduced into the plant cell by a technique comprising: pollen tube pathway, polyethylene glycol (PEG)-mediated gene transfer, electroporation, microinjection, microparticle bombardment, nanomaterial-mediated delivery, *Agrobacterium tumefaciens*-mediated transformation, or any combination thereof. The nanomaterial-mediated delivery can comprise: clay nanosheets, carbon nanotubes, carbon nanodots, self-assembled protein nanoparticles, peptides, DNA nanostructures, quantum dots, or any combination thereof. In some embodiments, the one

or more vectors are introduced into the plant cell via *Agrobacterium tumefaciens*-mediated transformation of the plant cell.

[0213] In some embodiments, the T-DNA region of the Ti vector comprises a bi-directional selection marker comprising a positive selection marker and a negative selection marker, wherein the identifying of step c) comprises: (i) generating a first filial generation (F1) plant comprising the T-DNA of the Ti vector by positive selection; and (ii) generating a second filial generation (F2) plant that does not comprise the T-DNA of the Ti vector from the first filial generation plant comprising the T-DNA of the Ti vector, by negative selection.

Kits

[0214] Disclosed herein include kits comprising a system or nucleic acid composition described herein, and a set of instructions for use. For example, the kit can comprise a nucleic acid composition comprising one or more polynucleotides of the disclosure (e.g., the first and/or second helper polynucleotide, the one or more helper accessory polynucleotides, and/or the donor polynucleotide).

[0215] In addition to above-mentioned components, a kit can further include instructions for using the components of the kit to practice the methods. The instructions for practicing the methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, flash drive, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

EXAMPLES

[0216] Some aspects of the embodiments discussed above are disclosed in further detail in the following examples, which are not in any way intended to limit the scope of the present disclosure.

Example 1

Targeted DNA Integration in Plants by CRISPR-Transposases

[0217] Provided in this example are methods and compositions of the disclosure related to targeted genome integration in plants using CRISPR-Transposases.

Establish CAST-Mediated Targeted DNA Integration in Plants

Construct Design of CAST Component

[0218] Vch and Pse CAST system was separated into three parts for modular reconstitution, as described below (FIG. **1**).

[0219] Helper plasmid (pHelper) contains all essential CAST proteins and guide RNA in monocistronic cassettes. The type and position of the nuclear localization signal (NLS) was designed based on the CAST expression system in the mammalian study. pCmYLCV911 (pCmY) promoter and AtHSP18.2t (tHSP) terminator were used for all protein expression, which is experimentally the strongest promoter-terminator pair in plants. U6 promoter and terminator were used for crRNA expression.

[0220] Donor plasmid (pDonor) contains the donor DNA, which is the cargo sequence flanked by transposon left and right ends (LE and RE). The donor DNA is incorporated in a geminivirus-derived replication vector, which after entering the plant cell will initiate autonomous replication. This approach is designed to increase the donor DNA copy number in the cells to increase the insertion efficiency.

[0221] Target plasmid (pTarget) contains the intended target site for integration, which is only used in the episomal integration experiments. In the chromosomal experiments, target sites are selected from native genomic sequences. In experiments discussed below, 35S promoter is selected as the target site as it is present in a widely used GFP-expressing transgenic *N. benthamiana* line (16c) to drive the expression of GFP. Three target sequences on this 35S promoter were chosen with a CC PAM. Adhering to the generic construct design described here, all experiments were done in parallel using the Vch and Pse CAST with crRNA targeting three different locations on the 35S promoter labeled as T1, T2, and T3.

Helper Protein Expression

[0222] To validate CAST helper protein expression, the HiBiT lytic assay was employed which requires tagging the protein of interest with the 1.3 kDa HiBiT peptide tag. The HiBiT peptide complexes with the larger LgBiT subunit to produce a complete luciferase enzyme. The resulting complex generates a bright luciferase signal in the presence of the furimazine substrate. CAST helper proteins from the following bacterial species, *Vibrio cholerae* (VchCAST), *Pseudoalteromonas* (PseCAST) and *Scytonema hofmanni* (ShoCAST), were tagged with the HiBiT tag (FIG. **2**A). HiBiT tagged pHelper proteins were transiently expressed using the *Agrobacterium* mediated T-DNA insertion. Each protein was expressed individually in plants. Initial attempts to detect protein expression failed because we were unable to lyse the plant nuclei and retrieve nuclear localized proteins. A modified protocol was devised to extract nuclei and detect the nuclear proteins separately which led to us detecting the majority of the CAST proteins (FIG. **2**B-FIG. **2**D).

[0223] A protocol for HiBiT lytic CAST protein detection includes the following steps: (1) Generate HiBiT tagged recombinant CAST proteins as shown in FIG. **2**A; (2) Transform construct into *Agrobacterium*; (3) Inoculate 5 mL cultures of transformed *Agrobacterium* overnight at 30° C. while shaking; (4) Centrifuge cultures at 4000 g, 4° C. for 10 minutes; (5) Resuspend *Agrobacterium* in induction media (10 mM MES pH 5.6, 10 mM $MgCl_2$, 100 μM acetosyringone) and incubate at room temperature while shaking for 4 hours; (6) Infiltrate *Agrobacterium* into 4 week old *Nicotiana benthamiana* plants; (7) 3 days post infiltration, grind 3 medium size infiltrated leaves to fine powder; (8) Extract the cytoplasmic protein fraction using lysis buffer (20 mM Tris-HCl, pH 7.4, 25% Glycerol, 20 mM KCl, 2 mM EDTA; 2.5 mM $MgCl_2$ 2.5 ml, 250 mM Sucrose, 0.1% PMSF); (9)

Centrifuge at 1500 g to pellet intact nuclei; (10) Wash nuclei up to 5 times (Wash buffer: 20 mM Tris-HCl, pH 7.4, 25% Glycerol, 2.5 mM $MgCl_2$, 0.2% Triton X-100); (11) Lyse nuclei with a high salt buffer (20 mM HEPES-KOH pH 7.9, 2.5 mM $MgCl_2$, 100 mM NaCl, 20% (v/v) Glycerol, 0.2 mM EDTA, 0.5 mM DTT, cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (1 tablet per 10 mL volume)); (12) Combine 30 μL of Furimazine and LgBiT with 30 μL of protein sample and detect luminescence signal using an integration time of 1000 ms.

Cast-Mediated Episomal DNA Integration in Protoplast Cells

[0224] Integration in episomal DNA is more efficient than chromosomal DNA, mainly attributed to the lack of complex chromatin structure. PEG-mediated plasmid transfection of protoplasts is the only practical approach for episomal experiments. Cargo DNA was designed to be a promoter-less YFP, ensuring exclusive YFP transcription and translation from the target plasmid 35S promoter only if the intended integration occurs at the target site. In consideration of plasmid size constraints in protoplast transfection, pHelper was separated into transposases and Cas effectors to either combine with the pDonor or pTarget, respectively. Fluorescent marker (dsRed or mTurq) was added to verify the transfection efficiency (FIG. **3**A).

[0225] At 48 hours post-transfection, DsRed and mTurq signal confirmed co-transfection into *A. thaliana* protoplasts. Of primary interest is the YFP signal, and a few instances were observed among all crRNA sequences tested (FIG. **4**). It should be noted that cells expressing YFP represent only a subset of cells with successful integration, as it requires the absence of premature stop codons in the sequence between the 35S promoter and the YFP start codon, which cannot be guaranteed due to the presence of transposon ends and the variable integration distance.

[0226] Molecular level analysis was performed next using nested PCR, where two consecutive PCR reactions are performed using two sets of primers to enhance the specificity and sensitivity of DNA amplification. To detect the integration junctions of the favored orientation Target-RE-Cargo-LE, OUT PCR reaction was done with 1F primer binding to p35S and 1R primer binding to YFP; followed by IN PCR reaction that was done with 2F primer binding to p35S and 2R primer binding to RE to specifically amplify the right end junction (FIG. **5**A). Simultaneously, OUT PCR reaction was done with 3F primer binding to YFP and 3R primer binding to plasmid backbone; followed by IN PCR reaction with 4F primer binding to RE and 4R primer binding to plasmid backbone to specifically amplify the left end junction (FIG. **5**A). The first OUT reaction typically yields no bands on the gel due to the low template amount. After the second round of amplification, the correct bands were detected for two out of three target choices in both Vch and Pse systems (FIG. **5**B). Similar results were observed both at the RE and LE junction (FIG. **5**C). The less-favored orientation is anticipated to have lower efficiency.

[0227] Sanger sequencing of the purified PCR bands confirmed correct integration junction products for Vch T2 & T3 and Pse T2 & T3. However, there were mixed peaks in the sequencing results originating from the variable insertion distances, given that the integration events happened around 50 bps downstream of the target site. TA-cloning was performed and single colonies selected from *E. coli* transformation to selectively sequence individual PCR products with set integration distance. Four colonies from Vch T2 RE exhibited different integration distances of 50, 51, 52 bp. Two colonies from Vch T2 LE exhibited an integration distance of 50 bp. These sequences also highlighted the creation of the hallmark 5-bp target site duplication (TSD) (FIG. **6**A). Integration distance of ~50 bp and TSDs were also observed for Vch T3, Pse T2, and Pse T3. (FIG. **6**B-FIG. **6**D). This provides the first demonstration of CAST functionality in plant cells.

[0228] Described below is an exemplary protocol for protoplast integration assays: (1) *Arabidopsis* protoplasts were collected and resuspended in W5 solution according to the tape method known in the art; (2) Change the buffer to MMG solution. Target for a protoplast concentration around 5×10^5 cell/mL; (3) Use round bottom culture tubes for protoplast transfection. Add total 50 μg of plasmid into the microcentrifuge tube. Then add 100 μL of protoplasts. Mix by tapping. Add equal volume of PEG solution. Mix by tapping again; (4) Incubate the mixture at room temperature for 15 min; (5) Add 2 mL W5 solution to the culture tubes and centrifuge at 200 g for 2 minutes with acceleration and deceleration set to minimum. Remove as much as supernatant without disturbing the pellet and repeat one time; (6) Cover the well plate with 0.1% BSA for several minutes and remove fully to prevent protoplast from sticking to the bottom of the plate; (7) Add 500 μl WI solution to the protoplasts and transfer to a 12 well-plate. Gently swirl the well-plate after transferring all the protoplasts samples; (8) Incubate at room temperature at dark for around 6 hours to overnight, the check the expression with fluorescence microscopy for reporter plasmids; (9) After a 48-hour incubation period, samples are subjected either to confocal imaging or DNA isolation.

[0229] Described below is an exemplary protocol for Crude DNA isolation and nested PCR to detect integration junctions: (1) Transfer appropriate volume of protoplasts into a 1.5 ml Eppendorf tube. Centrifuge at 200 g for 2 minutes and remove the supernatant without disturbing the pellet; (2) Add 1/10 volume TE buffer relative to the original protoplast volume. Disrupt the cells by pipetting vigorously for at least 10 times and vortexing vigorously for at least 30 seconds; (3) Before any downstream analysis, centrifuge at 12000 g for 1 minute and use the supernatant as the crude DNA solution; (4) Nested PCR is carried out by two consecutive PCR reactions using NEB Phusion® High-Fidelity PCR Master Mix with HF Buffer. The cycling number is 20 for OUT PCR and 30 for IN PCR. Annealing temperature and extension time is selected based on specific reaction, and the template for the IN nested PCR is prepared by a 50-fold dilution of the OUT nested PCR product; (5) Visualize the IN nested PCR product via electrophoresis using agarose gel.

[0230] Described below is an exemplary protocol for TA cloning and Sanger sequencing to confirm integration junctions: (1) Adding an A-overhang to the 3'-end of the purified PCR amplicon using Taq DNA polymerase and dATP; (2) Follow the instructions from pGEM®-T Easy Vector Systems—Promega kit for ligation reaction; (3) Transform the reaction mixture into competent *E. coli* cells using standard heat shock method and blue-white screening. Incubate the plates overnight at 37° C.; (4) Pick individual white bacterial colonies and inoculate them into LB media with the suitable antibiotic. Plasmids are isolated using QIAprep Spin Mini-

prep Kit; (5) Submit the purified plasmid DNA with appropriate primers for Sanger sequencing (Laragen®).

Cast-Mediated DNA Integration into the Plant Genome

[0231] Described below is RNA-guided DNA integration into the plant genome. The ability of genomic insertion is crucial for many fundamental and applied studies.

[0232] Two accessory proteins, ClpX and ClpP, were incorporated to helper plasmid because of the necessity for transpososome dissociation to initiate DNA repair. Cargo DNA was a full YFP expressed with a 35S promoter, thus, in this case, there was no positive phenotypic signal detection, as YFP can be expressed from the *Agrobacterium* inserted T-DNA. This can also complicate the primer design due to the presence of duplicated 35S elements (FIG. **7**A). The CAST machinery is designed to integrate the full YFP cargo into the endogenous 35S promoter within the GFP cassette of the transgenic *N. benthamiana*. Disrupting both alleles of GFP is expected to lead to the loss of GFP expression. All components were cloned into a single plasmid, transformed into *Agrobacterium*, and infiltrated into *N. benthamiana* leaves. Plants were grown for 7 days, and DNA samples were harvested from leaves for subsequent phenotypic and molecular analysis (FIG. **7**B).

[0233] Molecular analysis was performed using the nested PCR. To minimize off-target amplification from the 35S promoter, the 5F and 6F primers specifically target the unique 5' sequence upstream of the p35S target site, although at the expense of the long amplicon length. OUT PCR reaction was done with 5F primer and 5R primer binding to YFP; IN PCR reaction was done with 6F primer and 6R primer binding to RE to specifically amplify the right end junction (FIG. **8**A). The first OUT reaction typically yields no bands due to the low template amount; after the second round of amplification, the correct bands were detected for one out of three target sites in the Pse system (FIG. **8**B).

[0234] Then, the TA-cloning method was followed and individual PCR products from the Pse T2 RE PCR product were selectively sequenced. Two different integration distances of 49 and 50 bp were identified (FIG. **8**C).

[0235] Described below is an exemplary protocol for Leaf integration assays: (1) Inoculate 5 mL cultures of *Agrobacterium* harboring the construct, and incubate at 30° C. shaking overnight; (2) Centrifuge cultures down at 4000 g for 10 minutes at room temperature, and discard supernatant and resuspend pellet in MMA media (10 mM MES pH 5.6, 10 mM $MgCl_2$, 100 uM acetosyringone); (3) Dilute the resuspension in more MMA to reach an OD600 of 0.35-0.40; (4) Let the inoculated MMA media sit at room temperature for 2-4 hours; (5) Infiltrate leaf with syringe and mark the infiltrated area. *N. benthamiana* plants are kept in growth chamber for 7 days under a 16-h-light/8-h-dark cycle; (6) Plant DNA is harvested by following the DNeasy Plant Mini Kit instruction, and after elution with water, DNA concentration is adjusted to approximately 50 ng/μl for following molecular analysis.

TABLE 1

NESTED PCR PRIMERS - PROTOPLAST EXPERIMENT

Short Identifier	Sequence	Long Identifier	SEQ ID NO:
1F	gccgacagtggtcccaaagatggac	F p35S-2	47
1R	CTTGCCGGTGGTGCAGAG	R YPET-4	48
2F	ccacccacgaggagcatcgtg	F p35S-3	49
2R (Vch)	CTGAAGTTTAGACCATGAAGAGGC	R2 Vch RE amplicon	50
2R (Pse)	AGGGGCTTTATGTCAGAGTATGAC	R2 Pse RE amplicon	51
3F	GACTGCTGCTGGTATTACCGAGGG	F YPET-2	52
3R	GCGGATCTAACTAATAAAGTGTCTACcattagtatac	R for LE-dummy-1	53
4F (Vch)	TCCGTGACTGATGCAAGGTTATCGG	F2 Vch LE amplicon	54
4F (Pse)	AAACCAGTCTTATATCCGGGCTGG	F2 Pse LE amplicon	55
4R	GCGGATCTAACTAATAAAGTGTCTACcattagtatac	R for LE-dummy-1	56

TABLE 2

NESTED PCR PRIMERS - LEAF EXPERIMENT

Short Identifier	Sequence	Long Identifier	SEQ ID NO:
5F	tggcgaacagttcatacagagtctc	F pTarget p35S-1	57
5R	TCCAGCTCGACCAGGATGG	R YPET-1	58
6F	tcttcgtcaacatggtggagcacg	F pTarget p35S-2	59
6R (Vch)	CTGAAGTTTAGACCATGAAGAGGC	R2 Vch RE amplicon	60
6R (Pse)	AGGGGCTTTATGTCAGAGTATGAC	R2 Pse RE amplicon	61

TABLE 3

PLASMID SUMMARY - PROTOPLAST EXPERIMENT

Plasmid ID	Plasmid components
pGYW5CH0111	L5 pUAP4 pCk3 Vch pTarget 35S (A) + spc (B) + Helper Cas (C) + mTurq (E)
pGYW5CH0112	L5 pUAP4 pCk3 Pse pTarget 35S (A) + spc (B) + Helper Cas (C) + mTurq(E)
pGYW5CH0115 T001	L5 pUAP4 pCk2 Vch Donor YFP-spc-T001 (A) + Helper Tnp (B) + spc (C) + DsRed (E)
pGYW5CH0115 T002	L5 pUAP4 pCk2 Vch Donor YFP-spc-T002 (A) + Helper Tnp (B) + spc (C) + DsRed (E)
pGYW5CH0115 T003	L5 pUAP4 pCk2 Vch Donor YFP-spc-T003 (A) + Helper Tnp (B) + spc (C) + DsRed (E)
pGYW5CH0116 T001	L5 pUAP4 pCk2 Pse Donor YFP-spc-T001 (A) + Helper Tnp (B) + spc (C) + DsRed (E)

TABLE 3-continued

PLASMID SUMMARY - PROTOPLAST EXPERIMENT	
Plasmid ID	Plasmid components
pGYW5CH0116 T002	L5 pUAP4 pCk2 Pse Donor YFP-spc-T002 (A) + Helper Tnp (B) + spc (C) + DsRed (E)
pGYW5CH0116 T003	L5 pUAP4 pCk2 Pse Donor YFP-spc-T003 (A) + Helper Tnp (B) + spc (C) + DsRed (E)

TABLE 4

PLASMID SUMMARY - LEAF EXPERIMENT		
Plasmid ID	Plasmid detailed name	SEQ ID NO: (T-DNA sequence)
pGYW5KN0107 T001	L5 pCk2 Vch Donor pYFP-ClpXP-T001 (A) + Helper Tnp (B) + Helper Cas (C) + DsRed (E)	38
pGYW5KN0107 T002	L5 pCk2 Vch Donor pYFP-ClpXP-T002 (A) + Helper Tnp (B) + Helper Cas (C) + DsRed(E)	39
pGYW5KN0107 T003	L5 pCk2 Vch Donor pYFP-ClpXP-T003 (A) + Helper Tnp (B) + Helper Cas (C) + DsRed(E)	40
pGYW5KN0108 T001	L5 pCk2 Pse Donor pYFP-ClpXP-T001 (A) + Helper Tnp (B) + Helper Cas (C) + DsRed(E)	41
pGYW5KN0108 T002	L5 pCk2 Pse Donor pYFP-ClpXP-T002 (A) + Helper Tnp (B) + Helper Cas (C) + DsRed(E)	42
pGYW5KN0108 T003	L5 pCk2 Pse Donor pYFP-ClpXP-T003 (A) + Helper Tnp (B) + Helper Cas (C) + DsRed(E)	43

TABLE 5A

EXPRESSION DATA - VCH CAST (NUCLEAR FRACTION)*			
	Rep1	Rep2	Rep3
vchTnsA-B	958.182929	1793.462299	2113.894623
vchTnsC	36484.94021	26769.81795	48279.7851
vchTniQ	9606.525542	10173.125	16238.03773
vchCas6	22933.97474	30378.44578	14815.99063
vchCas7	17120.45531	8802.426611	15315.90506
vchCas8	32197.54667	49562.4393	64975.69638
Negative control	416.912023	783.9432465	599.6773796

*relative luminescence units normalized to protein concentration

TABLE 5B

EXPRESSION DATA - VCH CAST (CYTOPLASMIC FRACTION)*			
	Rep1	Rep2	Rep3
vchTnsA-B	670.2429549	156.3945564	390.0156227
vchTnsC	2275.723375	319.8470487	2200.177778
vchTniQ	778.8978805	1166.675387	1252.972136
vchCas6	1588.564004	1583.864819	2028.979321
vchCas7	1249.630374	2128.845234	4952.360336
vchCas8	625.5878633	982.0480473	1759.015744
Negative control	573.7433598	518.012268	492.6125331

*relative luminescence units normalized to protein concentration

TABLE 6A

EXPRESSION DATA - PSE CAST (NUCLEAR FRACTION)*			
	Rep1	Rep2	Rep3
vchTnsA-B	693.509481	865.4245229	579.5452139
vchTnsC	11840.68353	15471.35966	15593.2671
vchTniQ	15334.67665	17890.94761	20774.59751
vchCas6	6986.819874	13141.22572	22538.12285
vchCas7	5130.733098	9718.885697	3828.325613

TABLE 6A-continued

EXPRESSION DATA - PSE CAST (NUCLEAR FRACTION)*			
	Rep1	Rep2	Rep3
vchCas8	1427.339806	6009.792569	5744.788468
Negative control	127.1174404	147.6658962	172.6738303

*relative luminescence units normalized to protein concentration

TABLE 6B

EXPRESSION DATA - PSE CAST (CYTOPLASMIC FRACTION)*			
	Rep1	Rep2	Rep3
vchTnsA-B	91.82135127	294.0918765	102.4193156
vchTnsC	515.7902571	140.7168447	266.8300106
vchTniQ	5334.676646	3890.947609	3774.597515
vchCas6	4697.56809	6026.43677	5558.893581
vchCas7	40.2196634	21.6954773	81.07759961
vchCas8	24435.59284	10686.66526	10335.26801
Negative control	210.1881088	359.0901504	329.3213838

*relative luminescence units normalized to protein concentration

TABLE 7A

EXPRESSION DATA - SHO CAST (NUCLEAR FRACTION)*		
	Rep1	Rep2
shoTnsB	7998.311228	5687.729904
shoTnsC	21827.61629	20830.12271

TABLE 7A-continued

EXPRESSION DATA - SHO CAST (NUCLEAR FRACTION)*		
	Rep1	Rep2
shoTniQ	26810.86344	28024.77568
shoCas12k	13997.4974	12136.24733
Negative control	155.8554124	250.2784176

*relative luminescence units normalized to protein concentration

TABLE 7B

EXPRESSION DATA - SHO CAST (CYTOPLASMIC FRACTION)*		
	Rep1	Rep2
shoTnsB	11553.97761	11636.99973
shoTnsC	18673.85708	14342.24978
shoTniQ	50803.55048	49965.17347
shoCas12k	13216.10215	11452.42927
Negative control	355.2587132	321.1234335

*relative luminescence units normalized to protein concentration

Example 2

Selection and Screening

[0236] Described in this Example are exemplary use cases for the CAST methods of the disclosure.

Engineering a Stable Transgenic Line with CAST and Proper Selection

[0237] Described below is engineering a stable transgenic line with the developed CAST tool and selection. The cargo DNA is a complete FASTred cassette, expressing red fluorescence specifically in the seed coat, which simultaneously serves as the transformation marker. To select for CAST-mediated targeted integration events, a transgenic line harboring a selectable marker as the target site is employed. The options include 1) phytoene desaturase (PDS) gene, where knockout *Arabidopsis* exhibits an albino or mosaic phenotype corresponding to homozygous or heterozygous mutations, due to impaired chlorophyll and carotenoid synthesis, 2) codA gene, a cytosine deaminase that converts non-toxic 5-Fluorocytosine (5-FC) to toxic 5-FU causing cell death (FIG. **9**A).

[0238] All components are cloned into a single plasmid and transformed into *Agrobacterium*. Flowering *Arabidopsis* plants are dipped in *Agrobacterium* solution for transformation of genetic materials. Subsequently, seeds are collected and germinated on selection plates to observe the albino phenotype from the loss of PDS; or on 5-FC selection plates to enrich for the loss of codA. Germinated progeny are subjected to further molecular analysis (FIG. **9**B).

Apply CAST to Screen for Genomic Safe Harbor Loci

[0239] Genomic safe harbors (GSHs) are chromosomal regions that can accommodate transgenes without adverse effects on the host organism and possess high gene expression levels. In the pursuit of advancing precise breeding and molecular farming, reliable sites for gene integration need to be identified. However, to date, only two studies reported candidate GSH in rice via mutagenesis. Identifying ideal loci for transgene expression will be beneficial for conducting more rational and reproducible studies.

[0240] After establishing CAST as a potent targeted DNA insertion tool in plants, genome-wide screens are performed to identify much-needed genomic safe harbor sites. Initially, a crRNA library via oligo synthesis is designed. Then pooled crRNA construct cloning, and *Agrobacterium* transformation is performed to introduce crRNA and donor construct into a transgenic line expressing the transposition enzymes and CRISPR-Cas effectors. Importantly, T-DNA should be eliminated so that the phenotype is the pure result from CAST-mediated target integration. Two major strategies have been developed to remove or prevent the integration of T-DNA: 1) genetic segregation to eliminate transgenic sequences; 2) transient expression from DNA vectors or pre-assembled protein-guide RNA ribonucleoproteins. In the case of adapting the first approach, transformation experiments are carried out in *Arabidopsis* via floral dip. One strategy involves incorporating the T-DNA region with a bi-functional marker consisting of the negative selection marker CodA fused to the positive selection marker NPTII. In the first generation, transgenic mutant plants are selected by kanamycin, subsequently, transgene-free mutants will be selected by 5-FC in the next generation. In the case of adapting the second approach, experiments are carried out in *N. benthamiana* leaves using nanotechnology-enabled DNA delivery for transient expression. Afterwards, the edited plants are subjected to barcoded NGS and quantitative analysis of cargo gene expression. In the final phase, a bioinformatic analysis is undertaken to systematically identify safe harbor loci.

Example 3

Modifications and Assessments

[0241] Described in this Examples are various modifications and assessments of the systems, nucleic acid compositions, and methods described above in Example 1.

Verification of CAST Protein Expression in Plants

[0242] In the designs described above to express CAST proteins, they were expressed as monocistronic cassettes driven by the pCmYLCV911 (pCmY) promoter and the AtHSP18.2t (tHSP) terminator. The transposase proteins were combined into a fusion protein (TnsA-B), and a bipartite nuclear localization sequence was added (FIG. **11**A). To detect protein expression, the Nano-Glo HiBiT lytic assay was used, which enables the detection of recombinant proteins tagged with a HiBiT tag as a luminescence signal. The HiBiT tag forms a complex with LgBiT (the larger part of the luciferase enzyme) to reconstitute a complete luciferase enzyme, producing a luminescent signal in the presence of the substrate furimazine. To further validate protein localization in the nucleus, a yellow fluorescent protein reporter (Ypet) at the N-terminus, downstream of the HiBiT tag was included (FIG. **11**A). Constructs expressing individual CAST proteins were delivered as single plasmids for each protein using *Agrobacterium tumefaciens* into *Nicotiana benthamiana* plant leaves (FIG. **11**B).

[0243] A subset of proteins from the VchCAST system, including TnsC, TniQ, Cas6, Cas7, and Cas8, was selected for the initial expression detection experiments, with dCas9 serving as a positive control and non-infiltrated leaves

serving as a negative control. Most CAST proteins, while detectable using confocal microscopy (FIG. **11**C), were initially undetectable using the HiBiT lytic detection assay on whole-cell lysates (FIG. **11**D). Out of all proteins, Vch-Cas6—the smallest CAST protein with a size of 24.531 kDa, showed the highest expression levels in the HiBiT assay and was the only protein that showed cytoplasmic localization in confocal images. This finding allowed for the conclusion that undetected proteins were retained in the nucleus during the protein extraction step due to the bipartite nuclear localization peptide.

[0244] To address this, a protocol was developed to separate plant nuclei from the cytoplasmic fraction before nuclear lysis and protein detection. The Nano-Glo HiBiT Lytic Detection protocol was applied following the nuclear isolation and lysis (FIG. **12**A-FIG. **12**B). Using this protocol, the successful expression of all VchCAST and PseCAST proteins was detected, as indicated by the luminescent signal (FIG. **12**C-FIG. **12**D). Transposase proteins (TnsA-B fusion proteins) from both VchCAST and PseCAST exhibited significantly lower expression levels compared to the other CAST components (FIG. **12**C-FIG. **12**D). To enhance the heterologous expression of transposase proteins, geminivirus-derived replicons were incorporated into the constructs (FIG. **12**E). These elements initiate autonomous DNA replication within plant cells, thereby increasing DNA copy numbers and raising the transcriptional output of transposase proteins. The inclusion of geminiviral replicons resulted in at least a 10-fold increase in transposase protein expression compared to constructs without geminiviral elements (FIG. **12**F).

Cast-Mediated Episomal DNA Integration in Plant Protoplast Cells: Molecular Analysis Across Time Points

[0245] After the episomal DNA integration experiments described above, the temporal profile of integration, integration efficiency, and integration distance was further characterized, focusing on the PseCAST derived from *Pseudoalteromonas*.

[0246] For the following experiments, two different cargo DNAs were tested: a random 200 bp sequence and a 1200 bp promoterless mCherry coding sequence. Given plasmid size constraints in protoplast transfection, helper plasmid was separated into transposases and Cas effectors to either combine with the pDonor or pTarget, respectively. A fluorescent marker (mTurq or YPET) was added to verify the transfection efficiency. The integration target site was a 35S plant promoter and two crRNAs (T2 and T3) were designed and tested to target the 3' region in the 35S promoter.

[0247] Identifier and details of the PseCAST constructs are: #145, pTarget 35S-bds+Helper Cas+mTurq; #150, Helper Tnp+RE-200 bp-LE-crRNA T3+YPET; #151, Helper Tnp+RE-proless mCherry-LE-crRNA T3+YPET; #152, Helper Tnp+RE-p35S mCherry-LE-crRNA T3+YPET (positive control).

[0248] Samples were analyzed at 12-, 24-, and 48-hours post-transfection. 200 bp random cargo results were not expected to generate any fluorescence signal from targeted insertion, and were hence characterized at the molecular level. For the promoterless mCherry cargo, confocal images showed that at 12 and 24 hours, single transfection control groups did not show mCherry (as expected); and only the experimental dual transfection groups showed the gain of mCherry (FIG. **13**A-FIG. **13**B). However, at 48 hours, both the negative control and experiment groups showed mCherry expression (FIG. **13**C). This unexpected outcome could be explained by plasmid recombination events or long-distance interaction of the promoters present upstream of mCherry coding sequence that take place when long incubation periods are used. Full mCherry cassette delivery demonstrated expression as expected.

[0249] For all of the samples, molecular-level validation of targeted integration events was also conducted, in addition to the fluorescence imaging assays. One of the molecular-level analyses is junction PCR and sequencing, as correct junction sequences can confirm the CAST-mediated targeted insertion events. The samples were analyzed using nested PCR with primers 1F, 1R, 2F, 2R for the RE junction detection and 3F, 3R, 4F, 4R for the LE junction detection (FIG. **14**A). Positive nested PCR results were obtained for both RE and LE junctions, showing a correct increase in band intensity over time. After gel-purification, bulk Sanger sequencing followed by TA-cloning confirmed the presence of correct RE & LE junction sequences. Results showed that CAST-mediated episomal DNA integration can occur as early as 12 hours post-protoplast transfection (FIG. **14**B-FIG. **14**C). Extending the incubation period while maintaining cell viability results in the observation of additional integration events.

[0250] To obtain quantitative results, TaqMan probes and primers were designed to target the RE and LE junctions and the Cas8 reference gene (FIG. **15**A). All probes were labeled with FAM to allow for easy threshold adjustment if needed. A detectable threshold cycle number (Ct) of approximately 30 for both RE and LE junctions was observed, distinct from the negative controls with a high Ct of more than 38 or not detectable (FIG. **15**B, Table 8).

TABLE 8

THRESHOLD CYCLE (CT) VALUES

Template	Cas8-FAM	Pse RE-FAM	Pse LE-FAM
145 pTarget + CAS	11.1	NA	38.1
150 pDonor + TNP	34.6	NA	NA
145 + 150 (12 h)	11.4	Not tested	31.1
145 + 151 (12 h)	10.9	Not tested	30.8
145 + 150 (24 h)	11.8	35.3	31.3
145 + 151 (24 h)	11.1	30.1	29.4
145 + 150 (48 h)	15.3	Not tested	33.1
145 + 151 (48 h)	13.7	Not tested	29.2

[0251] An exemplary Taqman protocol is described below: (1) Design probes and primers using the IDT online tool; (2) Follow the protocol of Luna® Universal Probe qPCR Master Mix, and per 10 μl reaction, add 50~100 ng extracted DNA; (3) Use the cycling program: an initial denaturation at 95° C. for 60 seconds, followed by 40 cycles of denaturation at 95° C. for 15 seconds and extension at 60° C. for 30 seconds with a plate read during the extension step.

[0252] As the CAST system integrates cargo at varied distances downstream from the target site, it was next sought to determine the integration distance profile for the different crRNAs and cargo lengths tested. Primers were designed for pooled NGS sequencing, with the expected amplicon sizes as follows: RE: 198 bp (T2), 221 bp (T3); LE: 223 bp (T2), 196 bp (T3). The first round of PCR amplifies the target region while adding primer tags, and the second round introduces barcodes to the amplicons (FIG. **16**A).

[0253] Results show expected integration distance profiles ranging from 48 to 51 bp away from the target site (FIG. **16**B). Interestingly, variation of integration distance was observed between different crRNAs, but different cargo lengths did not affect this profile (FIG. **16**B).

[0254] An exemplary protocol for amplicon sequencing is described below: (1) PCR is carried out by two consecutive PCR reactions using NEB Phusion® High-Fidelity PCR Master Mix with HF Buffer. The cycling number is 35. Annealing temperature and extension time is selected based on specific reaction; (2) Purify and measure the DNA amplicon product from the first round, and normalize their amount and dilute 10 times, then use it for the second round of PCR; (3) Purify and measure the DNA amplicon product from the second round, and then combine all samples at this point to reach 35 μl at 40 ng/μl; (4) Submit for MGH CCIB CRISPR Sequencing service.

Cast-Mediated DNA Integration into the Plant Chromosome in Leaves

[0255] Described below is a focus on RNA-guided DNA integration into the plant genome, as the ability to precisely engineer genomic loci is essential for achieving heritable genetic edits. Successful genomic integration is critical for both fundamental research and applied studies.

[0256] Cargo DNA was designed as a 200 bp random sequence at this stage. The crRNA was designed to target the cargo into the endogenous 35S promoter within the GFP cassette of the transgenic 16C *N. benthamiana*. Components were designed that are either cloned into a single plasmid, where cloning is more cumbersome requiring the assembly of large constructs exceeding 30 kb with repetitive sequences, or split into two plasmids. These constructs were transformed into *Agrobacterium tumefaciens*, and infiltrated into *N. benthamiana* leaves. The plants were kept in a growth chamber for 7 days, with overnight 30° C. heat shock treatment applied during the last three days. DNA samples were then harvested from leaves for subsequent phenotypic and molecular analysis.

[0257] Identifier and details of the constructs are: #120, T2/T3 LIR-RE-200 bp-LE-SIR+ClpX/ClpP+crRNA T2/3+Tnp+Cas+DsRed (T-DNA, SEQ ID NO: 44); #158, Tnp+Cas+pVmV mTurq (T-DNA, SEQ ID NO: 45); #427 T2/T3 LIR-RE-200 bp-LE-SIR+ClpX+ClpP+crRNA T002/3 (T-DNA, SEQ ID NO: 46).

[0258] Multiple biological replicates (#1 to #4) were analyzed using nested PCR with primers 1F, 1R, 2F, 2R for the RE junction detection and 3F, 3R, 4F, 4R for the LE junction detection (FIG. **17**A). The first OUT reaction typically yields no bands due to the low template amount. After the second round of amplification, the correct bands were detected for LE junctions (FIG. **17**B). Bulk Sanger sequencing confirmed the LE junctions, with integration lengths of 48 to 50 bp, consistent with previous study in humans.

Improving CAST-Mediated DNA Integration into the Plant Genome in Protoplasts

[0259] To maximize integration efficiency, the CAST system was re-engineered where TnsA-B was included in the geminivirus-derived replicon to achieve high expression levels; and two accessory proteins, ClpX and ClpP, were incorporated to helper plasmid because of the necessity for transpososome dissociation to initiate DNA repair. This system was tested in protoplasts.

[0260] Cargo DNA was designed as a 200 bp random sequence at this stage. The crRNA was designed to target the cargo into the endogenous 35S promoter within the GFP cassette of the transgenic 16C *N. benthamiana*. All components were cloned into two separate plasmids, transfected into protoplasts and incubated for 72 hrs in dark and DNA samples were harvested for subsequent molecular analysis (FIG. **18**A).

[0261] Identifier and details of the constructs are: #626, Tnp+Cas+pUBQ4-YPET; #628, T2/T3GV-200 bp-crRNA T2/T3-TnsAB+ClpXP+pNOS-DsRed.

[0262] The samples were analyzed using nested PCR with primers 1F, 1R, 2F, 2R for the RE junction detection and 3F, 3R, 4F, 4R for the LE junction detection. Positive results for the RE junction were consistent with targeting by crRNA T2. The LE junction demonstrated integration distances of 49 or 50 bp as anticipated (FIG. **19**).

[0263] Disclosed herein is CAST-mediated targeted gene insertion in plants for the first time using *Vibrio cholera* and *Pseudoalteromonas* CAST. Successful expression of all CAST proteins is shown in plant leaves. Episomal integration of 200 bp and 1200 bp DNA cargo is achieved in *Arabidopsis thaliana* protoplast cells. This was validated using nested PCR, TaqMan probe-based qPCR, and next-generation sequencing of RE and LE integration junctions.

[0264] Chromosomal integration of 200 bp DNA cargo into the genome of *N. benthamiana* protoplast was achieved. This was validated using nested PCR followed by Sanger sequencing of RE and LE integration junctions. Chromosomal integration of 200 bp DNA cargo into the genome *Nicotiana benthamiana* leaves was achieved. This was validated using nested PCR followed by Sanger sequencing of LE junction.

[0265] Shown below in Table 9-Table 10 are exemplary sequences of the disclosure.

TABLE 9

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES

Description	Sequence	SEQ ID NO:
VchCas6 amino acid sequence	MVKWYYKTITFLPELCNNESLAAKCLRVLHGFNYQYETRNIGVSFPLWCDATV GKKISFVSKNKIELDLLLKQHYFVQMEQLQYFHISNTVLVPEDCTYVSFRRCQSI DKLTAAGLARKIRRLEKRALSRGEQFDPSSFAQKEHTAIAHYHSLGESSKQTNR NFRLNIRMLSEQPREGNSIFSSYGLSNSENSFQPVPLI*	8
PseCas6 amino acid sequence	MQRYYFTVHFLPKQANLALLTGRCISIMHGFILKHNIEGMGVTFPAWSDSSIGN EIAFVYTDKEILNTLKDQAYFVDMQDCGFFKVSQVLAVPDSCEEVRFIRNQAV AKIFTGESRRRLKRLQKRALARGEDFNPKKIEAPREIDIFHRVAMTSKSSQEDYI LHIQKQDVDCQAEPYFSNYGLASNEKFKGTVPDLSPSIDRN*	9

TABLE 9-continued

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES

Description	Sequence	SEQ ID NO:
VchCas7 amino acid sequence	MKLPTNLAYERSIDPSDVCFFVVWPDDRKTPLTYNSRTLLGQMEAASLAYDVS GQPIKSATAEALAQGNPHQVDFCHVPYGASHIECSFSVSFSSELRQPYKCNSSK VKQTLVQLVELYETKIGWTELATRYLMNICNGKWLWKNTRKAYCWNIVLTP WPWNGEKVGFEDIRTNYTSRQDFKNNKNWSAIVEMIKTAFSSTDGLAIFEVRA TLHLPTNAMVRPSQVFTEKESGSKSKSKTQNSRVFQSTTIDGERSPILGAFKTGA AIATIDDWYPEATEPLRVGRFGVHREDVTCYRHPSTGKDFFSILQQAEHYIEVLS ANKTPAQETINDMHFLMANLIKGGMFQHKGD*	10
PseCas7 amino acid sequence	MELCNILKYDRSLYPGKAVFFYKTADSDFVPLEADINKIRGPKSGFTEAFTPQFS PKNISPQDLTHNNILTLEECYVPPNVEHIFCRFSLRVQANSLVPSGCSDPEVFSLL KELAETFKECGGYKELAVRYCRNILIGTWLWRNQNTGNTQIEIKTSKGSCYLID NTRKLAWESKWASDDLKVLEELSNEIESALTDPNVFWSADITAKIEASFCQEIY PSQILNDKVKQGEASKQFVKAKCADGRYAVSFNSVKIGAALQSIDDWWDEDA SKRLRVHEFGADKEIGVARRPPDSEQNFYSIFKNTEWYLSALKNCITNKNEKIDP AIYYLFSVLIKGGMFQKKAEAKKA*	11
VchCas8 amino acid sequence	MQTLKELIASNPDDLTTELKRAFRPLTPHIAIDGNELDALTILVNLTDKTDDQKD LLDRAKCKQKLRDEKWWASCINCVNYRQSHNPKFPDIRSEGVIRTQALGELPSF LLSSSKIPPYHWSYSHDSKYVNKSAFLTNEFCWDGEISCLGELLKDADHPLWNT LKKLGCSQKTCKAMAKQLADITLTTINVTLAPNYLTQISLPDSDTSYISLSPVAS LSMQSHFHQRLQDENRHSAITRFSRTTNMGVTAMTCGGAFRMLKSGAKFSSPP HHRLNSKRSWLTSEHVQSLKQYQRLNKSLIPENSRIALRRKYKIELQNMVRSWF AMQDHTLDSNILIQHLNHDLSYLGATKRFAYDPAMTKLFTELLKRELSNSINNG EQHTNGSFLVLPNIRVCGATALSSPVTVGIPSLTAFFGFVHAFERNINRTTSSFRV ESFAICVHQLHVEKRGLTAEFVEKGDGTISAPATRDDWQCDVVFSLILNTNFAQ HIDQDTLVTSLPKRLARGSAKIAIDDFKHINSFSTLETAIESLPIEAGRWLSLYAQ SNNNLSDLLAAMTEDHQLMASCVGYHLLEEPKDKPNSLRGYKHAIAECIIGLIN SITFSSETDPNTIFWSLKNYQNYLVVQPRSINDETTDKSSL*	12
PseCas8 amino acid sequence	MHLKELLEITDTTERDRSLRRAFSPYTAMIDITGSEAVALIILLNLTYRKNQVDD LLDKKLAKQALKSEDHINKCIKEIAWFHTHNLKYPDIRVSKQNLAVEPPTLHSY VLSSANYPKAYGWSHNSAKVNFAKLFVSYFKWQNQVSWLAQVLATNSDNWK SAFTSLGLSVKAFKSLCVTVKNSLPEEAIPDSVDRYSRQIRMPYHDGYLAVTPVI SHVVQSKIQQAAIDKRARFSNVEFTRPAAVSMLAASLGGVINVLNYPPYIRSKY HGLSNSRAFKLNNGQTVFNVEALLKPELIKALEGIIFSNNALALKQRRQQKVKN IKELRNTLLEWFSPVFEWRLDAIENGYDLEQLESASERLEYKILSLPDNELPSLTI PLFRLLNEMLGGVSMTQRYAFHPKLMSPLKAALQWLLVNLTDQKHVLIEEDD EHYRYLHLSGIRVFDAQALSNPYCSGIPSLTAVWGMIHSYQRKLNEALGTNVRF TSFSWFIRNYSAVAGKKLPELSLQGAQQSRLKRPGIIDGKYCDLVFDLIIHIDGY EDDLQAVDSKPDILKAHFPSNFAGGVMHQPELNSNINWCCLYSNENQLFEKLR RLPLSGCWVMPTEHKIQDLDELLLLLNSDSKLSPSMMGYMLLTEPMARVGSLE RLHCYAEPAIGVVKYEAATSVRLKGIGNYFNSAFWMLDAQEKFMLMKKV*	13
VchTniQ amino acid sequence	MFLQRPKPYSDESLESFFIRVANKNGYGDVHRFLEATKRFLQDIDHNGYQTFPT DITRINPYSAKNSSSARTASFLKLAQLTFNEPPELLGLAINRTNMKYSPSTSAVV RGAEVFPRSLLRTHSIPCCPLCLRENGYASYLWHFQGYEYCHSHNVPLITTCSC GKEFDYRVSGLKGICCKCKEPITLTSRENGHEAACTVSNWLAGHESKPLPNLPK SYRWGLVHWWMGIKDSEFDHFSFVQFFSNWPRSFHSIIEDEVEFNLEHAVVSTS ELRLKDLLGRLFFGSIRLPERNLQHNIILGELLCYLENRLWQDKGLIANLKMNA LEATVMLNCSLDQIASMVEQRILKPNRKSKPNSPLDVTDYLFHFGDIFCLWLAE FQSDEFNRSFYVSRW*	14
PseTniQ amino acid sequence	MAFLFSPKARAFSDESLESYLLRVVSENFFDSYEGLSLAIREELHELDFEAHGAF PVDLKRLNVYHAKHNSHFRMRALGLLETLLDLPRYELQKLALLKSDIKFNSSV ALYNNGVDIPLRFIRHHAEEAVDSIPVCSQCLAEEAYIKQSWHIKWVNACTKHQ CALLHNCPECYAPINYIENESITHCSCGFELSCASTSPVNTLSIEHLNKLLDKGER NDSNPLFNNMTLTERFAALLWYQERYSQTDNFCLNDAVNYFSKWPAVFNTEL DELSKNAEMKLIDLFNKTEFKFIFGDAILACPSTQKQSESHFIYRALLDYLVTLV ESNPKTKKPNAADLLVSVLEAATLLGTSVEQVYRLYQNGILQTAFRHKMNQRI NPYKGAFFLRHVIEYKTSFGNDKARMYLSAWTG	15
VchTnsAB fusion amino acid sequence	MATSLPTPSAITTSALEYAFHTPARNLTKSRGKNIHRYVSVKMSKRITVESTLEC DACYHFDFEPSIVRFCAQPIRFLYYLNGQSHSYVPDFLVQFDTNEFVLYEVKSA YAKNKPDFDVEWEAKVKAATELGLELELVEESDIRDTVVLNNLKRMHRYASK DELNNVHNSLLKIIKYNGAQSARCLGEQLGLKGRTVLPILCDLLSRCLLDTRLD KPLSLESRFELASYGGSGSGKRTADGSEFESPKKKRKVGSGSGGAKKGFSSFHR KAVSSQDTLESIELVSSANCLESVTYQDISAFPETIAVEINFRLSILRF- LARKCETI VAKSIEPHRVELQQNYSRKIPSAITIYRWWLAFRKSDYNPISLAPNIKDRGNRET KVSTVVDSIMEQAVERVISGRKVNVSSAYKRVRRKVRQYNLTHGTKYTYPKY	16

TABLE 9-continued

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES

Description	Sequence	SEQ ID NO:
	ESVRKRVKKKTPFELLAAGKGERVAKREFRRMGKKILTSSVLERVEIDHTVVD LFAVHEEYRIPLGRPWLTQLVDCYSKAVIGFYLGFEPPSYVSVSLALKNAIQRK DDLISSYESIENEWLCYGIPDLLVTDNGKEFLSKAFDQACESLLINVHQNKVETP DNKPHVERNYGTINTSLLDDLPGKSFSQYLQREGYDSVGEATLTLNEIREIYLIW LVDIYHKKPNQRGTNCPNVAWKKGCQEWEPEEFSGSKDELDFKFAIVDYKQLT KVGITVYKELSYSNDRLAEYRGKKGNHKVQFKYNPECMAVIWVLDEDMNEYF TVNAIDYEYASRVSLWQHKYNMKYQAELNSAEYDEDKEIDAEIKIEEIADRSIV KTNKIRARRRGARHQENSARAKSISNANPASIQKHEDEIVSADNDDWDIDYV*	
PseTnsAB fusion amino acid sequence	MYIRNLRKPSPNKNVFKFASTKVSSVVMCESSLEFDACFHHEYNDLIESFGSQP EGFKYEFMGKSLPYTPDALISYTDKTQKYHEYKPYSKIASPLFRAEFAAKRAAS LKLGIDLVLVTDRQIRVNPILNNLKLLHRYSGVYGISGIQKELLSFIHKSGVIKLN DISSQVGIPIGETRSFLFGLMHKGLVKADLGCDDLTNNPTLWATPGSGSGKRTA DGSEFESPKKKRKVGSGSGGMTDFFNEFDESLVPLKPQTPTQYVKLDDANLIQR DLDTFSDTFKNQALQRYKLISTIDKKLSRGWTQRNLDPILDELFKGGDVVRPN WRTVARWRKKYIESNGDIASLADKNHKMGNRTNRIKGDDKFFDKALERFLDA KRPTIATAYQYYKDLIVIENESIVEGKIPIISYNAFNKRIKAIPPYAVAVARHGKF KADQWFAYCAAHVPPTRILERVEIDHTPLDLILLDDELLIPIGRPYLTLLIDVFSG CVLGFHLSYKSPSYVSAAKAITHAIKPKSLDALNIELQNDWPCFGKFENLVVDN GAEFWSKNLEHACQSAGINIQYNPVRKPWLKPFIERFFGVMNEYFLPELPGKTF SNILEKEEYKPEKDAIMRFSTFVEEFHRWIADVYHQDSNSRETRIPIKRWQQGFD AYPPLTMNEEEETRFSMLMRISDSRTLTRNGFKYQELMYDSTALADYRKHYPQ TKETVKKLIKVDPDDISKIYVYLEELESYLEVPCTDPTGYTDGLSIYEHKTIKKIN REVIRESKDSLGLAKARMAIHERVKQEQEVFIESKTKAKITAVKKQAQIADVSN TGTSTIKVSEESAAPVQKHISNDNSDDWDDDLEAFE*	17
VchTnsC amino acid sequence	MSETREARISRAKRAFVSTPSVRKILSYMDRCRDLSDLESEPTCMMVYGASGV GKTTVIKKYLNQNRRESEAGGDIIPVLHIELPDNAKPVDAARELLVEMGDPLAL YETDLARLTKRLTELIPAVGVKLIIIDEFQHLVEERSNRVLTQVGNWLKMILNKT KCPIVIFGMPYSKVVLQANSQLHGRFSIQVELRPFSYQGGRGVFKTFLEYLDKA LPFEKQAGLANESLQKKLYAFSQGNMRSLRNLIYQASIEAIDNQHETITEEDFVF ASKLTSGDKPNSWKNPFEEGVEVTEDMLRPPPKDIGWEDYLRHSTPRVSKPGR NKNFFE*	18
PseTnsC amino acid sequence	MGNALTEIQIEKLRNFSDCIVMHPQIKTIFNDFDELRLNRKFQSDQQCMLLIGDT GVGKSHTINHYKKRVLATQNYSRNTMPVLVSRISRGKGLDATLVQMLADLELF GSSQIKKRGYKTDLTKKLVESLIKAQVELLIINEFQELIEFKSVQERQQIANGLKF ISEEAKVPIVLVGMPWAAKIAEEPQWASRLVRKRKLEYFSLKNDSKYFRQYLM GLAKKMPFDVPPKLESKNTTIALFAACRGENRALKHLLLEALKLALSCNEYLE NKHFITAYDKFDFFNDKEKLKSKNPFKQDIKDIEIYEVIKNSSYNPNALDPEDML TDRVFAIVK	19
ClpP amino acid sequence	MSYSGERDNFAPHMALVPMVIEQTSRGERSFDIYSRLLKERVIFLTGQVEDHMA NLIVAQMLFLEAENPEKDIYLYINSPGGVITAGMSIYDTMQFIKPDVSTICMGQA ASMGAFLLTAGAKGKRFCLPNSRVMIHQPLGGYQGQATDIEIHAREILKVKGR MNELMALHTGQSLEQIERDTERDRFLSAPEAVEYGLVDSILTHRN*	20
ClpX amino acid sequence	MTDKRKDGSGKLLYCSFCGKSQHEVRKLIAGPSVYICDECVDLCNDIIREEIKE VAPHRERSALPTPHEIRNHLDDYVIGQEQAKKVLAVAVYNHYKRLRNGDTSN GVELGKSNILLIGPTGSGKTLLAETLARLLDVPFTMADATTLTEAGYVGEDVEN IIQKLLQKCDYDVQKAQRGIVYIDEIDKISRKSDNPSITRDVSGEGVQQALLKLIE GTVAAVPPQGGRKHPQQEFLQVDTSKILFICGGAFAGLDKVISHRVETGSGIGF GATVKAKSDKASEGELLAQVEPEDLIKFGLIPEFIGRLPVVATLNELSEEALIQIL KEPKNALTKQYQALFNLEGVDLEFRDEALDAIAKKAMARKTGARGLRSIVEAA LLDTMYDLPSMEDVEKVVIDESVIDGQSKPLLIYGKPEAQQASGE*	21
SV40 Nuclear Localization Signal (NLS) nucleotide sequence	CCAAAGAAGAAAAGAAAAGTT	22
Bipartite NLS nucleotide	ATGGGAAAGCGTACTGCTGATGGTTCAGAGTTTGAAAGTCCAAAGAAGAAA AGAAAAGTT	23
VchCas6 nucleotide sequence	ATGGTGAAATGGTACTACAAAACAATAACATTTCTACCTGAGCTCTGCAAC AATGAATCACTAGCAGCGAAATGTCTGCGTGTACTTCATGGATTCAACTACC AATATGAAACAAGAAACATTGGTGTTAGCTTCCCTCTTTGGTGTGACGCAAC TGTCGGAAAGAAGATATCCTTTGTGTCCAAGAACAAGATTGAGCTTGACTT	24

TABLE 9-continued

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES		
Description	Sequence	SEQ ID NO:
	GCTTCTCAAACAACATTACTTTGTTCAAATGGAGCAGCTTCAGTATTTCCAC ATCTCAAATACTGTGCTGGTTCCAGAAGATTGCACTTATGTCAGTTTCAGGA GATGTCAAAGTATTGATAAGCTTACTGCTGCTGGTTTAGCAAGAAAGATCC GTCGATTGGAGAAGAGAGCTTTGTCAAGAGGAGAACAGTTTGATCCCTCAA GTTTTGCTCAGAAAGAACACACCGCCATTGCTCATTATCATTCTCTCGGTGA ATCTTCAAAGCAAACAAATAGGAATTTTCGGTTAAATATAAGGATGTTGAG TGAACAACCAAGAGAAGGAAATTCCATCTTCTCTTCTTATGGGCTATCAAAT TCTGAGAATTCTTTTCAGCCTGTTCCGCTAATCTGA	
PseCas6 nucleotide sequence	ATGCAGAGATATTATTTTACCGTTCATTTTCTTCCTAAACAAGCTAATTTGG CCCTTCTCACCGGAAGATGCATATCTATCATGCATGGATTTATCTTGAAGCA TAACATTGAAGGAATGGGAGTAACATTTCCCGCATGGTCTGATTCAAGCAT TGGAAACGAGATTGCTTTTGTTTACACTGATAAAGAAATCCTAAACACATTG AAAGATCAGGCTTACTTTGTGGATATGCAAGACTGTGGTTTCTTCAAGGTTT CTCAGGTGTTGGCTGTTCCTGATTCTTGCGAGGAAGTCAGGTTCATCCGCAA TCAAGCCGTCGCAAAGATTTTCACTGGCGAATCGAGGAGGAGACTTAAACG GTTACAAAAGAGAGCCTTAGCAAGAGGAGAAGATTTCAATCCAAAGAAGA TAGAAGCTCCAAGAGAAATAGACATTTTTCACCGTGTAGCAATGACAAGCA AAAGTTCTCAAGAAGACTACATTCTGCACATCCAGAAGCAAGATGTTGATT GTCAAGCAGAGCCTTATTTTTCAAACTATGGTCTTGCGTCAAATGAGAAGTT CAAAGGAACTGTACCAGACTTGAGTCCGTCCATTGATAGAAATTGA	25
VchCas7 nucleotide sequence	ATGAAACTTCCAACAAATCTAGCTTACGAGAGGTCGATTGATCCTTCAGAT GTTTGTTTCTTTGTGGTCTGGCCAGATGATCGGAAAACACCGTTGACATACA ATTCACGTACCCTCCTAGGGCAGATGGAGGCTGCATCTCTGGCGTATGATGT AAGTGGTCAACCGATCAAGAGTGCCACAGCAGAAGCATTAGCTCAAGGAA ATCCTCATCAAGTTGACTTTTGTCATGTTCCTTATGGAGCATCTCACATCGA GTGCAGCTTCTCTGTTTCTTTTTCATCAGAGCTTCGCCAGCCTTACAAATGTA ACAGCAGTAAAGTGAAGCAAACTCTCGTTCAGCTCGTAGAGTTGTATGAAA CTAAGATTGGATGGACTGAACTAGCGACGAGATATTTGATGAACATATGCA ATGGTAAATGGCTCTGGAAAAATACAAGGAAGGCTTATTGTTGGAATATAG TCCTGACTCCATGGCCTTGGAACGGAGAGAAGGTTGGCTTTGAAGACATTA GGACAAACTACACTTCAAGACAAGATTTCAAGAACAACAAGAATTGGTCTG CCATTGTTGAGATGATTAAGACAGCTTTCAGTTCTACTGATGGCTTAGCGAT CTTCGAGGTTCGTGCGACTCTTCACTTACCAACCAATGCTATGGTTAGACCT TCCCAGGTATTCACAGAGAAAGAATCAGGGTCGAAATCAAAAAGCAAGAC TCAGAATTCTCGAGTGTTTCAATCAACAACCATAGATGGTGAAAGATCTCCC ATTCTTGGTGCATTCAAAACTGGTGCAGCTATAGCTACTATTGACGACTGGT ATCCTGAAGCTACAGAACCATTGCGTGTTGGAAGATTTGGTGTCCATAGAG AAGATGTGACTTGCTATAGACATCCTTCCACAGGGAAAGATTTTTTCTCCAT ACTTCAACAAGCTGAACATTACATTGAAGTACTTTCAGCAAACAAAACCCC AGCACAAGAAACCATTAATGATATGCATTTTCTTATGGCTAATCTCATCAAA GGTGGCATGTTTCAGCACAAAGGAGACTGA	26
PseCas7 nucleotide sequence	ATGGAGCTTTGCAATATACTCAAGTATGATCGAAGTCTTTACCCTGGGAAA GCGGTTTTCTTCTACAAAACGGCGGACAGCGATTTTGTTCCTTTGGAAGCTG ACATTAATAAGATAAGAGGACCAAAATCAGGCTTTACAGAGGCTTTCACTC CTCAATTCTCTCCGAAAAATATCTCTCCACAAGATCTAACTCACAACAACAT TCTTACACTCGAAGAGTGTTATGTTCCACCAAATGTCGAACATATTTTCTGC CGGTTTTCTCTGCGTGTCCAAGCTAATTCGTTAGTCCCCTCTGGTTGTTCCGA TCCTGAGGTTTTTAGCCTTCTCAAAGAGCTTGCAGAAACCTTCAAAGAATGT GGTGGTTACAAGGAGCTGGCAGTACGCTATTGTCGGAACATTTTGATTGGA ACATGGCTTTGGAGAAATCAGAATACCGGAAACACACAGATTGAGATCAAG ACTTCAAAAGGATCTTGCTACTTGATTGATAATACAAGAAAACTCGCTTGG GAAAGCAAATGGGCTTCAGACGACTTGAAAGTTTTGGAAGAATTAAGCAAT GAAATCGAATCTGCTTTAACAGATCCAAATGTGTTTTGGTCTGCAGATATAA CCGCCAAGATCGAAGCATCTTTCTGTCAAGAGATATATCCTTCACAGATATT GAATGATAAAGTGAAGCAAGGAGAAGCGTCAAAACAGTTTGTGAAAGCTA AATGTGCTGATGGAAGGTATGCTGTGAGTTTCAACAGTGTTAAGATCGGTG CTGCTTTGCAAAGTATTGATGATTGGTGGGATGAAGATGCTAGTAAAAGAC TTAGAGTTCATGAATTTGGTGCAGACAAGGAAATTGGCGTAGCGAGGAGGC CTCCCGATTCCGAACAAAACTTTTATTCAATCTTCAAGAACACTGAATGGTA TCTTTCCGCCTTAAAGAATTGCATCACTAACAAGAATGAGAAGATTGACCC AGCAATATACTATTTGTTCAGCGTGTTGATCAAAGGTGGAATGTTTCAGAAA AAGGCAGAGGCAAAGAAAGCATGA	27
VchCas8 nucleotide sequence	ATGCAGACTCTTAAAGAACTAATTGCTTCTAATCCGGATGACCTAACAACG GAGCTCAAAAGGGCATTTCGCCCCCTGACTCCTCACATTGCCATTGACGGTA ACGAGCTTGACGCTCTCACCATTCTTGTGAATCTCACAGACAAAACTGATGA TCAAAAGGATCTTCTCGATAGAGCTAAATGTAAACAGAAACTCAGAGATGA GAAATGGTGGGCTTCTTGCATCAATTGCGTTAATTATAGACAGTCTCACAAT	28

TABLE 9-continued

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES

Description	Sequence	SEQ ID NO:
	CCTAAATTCCCTGATATCAGGTCTGAAGGAGTGATTCGCACTCAAGCTTTGG GCGAACTTCCTTCATTTCTATTAAGCTCCAGTAAAATACCACCGTATCATTG GTCATATTCTCATGATTCAAAATATGTTAACAAGTCAGCCTTTTTAACAAAT GAATTCTGTTGGGACGGAGAGATAAGCTGCTTGGGGGAACTCCTCAAAGAT GCGGATCATCCACTTTGGAACACACTAAAGAAGCTTGGCTGCTCACAGAAG ACATGCAAAGCTATGGCGAAACAATTGGCTGATATCACATTAACAACGATT AATGTAACACTCGCTCCGAATTACTTGACTCAGATTTCTCTTCCGGATTCTG ATACAAGTTACATTTCTCTCTCGCCTGTCGCAAGTCTATCAATGCAAAGTCA TTTCCACCAACGTTTGCAAGACGAAAACAGACATTCTGCTATAACAAGATT CTCCCGAACAACTAACATGGGTGTGACTGCAATGACTTGTGGTGGAGCTTTC AGGATGTTAAAGTCTGGTGCAAAGTTCTCCAGTCCACCTCATCATAGATTAA ACTCGAAACGCTCTTGGTTGACCAGTGAGCATGTTCAATCATTGAAGCAGT ATCAACGTCTGAACAAATCATTGATCCCTGAGAATAGTAGGATTGCTCTTAG GCGCAAGTACAAGATTGAACTTCAAAATATGGTTCGATCATGGTTTGCCAT GCAAGATCACACTCTCGATTCAAATATCCTTATTCAACATCTCAATCACGAT TTGAGTTATCTCGGGGCGACCAAAAGATTTGCTTATGATCCAGCTATGACAA AATTGTTCACTGAGCTGCTAAAGAGGGAGCTGAGTAATTCTATCAACAACG GAGAACAACATACCAATGGCAGTTTTCTGGTCCTCCCAAACATACGGGTTT GTGGAGCAACTGCGTTATCTTCCCCAGTAACCGTTGGCATCCCTAGCTTAAC CGCCTTCTTTGGTTTTGTTCACGCCTTTGAGAGGAACATCAACAGGACCACT TCTTCCTTTCGTGTCGAGAGCTTTGCAATATGTGTCCATCAGCTTCATGTGG AGAAGCGTGGTTTGACAGCTGAGTTTGTAGAGAAGGGAGATGGGACAATTT CAGCACCAGCAACCAGAGATGATTGGCAATGTGATGTTGTGTTTTCTCTAAT TTTGAATACGAACTTCGCTCAACACATTGATCAGGACACACTGGTCACGAG TTTACCTAAGCGGCTGGCAAGGGGTTCTGCTAAGATCGCAATAGATGATTTC AAGCATATAAACTCATTCTCCACTCTTGAAACTGCAATTGAGTCTCTTCCCA TCGAAGCTGGAAGATGGCTTAGTCTGTATGCACAGAGTAATAATAATCTTTC GGACTTGTTAGCTGCTATGACCGAAGACCATCAACTTATGGCTAGTTGTGTT GGCTACCACCTTCTAGAAGAACCAAAAGATAAGCCAAATAGCCTTAGAGGA TACAAACATGCCATAGCTGAATGTATTATAGGATTAATTAATTCAATTACAT TTTCCTCGGAAACCGATCCTAATACTATATTCTGGTCCTTAAAGAACTACCA AAACTATCTTGTAGTTCAGCCCAGATCCATAAATGATGAAACAACAGACAA GAGCAGCCTGTGA	
PseCas8 nucleotide sequence	ATGCACCTGAAAGAGCTGTTAGAGATAACAGATACAACTGAGAGAGATAG ATCTCTCCGACGTGCCTTCTCGCCGTATACCGCCATGATCGACATTACAGGA TCAGAAGCAGTCGCTCTAATTATCCTACTCAATCTCACTTATAGAAAGAATC AGGTAGATGACCTGTTGGATAAGAAACTCGCGAAACAAGCCCTCAAGAGTG AGGATCATATCAACAAATGCATCAAGGAGATTGCTTGGTTTCATACTCACA ACCTTAAATATCCAGACATAAGGGTGAGTAAGCAAAATCTTGCAGTAGAAC CACCAACATTGCATTCTTATGTGCTTTCTTCTGCTAATTATCCCAAAGCTTAT GGTTGGTCCCATAATTCAGCTAAAGTGAATTTTGCTAAGTTGTTTGTCTCAT ACTTCAAATGGCAAAACCAAGTGAGCTGGCTTGCCCAAGTCTTGGCAACCA ACTCTGATAACTGGAAAAGTGCATTTACTTCCCTTGGTTTATCGGTGAAAGC CTTCAAATCTTTATGCGTTACAGTTAAGAATTCTTTACCTGAAGAGGCAATC CCGGACTCAGTGGATAGATATAGCAGGCAGATTAGAATGCCTTATCATGAT GGTTATCTTGCTGTCACTCCTGTGATATCTCACGTTGTTCAAAGCAAAATAC AACAAGCTGCTATTGACAAAAGAGCTAGGTTTTCTAATGTTGAGTTTACCAG ACCAGCTGCTGTTTCAATGCTGGCGGCAAGTCTAGGTGGTGTGATAAATGT ACTTAATTATCCTCCATACATTCGATCAAAGTACCATGGACTATCGAATTCA AGGGCTTTTAAGTTAAACAACGGACAGACTGTCTTCAACGTGGAAGCCCTC CTGAAGCCTGAGCTTATAAAAGCTCTTGAAGGAATTATATTCTCTAACAATG CATTGGCATTAAAACAGAGAAGGCAGCAGAAAGTAAAAAACATCAAAGAA CTGAGAAATACCCTCCTTGAGTGGTTTTCCCCTGTTTTCGAATGGAGATTGG ATGCAATTGAGAATGGTTATGATCTAGAACAATTGGAATCAGCGAGCGAAC GGTTAGAGTACAAGATTCTGAGTTTGCCGGATAATGAACTACCTTCTTTGAC TATACCCTTGTTTAGATTACTTAATGAGATGTTGGGAGGAGTTTCAATGACT CAGAGGTATGCATTCCATCCAAAGTTGATGTCTCCATTAAAGGCTGCTCTCC AATGGTTGCTCGTTAATTTAACTGATCAAAAGCATGTTTTAATTGAAGAAGA TGATGAGCATTACAGATACCTTCACCTGTCAGGGATTAGGGTTTTTGATGCC CAAGCATTGTCTAATCCTTACTGCTCTGGTATTCCTTCGCTTACGGCGGTTTG GGGAATGATCCACAGTTATCAACGCAAACTAAACGAAGCTTTGGGCACGAA CGTACGGTTCACAAGCTTTTCATGGTTCATCCGTAATTACTCTGCCGTTGCT GGTAAAAAGCTACCAGAACTTAGCCTTCAGGGTGCACAACAGTCTAGATTG AAAAGACCTGGGATAATAGACGGGAAGTATTGTGATTTGGTGTTTGACTTG ATTATTCATATCGATGGCTATGAAGACGATTTGCAGGCTGTTGATTCAAAGC CTGATATCCTCAAGGCTCATTTCCCAAGCAATTTCGCCGGCGGCGTGATGCA TCAACCCGAGTTGAATTCTAATATTAATTGGTGTTGTCTTTACTCCAACGAA AACCAGCTTTTTGAGAAGCTCCGTCGGCTTCCTCTTTCTGGATGTTGGGTTA TGCCCACAGAACATAAGATTCAAGATCTTGATGAACTGTTGTTGCTGTTGAA CTCGGATAGTAAACTAAGTCCAAGCATGATGGGATACATGCTTCTCACAGA	29

TABLE 9-continued

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES

Description	Sequence	SEQ ID NO:
	GCCAATGGCTCGTGTAGGATCGTTAGAGAGGCTTCACTGTTATGCAGAACC TGCTATTGGTGTTGTCAAATACGAGGCAGCAACTTCTGTTCGATTAAAAGGT ATCGGAAACTATTTCAATTCCGCGTTTTGGATGCTCGATGCTCAAGAAAAGT TCATGCTCATGAAGAAAGTGTGA	
VchTniQ nucleotide sequence	ATGTTTCTTCAAAGACCTAAACCGTATAGCGACGAAAGTCTGGAATCTTTCT TCATCCGTGTTGCCAACAAGAATGGTTATGGAGATGTTCACAGGTTTTTGGA AGCCACAAAGCGATTCCTTCAAGACATTGATCACAATGGATACCAAACATT TCCCACTGATATAACGCGTATCAATCCTTATTCAGCCAAAAACTCTTCATCA GCAAGAACAGCGAGTTTCTTGAAGCTAGCACAGCTCACCTTCAACGAACCA CCAGAATTGCTAGGGCTTGCAATAAACAGAACCAACATGAAGTACTCTCCT TCTACTAGTGCAGTGGTAAGAGGAGCAGAGGTTTTCCCAAGATCCCTCCTG CGTACTCATTCCATTCCATGTTGCCCTCTCTGCTTGAGGGAAAATGGGTATG CTAGTTACCTTTGGCATTTTCAAGGATACGAGTATTGTCACTCTCATAATGT CCCTCTAATTACCACTTGTTCTTGTGGTAAAGAATTCGATTACAGAGTATCT GGGTTAAAAGGAATTTGTTGTAAGTGCAAAGAGCCAATAACTTTGACATCG AGAGAGAATGGCCATGAAGCAGCTTGTACTGTTAGTAATTGGCTTGCCGGT CACGAGTCAAAACCATTGCCTAATTTACCCAAATCTTATAGATGGGGACTC GTCCATTGGTGGATGGGAATCAAGGACTCTGAATTTGATCATTTCAGCTTTG TGCAGTTTTTCTCAAACTGGCCTCGGTCGTTTCATAGTATCATAGAAGATGA AGTTGAATTTAATCTTGAACATGCTGTTGTTTCTACATCTGAGCTAAGGCTC AAAGATCTTCTTGGGAGACTGTTCTTTGGTAGCATCAGGCTTCCTGAGAGAA ATCTGCAACATAACATTATTTTGGGTGAGCTGTTATGTTATCTAGAAAACCG TTTGTGGCAAGATAAAGGCTTGATAGCCAATTTAAAGATGAATGCTCTTGA GGCAACAGTGATGCTTAACTGCTCCTTGGATCAGATTGCTTCAATGGTGGAG CAAAGAATTTTAAAGCCTAACCGAAAATCAAAGCCAAATTCACCGCTTGAT GTGACGGATTATTTGTTTCACTTTGGTGACATTTTCTGCTTATGGTTAGCTGA GTTTCAGTCTGATGAATTCAATCGAAGCTTTTATGTTTCTAGGTGGTGA	30
PseTniQ nucleotide sequence	ATGGCTTTTCTTTTCTCACCAAAAGCAAGAGCGTTTTCCGATGAATCGCTTG AAAGTTATTTACTTCGAGTTGTCTCAGAAAACTTCTTCGATTCATATGAAGG ACTATCTTTGGCCATCAGAGAAGAACTTCATGAGTTGGACTTCGAGGCTCAT GGAGCATTTCCCGTAGATTTAAAGAGACTTAATGTTTACCATGCAAAACAC AACTCACACTTTAGGATGAGAGCTCTTGGCCTATTAGAAACTCTTTTAGATC TTCCTCGTTATGAGCTTCAAAAGCTAGCACTCCTGAAGTCAGACATTAAATT CAATTCTTCCGTTGCCCTCTACAACAATGGTGTTGATATTCCTTTGAGATTCA TTAGACATCATGCTGAGGAAGCTGTTGACTCCATTCCGGTTTGTTCTCAATG CTTGGCTGAAGAAGCTTATATAAAGCAATCTTGGCATATCAAATGGGTTAA TGCTTGCACCAAACATCAATGTGCTTTGCTTCACAATTGTCCTGAATGTTAT GCACCGATCAATTACATTGAGAATGAATCAATCACTCATTGCTCTTGTGGTT TTGAATTGAGCTGCGCCTCTACAAGTCCTGTGAACACACTCTCGATTGAACA TTTGAACAAACTCCTGGACAAAGGAGAACGGAATGACTCAAATCCATTGTT TAACAACATGACGTTGACAGAAAGGTTTGCTGCTTTATTATGGTATCAAGA GAGGTATTCACAGACTGATAATTTCTGTTTGAATGATGCTGTGAACTACTTC AGTAAGTGGCCAGCAGTGTTTAATACTGAGCTAGATGAGCTCTCAAAGAAT GCCGAGATGAAGCTGATTGATCTCTTCAACAAGACAGAGTTTAAGTTTATCT TTGGAGATGCAATTCTTGCTTGTCCAAGTACTCAGAAGCAGAGTGAATCTCA TTTCATCTACCGGGCATTACTTGATTACCTTGTCACTTTAGTGGAGAGCAAT CCCAAAACAAAGAAGCCAAACGCAGCTGATTTGCTTGTTTCTGTCCTGGAA GCAGCAACTTTGTTGGGGACGAGTGTAGAACAAGTTTATAGGCTCTATCAG AATGGAATACTGCAAACCGCCTTTCGGCACAAGATGAATCAGAGAATAAAC CCTTACAAGGGTGCTTTCTTTCTCCGTCATGTGATAGAGTATAAAACCAGCT TTGGTAACGATAAAGCTAGAATGTATCTCTCAGCTTGGTG	31
VchTnsAB nucleotide sequence	ATGGCGACATCCCTTCCAACGCCATCAGCTATAACTACTTCGGCACTAGAGT ATGCTTTCCATACTCCGGCTCGCAATTTGACGAAGTCTAGGGGGAAGAACA TTCATCGATATGTTAGTGTCAAGATGTCTAAAAGAATTACTGTAGAATCAAC ACTAGAATGCGATGCTTGCTATCATTTTGACTTCGAGCCTTCTATCGTGAGA TTTTGTGCTCAGCCTATTCGCTTTCTGTATTATTTAAATGGGCAGTCACACTC CTATGTCCCAGATTTTCTTGTGCAATTCGACACAAACGAGTTTGTCTTGTAT GAAGTTAAAAGCGCTTATGCGAAAAACAAACCTGATTTCGATGTTGAGTGG GAAGCAAAGGTGAAGGCCGCAACAGAGCTGGGTTTGGAACTTGAACTTGTT GAGGAGAGCGATATAAGAGATACCGTTGTTTTGAACAATCTTAAGAGGATG CATAGATATGCGAGTAAGGACGAGCTTAATAACGTACACAATTCCCTACTT AAGATCATAAAGTATAACGGAGCTCAATCCGCCCGATGTTTGGGTGAACAG CTAGGGTTGAAAGGTCGTACGGTACTTCCTATTCTGTGTGATCTTCTTTCAA GATGCTTGCTCGACACCAGACTTGATAAGCCCCTCTCTCTGGAAAGTCGTTT TGAACTCGCCTCATATGGAGGATCAGGATCTGGCAAAAGAACCGCAGATGG TTCTGAGTTCGAAAGCCCAAAGAAAAAGAGAAAGGTGGGTTCCGGTTCTGG AGGAGCCAAGAAGGGTTTTTCTAGCTTTCACAGGAAGGCAGTTTCCTCACA GGACACTTTAGAGAGTATAGAACTAGTTTCCAGTGCGAATTGCCTCGAATC	32

TABLE 9-continued

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES

Description	Sequence	SEQ ID NO:
	AGTCACTTATCAAGATATATCTGCATTTCCTGAAACCATTGCAGTGGAGATC AACTTCAGATTGTCTATTCTCCGGTTTCTCGCAAGAAAATGTGAGACTATCG TTGCTAAATCTATTGAGCCCCATAGAGTTGAGCTTCAACAGAATTATTCTCG GAAGATTCCATCTGCAATAACAATATACAGATGGTGGTTGGCCTTTCGGAA GTCAGATTACAACCCGATTAGTCTGGCTCCTAATATCAAAGACAGAGGGAA CAGAGAAACTAAAGTATCTACTGTGGTTGACTCCATCATGGAACAAGCCGT TGAAAGGGTTATCTCTGGTCGTAAAGTCAATGTTTCTAGCGCGTACAAACG AGTCCGGAGGAAAGTTAGACAATATAATCTTACTCATGGAACAAAGTACAC ATATCCAAAATACGAGAGCGTACGAAAGCGCGTAAAGAAGAAAACACCGT TTGAGTTATTAGCTGCTGGTAAGGGAGAGAGGGTCGCCAAGCGTGAATTTA GGAGAATGGGAAAGAAGATTCTGACGTCAAGTGTTCTCGAGAGGGTGGAG ATAGATCACACTGTTGTGGATCTGTTCGCTGTTCATGAGGAATATAGGATTC CTCTCGGGAGGCCATGGCTGACTCAGTTAGTTGATTGTTACAGCAAAGCAG TCATTGGTTTCTACCTTGGCTTTGAACCTCCTAGTTATGTCTCCGTATCACTA GCTTTAAAGAATGCCATTCAACGCAAAGATGATTTGATTTCCTCTTACGAGT CAATAGAAAATGAATGGTTATGTTATGGTATACCAGATCTCCTTGTTACAGA CAATGGAAAAGAATTTTTGTCCAAAGCTTTTGATCAAGCTTGTGAGAGCTTA TTAATTAACGTGCATCAGAATAAGGTAGAAACCCCAGATAACAAGCCTCAT GTTGAAAGAAACTACGGGACAATAAATACTTCCTTATTGGACGATTTACCT GGTAAAAGTTTCTCTCAGTACTTGCAGAGGGAAGGATATGACTCTGTGGGA GAAGCTACATTGACTCTCAATGAGATCAGAGAAATCTACTTGATTTGGCTTG TCGACATATATCACAAGAAACCGAATCAAAGAGGAACTAACTGTCCAAATG TTGCTTGGAAAAAAGGTTGCCAAGAATGGGAGCCAGAAGAATTCTCCGGGT CAAAAGATGAGCTTGATTTTAAATTCGCAATTGTTGATTATAAACAACTTAC AAAAGTAGGAATCACTGTTTACAAGGAACTAAGCTACTCAAACGATCGTTT GGCAGAGTATAGAGGCAAAAAAGGAAACCACAAAGTTCAGTTCAAGTACA ATCCGGAATGTATGGCTGTGATATGGGTTCTTGATGAAGATATGAACGAAT ATTTCACAGTTAATGCTATCGACTATGAATACGCATCTAGGGTGTCACTGTG GCAACATAAATACAACATGAAGTATCAAGCAGAGCTAAATTCTGCGGAATA TGACGAAGATAAAGAGATTGATGCAGAAATTAAAATCGAGGAAATTGCTG ATAGATCTATAGTGAAAACCAATAAAATTAGAGCTCGTCGTCGTGGTGCTA GGCATCAAGAGAATAGTGCACGAGCTAAATCAATCTCTAATGCAAATCCTG CAAGTATTCAGAAGCATGAAGACGAAATAGTCAGTGCTGATAATGATGATT GGGATATCGATTACGTGTGA	
PseTnsAB nucleotide sequence	ATGTACATTAGGAACTTGAGAAAGCCTTCTCCAAATAAGAATGTTTTTAAGT TCGCATCTACCAAGGTCAGCTCTGTGGTCATGTGTGAATCCAGCCTTGAGTT TGACGCTTGCTTTCATCATGAATATAACGACTTGATTGAGAGTTTTGGTTCT CAGCCAGAAGGTTTCAAATACGAATTCATGGGTAAATCTCTCCCTTACACTC CAGATGCTTTGATCTCCTACACGGACAAAACTCAAAAGTATCACGAGTATA AACCATATTCAAAAATCGCATCACCTCTATTTAGAGCAGAGTTTGCTGCAAA GAGGGCTGCAAGTCTTAAGTTGGGTATTGATTTGGTACTCGTCACAGACAG ACAGATTAGAGTAAACCCAATACTCAACAATCTGAAGCTCCTTCATAGATA CTCGGGCGTTTATGGTATATCTGGTATCCAGAAGGAACTACTTAGCTTTATT CACAAAAGTGGCGTTATCAAACTCAATGATATCTCATCACAGGTTGGGATC CCCATTGGAGAAACTCGGTCATTTCTTTTTGGACTGATGCATAAGGGACTTG TAAAGGCTGACCTAGGTTGTGATGACCTGACAAACAACCCTACTCTTTGGG CTACTCCTGGAAGTGGCTCAGGCAAACGTACTGCTGACGGAAGTGAGTTCG AGTCCCCAAAGAAAAAAAGAAAAGTTGGTAGTGGATCTGGTGGAATGACT GATTTCTTCAATGAGTTTGATGAATCGCTTGTACCACTGAAACCACAAACTC CAACGCAGTATGTTAAGCTGGACGATGCTAATTTAATTCAGAGAGATTTAG ATACATTCTCAGATACTTTCAAGAACCAAGCCCTCCAAAGATATAAGCTAA TTTCTACAATTGATAAAAAGTTAAGCCGCGGATGGACTCAAAGAAATCTCG ACCCGATCCTCGACGAACTCTTCAAAGGCGGTGATGTAGTGAGGCCAAACT GGAGAACTGTGGCAAGATGGAGGAAGAAATATATAGAAAGCAATGGTGAC ATCGCTTCTCTTGCTGATAAAAACCACAAAATGGGAAACAGAACAAATCGG ATAAAAGGAGATGACAAATTTTTTGACAAGGCCTTGGAACGTTTCCTTGAC GCGAAAAGACCTACAATAGCAACAGCTTATCAATATTACAAGGACTTGATA GTGATTGAAAACGAGTCAATTGTTGAAGGGAAGATCCCCATCATCTCTTAT AATGCATTTAACAAACGAATTAAAGCCATTCCTCCTTATGCTGTTGCAGTCG CTAGGCATGGGAAATTCAAAGCTGATCAATGGTTTGCATATTGTGCAGCTC ACGTTCCACCAACGCGTATACTTGAACGAGTGGAGATTGATCATACTCCGTT GGATCTAATATTATTGGATGATGAGCTTCTGATTCCCATTGGCCGGCCATAT CTTACACTTCTCATAGATGTGTTTTCAGGTTGCGTTCTGGGATTCCATTTGTC ATACAAGAGCCCTTCTTATGTCTCTGCAGCAAAAGCAATCACTCACGCAAT AAAACCCAAAAGCTTAGATGCGCTTAATATTGAGCTTCAGAATGACTGGCC CTGTTTCGGGAAGTTTGAGAATTTAGTAGTTGACAATGGAGCTGAATTTTGG TCAAAAAATTTGGAACATGCTTGTCAATCTGCTGGTATAAATATTCAATACA ATCCGGTGCGTAAGCCATGGCTCAAGCCTTTTATAGAGAGGTTCTTTGGGGT GATGAATGAATACTTCTTGCCGGAACTCCCAGGAAAAACATTTTCTAACATT CTTGAGAAAGAAGAATACAAACCGGAGAAGGATGCAATAATGAGGTTCAG	33

TABLE 9-continued

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES

Description	Sequence	SEQ ID NO:
	CACCTTTGTTGAGGAGTTCCACCGTTGGATTGCCGACGTGTATCATCAAGAT AGTAATAGCCGTGAAACGCGTATTCCTATTAAGCGATGGCAACAAGGATTC GATGCTTATCCTCCTTTAACAATGAACGAAGAAGAGGAGACACGGTTTTCT ATGCTCATGAGAATCTCTGACTCTAGGACGCTAACCAGGAATGGGTTTAAG TACCAAGAATTGATGTATGATTCTACCGCCTTAGCAGATTACCGTAAGCATT ACCCTCAGACAAAAGAGACTGTGAAGAAGCTAATAAAGGTCGATCCAGAT GATATCTCCAAGATTTACGTTTATCTGGAAGAGTTGGAAAGTTACTTGGAGG TTCCTTGCACAGATCCAACTGGATACACTGATGGTTTATCCATATATGAGCA TAAAACCATTAAGAAGATCAACCGAGAAGTTATTAGAGAATCAAAAGATA GCCTTGGGTTGGCGAAAGCGAGAATGGCTATTCATGAGAGAGTGAAACAAG AGCAAGAAGTATTCATCGAAAGTAAGACCAAAGCTAAGATTACTGCAGTCA AGAAGCAGGCTCAGATCGCGGATGTTAGTAATACAGGCACATCAACAATCA AGGTGTCCGAGGAGTCTGCGGCTCCTGTCCAGAAACACATCTCAAATGATA ACAGTGATGATTGGGATGATGATCTGGAAGCTTTTGAATGA	
VchTnsC nucleotide sequence	ATGAGTGAAACACGAGAGGCTAGGATAAGCCGGGCCAAACGGGCTTTTGTT TCTACACCGTCTGTACGCAAGATCCTGAGCTACATGGATAGATGTAGGGAC TTATCAGATCTCGAATCAGAGCCAACTTGTATGATGGTTTATGGTGCCTCTG GTGTGGGAAAAACCACGGTCATCAAGAAGTATTTGAATCAAAACAGAAGA GAATCTGAGGCGGGTGGTGATATTATCCCAGTGTTACACATTGAGCTCCCTG ATAATGCAAAGCCTGTTGATGCCGCACGGGAGCTGTTGGTTGAGATGGGGG ATCCACTGGCCCTCTATGAAACTGATTTAGCAAGACTTACTAAGAGATTGAC TGAGCTGATCCCTGCTGTTGGGGTGAAGCTTATAATCATTGATGAATTTCAG CATCTTGTGGAAGAGAGATCCAATAGAGTGTTAACTCAAGTTGGAAACTGG CTTAAGATGATCTTAAACAAAACCAAATGCCCTATTGTAATATTTGGAATGC CTTACTCAAAGGTTGTTCTTCAAGCCAATTCGCAGCTTCACGGAAGGTTTTC AATACAAGTCGAACTACGTCCCTTCTCTTACCAAGGAGGAAGAGGAGTTTT CAAGACATTCTTGGAGTATCTTGATAAAGCTTTACCTTTCGAGAAGCAAGCA GGCTTAGCAAATGAAAGTCTTCAGAAAAAACTGTATGCTTTCTCTCAGGGC AACATGAGATCCTTAAGGAATCTCATATATCAGGCAAGTATTGAAGCTATT GACAACCAACATGAAACCATTACAGAGGAGGATTTTGTGTTTGCTTCAAAA TTGACAAGCGGAGACAAACCAAATTCATGGAAAAATCCTTTTGAAGAAGGT GTTGAAGTGACAGAAGACATGCTTCGTCCACCACCCAAAGATATTGGATGG GAAGATTACCTAAGGCATTCTACTCCCCGTGTATCCAAGCCGGGCAGAAAC AAGAATTTCTTCGAATGA	34
PseTnsC nucleotide sequence	ATGGGAAATGCTCTTACAGAGATTCAAATAGAGAAGTTACGGAATTTCAGT GATTGCATCGTGATGCATCCTCAGATAAAGACCATTTTCAATGATTTTGATG AACTAAGGCTTAATAGAAAGTTTCAGAGTGATCAGCAATGTATGCTTCTTAT AGGAGATACTGGTGTTGGCAAATCACACACCATCAACCATTACAAGAAAAG AGTCCTCGCAACACAAAACTATTCAAGAAACACAATGCCGGTCCTTGTTTCT CGCATATCACGTGGGAAAGGTTTGGATGCTACTTTGGTGCAGATGCTGGCG GATCTAGAGCTGTTTGGTTCTTCTCAGATCAAGAAGCGGGGCTACAAAACA GACCTTACCAAAAAATTAGTTGAGAGTTTGATCAAAGCTCAAGTGGAGCTG TTGATTATCAATGAATTCCAAGAGCTTATTGAGTTTAAAAGTGTTCAAGAAA GACAACAGATTGCAAATGGTCTAAAGTTTATTAGTGAAGAAGCCAAGGTTC CCATTGTTTTGGTTGGAATGCCATGGGCAGCAAAGATAGCTGAGGAACCAC AATGGGCTTCTCGACTCGTAAGAAAACGAAAGCTAGAATATTTTTCACTCA AGAATGACAGCAAGTACTTTAGGCAGTATTTGATGGGTCTTGCCAAAAAGA TGCCTTTCGACGTTCCTCCTAAATTAGAATCGAAGAACACTACGATTGCGCT ATTTGCAGCTTGTAGAGGGGAAAACAGAGCTTTGAAGCACCTTCTTTTAGA AGCACTGAAGCTCGCCCTTTCATGCAATGAGTATCTTGAGAACAAGCATTTC ATCACAGCTTATGACAAATTCGATTTCTTCAACGACAAGGAGAAACTGAAA TCCAAAAACCCTTTCAAGCAAGATATAAAAGATATTGAAATTTATGAAGTT ATTAAAAATTCTTCCTACAATCCAAATGCACTTGATCCAGAAGATATGTTGA CTGATCGAGTATTTGCTATTGTGAAA	35
ClpP nucleotide sequence	ATGAGTTATTCTGGAGAACGGGATAATTTTGCACCTCACATGGCATTAGTTC CAATGGTGATTGAACAAACAAGTAGAGGGGAGCGAAGCTTTGACATTTACT CTCGTCTATTAAAAGAAAGGGTCATCTTCTTAACCGGGCAAGTTGAAGATC ATATGGCAAATCTGATTGTAGCTCAGATGCTTTTTCTTGAAGCAGAGAATCC TGAGAAAGATATCTATCTCTACATTAACTCTCCTGGTGGTGTAATAACTGCT GGAATGTCTATATATGATACAATGCAGTTTATCAAACCTGATGTTTCAACAA TATGTATGGGTCAGGCTGCGTCAATGGGAGCTTTTCTTTTGACGGCTGGCGC AAAAGGAAAGCGGTTCTGCCTACCGAACTCAAGGGTGATGATCCATCAACC ACTTGGAGGATACCAAGGTCAAGCTACCGATATAGAGATTCATGCTAGAGA AATTTTGAAGGTGAAGGGTAGAATGAATGAACTGATGGCCCTTCACACTGG ACAATCGCTCGAGCAGATTGAGCGTGACACAGAGAGAGACAGGTTTTTATC CGCACCAGAAGCTGTTGAGTATGGTTTGGTTGATTCTATTCTTACTCATAGA AACTGA	36

TABLE 9-continued

HELPER PROTEIN AND NUCLEOTIDE SEQUENCES

Description	Sequence	SEQ ID NO:
ClpX nucleotide sequence	ATGACTGATAAACGGAAAGATGGTTCCGGCAAGCTTCTTTACTGCAGCTTTT GCGGGAAATCTCAACATGAAGTTAGGAAACTTATTGCAGGGCCTTCTGTTT ATATCTGTGACGAATGTGTTGATCTCTGCAATGATATTATTCGAGAAGAGAT CAAAGAAGTGGCTCCTCACAGAGAGAGGAGTGCGTTGCCGACACCACACG AGATTAGAAATCATCTTGATGATTATGTCATTGGTCAAGAACAAGCAAAGA AGGTCCTTGCTGTTGCAGTGTACAACCATTACAAGAGGCTCAGAAACGGAG ATACTTCTAATGGTGTGGAACTTGGCAAATCAAACATCCTGCTCATTGGACC CACTGGAAGTGGAAAGACACTCCTAGCAGAGACATTGGCAAGGCTCTTAGA CGTTCCTTTCACTATGGCTGATGCAACTACTCTTACTGAGGCTGGTTATGTT GGAGAAGATGTGGAGAATATTATTCAAAAGTTATTGCAGAAATGTGATTAC GATGTCCAGAAAGCTCAGAGAGGGATAGTCTACATAGATGAGATAGACAA GATCTCTAGGAAGAGTGATAATCCCTCGATAACACGCGACGTTTCTGGCGA AGGTGTTCAACAAGCTTTGTTGAAGCTCATTGAAGGCACCGTAGCGGCAGT TCCACCACAAGGAGGAAGAAAGCATCCTCAGCAAGAGTTTCTCCAAGTGGA TACCTCAAAAATCTTATTTATATGTGGTGGTGCCTTCGCTGGGCTTGACAAA GTTATCTCACACCGTGTAGAAACTGGTAGCGGGATTGGGTTTGGTGCAACG GTGAAGGCAAAATCTGATAAGGCTTCTGAAGGAGAGCTTCTGGCTCAAGTT GAACCAGAGGATCTTATCAAGTTTGGACTAATTCCTGAGTTCATCGGTAGAC TTCCTGTCGTTGCTACGTTGAACGAGTTAAGTGAGGAAGCCCTGATTCAAAT CTTGAAAGAGCCAAAGAATGCTCTGACAAAGCAATATCAGGCTCTTTTCAA TCTTGAGGGAGTTGACTTGGAATTTCGTGATGAAGCGCTGGATGCTATAGCT AAGAAAGCAATGGCGAGGAAAACCGGAGCTCGTGGTCTAAGATCGATTGTT GAAGCCGCTTTATTGGATACAATGTATGATTTGCCCTCCATGGAAGATGTTG AGAAAGTGGTCATCGACGAGTCTGTGATTGATGGACAGTCAAAGCCATTGC TCATATATGGTAAACCTGAAGCACAGCAGGCATCAGGAGAATGA	37

TABLE 10

DONOR CONSTRUCT SEQUENCES

Description	Sequence	SEQ ID NO:
Long intergenic region (LIR)	tagcagaaggcatgttgttgtgactccgag gggttgcctcaaactctatcttataaccgg cgtggaggcatggaggcaggggtattttgg tcattttaatagatagtggaaaatgacgtg gaatttacttaaagacgaagtctttgcgac aagggggggcccacgccgaatttgttaaag gtgttcacactataaaagcatatacgatgt gatggtatttgatggagcgtatattgtatc aggtatttccgttggatacgaataatatta ccggcgtggcccccccttatcgcgagtgct ttagcacgagcggtccagatttaaagtaga aaatttcccgcccactaggtattcgtacga ccctc	1
Short intergenic region (SIR)	aataaaatgattattttatgaatatatttc attgtgcaagtagatagaaattacatatgt tacataacacacgaaataaacaaaaaaaga caatccaaaaacaaacaccccaaaaaaaat aatcactttagataaactcgtatgaggaga ggcacgt	2
RepA coding sequence	tcagtgactcgacgattcccgagcaaaaaa agtctccccgtcacacatgtagtgggtgac gcaattatctttaaagtaatccttctgttg acttgtcattgataacatccagtcttcgtc aggattgcaaagaattatagaagggatccc accttttattttcttcttttttccatattt agggttgacagtgaaatcagactggcaacc tattaattgcttccacaatgggacgaactt gaaggggatgtcgtcgatgatattataggt ggcgtgttcatcgtagttggtgaaatcgat ggtaccgttccaatagttgtgtcgtccgag acttctagcccaggtggtctttccggtacg agttggtccgcagatgtagaggctggggtg	3

TABLE 10-continued

DONOR CONSTRUCT SEQUENCES

Description	Sequence	SEQ ID NO:
	tcggattccattccttccattgtccttgtt aaatcggccatccattcaaggtcagattga gcttgttggtatgagacaggatgtatgtaa gtataagcgtctatgcttacatggtataga tgggtttccctccaggagtgtagatcttcg tggcagcgaagatctgattctgtgaagggc gacacatacggttcaggttgtggagggaat aatttgttggctgaatattccagccattga agctttgttgcccattcatgagggaattct tccttgatcatgtcaagatattcctcctta gacgttgcagtctggataatagttctccat cgtgcgtcagatttgcgaggagaaacctta tgatctcggaaatctcctctggttttaata tctccgtcctttgatatgtaatcaaggact tgtttagagtttctagctggctggatatta gggtgatttccttcaaaatcgaaaaaagaa ggatccctaatacaaggtttttttatcaagc tggagGagagcatgatagtgggtagtgcca tcttgatgaagctcagaagcaacaccaagg aagaaaataagaaaaggtgtgagtttctcc cagagaaactggaataaatcatctctttga gatgagcacttgggataggtaaggaaaaca tatttagattggagtctgaagttcttacta gcagaaggcat	
2nd trans-posonend sequence (LE) for Pse	CCGTATAGGCAGCTGAAGATAAAAACAATC GATTCGATGTTTAAGTTAATGTAATTCTGC CGAAAAGGCAGTGAGTAGTAATTGAAAAAA CCAGTCTTATATCCGGGCTGGTTTTTTTGT TTTTAGCGTTATAACTTTTATTTCACTATG TCACTGTATGGTTGCATTTTATGTCAGTGA ATGGTTGCAGGTTTTTTGTAAGTTATTGAT ATGCCTATCCAGTGTATGTCACGCTTTGCT TGCAGCGACA	4

TABLE 10-continued

DONOR CONSTRUCT SEQUENCES		
Description	Sequence	SEQ ID NO:
2nd trans-posonend sequence (LE) for Vch	GTGTCCGTGACTGATGCAAGGTTATCGGAA CTGTAATGATTATATGTTTTCTAATTTTGA AGTGGTTCTTTTGATGAGTTTTGCCAGTTA CTGAAAGTAACTTTCTTACTGCAGTAGTTT TGCTGAAATACTCGATTCACAAAAATATCA ACTTATGGTTGTTTTGTGAGATATCAATAT ATGGTTGTTTTGTGGTTAAGTTGCTGATTA TAAATAATTATTAAATATCACTTTATGGTT GCATCAACA	5
1st trans-posonend sequence (RE) for Pse	TGTCGCTGCAACCATAGTTTGACATAATGC AACCATACCTTGACATAATTAAAGTCATAC TCTGACATAAAGCCCCTTATTAAAAATGAT GCTAAAAATAAATTCCATTATAATCAATTA TTTAAGAATTAATTAAACCATAGTGTGACA TAGTGGCTTTTTATTTGTATAGTATGTGTA CTCAAATAATAAAAATTTACTTATGTATAT ACGCAACTTAAGAAAGCCATCTCCAAATAA AAATGTATTT	6
1st trans-posonend sequence (RE) for Pse	TGTTGATACAACCATAAAATGATAATTACA CCCATAAATTGATAATTATCACACCCATAA ATTGATATTGCCTCTTCATGGTCTAAACTT CAGTAAGTTTACGACATTTTCATGAAGAGG TCATTTCATGACAAGCTTACCTACGCCCTC AGCAATTACGACTTCGGCGTTAGAGTATGC ATTCCATACTCCCGCTCGCAATCTAACGAA ATCTCGCGGAAAAAATATTCATCGTTATGT CAGTGTAAAG	7

[0266] In at least some of the previously described embodiments, one or more elements used in an embodiment can interchangeably be used in another embodiment unless such a replacement is not technically feasible. It will be appreciated by those skilled in the art that various other omissions, additions and modifications may be made to the methods and structures described above without departing from the scope of the claimed subject matter. All such modifications and changes are intended to fall within the scope of the subject matter, as defined by the appended claims.

[0267] With respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity. As used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural references unless the context clearly dictates otherwise. Any reference to "or" herein is intended to encompass "and/or" unless otherwise stated.

[0268] It will be understood by those within the art that, in general, terms used herein, and especially in the appended claims (e.g., bodies of the appended claims) are generally intended as "open" terms (e.g., the term "including" should be interpreted as "including but not limited to," the term "having" should be interpreted as "having at least," the term "includes" should be interpreted as "includes but is not limited to," etc.). It will be further understood by those within the art that if a specific number of an introduced claim recitation is intended, such an intent will be explicitly recited in the claim, and in the absence of such recitation no such intent is present. For example, as an aid to understanding, the following appended claims may contain usage of the introductory phrases "at least one" and "one or more" to introduce claim recitations. However, the use of such phrases should not be construed to imply that the introduction of a claim recitation by the indefinite articles "a" or "an" limits any particular claim containing such introduced claim recitation to embodiments containing only one such recitation, even when the same claim includes the introductory phrases "one or more" or "at least one" and indefinite articles such as "a" or "an" (e.g., "a" and/or "an" should be interpreted to mean "at least one" or "one or more"); the same holds true for the use of definite articles used to introduce claim recitations. In addition, even if a specific number of an introduced claim recitation is explicitly recited, those skilled in the art will recognize that such recitation should be interpreted to mean at least the recited number (e.g., the bare recitation of "two recitations," without other modifiers, means at least two recitations, or two or more recitations). Furthermore, in those instances where a convention analogous to "at least one of A, B, and C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (e.g., "a system having at least one of A, B, and C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). In those instances where a convention analogous to "at least one of A, B, or C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (e.g., "a system having at least one of A, B, or C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms.

[0269] In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0270] As will be understood by one skilled in the art, for any and all purposes, such as in terms of providing a written description, all ranges disclosed herein also encompass any and all possible sub-ranges and combinations of sub-ranges thereof. Any listed range can be easily recognized as sufficiently describing and enabling the same range being broken down into at least equal halves, thirds, quarters, fifths, tenths, etc. As a non-limiting example, each range discussed herein can be readily broken down into a lower third, middle third and upper third, etc. As will also be understood by one skilled in the art all language such as "up to," "at least," "greater than," "less than," and the like include the number recited and refer to ranges which can be subsequently broken down into sub-ranges as discussed above. Finally, as will be understood by one skilled in the art, a range includes each individual member. Thus, for example, a group having 1-3 articles refers to groups having 1, 2, or 3 articles.

Similarly, a group having 1-5 articles refers to groups having 1, 2, 3, 4, or 5 articles, and so forth.

[0271] While various aspects and embodiments have been disclosed herein, other aspects and embodiments will be apparent to those skilled in the art. The various aspects and embodiments disclosed herein are for purposes of illustration and are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

SEQUENCE LISTING

```
Sequence total quantity: 69
SEQ ID NO: 1            moltype = DNA  length = 365
FEATURE                 Location/Qualifiers
source                  1..365
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 1
tagcagaagg catgttgttg tgactccgag gggttgcctc aaactctatc ttataaccgg  60
cgtggaggca tggaggcagg ggtattttgg tcattttaat agatagtgga aaatgacgtg  120
gaatttactt aaagacgaag tctttgcgac aagggggggc ccacgccgaa tttaatatta  180
ccggcgtggc cccccttat cgcgagtgct ttagcacgag cggtccagat ttaaagtaga  240
aaatttcccg cccactaggg ttaaaggtgt tcacactata aaagcatata cgatgtgatg  300
gtatttgatg gagcgtatat tgtatcaggt atttccgttg gatacgaatt attcgtacga  360
ccctc                                                              365

SEQ ID NO: 2            moltype = DNA  length = 157
FEATURE                 Location/Qualifiers
source                  1..157
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 2
aataaaatga ttattttatg aatatatttc attgtgcaag tagatagaaa ttacatatgt  60
tacataacac acgaaataaa caaaaaaaga caatccaaaa acaaacaccc caaaaaaaat  120
aatcacttta gataaactcg tatgaggaga ggcacgt                           157

SEQ ID NO: 3            moltype = DNA  length = 1091
FEATURE                 Location/Qualifiers
source                  1..1091
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 3
tcagtgactc gacgattccc gagcaaaaaa agtctccccg tcacacatgt agtgggtgac  60
gcaattatct ttaaagtaat ccttctgttg acttgtcatt gataacatcc agtcttcgtc  120
aggattgcaa agaattatag aagggatccc acctttttatt ttcttctttt ttccatattt  180
agggttgaca gtgaaatcag actggcaacc tattaattgc ttccacaatg ggacgaactt  240
gaaggggatg tcgtcgatga tattataggt ggcgtgttca tcgtagttgg tgaaatcgat  300
ggtaccgttc caatagttgt gtcgtccgag acttctagcc caggtggtct ttccggtacg  360
agttggtccg cagatgtaga ggctggggtg tcggattcca ttccttccat tgtccttgtt  420
aaatcggcca tccattcaag gtcagattga gcttgttggt atgagacagg atgtatgtaa  480
gtataagcgt ctatgcttac atggtataga tgggtttccc tccaggagtg tagatcttcg  540
tggcagcgaa gatctgattc tgtgaagggc gacacatacg gttcaggttg tggagggaat  600
aatttgttgg ctgaatattc cagccattga agctttgttg cccattcatg agggaattct  660
tccttgatca tgtcaagata ttcctcctta gacgttgcag tctggataat agttctccat  720
cgtgcgtcag atttgcgagg agaaacctta tgatctcgga aatctcctct ggttttaata  780
tctccgtcct ttgatatgta atcaaggact tgtttagagt ttctagctgg ctggatatta  840
gggtgatttc cttcaaaatc gaaaaaagaa ggatccctaa tacaaggttt tttatcaagc  900
tggaggagag catgatagtg ggtagtgcca tcttgatgaa gctcagaagc aacaccaagg  960
aagaaaataa gaaaaggtgt gagtttctcc cagagaaact ggaataaatc atctctttga  1020
gatgagcact tgggataggt aaggaaaaca tatttagatt ggagtctgaa gttcttacta  1080
gcagaaggca t                                                       1091

SEQ ID NO: 4            moltype = DNA  length = 250
FEATURE                 Location/Qualifiers
source                  1..250
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 4
ccgtataggc agctgaagat aaaaacaatc gattcgatgt ttaagttaat gtaattctgc  60
cgaaaaggca gtgagtagta attgaaaaaa ccagtcttat atccgggctg gtttttttgt  120
ttttagcgtt ataactttta tttcactatg tcactgtatg gttgcatttt atgtcagtga  180
atggttgcag gtttttttgta agttattgat atgcctatcc agtgtatgtc acgctttgct  240
tgcagcgaca                                                         250

SEQ ID NO: 5            moltype = DNA  length = 249
FEATURE                 Location/Qualifiers
source                  1..249
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 5
```

-continued

```
gtgtccgtga ctgatgcaag gttatcggaa ctgtaatgat tatatgtttt ctaattttga  60
agtggttctt ttgatgagtt ttgccagtta ctgaaagtaa ctttcttact gcagtagttt  120
tgctgaaata ctcgattcac aaaaatatca acttatggtt gttttgtgag atatcaatat  180
atggttgttt tgtggttaag ttgctgatta taaataatta ttaaatatca ctttatggtt  240
gcatcaaca                                                          249

SEQ ID NO: 6            moltype = DNA  length = 250
FEATURE                 Location/Qualifiers
source                  1..250
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 6
tgtcgctgca accatagttt gacataatgc aaccatacct tgacataatt aaagtcatac  60
tctgacataa agccccttat taaaaatgat gctaaaaata aattccatta taatcaatta  120
tttaagaatt aattaaacca tagtgtgaca tagtggcttt ttatttgtat agtatgtgta  180
ctcaaataat aaaaatttac ttatgtatat acgcaactta agaaagccat ctccaaataa  240
aaatgtattt                                                         250

SEQ ID NO: 7            moltype = DNA  length = 250
FEATURE                 Location/Qualifiers
source                  1..250
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 7
tgttgataca accataaaat gataattaca cccataaatt gataattatc acacccataa  60
attgatattg cctcttcatg gtctaaactt cagtaagttt acgacatttt catgaagagg  120
tcatttcatg acaagcttac ctacgccctc agcaattacg acttcggcgt tagagtatgc  180
attccatact cccgctcgca atctaacgaa atctcgcgga aaaaatattc atcgttatgt  240
cagtgtaaag                                                         250

SEQ ID NO: 8            moltype = AA  length = 200
FEATURE                 Location/Qualifiers
source                  1..200
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 8
MVKWYYKTIT FLPELCNNES LAAKCLRVLH GFNYQYETRN IGVSFPLWCD ATVGKKISFV  60
SKNKIELDLL LKQHYFVQME QLQYFHISNT VLVPEDCTYV SFRRCQSIDK LTAAGLARKI  120
RRLEKRALSR GEQFDPSSFA QKEHTAIAHY HSLGESSKQT NRNFRLNIRM LSEQPREGNS  180
IFSSYGLSNS ENSFQPVPLI                                              200

SEQ ID NO: 9            moltype = AA  length = 203
FEATURE                 Location/Qualifiers
source                  1..203
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 9
MQRYYFTVHF LPKQANLALL TGRCISIMHG FILKHNIEGM GVTFPAWSDS SIGNEIAFVY  60
TDKEILNTLK DQAYFVDMQD CGFFKVSQVL AVPDSCEEVR FIRNQAVAKI FTGESRRRLK  120
RLQKRALARG EDFNPKKIEA PREIDIFHRV AMTSKSSQED YILHIQKQDV DCQAEPYFSN  180
YGLASNEKFK GTVPDLSPSI DRN                                          203

SEQ ID NO: 10           moltype = AA  length = 352
FEATURE                 Location/Qualifiers
source                  1..352
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 10
MKLPTNLAYE RSIDPSDVCF FVVWPDDRKT PLTYNSRTLL GQMEAASLAY DVSGQPIKSA  60
TAEALAQGNP HQVDFCHVPY GASHIECSFS VSFSSELRQP YKCNSSKVKQ TLVQLVELYE  120
TKIGWTELAT RYLMNICNGK WLWKNTRKAY CWNIVLTPWP WNGEKVGFED IRTNYTSRQD  180
FKNNKNWSAI VEMIKTAFSS TDGLAIFEVR ATLHLPTNAM VRPSQVFTEK ESGSKSKSKT  240
QNSRVFQSTT IDGERSPILG AFKTGAAIAT IDDWYPEATE PLRVGRFGVH REDVTCYRHP  300
STGKDFFSIL QQAEHYIEVL SANKTPAQET INDMHFLMAN LIKGGMFQHK GD          352

SEQ ID NO: 11           moltype = AA  length = 350
FEATURE                 Location/Qualifiers
source                  1..350
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 11
MELCNILKYD RSLYPGKAVF FYKTADSDFV PLEADINKIR GPKSGFTEAF TPQFSPKNIS  60
PQDLTHNNIL TLEECYVPPN VEHIFCRFSL RVQANSLVPS GCSDPEVFSL LKELAETFKE  120
CGGYKELAVR YCRNILIGTW LWRNQNTGNT QIEIKTSKGS CYLIDNTRKL AWESKWASDD  180
LKVLEELSNE IESALTDPNV FWSADITAKI EASFCQEIYP SQILNDKVKQ GEASKQFVKA  240
KCADGRYAVS FNSVKIGAAL QSIDDWWDED ASKRLRVHEF GADKEIGVAR RPPDSEQNFY  300
```

-continued

```
SIFKNTEWYL SALKNCITNK NEKIDPAIYY LFSVLIKGGM FQKKAEAKKA            350

SEQ ID NO: 12          moltype = AA  length = 640
FEATURE                Location/Qualifiers
source                 1..640
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 12
MQTLKELIAS NPDDLTTELK RAFRPLTPHI AIDGNELDAL TILVNLTDKT DDQKDLLDRA  60
KCKQKLRDEK WWASCINCVN YRQSHNPKFP DIRSEGVIRT QALGELPSFL LSSSKIPPYH  120
WSYSHDSKYV NKSAFLTNEF CWDGEISCLG ELLKDADHPL WNTLKKLGCS QKTCKAMAKQ  180
LADITLTTIN VTLAPNYLTQ ISLPDSDTSY ISLSPVASLS MQSHFHQRLQ DENRHSAITR  240
FSRTTNMGVT AMTCGGAFRM LKSGAKFSSP PHHRLNSKRS WLTSEHVQSL KQYQRLNKSL  300
IPENSRIALR RKYKIELQNM VRSWFAMQDH TLDSNILIQH LNHDLSYLGA TKRFAYDPAM  360
TKLFTELLKR ELSNSINNGE QHTNGSFLVL PNIRVCGATA LSSPVTVGIP SLTAFFGFVH  420
AFERNINRTT SSFRVESFAI CVHQLHVEKR GLTAEFVEKG DGTISAPATR DDWQCDVVFS  480
LILNTNFAQH IDQDTLVTSL PKRLARGSAK IAIDDFKHIN SFSTLETAIE SLPIEAGRWL  540
SLYAQSNNNL SDLLAAMTED HQLMASCVGY HLLEEPKDKP NSLRGYKHAI AECIIGLINS  600
ITFSSETDPN TIFWSLKNYQ NYLVVQPRSI NDETTDKSSL                       640

SEQ ID NO: 13          moltype = AA  length = 695
FEATURE                Location/Qualifiers
source                 1..695
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 13
MHLKELLEIT DTTERDRSLR RAFSPYTAMI DITGSEAVAL IILLNLTYRK NQVDDLLDKK  60
LAKQALKSED HINKCIKEIA WFHTHNLKYP DIRVSKQNLA VEPPTLHSYV LSSANYPKAY  120
GWSHNSAKVN FAKLFVSYFK WQNQVSWLAQ VLATNSDNWK SAFTSLGLSV KAFKSLCVTV  180
KNSLPEEAIP DSVDRYSRQI RMPYHDGYLA VTPVISHVVQ SKIQQAAIDK RARFSNVEFT  240
RPAAVSMLAA SLGGVINVLN YPPYIRSKYH GLSNSRAFKL NNGQTVFNVE ALLKPELIKA  300
LEGIIFSNNA LALKQRRQQK VKNIKELRNT LLEWFSPVFE WRLDAIENGY DLEQLESASE  360
RLEYKILSLP DNELPSLTIP LFRLLNEMLG GVSMTQRYAF HPKLMSPLKA ALQWLLVNLT  420
DQKHVLIEED DEHYRYLHLS GIRVFDAQAL SNPYCSGIPS LTAVWGMIHS YQRKLNEALG  480
TNVRFTSFSW FIRNYSAVAG KKLPELSLQG AQQSRLKRPG IIDGKYCDLV FDLIIHIDGY  540
EDDLQAVDSK PDILKAHFPS NFAGGVMHQP ELNSNINWCC LYSNENQLFE KLRRLPLSGC  600
WVMPTEHKIQ DLDELLLLLN SDSKLSPSMM GYMLLTEPMA RVGSLERLHC YAEPAIGVVK  660
YEAATSVRLK GIGNYFNSAF WMLDAQEKFM LMKKV                            695

SEQ ID NO: 14          moltype = AA  length = 394
FEATURE                Location/Qualifiers
source                 1..394
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 14
MFLQRPKPYS DESLESFFIR VANKNGYGDV HRFLEATKRF LQDIDHNGYQ TFPTDITRIN  60
PYSAKNSSSA RTASFLKLAQ LTFNEPPELL GLAINRTNMK YSPSTSAVVR GAEVFPRSLL  120
RTHSIPCCPL CLRENGYASY LWHFQGYEYC HSHNVPLITT CSCGKEFDYR VSGLKGICCK  180
CKEPITLTSR ENGHEAACTV SNWLAGHESK PLPNLPKSYR WGLVHWWMGI KDSEFDHFSF  240
VQFFSNWPRS FHSIIEDEVE FNLEHAVVST SELRLKDLLG RLFFGSIRLP ERNLQHNIIL  300
GELLCYLENR LWQDKGLIAN LKMNALEATV MLNCSLDQIA SMVEQRILKP NRKSKPNSPL  360
DVTDYLFHFG DIFCLWLAEF QSDEFNRSFY VSRW                             394

SEQ ID NO: 15          moltype = AA  length = 411
FEATURE                Location/Qualifiers
source                 1..411
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 15
MAFLFSPKAR AFSDESLESY LLRVVSENFF DSYEGLSLAI REELHELDFE AHGAFPVDLK  60
RLNVYHAKHN SHFRMRALGL LETLLDLPRY ELQKLALLKS DIKFNSSVAL YNNGVDIPLR  120
FIRHHAEEAV DSIPVCSQCL AEEAYIKQSW HIKWVNACTK HQCALLHNCP ECYAPINYIE  180
NESITHCSCG FELSCASTSP VNTLSIEHLN KLLDKGERND SNPLFNNMTL TERFAALLWY  240
QERYSQTDNF CLNDAVNYFS KWPAVFNTEL DELSKNAEMK LIDLFNKTEF KFIFGDAILA  300
CPSTQKQSES HFIYRALLDY LVTLVESNPK TKKPNAADLL VSVLEAATLL GTSVEQVYRL  360
YQNGILQTAF RHKMNQRINP YKGAFFLRHV IEYKTSFGND KARMYLSAWT G          411

SEQ ID NO: 16          moltype = AA  length = 861
FEATURE                Location/Qualifiers
source                 1..861
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 16
MATSLPTPSA ITTSALEYAF HTPARNLTKS RGKNIHRYVS VKMSKRITVE STLECDACYH  60
FDFEPSIVRF CAQPIRFLYY LNGQSHSYVP DFLVQFDTNE FVLYEVKSAY AKNKPDFDVE  120
WEAKVKAATE LGLELELVEE SDIRDTVVLN NLKRMHRYAS KDELNNVHNS LLKIIKYNGA  180
```

-continued

```
QSARCLGEQL GLKGRTVLPI LCDLLSRCLL DTRLDKPLSL ESRFELASYG GSGSGKRTAD  240
GSEFESPKKK RKVGSGSGGA KKGFSSFHRK AVSSQDTLES IELVSSANCL ESVTYQDISA  300
FPETIAVEIN FRLSILRFLA RKCETIVAKS IEPHRVELQQ NYSRKIPSAI TIYRWWLAFR  360
KSDYNPISLA PNIKDRGNRE TKVSTVVDSI MEQAVERVIS GRKVNVSSAY KRVRRKVRQY  420
NLTHGTKYTY PKYESVRKRV KKKTPFELLA AGKGERVAKR EFRRMGKKIL TSSVLERVEI  480
DHTVVDLFAV HEEYRIPLGR PWLTQLVDCY SKAVIGFYLG FEPPSYVSVS LALKNAIQRK  540
DDLISSYESI ENEWLCYGIP DLLVTDNGKE FLSKAFDQAC ESLLINVHQN KVETPDNKPH  600
VERNYGTINT SLLDDLPGKS FSQYLQREGY DSVGEATLTL NEIREIYLIW LVDIYHKKPN  660
QRGTNCPNVA WKKGCQEWEP EEFSGSKDEL DFKFAIVDYK QLTKVGITVY KELSYSNDRL  720
AEYRGKKGNH KVQFKYNPEC MAVIWVLDED MNEYFTVNAI DYEYASRVSL WQHKYNMKYQ  780
AELNSAEYDE DKEIDAEIKI EEIADRSIVK TNKIRARRRG ARHQENSARA KSISNANPAS  840
IQKHEDEIVS ADNDDWDIDY V                                            861

SEQ ID NO: 17          moltype = AA  length = 851
FEATURE                Location/Qualifiers
source                 1..851
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 17
MYIRNLRKPS PNKNVFKFAS TKVSSVVMCE SSLEFDACFH HEYNDLIESF GSQPEGFKYE  60
FMGKSLPYTP DALISYTDKT QKYHEYKPYS KIASPLFRAE FAAKRAASLK LGIDLVLVTD  120
RQIRVNPILN NLKLLHRYSG VYGISGIQKE LLSFIHKSGV IKLNDISSQV GIPIGETRSF  180
LFGLMHKGLV KADLGCDDLT NNPTLWATPG SGSGKRTADG SEFESPKKKR KVGSGSGGMT  240
DFFNEFDESL VPLKPQTPTQ YVKLDDANLI QRDLDTFSDT FKNQALQRYK LISTIDKKLS  300
RGWTQRNLDP ILDELFKGGD VVRPNWRTVA RWRKKYIESN GDIASLADKN HKMGNRTNRI  360
KGDDKFFDKA LERFLDAKRP TIATAYQYYK DLIVIENESI VEGKIPIISY NAFNKRIKAI  420
PPYAVAVARH GKFKADQWFA YCAAHVPPTR ILERVEIDHT PLDLILLDDE LLIPIGRPYL  480
TLLIDVFSGC VLGFHLSYKS PSYVSAAKAI THAIKPKSLD ALNIELQNDW PCFGKFENLV  540
VDNGAEFWSK NLEHACQSAG INIQYNPVRK PWLKPFIERF FGVMNEYFLP ELPGKTFSNI  600
LEKEEYKPEK DAIMRFSTFV EEFHRWIADV YHQDSNSRET RIPIKRWQQG FDAYPPLTMN  660
EEEETRFSML MRISDSRTLT RNGFKYQELM YDSTALADYR KHYPQTKETV KKLIKVDPDD  720
ISKIYVYLEE LESYLEVPCT DPTGYTDGLS IYEHKTIKKI NREVIRESKD SLGLAKARMA  780
IHERVKQEQE VFIESKTKAK ITAVKKQAQI ADVSNTGTST IKVSEESAAP VQKHISNDNS  840
DDWDDDLEAF E                                                       851

SEQ ID NO: 18          moltype = AA  length = 330
FEATURE                Location/Qualifiers
source                 1..330
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 18
MSETREARIS RAKRAFVSTP SVRKILSYMD RCRDLSDLES EPTCMMVYGA SGVGKTTVIK  60
KYLNQNRRES EAGGDIIPVL HIELPDNAKP VDAARELLVE MGDPLALYET DLARLTKRLT  120
ELIPAVGVKL IIIDEFQHLV EERSNRVLTQ VGNWLKMILN KTKCPIVIFG MPYSKVVLQA  180
NSQLHGRFSI QVELRPFSYQ GGRGVFKTFL EYLDKALPFE KQAGLANESL QKKLYAFSQG  240
NMRSLRNLIY QASIEAIDNQ HETITEEDFV FASKLTSGDK PNSWKNPFEE GVEVTEDMLR  300
PPPKDIGWED YLRHSTPRVS KPGRNKNFFE                                   330

SEQ ID NO: 19          moltype = AA  length = 334
FEATURE                Location/Qualifiers
source                 1..334
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 19
MGNALTEIQI EKLRNFSDCI VMHPQIKTIF NDFDELRLNR KFQSDQQCML LIGDTGVGKS  60
HTINHYKKRV LATQNYSRNT MPVLVSRISR GKGLDATLVQ MLADLELFGS SQIKKRGYKT  120
DLTKKLVESL IKAQVELLII NEFQELIEFK SVQERQQIAN GLKFISEEAK VPIVLVGMPW  180
AAKIAEEPQW ASRLVRKRKL EYFSLKNDSK YFRQYLMGLA KKMPFDVPPK LESKNTTIAL  240
FAACRGENRA LKHLLLEALK LALSCNEYLE NKHFITAYDK FDFFNDKEKL KSKNPFKQDI  300
KDIEIYEVIK NSSYNPNALD PEDMLTDRVF AIVK                              334

SEQ ID NO: 20          moltype = AA  length = 207
FEATURE                Location/Qualifiers
source                 1..207
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 20
MSYSGERDNF APHMALVPMV IEQTSRGERS FDIYSRLLKE RVIFLTGQVE DHMANLIVAQ  60
MLFLEAENPE KDIYLYINSP GGVITAGMSI YDTMQFIKPD VSTICMGQAA SMGAFLLTAG  120
AKGKRFCLPN SRVMIHQPLG GYQGQATDIE IHAREILKVK GRMNELMALH TGQSLEQIER  180
DTERDRFLSA PEAVEYGLVD SILTHRN                                      207

SEQ ID NO: 21          moltype = AA  length = 424
FEATURE                Location/Qualifiers
source                 1..424
                       mol_type = protein
```

-continued

```
                        organism = synthetic construct
SEQUENCE: 21
MTDKRKDGSG KLLYCSFCGK SQHEVRKLIA GPSVYICDEC VDLCNDIIRE EIKEVAPHRE  60
RSALPTPHEI RNHLDDYVIG QEQAKKVLAV AVYNHYKRLR NGDTSNGVEL GKSNILLIGP  120
TGSGKTLLAE TLARLLDVPF TMADATTLTE AGYVGEDVEN IIQKLLQKCD YDVQKAQRGI  180
VYIDEIDKIS RKSDNPSITR DVSGEGVQQA LLKLIEGTVA AVPPQGGRKH PQQEFLQVDT  240
SKILFICGGA FAGLDKVISH RVETGSGIGF GATVKAKSDK ASEGELLAQV EPEDLIKFGL  300
IPEFIGRLPV VATLNELSEE ALIQILKEPK NALTKQYQAL FNLEGVDLEF RDEALDAIAK  360
KAMARKTGAR GLRSIVEAAL LDTMYDLPSM EDVEKVVIDE SVIDGQSKPL LIYGKPEAQQ  420
ASGE                                                               424

SEQ ID NO: 22          moltype = DNA  length = 21
FEATURE                Location/Qualifiers
source                 1..21
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 22
ccaaagaaga aaagaaaagt t                                            21

SEQ ID NO: 23          moltype = DNA  length = 60
FEATURE                Location/Qualifiers
source                 1..60
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 23
atgggaaagc gtactgctga tggttcagag tttgaaagtc caaagaagaa aagaaaagtt  60

SEQ ID NO: 24          moltype = DNA  length = 603
FEATURE                Location/Qualifiers
source                 1..603
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 24
atggtgaaat ggtactacaa aacaataaca tttctacctg agctctgcaa caatgaatca  60
ctagcagcga aatgtctgcg tgtacttcat ggattcaact accaatatga aacaagaaac  120
attggtgtta gcttccctct ttggtgtgac gcaactgtcg gaaagaagat atcctttgtg  180
tccaagaaca agattgagct tgacttgctt ctcaaacaac attactttgt tcaaatggag  240
cagcttcagt atttccacat ctcaaatact gtgctggttc cagaagattg cacttatgtc  300
agtttcagga gatgtcaaag tattgataag cttactgctg ctggtttagc aagaaagatc  360
cgtcgattgg agaagagagc tttgtcaaga ggagaacagt ttgatccctc aagttttgct  420
cagaaagaac acaccgccat tgctcattat cattctctcg gtgaatcttc aaagcaaaca  480
aataggaatt ttcggttaaa tataaggatg ttgagtgaac aaccaagaga aggaaattcc  540
atcttctctt cttatgggct atcaaattct gagaattctt ttcagcctgt tccgctaatc  600
tga                                                                603

SEQ ID NO: 25          moltype = DNA  length = 612
FEATURE                Location/Qualifiers
source                 1..612
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 25
atgcagagat attattttac cgttcatttt cttcctaaac aagctaattt ggcccttctc  60
accggaagat gcatatctat catgcatgga tttatcttga agcataacat tgaaggaatg  120
ggagtaacat ttcccgcatg gtctgattca agcattggaa acgagattgc ttttgtttac  180
actgataaag aaatcctaaa cacattgaaa gatcaggctt actttgtgga tatgcaagac  240
tgtggtttct tcaaggtttc tcaggtgttg gctgttcctg attcttgcga ggaagtcagg  300
ttcatccgca atcaagccgt cgcaaagatt ttcactggcg aatcgaggag gagacttaaa  360
cggttacaaa agagagcctt agcaagagga gaagatttca atccaaagaa gatagaagct  420
ccaagagaaa tagacatttt tcaccgtgta gcaatgacaa gcaaaagttc tcaagaagac  480
tacattctgc acatccagaa gcaagatgtt gattgtcaag cagagcctta tttttcaaac  540
tatggtcttg cgtcaaatga gaagttcaaa ggaactgtac cagacttgag tccgtccatt  600
gatagaaatt ga                                                      612

SEQ ID NO: 26          moltype = DNA  length = 1059
FEATURE                Location/Qualifiers
source                 1..1059
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 26
atgaaacttc caacaaatct agcttacgag aggtcgattg atccttcaga tgtttgtttc  60
tttgtggtct ggccagatga tcggaaaaca ccgttgacat acaattcacg taccctccta  120
gggcagatgg aggctgcatc tctggcgtat gatgtaagtg gtcaaccgat caagagtgcc  180
acagcagaag cattagctca aggaaatcct catcaagttg acttttgtca tgttccttat  240
ggagcatctc acatcgagtg cagcttctct gtttcttttt catcagagct tcgccagcct  300
tacaaatgta acagcagtaa agtgaagcaa actctcgttc agctcgtaga gttgtatgaa  360
actaagattg gatggactga actagcgacg agatatttga tgaacatatg caatggtaaa  420
```

-continued

```
tggctctgga aaaatacaag gaaggcttat tgttggaata tagtcctgac tccatggcct  480
tggaacggag agaaggttgg ctttgaagac attaggacaa actacacttc aagacaagat  540
ttcaagaaca acaagaattg gtctgccatt gttgagatga ttaagacagc tttcagttct  600
actgatggct tagcgatctt cgaggttcgt gcgactcttc acttaccaac caatgctatg  660
gttagacctt cccaggtatt cacagagaaa gaatcagggt cgaaatcaaa aagcaagact  720
cagaattctc gagtgtttca atcaacaacc atagatggtg aaagatctcc cattcttggt  780
gcattcaaaa ctggtgcagc tatagctact attgacgact ggtatcctga agctacagaa  840
ccattgcgtg ttggaagatt tggtgtccat agagaagatg tgacttgcta tagacatcct  900
tccacaggga aagatttttt ctccatactt caacaagctg aacattacat tgaagtactt  960
tcagcaaaca aaaccccagc acaagaaacc attaatgata tgcattttct tatggctaat  1020
ctcatcaaag gtggcatgtt tcagcacaaa ggagactga                         1059
```

```
SEQ ID NO: 27          moltype = DNA  length = 1053
FEATURE                Location/Qualifiers
source                 1..1053
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 27
atggagcttt gcaatatact caagtatgat cgaagtcttt accctgggaa agcggttttc  60
ttctacaaaa cggcggacag cgattttgtt cctttggaag ctgacattaa taagataaga  120
ggaccaaaat caggctttac agaggctttc actcctcaat tctctccgaa aaatatctct  180
ccacaagatc taactcacaa caacattctt acactcgaag agtgttatgt tccaccaaat  240
gtcgaacata ttttctgccg gttttctctg cgtgtccaag ctaattcgtt agtcccctct  300
ggttgttccg atcctgaggt ttttagcctt ctcaaagagc ttgcagaaac cttcaaagaa  360
tgtggtggtt acaaggagct ggcagtacgc tattgtcgga acattttgat tggaacatgg  420
ctttggagaa atcagaatac cggaaacaca cagattgaga tcaagacttc aaaaggatct  480
tgctacttga ttgataatac aagaaaactc gcttgggaaa gcaaatgggc ttcagacgac  540
ttgaaagttt tggaagaatt aagcaatgaa atcgaatctg ctttaacaga tccaaatgtg  600
ttttggtctg cagatataac cgccaagatc gaagcatctt tctgtcaaga gatatatcct  660
tcacagatat tgaatgataa agtgaagcaa ggagaagcgt caaaacagtt tgtgaaagct  720
aaatgtgctg atggaaggta tgctgtgagt ttcaacagtg ttaagatcgg tgctgctttg  780
caaagtattg atgattggtg ggatgaagat gctagtaaaa gacttagagt tcatgaattt  840
ggtgcagaca aggaaattgg cgtagcgagg aggcctcccg attccgaaca aaacttttat  900
tcaatcttca agaacactga atggtatctt tccgccttaa agaattgcat cactaacaag  960
aatgagaaga ttgacccagc aatatactat ttgttcagcg tgttgatcaa aggtggaatg  1020
tttcagaaaa aggcagaggc aaagaaagca tga                               1053
```

```
SEQ ID NO: 28          moltype = DNA  length = 1923
FEATURE                Location/Qualifiers
source                 1..1923
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 28
atgcagactc ttaaagaact aattgcttct aatccggatg acctaacaac ggagctcaaa  60
agggcatttc gccccctgac tcctcacatt gccattgacg gtaacgagct tgacgctctc  120
accattcttg tgaatctcac agacaaaact gatgatcaaa aggatcttct cgatagagct  180
aaatgtaaac agaaactcag agatgagaaa tggtgggctt cttgcatcaa ttgcgttaat  240
tatagacagt ctcacaatcc taaattccct gatatcaggt ctgaaggagt gattcgcact  300
caagctttgg gcgaacttcc ttcatttcta ttaagctcca gtaaaatacc accgtatcat  360
tggtcatatt ctcatgattc aaaatatgtt aacaagtcag cctttttaac aaatgaattc  420
tgttgggacg gagagataag ctgcttgggg gaactcctca aagatgcgga tcatccactt  480
tggaacacac taaagaagct tggctgctca cagaagacat gcaaagctat ggcgaaacaa  540
ttggctgata tcacattaac aacgattaat gtaacactcg ctccgaatta cttgactcag  600
atttctcttc cggattctga tacaagttac atttctctct cgcctgtcgc aagtctatca  660
atgcaaagtc atttccacca acgtttgcaa gacgaaaaca gacattctgc tataacaaga  720
ttctcccgaa caactaacat gggtgtgact gcaatgactt gtggtggagc tttcaggatg  780
ttaaagtctg gtgcaaagtt ctccagtcca cctcatcata gattaaactc gaaacgctct  840
tggttgacca gtgagcatgt tcaatcattg aagcagtatc aacgtctgaa caaatcattg  900
atccctgaga atagtaggat tgctcttagg cgcaagtaca agattgaact tcaaaatatg  960
gttcgatcat ggtttgccat gcaagatcac actctcgatt caaatatcct tattcaacat  1020
ctcaatcacg atttgagtta tctcggggcg accaaaagat ttgcttatga tccagctatg  1080
acaaaattgt tcactgagct gctaaagagg gagctgagta attctatcaa caacggagaa  1140
caacatacca atggcagttt tctggtcctc ccaaacatac gggtttgtgg agcaactgcg  1200
ttatcttccc cagtaaccgt tggcatccct agcttaaccg ccttctttgg ttttgttcac  1260
gcctttgaga ggaacatcaa caggaccact tcttcctttc gtgtcgagag ctttgcaata  1320
tgtgtccatc agcttcatgt ggagaagcgt ggtttgacag ctgagtttgt agagaaggga  1380
gatgggacaa tttcagcacc agcaaccaga gatgattggc aatgtgatgt tgtgttttct  1440
ctaattttga atacgaactt cgctcaacac attgatcagg acacactggt cacgagttta  1500
cctaagcggc tggcaagggg ttctgctaag atcgcaatag atgatttcaa gcatataaac  1560
tcattctcca ctcttgaaac tgcaattgag tctcttccca tcgaagctgg aagatggctt  1620
agtctgtatg cacagagtaa taataatctt tcggacttgt tagctgctat gaccgaagac  1680
catcaactta tggctagttg tgttggctac caccttctag aagaaccaaa agataagcca  1740
aatagcctta gaggatacaa acatgccata gctgaatgta ttataggatt aattaattca  1800
attacatttt cctcggaaac cgatcctaat actatattct ggtccttaaa gaactaccaa  1860
aactatcttg tagttcagcc cagatccata aatgatgaaa caacagacaa gagcagcctg  1920
tga                                                                1923
```

-continued

```
SEQ ID NO: 29          moltype = DNA  length = 2088
FEATURE                Location/Qualifiers
source                 1..2088
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 29
atgcacctga aagagctgtt agagataaca gatacaactg agagagatag atctctccga  60
cgtgccttct cgccgtatac cgccatgatc gacattacag gatcagaagc agtcgctcta  120
attatcctac tcaatctcac ttatagaaag aatcaggtag atgacctgtt ggataagaaa  180
ctcgcgaaac aagccctcaa gagtgaggat catatcaaca aatgcatcaa ggagattgct  240
tggtttcata ctcacaacct taaatatcca gacataaggg tgagtaagca aaatcttgca  300
gtagaaccac caacattgca ttcttatgtg ctttcttctg ctaattatcc caaagcttat  360
ggttggtccc ataattcagc taaagtgaat tttgctaagt tgtttgtctc atacttcaaa  420
tggcaaaacc aagtgagctg gcttgcccaa gtcttggcaa ccaactctga taactggaaa  480
agtgcattta cttcccttgg tttatcggtg aaagccttca aatctttatg cgttacagtt  540
aagaattctt tacctgaaga ggcaatcccg gactcagtgg atagatatag caggcagatt  600
agaatgcctt atcatgatgg ttatcttgct gtcactcctg tgatatctca cgttgttcaa  660
agcaaaatac aacaagctgc tattgacaaa agagctaggt tttctaatgt tgagtttacc  720
agaccagctg ctgtttcaat gctggcggca agtctaggtg gtgtgataaa tgtacttaat  780
tatcctccat acattcgatc aaagtaccat ggactatcga attcaagggc ttttaagtta  840
aacaacggac agactgtctt caacgtggaa gccctcctga agcctgagct tataaaagct  900
cttgaaggaa ttatattctc taacaatgca ttggcattaa aacagagaag gcagcagaaa  960
gtaaaaaaca tcaaagaact gagaaataacc ctccttgagt ggttttcccc tgttttcgaa  1020
tggagattgg atgcaattga gaatggttat gatctagaac aattggaatc agcgagcgaa  1080
cggttagagt acaagattct gagtttgccg gataatgaac taccttcttt gactataccc  1140
ttgtttagat tacttaatga gatgttggga ggagtttcaa tgactcagag gtatgcattc  1200
catccaaagt tgatgtctcc attaaaggct gctctccaat ggttgctcgt taatttaact  1260
gatcaaaagc atgttttaat tgaagaagat gatgagcatt acagatacct tcacctgtca  1320
gggattaggg tttttgatgc ccaagcattg tctaatcctt actgctctgg tattccttcg  1380
cttacggcgg tttggggaat gatccacagt tatcaacgca aactaaacga agctttgggc  1440
acgaacgtac ggttcacaag cttttcatgg ttcatccgta attactctgc cgttgctggt  1500
aaaaagctac cagaacttag ccttcagggt gcacaacagt ctagattgaa aagacctggg  1560
ataatagacg ggaagtattg tgatttggtg tttgacttga ttattcatat cgatggctat  1620
gaagacgatt tgcaggctgt tgattcaaag cctgatatcc tcaaggctca tttcccaagc  1680
aatttcgccg gcggcgtgat gcatcaaccc gagttgaatt ctaatattaa ttggtgttgt  1740
ctttactcca acgaaaacca gctttttgag aagctccgtc ggcttcctct ttctggatgt  1800
tgggttatgc ccacagaaca taagattcaa gatcttgatg aactgttgtt gctgttgaac  1860
tcggatagta aactaagtcc aagcatgatg ggatacatgc ttctcacaga gccaatggct  1920
cgtgtaggat cgttagagag gcttcactgt tatgcagaac ctgctattgg tgttgtcaaa  1980
tacgaggcag caacttctgt tcgattaaaa ggtatcggaa actatttcaa ttccgcgttt  2040
tggatgctcg atgctcaaga aaagttcatg ctcatgaaga aagtgtga               2088

SEQ ID NO: 30          moltype = DNA  length = 1185
FEATURE                Location/Qualifiers
source                 1..1185
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 30
atgtttcttc aaagacctaa accgtatagc gacgaaagtc tggaatcttt cttcatccgt  60
gttgccaaca agaatggtta tggagatgtt cacaggtttt tggaagccac aaagcgattc  120
cttcaagaca ttgatcacaa tggataccaa acatttccca ctgatataac gcgtatcaat  180
ccttattcag ccaaaaactc ttcatcagca agaacagcga gtttcttgaa gctagcacag  240
ctcaccttca acgaaccacc agaattgcta gggcttgcaa taaacagaac caacatgaag  300
tactctcctt ctactagtgc agtggtaaga ggagcagagg ttttcccaag atccctcctg  360
cgtactcatt ccattccatg ttgccctctc tgcttgaggg aaaatgggta tgctagttac  420
ctttggcatt ttcaaggata cgagtattgt cactctcata atgtccctct aattaccact  480
tgttcttgtg gtaaagaatt cgattacaga gtatctgggt taaaaggaat ttgttgtaag  540
tgcaaagagc caataacttt gacatcgaga gagaatggcc atgaagcagc ttgtactgtt  600
agtaattggc ttgccggtca cgagtcaaaa ccattgccta atttacccaa atcttataga  660
tggggactcg tccattggtg gatgggaatc aaggactctg aatttgatca tttcagcttt  720
gtgcagtttt tctcaaactg gcctcggtcg tttcatagta tcatagaaga tgaagttgaa  780
tttaatcttg aacatgctgt tgtttctaca tctgagctaa ggctcaaaga tcttcttggg  840
agactgttct ttggtagcat caggcttcct gagagaaatc tgcaacataa cattattttg  900
ggtgagctgt tatgttatct agaaaaccgt ttgtggcaag ataaaggctt gatagccaat  960
ttaaagatga atgctcttga ggcaacagtg atgcttaact gctccttgga tcagattgct  1020
tcaatggtgg agcaaagaat tttaaagcct aaccgaaaat caaagccaaa ttcaccgctt  1080
gatgtgacgg attatttgtt tcactttggt gacattttct gcttatggtt agctgagttt  1140
cagtctgatg aattcaatcg aagcttttat gtttctaggt ggtga                  1185

SEQ ID NO: 31          moltype = DNA  length = 1229
FEATURE                Location/Qualifiers
source                 1..1229
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 31
```

-continued

```
atggcttttc ttttctcacc aaaagcaaga gcgttttccg atgaatcgct tgaaagttat  60
ttacttcgag ttgtctcaga aaacttcttc gattcatatg aaggactatc tttggccatc  120
agagaagaac ttcatgagtt ggacttcgag gctcatggag catttcccgt agatttaaag  180
agacttaatg tttaccatgc aaaacacaac tcacacttta ggatgagagc tcttggccta  240
ttagaaactc ttttagatct tcctcgttat gagcttcaaa agctagcact cctgaagtca  300
gacattaaat tcaattcttc cgttgccctc tacaacaatg gtgttgatat tcctttgaga  360
ttcattagac atcatgctga ggaagctgtt gactccattc cggtttgttc tcaatgcttg  420
gctgaagaag cttatataaa gcaatcttgg catatcaaat gggttaatgc ttgcaccaaa  480
catcaatgtg ctttgcttca caattgtcct gaatgttatg caccgatcaa ttacattgag  540
aatgaatcaa tcactcattg ctcttgtggt tttgaattga gctgcgcctc tacaagtcct  600
gtgaacacac tctcgattga acatttgaac aaactcctgg acaaaggaga acggaatgac  660
tcaaatccat tgtttaacaa catgacgttg acagaaaggt ttgctgcttt attatggtat  720
caagagaggt attcacagac tgataatttc tgtttgaatg atgctgtgaa ctacttcagt  780
aagtggccag cagtgtttaa tactgagcta gatgagctct caaagaatgc cgagatgaag  840
ctgattgatc tcttcaacaa gacagagttt aagtttatct ttggagatgc aattcttgct  900
tgtccaagta ctcagaagca gagtgaatct catttcatct accgggcatt acttgattac  960
cttgtcactt tagtggagag caatcccaaa acaaagaagc caaacgcagc tgatttgctt  1020
gtttctgtcc tggaagcagc aactttgttg gggacgagtg tagaacaagt ttataggctc  1080
tatcagaatg gaatactgca aaccgccttt cggcacaaga tgaatcagag aataaaccct  1140
tacaagggtg ctttctttct ccgtcatgtg atagagtata aaaccagctt tggtaacgat  1200
aaagctagaa tgtatctctc agcttggtg                                    1229
```

```
SEQ ID NO: 32          moltype = DNA  length = 2586
FEATURE                Location/Qualifiers
source                 1..2586
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 32
atggcgacat cccttccaac gccatcagct ataactactt cggcactaga gtatgctttc  60
catactccgg ctcgcaattt gacgaagtct agggggaaga acattcatcg atatgttagt  120
gtcaagatgt ctaaaagaat tactgtagaa tcaacactag aatgcgatgc ttgctatcat  180
tttgacttcg agccttctat cgtgagattt tgtgctcagc ctattcgctt tctgtattat  240
ttaaatgggc agtcacactc ctatgtccca gattttcttg tgcaattcga cacaaacgag  300
tttgtcttgt atgaagttaa aagcgcttat gcgaaaaaca aacctgattt cgatgttgag  360
tgggaagcaa aggtgaaggc cgcaacagag ctgggtttgg aacttgaact tgttgaggag  420
agcgatataa gagataccgt tgttttgaac aatcttaaga ggatgcatag atatgcgagt  480
aaggacgagc ttaataacgt acacaattcc ctacttaaga tcataaagta taacggagct  540
caatccgccc gatgtttggg tgaacagcta gggttgaaag gtcgtacggt acttcctatt  600
ctgtgtgatc ttctttcaag atgcttgctc gacaccagac ttgataagcc cctctctctg  660
gaaagtcgtt ttgaactcgc ctcatatgga ggatcaggat ctggcaaaag aaccgcagat  720
ggttctgagt tcgaaagccc aaagaaaaag agaaaggtgg gttccggttc tggaggagcc  780
aagaagggtt tttctagctt tcacaggaag gcagtttcct cacaggacac tttagagagt  840
atagaactag tttccagtgc gaattgcctc gaatcagtca cttatcaaga tatatctgca  900
tttcctgaaa ccattgcagt ggagatcaac ttcagattgt ctattctccg gtttctcgca  960
agaaaatgtg agactatcgt tgctaaatct attgagcccc atagagttga gcttcaacag  1020
aattattctc ggaagattcc atctgcaata acaatataca gatggtggtt ggcctttcgg  1080
aagtcagatt acaacccgat tagtctggct cctaatatca aagacagagg gaacagagaa  1140
actaaagtat ctactgtggt tgactccatc atggaacaag ccgttgaaag ggttatctct  1200
ggtcgtaaag tcaatgtttc tagcgcgtac aaacgagtcc ggaggaaagt tagacaatat  1260
aatcttactc atggaacaaa gtacacatat ccaaaatacg agagcgtacg aaagcgcgta  1320
aagaagaaaa caccgtttga gttattagct gctggtaagg gagagagggt cgccaagcgt  1380
gaatttagga gaatgggaaa gaagattctg acgtcaagtg ttctcgagag ggtggagata  1440
gatcacactg ttgtggatct gttcgctgtt catgaggaat ataggattcc tctcgggagg  1500
ccatggctga ctcagttagt tgattgttac agcaaagcag tcattggttt ctaccttggc  1560
tttgaacctc ctagttatgt ctccgtatca ctagctttaa agaatgccat tcaacgcaaa  1620
gatgatttga tttcctctta cgagtcaata gaaaatgaat ggttatgtta tggtatacca  1680
gatctccttg ttacagacaa tggaaaagaa tttttgtcca aagcttttga tcaagcttgt  1740
gagagcttat taattaacgt gcatcagaat aaggtagaaa ccccagataa caagcctcat  1800
gttgaaagaa actacgggac aataaatact tccttattgg acgatttacc tggtaaaagt  1860
ttctctcagt acttgcagag ggaaggatat gactctgtgg gagaagctac attgactctc  1920
aatgagatca gagaaatcta cttgatttgg cttgtcgaca tatatcacaa gaaaccgaat  1980
caaagaggaa ctaactgtcc aaatgttgct tggaaaaaag gttgccaaga atgggagcca  2040
gaagaattct ccgggtcaaa agatgagctt gattttaaat tcgcaattgt tgattataaa  2100
caacttacaa aagtaggaat cactgtttac aaggaactaa gctactcaaa cgatcgtttg  2160
gcagagtata gaggcaaaaa aggaaaccac aaagttcagt tcaagtacaa tccggaatgt  2220
atggctgtga tatgggttct tgatgaagat atgaacgaat atttcacagt taatgctatc  2280
gactatgaat acgcatctag ggtgtcactg tggcaacata aatacaacat gaagtatcaa  2340
gcagagctaa attctgcgga atatgacgaa gataaagaga ttgatgcaga aattaaaatc  2400
gaggaaattg ctgatagatc tatagtgaaa accaataaaa ttagagctcg tcgtcgtggt  2460
gctaggcatc aagagaatag tgcacgagct aaatcaatct ctaatgcaaa tcctgcaagt  2520
attcagaagc atgaagacga aatagtcagt gctgataatg atgattggga tatcgattac  2580
gtgtga                                                             2586
```

```
SEQ ID NO: 33          moltype = DNA  length = 2556
FEATURE                Location/Qualifiers
source                 1..2556
```

-continued

```
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 33
atgtacatta ggaacttgag aaagccttct ccaaataaga atgtttttaa gttcgcatct  60
accaaggtca gctctgtggt catgtgtgaa tccagccttg agtttgacgc ttgctttcat  120
catgaatata acgacttgat tgagagtttt ggttctcagc cagaaggttt caaatacgaa  180
ttcatgggta aatctctccc ttacactcca gatgctttga tctcctacac ggacaaaact  240
caaaagtatc acgagtataa accatattca aaaatcgcat cacctctatt tagagcagag  300
tttgctgcaa agagggctgc aagtcttaag ttgggtattg atttggtact cgtcacagac  360
agacagatta gagtaaaccc aatactcaac aatctgaagc tccttcatag atactcgggc  420
gtttatggta tatctggtat ccagaaggaa ctacttagct ttattcacaa aagtggcgtt  480
atcaaactca atgatatctc atcacaggtt gggatcccca ttggagaaac tcggtcattt  540
ctttttggac tgatgcataa gggacttgta aaggctgacc taggttgtga tgacctgaca  600
aacaacccta ctctttgggc tactcctgga agtggctcag gcaaacgtac tgctgacgga  660
agtgagttcg agtccccaaa gaaaaaaaga aaagttggta gtggatctgg tggaatgact  720
gatttcttca atgagtttga tgaatcgctt gtaccactga aaccacaaac tccaacgcag  780
tatgttaagc tggacgatgc taatttaatt cagagagatt tagatacatt ctcagatact  840
ttcaagaacc aagccctcca aagatataag ctaatttcta caattgataa aaagttaagc  900
cgcggatgga ctcaaagaaa tctcgacccg atcctcgacg aactcttcaa aggcggtgat  960
gtagtgaggc caaactggag aactgtggca agatggagga agaaatatat agaaagcaat  1020
ggtgacatcg cttctcttgc tgataaaaac cacaaaatgg gaaacagaac aaatcggata  1080
aaaggagatg acaaattttt tgacaaggcc ttggaacgtt tccttgacgc gaaaagacct  1140
acaatagcaa cagcttatca atattacaag gacttgatag tgattgaaaa cgagtcaatt  1200
gttgaaggga agatccccat catctcttat aatgcattta acaaacgaat taaagccatt  1260
cctccttatg ctgttgcagt cgctaggcat gggaaattca aagctgatca atggtttgca  1320
tattgtgcag ctcacgttcc accaacgcgt atacttgaac gagtggagat tgatcatact  1380
ccgttggatc taatattatt ggatgatgag cttctgattc ccattggccg gccatatctt  1440
acacttctca tagatgtgtt ttcaggttgc gttctgggat tccatttgtc atacaagagc  1500
ccttcttatg tctctgcagc aaaagcaatc actcacgcaa taaaacccaa aagcttagat  1560
gcgcttaata ttgagcttca gaatgactgg ccctgtttcg ggaagtttga gaatttagta  1620
gttgacaatg gagctgaatt ttggtcaaaa aatttggaac atgcttgtca atctgctggt  1680
ataaatattc aatacaatcc ggtgcgtaag ccatggctca agccttttat agagaggttc  1740
tttggggtga tgaatgaata cttcttgccg gaactcccag gaaaaacatt ttctaacatt  1800
cttgagaaag aagaatacaa accggagaag gatgcaataa tgaggttcag cacctttgtt  1860
gaggagttcc accgttggat tgccgacgtg tatcatcaag atagtaatag ccgtgaaacg  1920
cgtattccta ttaagcgatg gcaacaagga ttcgatgctt atcctccttt aacaatgaac  1980
gaagaagagg agacacggtt ttctatgctc atgagaatct ctgactctag gacgctaacc  2040
aggaatgggt ttaagtacca agaattgatg tatgattcta ccgccttagc agattaccgt  2100
aagcattacc ctcagacaaa agagactgtg aagaagctaa taaaggtcga tccagatgat  2160
atctccaaga tttacgttta tctggaagag ttggaaagtt acttggaggt tccttgcaca  2220
gatccaactg gatacactga tggtttatcc atatatgagc ataaaaccat taagaagatc  2280
aaccgagaag ttattagaga atcaaaagat agccttgggt tggcgaaagc gagaatggct  2340
attcatgaga gagtgaaaca agagcaagaa gtattcatcg aaagtaagac caaagctaag  2400
attactgcag tcaagaagca ggctcagatc gcggatgtta gtaatacagg cacatcaaca  2460
atcaaggtgt ccgaggagtc tgcggctcct gtccagaaac acatctcaaa tgataacagt  2520
gatgattggg atgatgatct ggaagctttt gaatga                            2556

SEQ ID NO: 34           moltype = DNA  length = 993
FEATURE                 Location/Qualifiers
source                  1..993
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 34
atgagtgaaa cacgagaggc taggataagc cgggccaaac gggcttttgt ttctacaccg  60
tctgtacgca agatcctgag ctacatggat agatgtaggg acttatcaga tctcgaatca  120
gagccaactt gtatgatggt ttatggtgcc tctggtgtgg gaaaaaccac ggtcatcaag  180
aagtatttga atcaaaacag aagagaatct gaggcgggtg gtgatattat cccagtgtta  240
cacattgagc tccctgataa tgcaaagcct gttgatgccg cacgggagct gttggttgag  300
atgggggatc cactggccct ctatgaaact gatttagcaa gacttactaa gagattgact  360
gagctgatcc ctgctgttgg ggtgaagctt ataatcattg atgaatttca gcatcttgtg  420
gaagagagat ccaatagagt gttaactcaa gttggaaact ggcttaagat gatcttaaac  480
aaaaccaaat gccctattgt aatatttgga atgccttact caaaggttgt tcttcaagcc  540
aattcgcagc ttcacggaag gttttcaata caagtcgaac tacgtccctt ctcttaccaa  600
ggaggaagag gagttttcaa gacattcttg gagtatcttg ataaagcttt acctttcgag  660
aagcaagcag gcttagcaaa tgaaagtctt cagaaaaaac tgtatgcttt ctctcagggc  720
aacatgagat ccttaaggaa tctcatatat caggcaagta ttgaagctat tgacaaccaa  780
catgaaacca ttacagagga ggattttgtg tttgcttcaa aattgacaag cggagacaaa  840
ccaaattcat ggaaaaatcc ttttgaagaa ggtgttgaag tgacagaaga catgcttcgt  900
ccaccaccca aagatattgg atgggaagat tacctaaggc attctactcc ccgtgtatcc  960
aagccgggca gaaacaagaa tttcttcgaa tga                               993

SEQ ID NO: 35           moltype = DNA  length = 1002
FEATURE                 Location/Qualifiers
source                  1..1002
                        mol_type = other DNA
                        organism = synthetic construct
```

-continued

```
SEQUENCE: 35
atgggaaatg ctcttacaga gattcaaata gagaagttac ggaatttcag tgattgcatc  60
gtgatgcatc ctcagataaa gaccattttc aatgattttg atgaactaag gcttaataga  120
aagtttcaga gtgatcagca atgtatgctt cttataggag atactggtgt tggcaaatca  180
cacaccatca accattacaa gaaaagagtc ctcgcaacac aaaactattc aagaaacaca  240
atgccggtcc ttgtttctcg catatcacgt gggaaaggtt tggatgctac tttggtgcag  300
atgctggcgg atctagagct gtttggttct tctcagatca agaagcgggg ctacaaaaca  360
gaccttacca aaaaattagt tgagagtttg atcaaagctc aagtggagct gttgattatc  420
aatgaattcc aagagcttat tgagtttaaa agtgttcaag aaagacaaca gattgcaaat  480
ggtctaaagt ttattagtga agaagccaag gttcccattg ttttggttgg aatgccatgg  540
gcagcaaaga tagctgagga accacaatgg gcttctcgac tcgtaagaaa acgaaagcta  600
gaatattttt cactcaagaa tgacagcaag tactttaggc agtatttgat gggtcttgcc  660
aaaaagatgc ctttcgacgt tcctcctaaa ttagaatcga agaacactac gattgcgcta  720
tttgcagctt gtagagggga aaacagagct ttgaagcacc ttcttttaga agcactgaag  780
ctcgcccttt catgcaatga gtatcttgag aacaagcatt tcatcacagc ttatgacaaa  840
ttcgatttct tcaacgacaa ggagaaactg aaatccaaaa acccttttcaa gcaagatata  900
aaagatattg aaatttatga agttattaaa aattcttcct acaatccaaa tgcacttgat  960
ccagaagata tgttgactga tcgagtattt gctattgtga aa                     1002

SEQ ID NO: 36          moltype = DNA  length = 624
FEATURE                Location/Qualifiers
source                 1..624
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 36
atgagttatt ctggagaacg ggataatttt gcacctcaca tggcattagt tccaatggtg  60
attgaacaaa caagtagagg ggagcgaagc tttgacattt actctcgtct attaaaagaa  120
agggtcatct tcttaaccgg gcaagttgaa gatcatatgg caaatctgat tgtagctcag  180
atgctttttc ttgaagcaga gaatcctgag aaagatatct atctctacat taactctcct  240
ggtggtgtaa taactgctgg aatgtctata tatgatacaa tgcagtttat caaacctgat  300
gtttcaacaa tatgtatggg tcaggctgcg tcaatgggag cttttctttt gacggctggc  360
gcaaaaggaa agcggttctg cctaccgaac tcaagggtga tgatccatca accacttgga  420
ggataccaag gtcaagctac cgatatagag attcatgcta gagaaatttt gaaggtgaag  480
ggtagaatga atgaactgat ggcccttcac actggacaat cgctcgagca gattgagcgt  540
gacacagaga gagacaggtt tttatccgca ccagaagctg ttgagtatgg tttggttgat  600
tctattctta ctcatagaaa ctga                                          624

SEQ ID NO: 37          moltype = DNA  length = 1275
FEATURE                Location/Qualifiers
source                 1..1275
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 37
atgactgata aacggaaaga tggttccggc aagcttcttt actgcagctt ttgcgggaaa  60
tctcaacatg aagttaggaa acttattgca gggccttctg tttatatctg tgacgaatgt  120
gttgatctct gcaatgatat tattcgagaa gagatcaaag aagtggctcc tcacagagag  180
aggagtgcgt tgccgacacc acacgagatt agaaatcatc ttgatgatta tgtcattggt  240
caagaacaag caaagaaggt ccttgctgtt gcagtgtaca accattacaa gaggctcaga  300
aacggagata cttctaatgg tgtggaactt ggcaaatcaa acatcctgct cattggaccc  360
actggaagtg gaaagacact cctagcagag acattggcaa ggctcttaga cgttcctttc  420
actatggctg atgcaactac tcttactgag gctggttatg ttggagaaga tgtggagaat  480
attattcaaa agttattgca gaaatgtgat tacgatgtcc agaaagctca gagagggata  540
gtctacatag atgagataga caagatctct aggaagagtg ataatccctc gataacacgc  600
gacgtttctg gcgaaggtgt tcaacaagct ttgttgaagc tcattgaagg caccgtagcg  660
gcagttccac cacaaggagg aagaaagcat cctcagcaag agtttctcca agtggatacc  720
tcaaaaatct tatttatatg tggtggtgcc ttcgctgggc ttgacaaagt tatctcacac  780
cgtgtagaaa ctggtagcgg gattgggttt ggtgcaacgg tgaaggcaaa atctgataag  840
gcttctgaag gagagcttct ggctcaagtt gaaccagagg atcttatcaa gtttggacta  900
attcctgagt tcatcggtag acttcctgtc gttgctacgt tgaacgagtt aagtgaggaa  960
gccctgattc aaatcttgaa agagccaaag aatgctctga caaagcaata tcaggctctt  1020
ttcaatcttg agggagttga cttggaattt cgtgatgaag cgctggatgc tatagctaag  1080
aaagcaatgg cgaggaaaac cggagctcgt ggtctaagat cgattgttga agccgcttta  1140
ttggatacaa tgtatgattt gccctccatg gaagatgttg agaaagtggt catcgacgag  1200
tctgtgattg atggacagtc aaagccattg ctcatatatg gtaaacctga agcacagcag  1260
gcatcaggag aatga                                                    1275

SEQ ID NO: 38          moltype = DNA  length = 29807
FEATURE                Location/Qualifiers
source                 1..29807
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 38
tggcaggata tattgtggtg taaacataac gaattgctct tcagcaagga gaatgaggag  60
aatgtagcag aaggcatgtt gttgtgactc cgaggggttg cctcaaactc tatcttataa  120
ccggcgtgga ggcatggagg caggggtatt ttggtcattt taatagatag tggaaaatga  180
cgtggaattt acttaaagac gaagtctttg cgacaagggg gggcccacgc cgaatttaat  240
```

-continued

```
attaccggcg tggccccccc ttatcgcgag tgctttagca cgagcggtcc agatttaaag  300
tagaaaattt cccgcccact agggttaaag gtgttcacac tataaaagca tatacgatgt  360
gatggtattt gatggagcgt atattgtatc aggtatttcc gttggatacg aattattcgt  420
acgaccctcg gtatactaat gaggagtgtt gatacaacca taaaatgata attacaccca  480
taaattgata attatcacac ccataaattg atattgcctc ttcatggtct aaacttcagt  540
aagtttacga cattttcatg aagaggtcat ttcatgacaa gcttacctac gccctcagca  600
attacgactt cggcgttaga gtatgcattc catactcccg ctcgcaatct aacgaaatct  660
cgcggaaaaa atattcatcg ttatgtcagt gtaaagcgct agcaaggagt gagacttttc  720
aacaaagggt aatatcggga aacctcctcg gattccattg cccagctatc tgtcacttca  780
tcaaaaggac agtagaaaag gaaggtggca cctacaaatg ccatcattgc gataaaggaa  840
aggctatcgt tcaagatgcc tctgccgaca gtggtcccaa agatggaccc ccacccacga  900
ggagcatcgt ggaaaaagaa gacgttccaa ccacgtcttc aaagcaagtg gattgatgtg  960
atatctccac tgacgtaagg gatgacgcac aatcccacta tccttcgcaa gacccttcct  1020
ctatataagg aagttcattt catttggaga ggacacgctg acggccgcgt atttttacaa  1080
caattaccaa caacaacaaa caacaaacaa cattacaatt actatttaca attacagcgg  1140
ccgccccctt caccaaaaaa atggctgcag ccaagggcga ggagctgttc accggggtgg  1200
tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg  1260
agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gctgctctgc accaccggca  1320
agctgcccgt gccctggccc accctcgtga ccaccctggg ctacggcgtg cagtgcttcg  1380
cccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg cccgaaggct  1440
acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg  1500
tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcatc gacttcaagg  1560
aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac aacgtctata  1620
tcaccgccga caagcagaag aacggcatca aggccaactt caagatccgc cacaacatcg  1680
aggacggcgg cgtgcagctc gccgaccact accagcagaa cacccccatc ggcgacggcc  1740
ccgtgctgct gcccgacaac cactacctga gctaccagtc caaactgttc aaagacccca  1800
acgagaagcg cgatcacatg gtcttgttag aatttttgac tgctgctggt attaccgagg  1860
gtatgaatga attgtacaaa ggctccggac cgaagaagaa gaggaaggtt taaatatgaa  1920
gatgaagatg aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt  1980
tttcttggct tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat  2040
ctcattataa tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc  2100
ttctccagca tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat  2160
aagcgctata caggagagtg tccgtgactg atgcaaggtt atcggaactg taatgattat  2220
atgttttcta attttgaagt ggttcttttg atgagttttg ccagttactg aaagtaactt  2280
tcttactgca gtagttttgc tgaaatactc gattcacaaa aatatcaact tatggttgtt  2340
ttgtgagata tcaatatatg gttgttttgt ggttaagttg ctgattataa ataattatta  2400
aatatcactt tatggttgca tcaacacgct acagaggagt ctgagcagga caactcgcgt  2460
agtgagagtt acatgttcgt tgggttcttc cgacacggac ctgagttggc caacgtccca  2520
cctgaggtct gtgccccggt gatgagaagt gtgcatctcg ttcttgcagc tcgtcagtac  2580
tttcagaatc atggcgtgca tggtagaatg acccttataa cggacttcga catggcaatc  2640
gctaggtaaa tgaatggaaa aagtggcacc gagtcggtgc tttttttccc ggggcgcgcc  2700
gatatcgagc tctcccggcg cgccgatatc gagctcagtg tttgatcgcc ggcggtaccg  2760
agtgtacttc aagtcagtgg gaaatcaata aaatgattat tttatgaata tatttcattg  2820
tgcaagtaga tagaaattac atatgttaca taacacacga aataaacaaa aaaagacaat  2880
ccaaaaacaa acaccccaaa aaaaataatc actttagata aactcgtatg aggagaggca  2940
cgttcagtga ctcgacgatt cccgagcaaa aaaagtctcc ccgtcacaca tgtagtgggt  3000
gacgcaatta tctttaaagt aatccttctg ttgacttgtc attgataaca tccagtcttc  3060
gtcaggattg caaagaatta tagaagggat cccacctttt attttcttct tttttccata  3120
tttagggttg acagtgaaat cagactggca acctattaat tgcttccaca atgggacgaa  3180
cttgaagggg atgtcgtcga tgatattata ggtggcgtgt tcatcgtagt tggtgaaatc  3240
gatggtaccg ttccaatagt tgtgtcgtcc gagacttcta gcccaggtgg tctttccggt  3300
acgagttggt ccgcagatgt agaggctggg gtgtcggatt ccattccttc cattgtcctt  3360
gttaaatcgg ccatccattc aaggtcagat tgagcttgtt ggtatgagac aggatgtatg  3420
taagtataag cgtctatgct tacatggtat agatgggttt ccctccagga gtgtagatct  3480
tcgtggcagc gaagatctga ttctgtgaag ggcgacacat acggttcagg ttgtggaggg  3540
aataatttgt tggctgaata ttccagccat tgaagctttg ttgcccattc atgagggaat  3600
tcttccttga tcatgtcaag atattcctcc ttagacgttg cagtctggat aatagttctc  3660
catcgtgcgt cagatttgcg aggagaaacc ttatgatctc ggaaatctcc tctggtttta  3720
atatctccgt cctttgatat gtaatcaagg acttgtttag agtttctagc tggctggata  3780
ttagggtgat ttccttcaaa atcgaaaaaa gaaggatccc taatacaagg ttttttatca  3840
agctggagga gagcatgata gtgggtagtg ccatcttgat gaagctcaga agcaacacca  3900
aggaagaaaa taagaaaagg tgtgagtttc tcccagagaa actggaataa atcatctctt  3960
tgagatgagc acttgggata ggtaaggaaa acatatttag attggagtct gaagttctta  4020
ctagcagaag gcatgttgtt gtgactccga ggggttgcct caaactctat cttataaccg  4080
gcgtggaggc atggaggcag ggtatttttg gtcattttaa tagatagtgg aaaatgacgt  4140
ggaatttact taaagacgaa gtctttgcga caaggggggg cccacgccga atttaatatt  4200
accggcgtgg cccccccttа tcgcgagtgc tttagcacga gcggtccaga tttaaagtag  4260
aaaatttccc gcccactagg gttaaaggtg ttcacactat aaaagcatat acgatgtgat  4320
ggtatttgat ggagcgtata ttgtatcagg tatttccgtt ggatacgaat tattcgtacg  4380
accctcatag tttaaactat cagtgggtag cttaatggta gacactttat tagttagatc  4440
cgctgttatt ctgcacgatt gcgcttgccg gtggaaattt aaggacgagt gtagtgcact  4500
cacccttagc cttcagcgtg acaagattag ggtagcaccg gaaacagcga ttttacggac  4560
agatagtaca gtctgagact cagtcactgg tccctaaacg ggagccggta cgtcggcggt  4620
tcgctagcaa ggagtggcag acatactgtc ccacaaatga agatggaatc tgtaaaagaa  4680
aacgcgtgaa ataatgcgtc tgacaaaggt taggtcggct gcctttaatc aataccaaag  4740
tggtccctac cacgatggaa aaactgtgca gtcggtttgg ctttttctga cgaacaaata  4800
```

-continued

```
agattcgtgg ccgacaggtg ggggtccacc atgtgaaggc atcttcagac tccaataatg  4860
gagcaatgac gtaagggctt acgaaataag taagggtagt ttgggaaatg tccactcacc  4920
cgtcagtcta taaatactta gcccctccct cattgttaag ggagcaaaat ctcagagaga  4980
tagtcctaga gagagaaaga gagcaagtag cctagaagta gtcaaggcgg cgaagtattc  5040
aggcacgtgg ccaggaagaa gaaaagccaa gacgacgaaa acaggtaaga gctaagctta  5100
atgggaaagc gtactgctga tggttcagag tttgaaagtc caaagaagaa aagaaaagtt  5160
ggtagtggta tgactgataa acggaaagat ggttccggca agcttcttta ctgcagcttt  5220
tgcgggaaat ctcaacatga agttaggaaa cttattgcag ggccttctgt ttatatctgt  5280
gacgaatgtg ttgatctctg caatgatatt attcgagaag agatcaaaga agtggctcct  5340
cacagagaga ggagtgcgtt gccgacacca cacgagatta gaaatcatct tgatgattat  5400
gtcattggtc aagaacaagc aaagaaggtc cttgctgttg cagtgtacaa ccattacaag  5460
aggctcagaa acggagatac ttctaatggt gtggaacttg gcaaatcaaa catcctgctc  5520
attggaccca ctggaagtgg aaagacactc ctagcagaga cattggcaag gctcttagac  5580
gttcctttca ctatggctga tgcaactact cttactgagg ctggttatgt tggagaagat  5640
gtggagaata ttattcaaaa gttattgcag aaatgtgatt acgatgtcca gaaagctcag  5700
agagggatag tctacataga tgagatagac aagatctcta ggaagagtga taatccctcg  5760
ataacacgcg acgtttctgg cgaaggtgtt caacaagctt tgttgaagct cattgaaggc  5820
accgtagcgg cagttccacc acaaggagga agaaagcatc ctcagcaaga gtttctccaa  5880
gtggatacct caaaaatctt atttatatgt ggtggtgcct tcgctgggct tgacaaagtt  5940
atctcacacc gtgtagaaac tggtagcggg attgggtttg gtgcaacggt gaaggcaaaa  6000
tctgataagg cttctgaagg agagcttctg gctcaagttg aaccagagga tcttatcaag  6060
tttggactaa ttcctgagtt catcggtaga cttcctgtcg ttgctacgtt gaacgagtta  6120
agtgaggaag ccctgattca aatcttgaaa gagccaaaga atgctctgac aaagcaatat  6180
caggctcttt tcaatcttga gggagttgac ttggaatttc gtgatgaagc gctggatgct  6240
atagctaaga aagcaatggc gaggaaaacc ggagctcgtg gtctaagatc gattgttgaa  6300
gccgctttat tggatacaat gtatgatttg ccctccatgg aagatgttga gaaagtggtc  6360
atcgacgagt ctgtgattga tggacagtca aagccattgc tcatatatgg taaacctgaa  6420
gcacagcagg catcaggaga atgaatatga agatgaagat gaaatatttg gtgtgtcaaa  6480
taaaaaggtt gtgtgcttaa gtttgtgttt ttttcttggc ttgttgtgtt atgaatttgt  6540
ggctttttct aatattaaat gaatgtaaca tctcattata atgaataaac aaatgtttct  6600
ataatccatt gtgaatgttt tgttggatct cttctccagc atataactac tgtatgtgct  6660
atggtatgga ctatggaata tgattaaaga taagcgctat acaggagtgg cagacatact  6720
gtcccacaaa tgaagatgga atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa  6780
ggttaggtcg gctgccttta atcaatacca aagtggtccc taccacgatg gaaaaactgt  6840
gcagtcggtt tggctttttc tgacgaacaa ataagattcg tggccgacag gtgggggtcc  6900
accatgtgaa ggcatcttca gactccaata atggagcaat gacgtaaggg cttacgaaat  6960
aagtaagggt agtttgggaa atgtccactc acccgtcagt ctataaatac ttagcccctc  7020
cctcattgtt aagggagcaa aatctcagag agatagtcct agagagagaa agagagcaag  7080
tagcctagaa gtagtcaagg cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc  7140
caagacgacg aaaacaggta agagctaagc ttaatgggaa agcgtactgc tgatggttca  7200
gagtttgaaa gtccaaagaa gaaaagaaaa gttggtagtg gtatgagtta ttctggagaa  7260
cgggataatt ttgcacctca catggcatta gttccaatgg tgattgaaca aacaagtaga  7320
ggggagcgaa gctttgacat ttactctcgt ctattaaaag aaagggtcat cttcttaacc  7380
gggcaagttg aagatcatat ggcaaatctg attgtagctc agatgctttt tcttgaagca  7440
gagaatcctg agaaagatat ctatctctac attaactctc ctggtggtgt aataactgct  7500
ggaatgtcta tatatgatac aatgcagttt atcaaacctg atgtttcaac aatatgtatg  7560
ggtcaggctg cgtcaatggg agcttttctt ttgacggctg gcgcaaaagg aaagcggttc  7620
tgcctaccga actcaagggt gatgatccat caaccacttg gaggatacca aggtcaagct  7680
accgatatag agattcatgc tagagaaatt ttgaaggtga agggtagaat gaatgaactg  7740
atggcccttc acactggaca atcgctcgag cagattgagc gtgacacaga gagagacagg  7800
tttttatccg caccagaagc tgttgagtat ggtttggttg attctattct tactcataga  7860
aactgaatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  7920
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  7980
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  8040
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  8100
tatgattaaa gataagcgct acagaggagc atcttcattc ttaagatatg aagataatct  8160
tcaaaaggcc cctgggaatc tgaaagaaga gaagcaggcc catttatatg ggaaagaaca  8220
atagtatttc ttatataggc ccatttaagt tgaaaacaat cttcaaaagt cccacatcgc  8280
ttagataaga aaacgaagct gagtttatat acagctagag tcgaagtagt gataatgtca  8340
ttactactgc aaagtagctg ataacttcct ctatataagg aagttcattt catttggctt  8400
tactgctgaa taagtagata actacgcttt tttttttgca aaattttcca gatcgatttc  8460
ttcttcctct gttcttcggc gttcaatttc tggggttttc tcttcgtttt ctgtaactga  8520
aacctaaaat ttgacctaaa aaaaatctca aataatatga ttcagtggtt ttgtactttt  8580
cagttagttg agttttgcag ttccgatgag ataaaccaat acgctaggta tactaatgag  8640
gagtggcaga catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa  8700
taatgcgtct gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc  8760
acgatggaaa aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc  8820
cgacaggtgg gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg  8880
taagggctta cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat  8940
aaatacttag cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag  9000
agagaaagag agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc  9060
caggaagaag aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tggcgacatc  9120
ccttccaacg ccatcagcta taactacttc ggcactagag tatgctttcc atactccggc  9180
tcgcaatttg acgaagtcta gggggaagaa cattcatcga tatgttagtg tcaagatgtc  9240
taaaagaatt actgtagaat caacactaga atgcgatgct tgctatcatt ttgacttcga  9300
gccttctatc gtgagatttt gtgctcagcc tattcgcttt ctgtattatt taaatgggca  9360
```

-continued

```
gtcacactcc tatgtcccag attttcttgt gcaattcgac acaaacgagt ttgtcttgta  9420
tgaagttaaa agcgcttatg cgaaaaacaa acctgatttc gatgttgagt gggaagcaaa  9480
ggtgaaggcc gcaacagagc tgggtttgga acttgaactt gttgaggaga gcgatataag  9540
agataccgtt gttttgaaca atcttaagag gatgcataga tatgcgagta aggacgagct  9600
taataacgta cacaattccc tacttaagat cataaagtat aacggagctc aatccgcccg  9660
atgtttgggt gaacagctag ggttgaaagg tcgtacggta cttcctattc tgtgtgatct  9720
tctttcaaga tgcttgctcg acaccagact tgataagccc ctctctctgg aaagtcgttt  9780
tgaactcgcc tcatatggag gatcaggatc tggcaaaaga accgcagatg gttctgagtt  9840
cgaaagccca aagaaaaaga gaaaggtggg ttccggttct ggaggagcca agaagggttt  9900
ttctagcttt cacaggaagg cagtttcctc acaggacact ttagagagta tagaactagt  9960
ttccagtgcg aattgcctcg aatcagtcac ttatcaagat atatctgcat ttcctgaaac 10020
cattgcagtg gagatcaact tcagattgtc tattctccgg tttctcgcaa gaaaatgtga 10080
gactatcgtt gctaaatcta ttgagcccca tagagttgag cttcaacaga attattctcg 10140
gaagattcca tctgcaataa caatatacag atggtggttg gcctttcgga agtcagatta 10200
caacccgatt agtctggctc ctaatatcaa agacagaggg aacagagaaa ctaaagtatc 10260
tactgtggtt gactccatca tggaacaagc cgttgaaagg gttatctctg gtcgtaaagt 10320
caatgtttct agcgcgtaca aacgagtccg gaggaaagtt agacaatata atcttactca 10380
tggaacaaag tacacatatc caaaatacga gagcgtacga aagcgcgtaa agaagaaaac 10440
accgtttgag ttattagctg ctggtaaggg agagagggtc gccaagcgtg aatttaggag 10500
aatgggaaag aagattctga cgtcaagtgt tctcgagagg gtggagatag atcacactgt 10560
tgtggatctg ttcgctgttc atgaggaata taggattcct ctcgggaggc catggctgac 10620
tcagttagtt gattgttaca gcaaagcagt cattggtttc taccttggct ttgaacctcc 10680
tagttatgtc tccgtatcac tagctttaaa gaatgccatt caacgcaaag atgatttgat 10740
ttcctcttac gagtcaatag aaaatgaatg gttatgttat ggtataccag atctccttgt 10800
tacagacaat ggaaaagaat ttttgtccaa agcttttgat caagcttgtg agagcttatt 10860
aattaacgtg catcagaata aggtagaaac cccagataac aagcctcatg ttgaaagaaa 10920
ctacgggaca ataaatactt ccttattgga cgatttacct ggtaaaagtt tctctcagta 10980
cttgcagagg gaaggatatg actctgtggg agaagctaca ttgactctca atgagatcag 11040
agaaatctac ttgatttggc ttgtcgacat atatcacaag aaaccgaatc aaagaggaac 11100
taactgtcca aatgttgctt ggaaaaaagg ttgccaagaa tgggagccag aagaattctc 11160
cgggtcaaaa gatgagcttg attttaaatt cgcaattgtt gattataaac aacttacaaa 11220
agtaggaatc actgtttaca aggaactaag ctactcaaac gatcgtttgg cagagtatag 11280
aggcaaaaaa ggaaaccaca aagttcagtt caagtacaat ccggaatgta tggctgtgat 11340
atgggttctt gatgaagata tgaacgaata tttcacagtt aatgctatcg actatgaata 11400
cgcatctagg gtgtcactgt ggcaacataa atacaacatg aagtatcaag cagagctaaa 11460
ttctgcggaa tatgacgaag ataaagagat tgatgcagaa attaaaatcg aggaaattgc 11520
tgatagatct atagtgaaaa ccaataaaat tagagctcgt cgtcgtggtg ctaggcatca 11580
agagaatagt gcacgagcta aatcaatctc taatgcaaat cctgcaagta ttcagaagca 11640
tgaagacgaa atagtcagtg ctgataatga tgattgggat atcgattacg tgtgagctta 11700
tatgaagatg aagatgaaat atttggtgtg tcaaataaaa aggttgtgtg cttaagtttg 11760
tgtttttttc ttggcttgtt gtgttatgaa tttgtggctt tttctaatat taaatgaatg 11820
taacatctca ttataatgaa taaacaaatg tttctataat ccattgtgaa tgttttgttg 11880
gatctcttct ccagcatata actactgtat gtgctatggt atggactatg gaatatgatt 11940
aaagataagc gctagcaagg agtggcagac atactgtccc acaaatgaag atggaatctg 12000
taaaagaaaa cgcgtgaaat aatgcgtctg acaaaggtta ggtcggctgc ctttaatcaa 12060
taccaaagtg gtccctacca cgatggaaaa actgtgcagt cggtttggct ttttctgacg 12120
aacaaataag attcgtggcc gacaggtggg ggtccaccat gtgaaggcat cttcagactc 12180
caataatgga gcaatgacgt aagggcttac gaaataagta agggtagttt gggaaatgtc 12240
cactcacccg tcagtctata aatacttagc ccctccctca ttgttaaggg agcaaaatct 12300
cagagagata gtcctagaga gagaaagaga gcaagtagcc tagaagtagt caaggcggcg 12360
aagtattcag gcacgtggcc aggaagaaga aaagccaaga cgacgaaaac aggtaagagc 12420
taagcttaat gggaaagcgt actgctgatg gttcagagtt tgaaagtcca aagaagaaaa 12480
gaaaagttgg tagtggtatg agtgaaacac gagaggctag gataagccgg gccaaacggg 12540
cttttgtttc tacaccgtct gtacgcaaga tcctgagcta catggataga tgtagggact 12600
tatcagatct cgaatcagag ccaacttgta tgatggttta tggtgcctct ggtgtgggaa 12660
aaaccacggt catcaagaag tatttgaatc aaaacagaag agaatctgag gcgggtggtg 12720
atattatccc agtgttacac attgagctcc ctgataatgc aaagcctgtt gatgccgcac 12780
gggagctgtt ggttgagatg ggggatccac tggccctcta tgaaactgat ttagcaagac 12840
ttactaagag attgactgag ctgatccctg ctgttggggt gaagcttata atcattgatg 12900
aatttcagca tcttgtggaa gagagatcca atagagtgtt aactcaagtt ggaaactggc 12960
ttaagatgat cttaaacaaa accaaatgcc ctattgtaat atttggaatg ccttactcaa 13020
aggttgttct tcaagccaat tcgcagcttc acggaaggtt ttcaatacaa gtcgaactac 13080
gtcccttctc ttaccaagga ggaagaggag ttttcaagac attcttggag tatcttgata 13140
aagctttacc tttcgagaag caagcaggct tagcaaatga aagtcttcag aaaaaactgt 13200
atgctttctc tcagggcaac atgagatcct taaggaatct catatatcag gcaagtattg 13260
aagctattga caaccaacat gaaaccatta cagaggagga ttttgtgttt gcttcaaaat 13320
tgacaagcgg agacaaacca aattcatgga aaaatccttt tgaagaaggt gttgaagtga 13380
cagaagacat gcttcgtcca ccacccaaag atattggatg ggaagattac ctaaggcatt 13440
ctactccccg tgtatccaag ccgggcagaa acaagaattt cttcgaatga gcttatatga 13500
agatgaagat gaaatatttg gtgtgtcaaa taaaaaggtt gtgtgcttaa gtttgtgttt 13560
ttttcttggc ttgttgtgtt atgaatttgt ggctttttct aatattaaat gaatgtaaca 13620
tctcattata atgaataaac aaatgtttct ataatccatt gtgaatgttt tgttggatct 13680
cttctccagc atataactac tgtatgtgct atggtatgga ctatggaata tgattaaaga 13740
taagcgctat acaggagtgg cagacatact gtcccacaaa tgaagatgga atctgtaaaa 13800
gaaaacgcgt gaaataatgc gtctgacaaa ggttaggtcg gctgccttta atcaatacca 13860
aagtggtccc taccacgatg gaaaaactgt gcagtcggtt tggctttttc tgacgaacaa 13920
```

-continued

```
ataagattcg tggccgacag gtgggggtcc accatgtgaa ggcatcttca gactccaata  13980
atggagcaat gacgtaaggg cttacgaaat aagtaagggt agtttgggaa atgtccactc  14040
acccgtcagt ctataaatac ttagcccctc cctcattgtt aagggagcaa aatctcagag  14100
agatagtcct agagagagaa agagagcaag tagcctagaa gtagtcaagg cggcgaagta  14160
ttcaggcacg tggccaggaa gaagaaaagc caagacgacg aaaacaggta agagctaagc  14220
ttaatgggaa agcgtactgc tgatggttca gagtttgaaa gtccaaagaa gaaaagaaaa  14280
gttggtagtg gtatgtttct tcaaagacct aaaccgtata gcgacgaaag tctggaatct  14340
ttcttcatcc gtgttgccaa caagaatggt tatggagatg ttcacaggtt tttggaagcc  14400
acaaagcgat tccttcaaga cattgatcac aatggatacc aaacatttcc cactgatata  14460
acgcgtatca atccttattc agccaaaaac tcttcatcag caagaacagc gagtttcttg  14520
aagctagcac agctcacctt caacgaacca ccagaattgc tagggcttgc aataaacaga  14580
accaacatga agtactctcc ttctactagt gcagtggtaa gaggagcaga ggttttccca  14640
agatccctcc tgcgtactca ttccattcca tgttgccctc tctgcttgag ggaaaatggg  14700
tatgctagtt acctttggca ttttcaagga tacgagtatt gtcactctca taatgtccct  14760
ctaattacca cttgttcttg tggtaaagaa ttcgattaca gagtatctgg gttaaaagga  14820
atttgttgta agtgcaaaga gccaataact ttgacatcga gagagaatgg ccatgaagca  14880
gcttgtactg ttagtaattg gcttgccggt cacgagtcaa aaccattgcc taatttaccc  14940
aaatcttata gatggggact cgtccattgg tggatgggaa tcaaggactc tgaatttgat  15000
catttcagct ttgtgcagtt tttctcaaac tggcctcggt cgtttcatag tatcatagaa  15060
gatgaagttg aatttaatct tgaacatgct gttgtttcta catctgagct aaggctcaaa  15120
gatcttcttg ggagactgtt ctttggtagc atcaggcttc ctgagagaaa tctgcaacat  15180
aacattattt tgggtgagct gttatgttat ctagaaaacc gtttgtggca agataaaggc  15240
ttgatagcca atttaaagat gaatgctctt gaggcaacag tgatgcttaa ctgctccttg  15300
gatcagattg cttcaatggt ggagcaaaga attttaaagc ctaaccgaaa atcaaagcca  15360
aattcaccgc ttgatgtgac ggattatttg tttcactttg gtgacatttt ctgcttatgg  15420
ttagctgagt ttcagtctga tgaattcaat cgaagctttt atgtttctag gtggtgagct  15480
tatatgaaga tgaagatgaa atatttggtg tgtcaaataa aaaggttgtg tgcttaagtt  15540
tgtgtttttt tcttggcttg ttgtgttatg aatttgtggc tttttctaat attaaatgaa  15600
tgtaacatct cattataatg aataaacaaa tgtttctata atccattgtg aatgttttgt  15660
tggatctctt ctccagcata taactactgt atgtgctatg gtatggacta tggaatatga  15720
ttaaagataa gcgctacaga ggagtctgag caggacaact cgcgtagtga gagttacatg  15780
ttcgttgggt tcttccgaca cggacctgag ttggccaacg tcccacctga ggtctgtgcc  15840
ccggtgatga gaagtgtgca tctcgttctt gcagctcgtc agtactttca gaatcatggc  15900
gtgcatggta gaatgaccct tataacggac ttcgacatgg caatcgctag gtaaatgaat  15960
gaggagtggc agacatactg tcccacaaat gaagatggaa tctgtaaaag aaaacgcgtg  16020
aaataatgcg tctgacaaag gttaggtcgg ctgcctttaa tcaataccaa agtggtccct  16080
accacgatgg aaaaactgtg cagtcggttt ggctttttct gacgaacaaa taagattcgt  16140
ggccgacagg tgggggtcca ccatgtgaag gcatcttcag actccaataa tggagcaatg  16200
acgtaagggc ttacgaaata agtaagggta gtttgggaaa tgtccactca cccgtcagtc  16260
tataaatact tagcccctcc ctcattgtta agggagcaaa atctcagaga gatagtccta  16320
gagagagaaa gagagcaagt agcctagaag tagtcaaggc ggcgaagtat tcaggcacgt  16380
ggccaggaag aagaaaagcc aagacgacga aaacaggtaa gagctaagct taatgggaaa  16440
gcgtactgct gatggttcag agtttgaaag tccaaagaag aaaagaaaag ttggtagtgg  16500
tgtgaaatgg tactacaaaa caataacatt tctacctgag ctctgcaaca atgaatcact  16560
agcagcgaaa tgtctgcgtg tacttcatgg attcaactac caatatgaaa caagaaacat  16620
tggtgttagc ttccctcttt ggtgtgacgc aactgtcgga aagaagatat cctttgtgtc  16680
caagaacaag attgagcttg acttgcttct caaacaacat tactttgttc aaatggagca  16740
gcttcagtat ttccacatct caaatactgt gctggttcca gaagattgca cttatgtcag  16800
tttcaggaga tgtcaaagta ttgataagct tactgctgct ggtttagcaa gaaagatccg  16860
tcgattggag aagagagctt tgtcaagagg agaacagttt gatccctcaa gttttgctca  16920
gaaagaacac accgccattg ctcattatca ttctctcggt gaatcttcaa agcaaacaaa  16980
taggaatttt cggttaaata taaggatgtt gagtgaacaa ccaagagaag gaaattccat  17040
cttctcttct tatgggctat caaattctga gaattctttt cagcctgttc cgctaatctg  17100
agcttatatg aagatgaaga tgaaatattt ggtgtgtcaa ataaaaaggt tgtgtgctta  17160
agtttgtgtt tttttcttgg cttgttgtgt tatgaatttg tggctttttc taatattaaa  17220
tgaatgtaac atctcattat aatgaataaa caaatgtttc tataatccat tgtgaatgtt  17280
ttgttggatc tcttctccag catataacta ctgtatgtgc tatggtatgg actatggaat  17340
atgattaaag ataagcgcta gcaaggagtg gcagacatac tgtcccacaa atgaagatgg  17400
aatctgtaaa agaaaacgcg tgaaataatg cgtctgacaa aggttaggtc ggctgccttt  17460
aatcaatacc aaagtggtcc ctaccacgat ggaaaaactg tgcagtcggt ttggcttttt  17520
ctgacgaaca aataagattc gtggccgaca ggtgggggtc caccatgtga aggcatcttc  17580
agactccaat aatggagcaa tgacgtaagg gcttacgaaa taagtaaggg tagtttggga  17640
aatgtccact cacccgtcag tctataaata cttagcccct ccctcattgt taagggagca  17700
aaatctcaga gagatagtcc tagagagaga aagagagcaa gtagcctaga agtagtcaag  17760
gcggcgaagt attcaggcac gtggccagga agaagaaaag ccaagacgac gaaaacaggt  17820
aagagctaag cttaatggga aagcgtactg ctgatggttc agagtttgaa agtccaaaga  17880
agaaaagaaa agttggtagt ggtatgaaac ttccaacaaa tctagcttac gagaggtcga  17940
ttgatccttc agatgtttgt ttctttgtgg tctggccaga tgatcggaaa acaccgttga  18000
catacaattc acgtaccctc ctagggcaga tggaggctgc atctctggcg tatgatgtaa  18060
gtggtcaacc gatcaagagt gccacagcag aagcattagc tcaaggaaat cctcatcaag  18120
ttgacttttg tcatgttcct tatggagcat ctcacatcga gtgcagcttc tctgtttctt  18180
tttcatcaga gcttcgccag ccttacaaat gtaacagcag taaagtgaag caaactctcg  18240
ttcagctcgt agagttgtat gaaactaaga ttggatggac tgaactagcg acgagatatt  18300
tgatgaacat atgcaatggt aaatggctct ggaaaaatac aaggaaggct tattgttgga  18360
atatagtcct gactccatgg ccttggaacg gagagaaggt tggctttgaa gacattagga  18420
caaactacac ttcaagacaa gatttcaaga acaacaagaa ttggtctgcc attgttgaga  18480
```

-continued

```
tgattaagac agctttcagt tctactgatg gcttagcgat cttcgaggtt cgtgcgactc  18540
ttcacttacc aaccaatgct atggttagac cttcccaggt attcacagag aaagaatcag  18600
ggtcgaaatc aaaaagcaag actcagaatt ctcgagtgtt tcaatcaaca accatagatg  18660
gtgaaagatc tcccattctt ggtgcattca aaactggtgc agctatagct actattgacg  18720
actggtatcc tgaagctaca gaaccattgc gtgttggaag atttggtgtc catagagaag  18780
atgtgacttg ctatagacat ccttccacag ggaaagattt tttctccata cttcaacaag  18840
ctgaacatta cattgaagta ctttcagcaa acaaaacccc agcacaagaa accattaatg  18900
atatgcattt tcttatggct aatctcatca aaggtggcat gtttcagcac aaaggagact  18960
gagcttatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  19020
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  19080
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  19140
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  19200
tatgattaaa gataagcgct atacaggagt ggcagacata ctgtcccaca aatgaagatg  19260
gaatctgtaa aagaaaacgc gtgaaataat gcgtctgaca aaggttaggt cggctgcctt  19320
taatcaatac caaagtggtc cctaccacga tggaaaaact gtgcagtcgg tttggctttt  19380
tctgacgaac aaataagatt cgtggccgac aggtgggggt ccaccatgtg aaggcatctt  19440
cagactccaa taatggagca atgacgtaag ggcttacgaa ataagtaagg gtagtttggg  19500
aaatgtccac tcacccgtca gtctataaat acttagcccc tccctcattg ttaagggagc  19560
aaaatctcag agagatagtc ctagagagag aaagagagca agtagcctag aagtagtcaa  19620
ggcggcgaag tattcaggca cgtggccagg aagaagaaaa gccaagacga cgaaaacagg  19680
taagagctaa gcttaatggg aaagcgtact gctgatggtt cagagtttga aagtccaaag  19740
aagaaaagaa aagttggtag tggtatgcag actcttaaag aactaattgc ttctaatccg  19800
gatgacctaa caacggagct caaaagggca tttcgccccc tgactcctca cattgccatt  19860
gacggtaacg agcttgacgc tctcaccatt cttgtgaatc tcacagacaa aactgatgat  19920
caaaaggatc ttctcgatag agctaaatgt aaacagaaac tcagagatga gaaatggtgg  19980
gcttcttgca tcaattgcgt taattataga cagtctcaca atcctaaatt ccctgatatc  20040
aggtctgaag gagtgattcg cactcaagct ttgggcgaac ttccttcatt tctattaagc  20100
tccagtaaaa taccaccgta tcattggtca tattctcatg attcaaaata tgttaacaag  20160
tcagcctttt taacaaatga attctgttgg gacggagaga taagctgctt gggggaactc  20220
ctcaaagatg cggatcatcc actttggaac acactaaaga agcttggctg ctcacagaag  20280
acatgcaaag ctatggcgaa acaattggct gatatcacat taacaacgat taatgtaaca  20340
ctcgctccga attacttgac tcagatttct cttccggatt ctgatacaag ttacatttct  20400
ctctcgcctg tcgcaagtct atcaatgcaa agtcatttcc accaacgttt gcaagacgaa  20460
aacagacatt ctgctataac aagattctcc cgaacaacta acatgggtgt gactgcaatg  20520
acttgtggtg gagctttcag gatgttaaag tctggtgcaa agttctccag tccacctcat  20580
catagattaa actcgaaacg ctcttggttg accagtgagc atgttcaatc attgaagcag  20640
tatcaacgtc tgaacaaatc attgatccct gagaatagta ggattgctct taggcgcaag  20700
tacaagattg aacttcaaaa tatggttcga tcatggtttg ccatgcaaga tcacactctc  20760
gattcaaata tccttattca acatctcaat cacgatttga gttatctcgg ggcgaccaaa  20820
agatttgctt atgatccagc tatgacaaaa ttgttcactg agctgctaaa gagggagctg  20880
agtaattcta tcaacaacgg agaacaacat accaatggca gttttctggt cctcccaaac  20940
atacgggttt gtggagcaac tgcgttatct tccccagtaa ccgttggcat ccctagctta  21000
accgccttct ttggttttgt tcacgccttt gagaggaaca tcaacaggac cacttcttcc  21060
tttcgtgtcg agagctttgc aatatgtgtc catcagcttc atgtggagaa gcgtggtttg  21120
acagctgagt ttgtagagaa gggagatggg acaatttcag caccagcaac cagagatgat  21180
tggcaatgtg atgttgtgtt ttctctaatt ttgaatacga acttcgctca acacattgat  21240
caggacacac tggtcacgag tttacctaag cggctggcaa ggggttctgc taagatcgca  21300
atagatgatt tcaagcatat aaactcattc tccactcttg aaactgcaat tgagtctctt  21360
cccatcgaag ctggaagatg gcttagtctg tatgcacaga gtaataataa tctttcggac  21420
ttgttagctg ctatgaccga agaccatcaa cttatggcta gttgtgttgg ctaccacctt  21480
ctagaagaac caaaagataa gccaaatagc cttagaggat acaaacatgc catagctgaa  21540
tgtattatag gattaattaa ttcaattaca ttttcctcgg aaaccgatcc taatactata  21600
ttctggtcct taaagaacta ccaaaactat cttgtagttc agcccagatc cataaatgat  21660
gaaacaacag acaagagcag cctgtgagct tatatgaaga tgaagatgaa atatttggtg  21720
tgtcaaataa aaaggttgtg tgcttaagtt tgtgtttttt tcttggcttg ttgtgttatg  21780
aatttgtggc tttttctaat attaaatgaa tgtaacatct cattataatg aataaacaaa  21840
tgtttctata atccattgtg aatgttttgt tggatctctt ctccagcata taactactgt  21900
atgtgctatg gtatggacta tggaatatga ttaaagataa gcgctacaga ggagtctgag  21960
caggacaact cgcgtagtga gagttacatg ttcgttgggt tcttccgaca cggacctgag  22020
ttggccaacg tcccacctga ggtctgtgcc ccggtgatga gaagtgtgca tctcgttctt  22080
gcagctcgtc agtactttca gaatcatggc gtgcatggta gaatgaccct tataacggac  22140
ttcgacatgg caatcgctag gtagcttaat gaataggcgt atcacgaggc cctttcgtct  22200
tcaagaatta gcttggcaat tcccgatcta gtaacataga tgacaccgcg cgcgataatt  22260
tatcctagtt tgcgcgctat attttgtttt ctatcgcgta ttaaatgtat aattgcggga  22320
ctctaatcat aaaaacccat ctcataaata acgtcatgca ttacatgtta attattacat  22380
gcttaacgta attcaacaga aattatatga taatcgtcgc aagaccggca acaggattca  22440
atcttaagaa actttattgc caaatgtttg aacgatcggg atcgaagatc cattccaagc  22500
ctacaggaac aggtggtggc ggccctcggt gcgctcgtac tgctccacga tggtgtagtc  22560
ctcgttgtgg gaggtgatgt ccagcttggc gtccacgtag tagtagccgg gcagctgcac  22620
gggcttcttg gccatgtaga tggacttgaa ctccaccagg tagtggccgc cgtccttcag  22680
cttcagggcc ttgtgtgtct cgcccttcag cacgccgtcg cgggggtaca ggcgctcggt  22740
ggaggcctcc cagcccatcg tcttcttctg catcacgggg ccgtcggagg ggaagttcac  22800
gccgatgaac ttcaccttgt agatgaagca gccgtcctgc agggaggagt cctgggtcac  22860
ggtcgccacg ccgccgtcct cgaagttcat cacgcgctcc cacttgaagc cctcggggaa  22920
ggacagcttc ttgtagtcgg ggatgtcggc ggggtgcttc acgtacacct tggagccgta  22980
ctggaactgg ggggacagga tgtcccaggc gaagggcagg gggccgccct tggtcacctt  23040
```

-continued

```
cagcttcacg gtgttgtggc cctcgtaggg gcggccctcg ccctcgccct cgatctcgaa  23100
ctcgtggccg ttcacggtgc cctccatgcg caccttgaag cgcatgaact cggtgatgac  23160
gttctcggag gaggccattt cgcctcgatc gagttgagag tgaatatgag actctaattg  23220
gataccgagg ggaatttatg gaacgtcagt ggagcatttt tgacaagaaa tatttgctag  23280
ctgatagtga ccttaggcga cttttgaacg cgcaataatg gtttctgacg tatgtgctta  23340
gctcattaaa ctccagaaac ccgcggctca gtggctcctt caacgttgcg gttctgtcag  23400
ttccaaacgt aaaacggctt gtcccgcgtc atcggcgggg gtcataccgt gactccctta  23460
attctcatgt atggttcgct atactgaaga gcaattctga caggatatat tggcgggtaa  23520
acctaagaga aaagagcgtt tattagaata atcggatatt taaaagggcg tgaaaaggtt  23580
tatccgttcg tccatttgta tgtgcatgcc aaccacaggg ttcccctcgg gatcaaagta  23640
ctttgatcca acccctccgc tgctatagtg cagtcggctt ctgacgttca gtgcagccgt  23700
catctgaaaa cgacatgtcg cacaagtcct aagttacgcg acaggctgcc gccctgccct  23760
tttcctggcg ttttcttgtc gcgtgtttta gtcgcataaa gtagaatact tgcgactaga  23820
accggagaca ttacgccatg aacaagagcg ccgccgctgg cctgctgggc tatgcccgcg  23880
tcagcaccga cgaccaggac ttgaccaacc aacgggccga actgcacgcg gccggctgca  23940
ccaagctgtt ttccgagaag atcaccggca ccaggcgcga ccgcccggag ctggccagga  24000
tgcttgacca cctacgccct ggcgacgttg tgacagtgac caggctagac cgcctggccc  24060
gcagcacccg cgacctactg gacattgccg agcgcatcca ggaggccggc gcgggcctgc  24120
gtagcctggc agagccgtgg gccgacacca ccacgccggc cggccgcatg gtgttgaccg  24180
tgttcgccgg cattgccgag ttcgagcgtt ccctaatcat cgaccgcacc cggagcgggc  24240
gcgaggccgc caaggcccga ggcgtgaagt ttggcccccg ccctaccctc accccggcac  24300
agatcgcgca cgcccgcgag ctgatcgacc aggaaggccg caccgtgaaa gaggcggctg  24360
cactgcttgg cgtgcatcgc tcgaccctgt accgcgcact tgagcgcagc gaggaagtga  24420
cgcccaccga ggccaggcgg cgcggtgcct tccgtgagga cgcattgacc gaggccgacg  24480
ccctggcggc cgccgagaat gaacgccaag aggaacaagc atgaaaccgc accaggacgg  24540
ccaggacgaa ccgtttttca ttaccgaaga gatcgaggcg gagatgatcg cggccgggta  24600
cgtgttcgag ccgcccgcgc acctctcaac cgtgcggctg catgaaatcc tggccggttt  24660
gtctgatgcc aagctggcgg cctggccggc cagcttggcc gctgaagaaa ccgagcgccg  24720
ccgtctaaaa aggtgatgtg tatttgagta aaacagcttg cgtcatgcgg tcgctgcgta  24780
tatgatgcga tgagtaaata aacaaatacg caaggggaac gcatgaaggt tatcgctgta  24840
cttaaccaga aaggcgggtc aggcaagacg accatcgcaa cccatctagc ccgcgccctg  24900
caactcgccg gggccgatgt tctgttagtc gattccgatc cccagggcag tgcccgcgat  24960
tgggcggccg tgcgggaaga tcaaccgcta accgttgtcg gcatcgaccg cccgacgatt  25020
gaccgcgacg tgaaggccat cggccggcgc gacttcgtag tgatcgacgg agcgccccag  25080
gcggcggact tggctgtgtc cgcgatcaag gcagccgact tcgtgctgat tccggtgcag  25140
ccaagccctt acgacatatg ggccaccgcc gacctggtgg agctggttaa gcagcgcatt  25200
gaggtcacgg atggaaggct acaagcggcc tttgtcgtgt cgcgggcgat caaaggcacg  25260
cgcatcggcg gtgaggttgc cgaggcgctg gccgggtacg agctgcccat tcttgagtcc  25320
cgtatcacgc agcgcgtgag ctacccaggc actgccgccg ccggcacaac cgttcttgaa  25380
tcagaacccg agggcgacgc tgcccgcgag gtccaggcgc tggccgctga aattaaatca  25440
aaactcattt gagttaatga ggtaaagaga aaatgagcaa aagcacaaac acgctaagtg  25500
ccggccgtcc gagcgcacgc agcagcaagg ctgcaacgtt ggccagcctg gcagacacgc  25560
cagccatgaa gcgggtcaac tttcagttgc cggcggagga tcacaccaag ctgaagatgt  25620
acgcggtacg ccaaggcaag accattaccg agctgctatc tgaatacatc gcgcagctac  25680
cagagtaaat gagcaaatga ataaatgagt agatgaattt tagcggctaa aggaggcggc  25740
atggaaaatc aagaacaacc aggcaccgac gccgtggaat gccccatgtg tggaggaacg  25800
ggcggttggc caggcgtaag cggctgggtt gcctgccggc cctgcaatgg cactggaacc  25860
cccaagcccg aggaatcggc gtgagcggtc gcaaaccatc cggcccggta caaatcggcg  25920
cggcgctggg tgatgacctg gtggagaagt tgaaggccgc gcaggccgcc cagcggcaac  25980
gcatcgaggc agaagcacgc cccggtgaat cgtggcaagc ggccgctgat cgaatccgca  26040
aagaatcccg gcaaccgccg gcagccggtg cgccgtcgat taggaagccg cccaagggcg  26100
acgagcaacc agatttttc gttccgatgc tctatgacgt gggcacccgc gatagtcgca  26160
gcatcatgga cgtggccgtt ttccgtctgt cgaagcgtga ccgacgagct ggcgaggtga  26220
tccgctacga gcttccagac gggcacgtag aggtttccgc agggccggcc ggcatggcca  26280
gtgtgtggga ttacgacctg gtactgatgg cggtttccca tctaaccgaa tccatgaacc  26340
gataccggga agggaaggga gacaagcccg gccgcgtgtt ccgtccacac gttgcggacg  26400
tactcaagtt ctgccggcga gccgatggcg gaaagcagaa agacgacctg gtagaaacct  26460
gcattcggtt aaacaccacg cacgttgcca tgcagcgtac gaagaaggcc aagaacggcc  26520
gcctggtgac ggtatccgag ggtgaagcct tgattagccg ctacaagatc gtaaagagcg  26580
aaaccgggcg gccggagtac atcgagatcg agctagctga ttggatgtac cgcgagatca  26640
cagaaggcaa gaacccggac gtgctgacgg ttcaccccga ttactttttg atcgatcccg  26700
gcatcggccg ttttctctac cgcctggcac gccgcgccgc aggcaaggca gaagccagat  26760
ggttgttcaa gacgatctac gaacgcagtg gcagcgccgg agagttcaag aagttctgtt  26820
tcaccgtgcg caagctgatc gggtcaaatg acctgccgga gtacgatttg aaggaggagg  26880
cggggcaggc tggcccgatc ctagtcatgc gctaccgcaa cctgatcgag ggcgaagcat  26940
ccgccggttc ctaatgtacg gagcagatgc tagggcaaat tgccctagca ggggaaaaag  27000
gtcgaaaagg actctttcct gtggatagca cgtacattgg gaacccaaag ccgtacattg  27060
ggaaccggaa cccgtacatt gggaacccaa agccgtacat tgggaaccgg tcacacatgt  27120
aagtgactga tataaaagag aaaaaaggcg atttttccgc ctaaaactct ttaaaactta  27180
ttaaaactct taaaacccgc ctggcctgtg cataactgtc tggccagcgc acagccgaag  27240
tgctgcaaaa agcgcctacc cttcggtcgc tgcgctccct acgccccgcc gcttcgcgtc  27300
ggcctatcgc ggccgctggc cgctcaaaaa tggctggcct acggccaggc aatctaccag  27360
ggcgcggaca agccgcgccg tcgccactcg accgccggcg cccacatcaa ggcaccctgc  27420
ctcgcgcgtt tcggtgatga cggtgaaaac ctctgacaca tgcagctccc ggtgacggtc  27480
acagcttgtc tgtaagcgga tgccgggagc agacaagccc gtcagggcgc gtcagcgggt  27540
gttggcgggt gtcggggcgc agccatgacc cagtcacgta gcgatagcgg agtgtatact  27600
```

-continued

```
ggcttaacta tgcggcatca gagcagattg tactgagagt gcaccatatg cggtgtgaaa  27660
taccgcacag atgcgtaagg agaaaatacc gcatcaggcg ctcatccgct tcctcgctca  27720
ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg  27780
taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc  27840
agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc  27900
cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac  27960
tataaagata ccaggcgttt ccccctggaa gctccctcgt gcgctctcct gttccgaccc  28020
tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata  28080
gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc  28140
acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca  28200
acccggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag  28260
cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta  28320
gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg  28380
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggtttttt gtttgcaagc  28440
agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt  28500
ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgcat tctaggtgat  28560
tagaaaaact catcgagcat caaatgaaac tgcaatttat tcatatcagg attatcaata  28620
ccatattttt gaaaaagccg tttctgtaat gaaggagaaa actcaccgag gcagttccat  28680
aggatggcaa gatcctggta tcggtctgcg attccgactc gtccaacatc aatacaacct  28740
attaatttcc cctcgtcaaa aataaggtta tcaagtgaga aatcaccatg agtgacgact  28800
gaatccggtg agaatggcaa aagtttatgc atttctttcc agacttgttc aacaggccag  28860
ccattacgct cgtcatcaaa atcactcgca tcaaccaaac cgttattcat tcgtgattgc  28920
gcctgagcga gtcgaaatac gcgatcgctg ttaaaaggac aattacaaac aggaatcgaa  28980
tgcaaccggc gcaggaacac tgccagcgca tcaacaatat tttcacctga atcaggatat  29040
tcttctaata cctggaatgc tgttttccct gggatcgcag tggtgagtaa ccatgcatca  29100
tcaggagtac ggataaaatg cttgatggtc ggaagaggca taaattccgt cagccagttt  29160
agtctgacca tctcatctgt aacatcattg gcaacgctac ctttgccatg tttcagaaac  29220
aactctggcg catcgggctt cccatacaat cggtagattg tcgcacctga ttgcccgaca  29280
ttatcgcgag cccatttata cccatataaa tcagcatcca tgttggaatt taatcgcggc  29340
cttgagcaag acgtttcccg ttgaatatgg ctcataacag aacttattat ttccttcctc  29400
ttttctacag tatttaaaga taccccaaga agctaattat aacaagacga actccaattc  29460
actgttcctt gcattctaaa accttaaata ccagaaaaca gctttttcaa agttgttttc  29520
aaagttggcg tataacatag tatcgacgga gccgattttg aaaccgcggt gatcacaggc  29580
agcaacgctc tgtcatcgtt acaatcaaca tgctaccctc cgcgagatca tccgtgtttc  29640
aaacccggca gcttagttgc cgttcttccg aatagcatcg gtaacatgag caaagtctgc  29700
cgccttacaa cggctctccc gctgacgccg tcccggactg atgggctgcc tgtatcgagt  29760
ggtgattttg tgccgagctg ccggtcgggg agctgttggc tggctgg                29807
```

```
SEQ ID NO: 39          moltype = DNA  length = 29807
FEATURE                Location/Qualifiers
source                 1..29807
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 39
tggcaggata tattgtggtg taaacataac gaattgctct tcagcaagga gaatgaggag  60
aatgtagcag aaggcatgtt gttgtgactc cgaggggttg cctcaaactc tatcttataa  120
ccggcgtgga ggcatggagg caggggtatt ttggtcattt taatagatag tggaaaatga  180
cgtggaattt acttaaagac gaagtctttg cgacaagggg gggcccacgc cgaatttaat  240
attaccggcg tggcccccc ttatcgcgag tgctttagca cgagcggtcc agatttaaag  300
tagaaaattt cccgcccact agggttaaag gtgttcacac tataaaagca tatacgatgt  360
gatggtattt gatggagcgt atattgtatc aggtatttcc gttggatacg aattattcgt  420
acgaccctcg gtatactaat gaggagtgtt gatacaacca taaaatgata attacaccca  480
taaattgata attatcacac ccataaattg atattgcctc ttcatggtct aaacttcagt  540
aagtttacga cattttcatg aagaggtcat ttcatgacaa gcttacctac gccctcagca  600
attacgactt cggcgttaga gtatgcattc catactcccg ctcgcaatct aacgaaatct  660
cgcggaaaaa atattcatcg ttatgtcagt gtaaagcgct agcaaggagt gagacttttc  720
aacaaagggt aatatcggga aacctcctcg gattccattg cccagctatc tgtcacttca  780
tcaaaaggac agtagaaaag gaaggtggca cctacaaatg ccatcattgc gataaaggaa  840
aggctatcgt tcaagatgcc tctgccgaca gtggtcccaa agatggaccc ccacccacga  900
ggagcatcgt ggaaaaagaa gacgttccaa ccacgtcttc aaagcaagtg gattgatgtg  960
atatctccac tgacgtaagg gatgacgcac aatcccacta tccttcgcaa gacccttcct  1020
ctatataagg aagttcattt catttggaga ggacacgctg acggccgcgt atttttacaa  1080
caattaccaa caacaacaaa caacaaacaa cattacaatt actatttaca attacagcgg  1140
ccgccccctt caccaaaaaa atggctgcag ccaagggcga ggagctgttc accggggtgg  1200
tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg  1260
agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gctgctctgc accaccggca  1320
agctgcccgt gccctggccc accctcgtga ccaccctggg ctacggcgtg cagtgcttcg  1380
cccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg cccgaaggct  1440
acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg  1500
tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcatc gacttcaagg  1560
aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac aacgtctata  1620
tcaccgccga caagcagaag aacggcatca aggccaactt caagatccgc cacaacatcg  1680
aggacggcgg cgtgcagctc gccgaccact accagcagaa cacccccatc ggcgacggcc  1740
ccgtgctgct gcccgacaac cactacctga gctaccagtc caaactgttc aaagacccca  1800
acgagaagcg cgatcacatg gtcttgttag aatttttgac tgctgctggt attaccgagg  1860
gtatgaatga attgtacaaa ggctccggac cgaagaagaa gaggaaggtt taaatatgaa  1920
```

-continued

```
gatgaagatg aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt  1980
tttcttggct tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat  2040
ctcattataa tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc  2100
ttctccagca tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat  2160
aagcgctata caggagagtg tccgtgactg atgcaaggtt atcggaactg taatgattat  2220
atgttttcta attttgaagt ggttcttttg atgagttttg ccagttactg aaagtaactt  2280
tcttactgca gtagttttgc tgaaatactc gattcacaaa aatatcaact tatggttgtt  2340
ttgtgagata tcaatatatg gttgttttgt ggttaagttg ctgattataa ataattatta  2400
aatatcactt tatggttgca tcaacacgct acagaggagt ctgagcagga caactcgcgt  2460
agtgagagtt acatgttcgt tgggttcttc cgacacggac ctgagttggc caacgtccca  2520
cctgaggtct gtgccccggt gatgagaagt gtgcatctcg ttcttgcagc tcgtcagtac  2580
tttcagaatc atggcgtgca tggtagaatg acccttataa cggacttcga catggcaatc  2640
gctaggtaaa tgaatggaaa aagtggcacc gagtcggtgc tttttttccc ggggcgcgcc  2700
gatatcgagc tctcccggcg cgccgatatc gagctcagtg tttgatcgcc ggcggtaccg  2760
agtgtacttc aagtcagtgg gaaatcaata aaatgattat tttatgaata tatttcattg  2820
tgcaagtaga tagaaattac atatgttaca taacacacga aataaacaaa aaaagacaat  2880
ccaaaaacaa acaccccaaa aaaaataatc actttagata aactcgtatg aggagaggca  2940
cgttcagtga ctcgacgatt cccgagcaaa aaaagtctcc ccgtcacaca tgtagtgggt  3000
gacgcaatta tctttaaagt aatccttctg ttgacttgtc attgataaca tccagtcttc  3060
gtcaggattg caaagaatta tagaagggat cccacctttt attttcttct tttttccata  3120
tttagggttg acagtgaaat cagactggca acctattaat tgcttccaca atgggacgaa  3180
cttgaagggg atgtcgtcga tgatattata ggtggcgtgt tcatcgtagt tggtgaaatc  3240
gatggtaccg ttccaatagt tgtgtcgtcc gagacttcta gcccaggtgg tctttccggt  3300
acgagttggt ccgcagatgt agaggctggg gtgtcggatt ccattccttc cattgtcctt  3360
gttaaatcgg ccatccattc aaggtcagat tgagcttgtt ggtatgagac aggatgtatg  3420
taagtataag cgtctatgct tacatggtat agatgggttt ccctccagga gtgtagatct  3480
tcgtggcagc gaagatctga ttctgtgaag ggcgacacat acggttcagg ttgtggaggg  3540
aataatttgt tggctgaata ttccagccat tgaagctttg ttgcccattc atgagggaat  3600
tcttccttga tcatgtcaag atattcctcc ttagacgttg cagtctggat aatagttctc  3660
catcgtgcgt cagatttgcg aggagaaacc ttatgatctc ggaaatctcc tctggtttta  3720
atatctccgt cctttgatat gtaatcaagg acttgtttag agtttctagc tggctggata  3780
ttagggtgat ttccttcaaa atcgaaaaaa gaaggatccc taatacaagg ttttttatca  3840
agctggagga gagcatgata gtgggtagtg ccatcttgat gaagctcaga agcaacacca  3900
aggaagaaaa taagaaaagg tgtgagtttc tcccagagaa actggaataa atcatctctt  3960
tgagatgagc acttgggata ggtaaggaaa acatatttag attggagtct gaagttctta  4020
ctagcagaag gcatgttgtt gtgactccga ggggttgcct caaactctat cttataaccg  4080
gcgtggaggc atggaggcag gggtattttg gtcattttaa tagatagtgg aaaatgacgt  4140
ggaatttact taaagacgaa gtctttgcga caaggggggg cccacgccga atttaatatt  4200
accggcgtgg cccccccttta tcgcgagtgc tttagcacga gcggtccaga tttaaagtag  4260
aaaatttccc gcccactagg gttaaaggtg ttcacactat aaaagcatat acgatgtgat  4320
ggtatttgat ggagcgtata ttgtatcagg tatttccgtt ggatacgaat tattcgtacg  4380
accctcatag tttaaactat cagtgggtag cttaatggta gacactttat tagttagatc  4440
cgctgttatt ctgcacgatt gcgcttgccg gtggaaattt aaggacgagt gtagtgcact  4500
cacccttagc cttcagcgtg acaagattag ggtagcaccg gaaacagcga ttttacggac  4560
agatagtaca gtctgagact cagtcactgg tccctaaacg ggagccggta cgtcggcggt  4620
tcgctagcaa ggagtggcag acatactgtc ccacaaatga agatggaatc tgtaaaagaa  4680
aacgcgtgaa ataatgcgtc tgacaaaggt taggtcggct gcctttaatc aataccaaag  4740
tggtccctac cacgatggaa aaactgtgca gtcggtttgg ctttttctga cgaacaaata  4800
agattcgtgg ccgacaggtg ggggtccacc atgtgaaggc atcttcagac tccaataatg  4860
gagcaatgac gtaagggctt acgaaataag taagggtagt ttgggaaatg tccactcacc  4920
cgtcagtcta taaatactta gcccctccct cattgttaag ggagcaaaat ctcagagaga  4980
tagtcctaga gagagaaaga gagcaagtag cctagaagta gtcaaggcgg cgaagtattc  5040
aggcacgtgg ccaggaagaa gaaaagccaa gacgacgaaa acaggtaaga gctaagctta  5100
atgggaaagc gtactgctga tggttcagag tttgaaagtc caaagaagaa aagaaaagtt  5160
ggtagtggta tgactgataa acggaaagat ggttccggca agcttcttta ctgcagcttt  5220
tgcgggaaat ctcaacatga agttaggaaa cttattgcag ggccttctgt ttatatctgt  5280
gacgaatgtg ttgatctctg caatgatatt attcgagaag agatcaaaga agtggctcct  5340
cacagagaga ggagtgcgtt gccgacacca cacgagatta gaaatcatct tgatgattat  5400
gtcattggtc aagaacaagc aaagaaggtc cttgctgttg cagtgtacaa ccattacaag  5460
aggctcagaa acggagatac ttctaatggt gtggaacttg gcaaatcaaa catcctgctc  5520
attggaccca ctggaagtgg aaagacactc ctagcagaga cattggcaag gctcttagac  5580
gttcctttca ctatggctga tgcaactact cttactgagg ctggttatgt tggagaagat  5640
gtggagaata ttattcaaaa gttattgcag aaatgtgatt acgatgtcca gaaagctcag  5700
agagggatag tctacataga tgagatagac aagatctcta ggaagagtga taatccctcg  5760
ataacacgcg acgtttctgg cgaaggtgtt caacaagctt tgttgaagct cattgaaggc  5820
accgtagcgg cagttccacc acaaggagga agaaagcatc ctcagcaaga gtttctccaa  5880
gtggatacct caaaaatctt atttatatgt ggtggtgcct tcgctgggct tgacaaagtt  5940
atctcacacc gtgtagaaac tggtagcggg attgggtttg gtgcaacggt gaaggcaaaa  6000
tctgataagg cttctgaagg agagcttctg gctcaagttg aaccagagga tcttatcaag  6060
tttggactaa ttcctgagtt catcggtaga cttcctgtcg ttgctacgtt gaacgagtta  6120
agtgaggaag ccctgattca aatcttgaaa gagccaaaga atgctctgac aaagcaatat  6180
caggctcttt tcaatcttga gggagttgac ttggaatttc gtgatgaagc gctggatgct  6240
atagctaaga aagcaatggc gaggaaaacc ggagctcgtg gtctaagatc gattgttgaa  6300
gccgctttat tggatacaat gtatgatttg ccctccatgg aagatgttga gaaagtggtc  6360
atcgacgagt ctgtgattga tggacagtca aagccattgc tcatatatgg taaacctgaa  6420
gcacagcagg catcaggaga atgaatatga agatgaagat gaaatatttg gtgtgtcaaa  6480
```

-continued

```
taaaaaggtt gtgtgcttaa gtttgtgttt ttttcttggc ttgttgtgtt atgaatttgt  6540
ggctttttct aatattaaat gaatgtaaca tctcattata atgaataaac aaatgtttct  6600
ataatccatt gtgaatgttt tgttggatct cttctccagc atataactac tgtatgtgct  6660
atggtatgga ctatggaata tgattaaaga taagcgctat acaggagtgg cagacatact  6720
gtcccacaaa tgaagatgga atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa  6780
ggttaggtcg gctgccttta atcaatacca aagtggtccc taccacgatg gaaaaactgt  6840
gcagtcggtt tggctttttc tgacgaacaa ataagattcg tggccgacag gtgggggtcc  6900
accatgtgaa ggcatcttca gactccaata atggagcaat gacgtaaggg cttacgaaat  6960
aagtaagggt agtttgggaa atgtccactc acccgtcagt ctataaatac ttagcccctc  7020
cctcattgtt aagggagcaa aatctcagag agatagtcct agagagagaa agagagcaag  7080
tagcctagaa gtagtcaagg cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc  7140
caagacgacg aaaacaggta agagctaagc ttaatgggaa agcgtactgc tgatggttca  7200
gagtttgaaa gtccaaagaa gaaaagaaaa gttggtagtg gtatgagtta ttctggagaa  7260
cgggataatt ttgcacctca catggcatta gttccaatgg tgattgaaca aacaagtaga  7320
ggggagcgaa gctttgacat ttactctcgt ctattaaaag aaagggtcat cttcttaacc  7380
gggcaagttg aagatcatat ggcaaatctg attgtagctc agatgctttt tcttgaagca  7440
gagaatcctg agaaagatat ctatctctac attaactctc ctggtggtgt aataactgct  7500
ggaatgtcta tatatgatac aatgcagttt atcaaacctg atgtttcaac aatatgtatg  7560
ggtcaggctg cgtcaatggg agcttttctt ttgacggctg gcgcaaaagg aaagcggttc  7620
tgcctaccga actcaagggt gatgatccat caaccacttg gaggatacca aggtcaagct  7680
accgatatag agattcatgc tagagaaatt ttgaaggtga agggtagaat gaatgaactg  7740
atggcccttc acactggaca atcgctcgag cagattgagc gtgacacaga gagagacagg  7800
tttttatccg caccagaagc tgttgagtat ggtttggttg attctattct tactcataga  7860
aactgaatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  7920
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  7980
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  8040
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  8100
tatgattaaa gataagcgct acagaggagc atcttcattc ttaagatatg aagataatct  8160
tcaaaaggcc cctgggaatc tgaaagaaga gaagcaggcc catttatatg ggaaagaaca  8220
atagtatttc ttatataggc ccatttaagt tgaaaacaat cttcaaaagt cccacatcgc  8280
ttagataaga aaacgaagct gagtttatat acagctagag tcgaagtagt gataatgtca  8340
ttactactgc aaagtagctg ataacactga cgtaagggat gacgcacaat cccactactt  8400
tactgctgaa taagtagata actacgcttt tttttttgca aaattttcca gatcgatttc  8460
ttcttcctct gttcttcggc gttcaatttc tggggttttc tcttcgtttt ctgtaactga  8520
aacctaaaat ttgacctaaa aaaaatctca aataatatga ttcagtggtt ttgtactttt  8580
cagttagttg agttttgcag ttccgatgag ataaaccaat acgctaggta tactaatgag  8640
gagtggcaga catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa  8700
taatgcgtct gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc  8760
acgatggaaa aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc  8820
cgacaggtgg gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg  8880
taagggctta cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat  8940
aaatacttag cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag  9000
agagaaagag agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc  9060
caggaagaag aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tggcgacatc  9120
ccttccaacg ccatcagcta taactacttc ggcactagag tatgctttcc atactccggc  9180
tcgcaatttg acgaagtcta gggggaagaa cattcatcga tatgttagtg tcaagatgtc  9240
taaaagaatt actgtagaat caacactaga atgcgatgct tgctatcatt ttgacttcga  9300
gccttctatc gtgagatttt gtgctcagcc tattcgcttt ctgtattatt taaatgggca  9360
gtcacactcc tatgtcccag attttcttgt gcaattcgac acaaacgagt ttgtcttgta  9420
tgaagttaaa agcgcttatg cgaaaaacaa acctgatttc gatgttgagt gggaagcaaa  9480
ggtgaaggcc gcaacagagc tgggtttgga acttgaactt gttgaggaga gcgatataag  9540
agataccgtt gttttgaaca atcttaagag gatgcataga tatgcgagta aggacgagct  9600
taataacgta cacaattccc tacttaagat cataaagtat aacggagctc aatccgcccg  9660
atgtttgggt gaacagctag ggttgaaagg tcgtacggta cttcctattc tgtgtgatct  9720
tctttcaaga tgcttgctcg acaccagact tgataagccc ctctctctgg aaagtcgttt  9780
tgaactcgcc tcatatggag gatcaggatc tggcaaaaga accgcagatg gttctgagtt  9840
cgaaagccca aagaaaaaga gaaaggtggg ttccggttct ggaggagcca agaagggttt  9900
ttctagcttt cacaggaagg cagtttcctc acaggacact ttagagagta tagaactagt  9960
ttccagtgcg aattgcctcg aatcagtcac ttatcaagat atatctgcat ttcctgaaac  10020
cattgcagtg gagatcaact tcagattgtc tattctccgg tttctcgcaa gaaaatgtga  10080
gactatcgtt gctaaatcta ttgagcccca tagagttgag cttcaacaga attattctcg  10140
gaagattcca tctgcaataa caatatacag atggtggttg gcctttcgga agtcagatta  10200
caacccgatt agtctggctc ctaatatcaa agacagaggg aacagagaaa ctaaagtatc  10260
tactgtggtt gactccatca tggaacaagc cgttgaaagg gttatctctg gtcgtaaagt  10320
caatgtttct agcgcgtaca aacgagtccg gaggaaagtt agacaatata atcttactca  10380
tggaacaaag tacacatatc caaaatacga gagcgtacga aagcgcgtaa agaagaaaac  10440
accgtttgag ttattagctg ctggtaaggg agagagggtc gccaagcgtg aatttaggag  10500
aatgggaaag aagattctga cgtcaagtgt tctcgagagg gtggagatag atcacactgt  10560
tgtggatctg ttcgctgttc atgaggaata taggattcct ctcgggaggc catggctgac  10620
tcagttagtt gattgttaca gcaaagcagt cattggtttc taccttggct ttgaacctcc  10680
tagttatgtc tccgtatcac tagctttaaa gaatgccatt caacgcaaag atgatttgat  10740
ttcctcttac gagtcaatag aaaatgaatg gttatgttat ggtataccag atctccttgt  10800
tacagacaat ggaaaagaat ttttgtccaa agcttttgat caagcttgtg agagcttatt  10860
aattaacgtg catcagaata aggtagaaac cccagataac aagcctcatg ttgaaagaaa  10920
ctacgggaca ataaatactt ccttattgga cgatttacct ggtaaaagtt tctctcagta  10980
cttgcagagg gaaggatatg actctgtggg agaagctaca ttgactctca atgagatcag  11040
```

-continued

```
agaaatctac ttgatttggc ttgtcgacat atatcacaag aaaccgaatc aaagaggaac  11100
taactgtcca aatgttgctt ggaaaaaagg ttgccaagaa tgggagccag aagaattctc  11160
cgggtcaaaa gatgagcttg attttaaatt cgcaattgtt gattataaac aacttacaaa  11220
agtaggaatc actgtttaca aggaactaag ctactcaaac gatcgtttgg cagagtatag  11280
aggcaaaaaa ggaaaccaca aagttcagtt caagtacaat ccggaatgta tggctgtgat  11340
atgggttctt gatgaagata tgaacgaata tttcacagtt aatgctatcg actatgaata  11400
cgcatctagg gtgtcactgt ggcaacataa atacaacatg aagtatcaag cagagctaaa  11460
ttctgcggaa tatgacgaag ataaagagat tgatgcagaa attaaaatcg aggaaattgc  11520
tgatagatct atagtgaaaa ccaataaaat tagagctcgt cgtcgtggtg ctaggcatca  11580
agagaatagt gcacgagcta aatcaatctc taatgcaaat cctgcaagta ttcagaagca  11640
tgaagacgaa atagtcagtg ctgataatga tgattgggat atcgattacg tgtgagctta  11700
tatgaagatg aagatgaaat atttggtgtg tcaaataaaa aggttgtgtg cttaagtttg  11760
tgtttttttc ttggcttgtt gtgttatgaa tttgtggctt tttctaatat taaatgaatg  11820
taacatctca ttataatgaa taaacaaatg tttctataat ccattgtgaa tgttttgttg  11880
gatctcttct ccagcatata actactgtat gtgctatggt atggactatg gaatatgatt  11940
aaagataagc gctagcaagg agtggcagac atactgtccc acaaatgaag atggaatctg  12000
taaaagaaaa cgcgtgaaat aatgcgtctg acaaaggtta ggtcggctgc ctttaatcaa  12060
taccaaagtg gtccctacca cgatggaaaa actgtgcagt cggtttggct ttttctgacg  12120
aacaaataag attcgtggcc gacaggtggg ggtccaccat gtgaaggcat cttcagactc  12180
caataatgga gcaatgacgt aagggcttac gaaataagta agggtagttt gggaaatgtc  12240
cactcacccg tcagtctata aatacttagc ccctccctca ttgttaaggg agcaaaatct  12300
cagagagata gtcctagaga gagaaagaga gcaagtagcc tagaagtagt caaggcggcg  12360
aagtattcag gcacgtggcc aggaagaaga aaagccaaga cgacgaaaac aggtaagagc  12420
taagcttaat gggaaagcgt actgctgatg gttcagagtt tgaaagtcca aagaagaaaa  12480
gaaaagttgg tagtggtatg agtgaaacac gagaggctag gataagccgg gccaaacggg  12540
cttttgtttc tacaccgtct gtacgcaaga tcctgagcta catggataga tgtagggact  12600
tatcagatct cgaatcagag ccaacttgta tgatggttta tggtgcctct ggtgtgggaa  12660
aaaccacggt catcaagaag tatttgaatc aaaacagaag agaatctgag gcgggtggtg  12720
atattatccc agtgttacac attgagctcc ctgataatgc aaagcctgtt gatgccgcac  12780
gggagctgtt ggttgagatg ggggatccac tggccctcta tgaaactgat ttagcaagac  12840
ttactaagag attgactgag ctgatccctg ctgttggggt gaagcttata atcattgatg  12900
aatttcagca tcttgtggaa gagagatcca atagagtgtt aactcaagtt ggaaactggc  12960
ttaagatgat cttaaacaaa accaaatgcc ctattgtaat atttggaatg ccttactcaa  13020
aggttgttct tcaagccaat tcgcagcttc acggaaggtt ttcaatacaa gtcgaactac  13080
gtcccttctc ttaccaagga ggaagaggag ttttcaagac attcttggag tatcttgata  13140
aagctttacc tttcgagaag caagcaggct tagcaaatga aagtcttcag aaaaaactgt  13200
atgctttctc tcagggcaac atgagatcct taaggaatct catatatcag gcaagtattg  13260
aagctattga caaccaacat gaaaccatta cagaggagga ttttgtgttt gcttcaaaat  13320
tgacaagcgg agacaaacca aattcatgga aaaatccttt tgaagaaggt gttgaagtga  13380
cagaagacat gcttcgtcca ccacccaaag atattggatg ggaagattac ctaaggcatt  13440
ctactccccg tgtatccaag ccgggcagaa acaagaattt cttcgaatga gcttatatga  13500
agatgaagat gaaatatttg gtgtgtcaaa taaaaaggtt gtgtgcttaa gtttgtgttt  13560
ttttcttggc ttgttgtgtt atgaatttgt ggctttttct aatattaaat gaatgtaaca  13620
tctcattata atgaataaac aaatgtttct ataatccatt gtgaatgttt tgttggatct  13680
cttctccagc atataactac tgtatgtgct atggtatgga ctatggaata tgattaaaga  13740
taagcgctat acaggagtgg cagacatact gtcccacaaa tgaagatgga atctgtaaaa  13800
gaaaacgcgt gaaataatgc gtctgacaaa ggttaggtcg gctgccttta atcaatacca  13860
aagtggtccc taccacgatg gaaaaactgt gcagtcggtt tggctttttc tgacgaacaa  13920
ataagattcg tggccgacag gtgggggtcc accatgtgaa ggcatcttca gactccaata  13980
atggagcaat gacgtaaggg cttacgaaat aagtaagggt agtttgggaa atgtccactc  14040
acccgtcagt ctataaatac ttagcccctc cctcattgtt aagggagcaa aatctcagag  14100
agatagtcct agagagagaa agagagcaag tagcctagaa gtagtcaagg cggcgaagta  14160
ttcaggcacg tggccaggaa gaagaaaagc caagacgacg aaaacaggta agagctaagc  14220
ttaatgggaa agcgtactgc tgatggttca gagtttgaaa gtccaaagaa gaaaagaaaa  14280
gttggtagtg gtatgtttct tcaaagacct aaaccgtata gcgacgaaag tctggaatct  14340
ttcttcatcc gtgttgccaa caagaatggt tatggagatg ttcacaggtt tttggaagcc  14400
acaaagcgat tccttcaaga cattgatcac aatggatacc aaacatttcc cactgatata  14460
acgcgtatca atccttattc agccaaaaac tcttcatcag caagaacagc gagtttcttg  14520
aagctagcac agctcacctt caacgaacca ccagaattgc tagggcttgc aataaacaga  14580
accaacatga agtactctcc ttctactagt gcagtggtaa gaggagcaga ggttttccca  14640
agatccctcc tgcgtactca ttccattcca tgttgccctc tctgcttgag ggaaaatggg  14700
tatgctagtt acctttggca ttttcaagga tacgagtatt gtcactctca taatgtccct  14760
ctaattacca cttgttcttg tggtaaagaa ttcgattaca gagtatctgg gttaaaagga  14820
atttgttgta agtgcaaaga gccaataact ttgacatcga gagagaatgg ccatgaagca  14880
gcttgtactg ttagtaattg gcttgccggt cacgagtcaa aaccattgcc taatttaccc  14940
aaatcttata gatggggact cgtccattgg tggatgggaa tcaaggactc tgaatttgat  15000
catttcagct ttgtgcagtt tttctcaaac tggcctcggt cgtttcatag tatcatagaa  15060
gatgaagttg aatttaatct tgaacatgct gttgtttcta catctgagct aaggctcaaa  15120
gatcttcttg ggagactgtt ctttggtagc atcaggcttc ctgagagaaa tctgcaacat  15180
aacattattt tgggtgagct gttatgttat ctagaaaacc gtttgtggca agataaaggc  15240
ttgatagcca atttaaagat gaatgctctt gaggcaacag tgatgcttaa ctgctccttg  15300
gatcagattg cttcaatggt ggagcaaaga attttaaagc ctaaccgaaa atcaaagcca  15360
aattcaccgc ttgatgtgac ggattatttg tttcactttg gtgacatttt ctgcttatgg  15420
ttagctgagt ttcagtctga tgaattcaat cgaagctttt atgtttctag gtggtgagct  15480
tatatgaaga tgaagatgaa atatttggtg tgtcaaataa aaaggttgtg tgcttaagtt  15540
tgtgtttttt tcttggcttg ttgtgttatg aatttgtggc tttttctaat attaaatgaa  15600
```

-continued

```
tgtaacatct cattataatg aataaacaaa tgtttctata atccattgtg aatgttttgt  15660
tggatctctt ctccagcata taactactgt atgtgctatg gtatggacta tggaatatga  15720
ttaaagataa gcgctacaga ggagtctgag caggacaact cgcgtagtga gagttacatg  15780
ttcgttgggt tcttccgaca cggacctgag ttggccaacg tcccacctga ggtctgtgcc  15840
ccggtgatga gaagtgtgca tctcgttctt gcagctcgtc agtactttca gaatcatggc  15900
gtgcatggta gaatgaccct tataacggac ttcgacatgg caatcgctag gtaaatgaat  15960
gaggagtggc agacatactg tcccacaaat gaagatggaa tctgtaaaag aaaacgcgtg  16020
aaataatgcg tctgacaaag gttaggtcgg ctgcctttaa tcaataccaa agtggtccct  16080
accacgatgg aaaaactgtg cagtcggttt ggctttttct gacgaacaaa taagattcgt  16140
ggccgacagg tgggggtcca ccatgtgaag gcatcttcag actccaataa tggagcaatg  16200
acgtaagggc ttacgaaata agtaagggta gtttgggaaa tgtccactca cccgtcagtc  16260
tataaatact tagcccctcc ctcattgtta agggagcaaa atctcagaga gatagtccta  16320
gagagagaaa gagagcaagt agcctagaag tagtcaaggc ggcgaagtat tcaggcacgt  16380
ggccaggaag aagaaaagcc aagacgacga aaacaggtaa gagctaagct taatgggaaa  16440
gcgtactgct gatggttcag agtttgaaag tccaaagaag aaaagaaaag ttggtagtgg  16500
tgtgaaatgg tactacaaaa caataacatt tctacctgag ctctgcaaca atgaatcact  16560
agcagcgaaa tgtctgcgtg tacttcatgg attcaactac caatatgaaa caagaaacat  16620
tggtgttagc ttccctcttt ggtgtgacgc aactgtcgga aagaagatat cctttgtgtc  16680
caagaacaag attgagcttg acttgcttct caaacaacat tactttgttc aaatggagca  16740
gcttcagtat ttccacatct caaatactgt gctggttcca gaagattgca cttatgtcag  16800
tttcaggaga tgtcaaagta ttgataagct tactgctgct ggtttagcaa gaaagatccg  16860
tcgattggag aagagagctt tgtcaagagg agaacagttt gatccctcaa gttttgctca  16920
gaaagaacac accgccattg ctcattatca ttctctcggt gaatcttcaa agcaaacaaa  16980
taggaatttt cggttaaata taaggatgtt gagtgaacaa ccaagagaag gaaattccat  17040
cttctcttct tatgggctat caaattctga gaattctttt cagcctgttc cgctaatctg  17100
agcttatatg aagatgaaga tgaaatattt ggtgtgtcaa ataaaaaggt tgtgtgctta  17160
agtttgtgtt tttttcttgg cttgttgtgt tatgaatttg tggctttttc taatattaaa  17220
tgaatgtaac atctcattat aatgaataaa caaatgtttc tataatccat tgtgaatgtt  17280
ttgttggatc tcttctccag catataacta ctgtatgtgc tatggtatgg actatggaat  17340
atgattaaag ataagcgcta gcaaggagtg gcagacatac tgtcccacaa atgaagatgg  17400
aatctgtaaa agaaaacgcg tgaaataatg cgtctgacaa aggttaggtc ggctgccttt  17460
aatcaatacc aaagtggtcc ctaccacgat ggaaaaactg tgcagtcggt ttggcttttt  17520
ctgacgaaca aataagattc gtggccgaca ggtgggggtc caccatgtga aggcatcttc  17580
agactccaat aatggagcaa tgacgtaagg gcttacgaaa taagtaaggg tagtttggga  17640
aatgtccact cacccgtcag tctataaata cttagcccct ccctcattgt taagggagca  17700
aaatctcaga gagatagtcc tagagagaga aagagagcaa gtagcctaga agtagtcaag  17760
gcggcgaagt attcaggcac gtggccagga agaagaaaag ccaagacgac gaaaacaggt  17820
aagagctaag cttaatggga aagcgtactg ctgatggttc agagtttgaa agtccaaaga  17880
agaaaagaaa agttggtagt ggtatgaaac ttccaacaaa tctagcttac gagaggtcga  17940
ttgatccttc agatgtttgt ttctttgtgg tctggccaga tgatcggaaa acaccgttga  18000
catacaattc acgtaccctc ctagggcaga tggaggctgc atctctggcg tatgatgtaa  18060
gtggtcaacc gatcaagagt gccacagcag aagcattagc tcaaggaaat cctcatcaag  18120
ttgacttttg tcatgttcct tatggagcat ctcacatcga gtgcagcttc tctgtttctt  18180
tttcatcaga gcttcgccag ccttacaaat gtaacagcag taaagtgaag caaactctcg  18240
ttcagctcgt agagttgtat gaaactaaga ttggatggac tgaactagcg acgagatatt  18300
tgatgaacat atgcaatggt aaatggctct ggaaaaatac aaggaaggct tattgttgga  18360
atatagtcct gactccatgg ccttggaacg gagagaaggt tggctttgaa gacattagga  18420
caaactacac ttcaagacaa gatttcaaga acaacaagaa ttggtctgcc attgttgaga  18480
tgattaagac agctttcagt tctactgatg gcttagcgat cttcgaggtt cgtgcgactc  18540
ttcacttacc aaccaatgct atggttagac cttcccaggt attcacagag aaagaatcag  18600
ggtcgaaatc aaaaagcaag actcagaatt ctcgagtgtt tcaatcaaca accatagatg  18660
gtgaaagatc tcccattctt ggtgcattca aaactggtgc agctatagct actattgacg  18720
actggtatcc tgaagctaca gaaccattgc gtgttggaag atttggtgtc catagagaag  18780
atgtgacttg ctatagacat ccttccacag ggaaagattt tttctccata cttcaacaag  18840
ctgaacatta cattgaagta ctttcagcaa acaaaacccc agcacaagaa accattaatg  18900
atatgcattt tcttatggct aatctcatca aaggtggcat gtttcagcac aaaggagact  18960
gagcttatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  19020
aagtttgtgt tttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  19080
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  19140
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  19200
tatgattaaa gataagcgct atacaggagt ggcagacata ctgtcccaca aatgaagatg  19260
gaatctgtaa aagaaaacgc gtgaaataat gcgtctgaca aaggttaggt cggctgcctt  19320
taatcaatac caaagtggtc cctaccacga tggaaaaact gtgcagtcgg tttggctttt  19380
tctgacgaac aaataagatt cgtggccgac aggtgggggt ccaccatgtg aaggcatctt  19440
cagactccaa taatggagca atgacgtaag ggcttacgaa ataagtaagg gtagtttggg  19500
aaatgtccac tcacccgtca gtctataaat acttagcccc tccctcattg ttaagggagc  19560
aaaatctcag agagatagtc ctagagagag aaagagagca agtagcctag aagtagtcaa  19620
ggcggcgaag tattcaggca cgtggccagg aagaagaaaa gccaagacga cgaaaacagg  19680
taagagctaa gcttaatggg aaagcgtact gctgatggtt cagagtttga aagtccaaag  19740
aagaaaagaa aagttggtag tggtatgcag actcttaaag aactaattgc ttctaatccg  19800
gatgacctaa caacggagct caaaagggca tttcgccccc tgactcctca cattgccatt  19860
gacggtaacg agcttgacgc tctcaccatt cttgtgaatc tcacagacaa aactgatgat  19920
caaaaggatc ttctcgatag agctaaatgt aaacagaaac tcagagatga gaaatggtgg  19980
gcttcttgca tcaattgcgt taattataga cagtctcaca atcctaaatt ccctgatatc  20040
aggtctgaag gagtgattcg cactcaagct ttgggcgaac ttccttcatt tctattaagc  20100
tccagtaaaa taccaccgta tcattggtca tattctcatg attcaaaata tgttaacaag  20160
```

-continued

```
tcagcctttt taacaaatga attctgttgg gacggagaga taagctgctt gggggaactc  20220
ctcaaagatg cggatcatcc actttggaac acactaaaga agcttggctg ctcacagaag  20280
acatgcaaag ctatggcgaa acaattggct gatatcacat taacaacgat taatgtaaca  20340
ctcgctccga attacttgac tcagatttct cttccggatt ctgatacaag ttacatttct  20400
ctctcgcctg tcgcaagtct atcaatgcaa agtcatttcc accaacgttt gcaagacgaa  20460
aacagacatt ctgctataac aagattctcc cgaacaacta acatgggtgt gactgcaatg  20520
acttgtggtg gagctttcag gatgttaaag tctggtgcaa agttctccag tccacctcat  20580
catagattaa actcgaaacg ctcttggttg accagtgagc atgttcaatc attgaagcag  20640
tatcaacgtc tgaacaaatc attgatccct gagaatagta ggattgctct taggcgcaag  20700
tacaagattg aacttcaaaa tatggttcga tcatggtttg ccatgcaaga tcacactctc  20760
gattcaaata tccttattca acatctcaat cacgatttga gttatctcgg ggcgaccaaa  20820
agatttgctt atgatccagc tatgacaaaa ttgttcactg agctgctaaa gagggagctg  20880
agtaattcta tcaacaacgg agaacaacat accaatggca gttttctggt cctcccaaac  20940
atacgggttt gtggagcaac tgcgttatct tccccagtaa ccgttggcat ccctagctta  21000
accgccttct ttggttttgt tcacgccttt gagaggaaca tcaacaggac cacttcttcc  21060
tttcgtgtcg agagctttgc aatatgtgtc catcagcttc atgtggagaa gcgtggtttg  21120
acagctgagt ttgtagagaa gggagatggg acaatttcag caccagcaac cagagatgat  21180
tggcaatgtg atgttgtgtt ttctctaatt ttgaatacga acttcgctca acacattgat  21240
caggacacac tggtcacgag tttacctaag cggctggcaa ggggttctgc taagatcgca  21300
atagatgatt tcaagcatat aaactcattc tccactcttg aaactgcaat tgagtctctt  21360
cccatcgaag ctggaagatg gcttagtctg tatgcacaga gtaataataa tctttcggac  21420
ttgttagctg ctatgaccga agaccatcaa cttatggcta gttgtgttgg ctaccacctt  21480
ctagaagaac caaaagataa gccaaatagc cttagaggat acaaacatgc catagctgaa  21540
tgtattatag gattaattaa ttcaattaca ttttcctcgg aaaccgatcc taatactata  21600
ttctggtcct taaagaacta ccaaaactat cttgtagttc agcccagatc cataaatgat  21660
gaaacaacag acaagagcag cctgtgagct tatatgaaga tgaagatgaa atatttggtg  21720
tgtcaaataa aaaggttgtg tgcttaagtt tgtgtttttt tcttggcttg ttgtgttatg  21780
aatttgtggc tttttctaat attaaatgaa tgtaacatct cattataatg aataaacaaa  21840
tgtttctata atccattgtg aatgttttgt tggatctctt ctccagcata taactactgt  21900
atgtgctatg gtatggacta tggaatatga ttaaagataa gcgctacaga ggagtctgag  21960
caggacaact cgcgtagtga gagttacatg ttcgttgggt tcttccgaca cggacctgag  22020
ttggccaacg tcccacctga ggtctgtgcc ccggtgatga gaagtgtgca tctcgttctt  22080
gcagctcgtc agtactttca gaatcatggc gtgcatggta gaatgaccct tataacggac  22140
ttcgacatgg caatcgctag gtagcttaat gaataggcgt atcacgaggc cctttcgtct  22200
tcaagaatta gcttggcaat tcccgatcta gtaacataga tgacaccgcg cgcgataatt  22260
tatcctagtt tgcgcgctat attttgtttt ctatcgcgta ttaaatgtat aattgcggga  22320
ctctaatcat aaaaacccat ctcataaata acgtcatgca ttacatgtta attattacat  22380
gcttaacgta attcaacaga aattatatga taatcgtcgc aagaccggca acaggattca  22440
atcttaagaa actttattgc caaatgtttg aacgatcggg atcgaagatc cattccaagc  22500
ctacaggaac aggtggtggc ggccctcggt gcgctcgtac tgctccacga tggtgtagtc  22560
ctcgttgtgg gaggtgatgt ccagcttggc gtccacgtag tagtagccgg gcagctgcac  22620
gggcttcttg gccatgtaga tggacttgaa ctccaccagg tagtggccgc cgtccttcag  22680
cttcagggcc ttgtgtgtct cgcccttcag cacgccgtcg cgggggtaca ggcgctcggt  22740
ggaggcctcc cagcccatcg tcttcttctg catcacgggg ccgtcggagg ggaagttcac  22800
gccgatgaac ttcaccttgt agatgaagca gccgtcctgc agggaggagt cctgggtcac  22860
ggtcgccacg ccgccgtcct cgaagttcat cacgcgctcc cacttgaagc cctcggggaa  22920
ggacagcttc ttgtagtcgg ggatgtcggc ggggtgcttc acgtacacct tggagccgta  22980
ctggaactgg ggggacagga tgtcccaggc gaagggcagg gggccgccct tggtcacctt  23040
cagcttcacg gtgttgtggc cctcgtaggg gcggccctcg ccctcgccct cgatctcgaa  23100
ctcgtggccg ttcacggtgc cctccatgcg caccttgaag cgcatgaact cggtgatgac  23160
gttctcggag gaggccattt cgcctcgatc gagttgagag tgaatatgag actctaattg  23220
gataccgagg ggaatttatg gaacgtcagt ggagcatttt tgacaagaaa tatttgctag  23280
ctgatagtga ccttaggcga cttttgaacg cgcaataatg gtttctgacg tatgtgctta  23340
gctcattaaa ctccagaaac ccgcggctca gtggctcctt caacgttgcg gttctgtcag  23400
ttccaaacgt aaaacggctt gtcccgcgtc atcggcgggg gtcataccgt gactccctta  23460
attctcatgt atggttcgct atactgaaga gcaattctga caggatatat tggcgggtaa  23520
acctaagaga aaagagcgtt tattagaata atcggatatt taaaagggcg tgaaaaggtt  23580
tatccgttcg tccatttgta tgtgcatgcc aaccacaggg ttcccctcgg gatcaaagta  23640
ctttgatcca acccctccgc tgctatagtg cagtcggctt ctgacgttca gtgcagccgt  23700
catctgaaaa cgacatgtcg cacaagtcct aagttacgcg acaggctgcc gccctgccct  23760
tttcctggcg ttttcttgtc gcgtgtttta gtcgcataaa gtagaatact tgcgactaga  23820
accggagaca ttacgccatg aacaagagcg ccgccgctgg cctgctgggc tatgcccgcg  23880
tcagcaccga cgaccaggac ttgaccaacc aacgggccga actgcacgcg gccggctgca  23940
ccaagctgtt ttccgagaag atcaccggca ccaggcgcga ccgcccggag ctggccagga  24000
tgcttgacca cctacgccct ggcgacgttg tgacagtgac caggctagac cgcctggccc  24060
gcagcacccg cgacctactg gacattgccg agcgcatcca ggaggccggc gcgggcctgc  24120
gtagcctggc agagccgtgg gccgacacca ccacgccggc cggccgcatg gtgttgaccg  24180
tgttcgccgg cattgccgag ttcgagcgtt ccctaatcat cgaccgcacc cggagcgggc  24240
gcgaggccgc caaggcccga ggcgtgaagt ttggcccccg ccctaccctc accccggcac  24300
agatcgcgca cgcccgcgag ctgatcgacc aggaaggccg caccgtgaaa gaggcggctg  24360
cactgcttgg cgtgcatcgc tcgaccctgt accgcgcact tgagcgcagc gaggaagtga  24420
cgcccaccga ggccaggcgg cgcggtgcct tccgtgagga cgcattgacc gaggccgacg  24480
ccctggcggc cgccgagaat gaacgccaag aggaacaagc atgaaaccgc accaggacgg  24540
ccaggacgaa ccgtttttca ttaccgaaga gatcgaggcg gagatgatcg cggccgggta  24600
cgtgttcgag ccgcccgcgc acctctcaac cgtgcggctg catgaaatcc tggccggttt  24660
gtctgatgcc aagctggcgg cctggccggc cagcttggcc gctgaagaaa ccgagcgccg  24720
```

-continued

```
ccgtctaaaa aggtgatgtg tatttgagta aaacagcttg cgtcatgcgg tcgctgcgta  24780
tatgatgcga tgagtaaata aacaaatacg caaggggaac gcatgaaggt tatcgctgta  24840
cttaaccaga aaggcgggtc aggcaagacg accatcgcaa cccatctagc ccgcgccctg  24900
caactcgccg gggccgatgt tctgttagtc gattccgatc cccagggcag tgcccgcgat  24960
tgggcggccg tgcgggaaga tcaaccgcta accgttgtcg gcatcgaccg cccgacgatt  25020
gaccgcgacg tgaaggccat cggccggcgc gacttcgtag tgatcgacgg agcgccccag  25080
gcggcggact tggctgtgtc cgcgatcaag gcagccgact tcgtgctgat tccggtgcag  25140
ccaagccctt acgacatatg ggccaccgcc gacctggtgg agctggttaa gcagcgcatt  25200
gaggtcacgg atggaaggct acaagcggcc tttgtcgtgt cgcgggcgat caaaggcacg  25260
cgcatcggcg gtgaggttgc cgaggcgctg gccgggtacg agctgcccat tcttgagtcc  25320
cgtatcacgc agcgcgtgag ctacccaggc actgccgccg ccggcacaac cgttcttgaa  25380
tcagaacccg agggcgacgc tgcccgcgag gtccaggcgc tggccgctga aattaaatca  25440
aaactcattt gagttaatga ggtaaagaga aaatgagcaa aagcacaaac acgctaagtg  25500
ccggccgtcc gagcgcacgc agcagcaagg ctgcaacgtt ggccagcctg gcagacacgc  25560
cagccatgaa gcgggtcaac tttcagttgc cggcggagga tcacaccaag ctgaagatgt  25620
acgcggtacg ccaaggcaag accattaccg agctgctatc tgaatacatc gcgcagctac  25680
cagagtaaat gagcaaatga ataaatgagt agatgaattt tagcggctaa aggaggcggc  25740
atggaaaatc aagaacaacc aggcaccgac gccgtggaat gccccatgtg tggaggaacg  25800
ggcggttggc caggcgtaag cggctgggtt gcctgccggc cctgcaatgg cactggaacc  25860
cccaagcccg aggaatcggc gtgagcggtc gcaaaccatc cggcccggta caaatcggcg  25920
cggcgctggg tgatgacctg gtggagaagt tgaaggccgc gcaggccgcc cagcggcaac  25980
gcatcgaggc agaagcacgc cccggtgaat cgtggcaagc ggccgctgat cgaatccgca  26040
aagaatcccg gcaaccgccg gcagccggtg cgccgtcgat taggaagccg cccaagggcg  26100
acgagcaacc agatttttc gttccgatgc tctatgacgt gggcacccgc gatagtcgca  26160
gcatcatgga cgtggccgtt ttccgtctgt cgaagcgtga ccgacgagct ggcgaggtga  26220
tccgctacga gcttccagac gggcacgtag aggtttccgc agggccggcc ggcatggcca  26280
gtgtgtggga ttacgacctg gtactgatgg cggtttccca tctaaccgaa tccatgaacc  26340
gataccggga agggaaggga gacaagcccg gccgcgtgtt ccgtccacac gttgcggacg  26400
tactcaagtt ctgccggcga gccgatggcg gaaagcagaa agacgacctg gtagaaacct  26460
gcattcggtt aaacaccacg cacgttgcca tgcagcgtac gaagaaggcc aagaacggcc  26520
gcctggtgac ggtatccgag ggtgaagcct tgattagccg ctacaagatc gtaaagagcg  26580
aaaccgggcg gccggagtac atcgagatcg agctagctga ttggatgtac cgcgagatca  26640
cagaaggcaa gaacccggac gtgctgacgg ttcaccccga ttactttttg atcgatcccg  26700
gcatcggccg ttttctctac cgcctggcac gccgcgccgc aggcaaggca gaagccagat  26760
ggttgttcaa gacgatctac gaacgcagtg gcagcgccgg agagttcaag aagttctgtt  26820
tcaccgtgcg caagctgatc gggtcaaatg acctgccgga gtacgatttg aaggaggagg  26880
cggggcaggc tggcccgatc ctagtcatgc gctaccgcaa cctgatcgag ggcgaagcat  26940
ccgccggttc ctaatgtacg gagcagatgc tagggcaaat tgccctagca ggggaaaaag  27000
gtcgaaaagg actctttcct gtggatagca cgtacattgg gaacccaaag ccgtacattg  27060
ggaaccggaa cccgtacatt gggaacccaa agccgtacat tgggaaccgg tcacacatgt  27120
aagtgactga tataaaagag aaaaaaggcg atttttccgc ctaaaactct ttaaaactta  27180
ttaaaactct taaaacccgc ctggcctgtg cataactgtc tggccagcgc acagccgaag  27240
tgctgcaaaa agcgcctacc cttcggtcgc tgcgctccct acgccccgcc gcttcgcgtc  27300
ggcctatcgc ggccgctggc cgctcaaaaa tggctggcct acggccaggc aatctaccag  27360
ggcgcggaca agccgcgccg tcgccactcg accgccggcg cccacatcaa ggcaccctgc  27420
ctcgcgcgtt tcggtgatga cggtgaaaac ctctgacaca tgcagctccc ggtgacggtc  27480
acagcttgtc tgtaagcgga tgccgggagc agacaagccc gtcagggcgc gtcagcgggt  27540
gttggcgggt gtcggggcgc agccatgacc cagtcacgta gcgatagcgg agtgtatact  27600
ggcttaacta tgcggcatca gagcagattg tactgagagt gcaccatatg cggtgtgaaa  27660
taccgcacag atgcgtaagg agaaaatacc gcatcaggcg ctcatccgct tcctcgctca  27720
ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg  27780
taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc  27840
agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc  27900
cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac  27960
tataaagata ccaggcgttt ccccctggaa gctccctcgt gcgctctcct gttccgaccc  28020
tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata  28080
gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc  28140
acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca  28200
acccggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag  28260
cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta  28320
gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg  28380
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggtttttt gtttgcaagc  28440
agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt  28500
ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgcat tctaggtgat  28560
tagaaaaact catcgagcat caaatgaaac tgcaatttat tcatatcagg attatcaata  28620
ccatattttt gaaaaagccg tttctgtaat gaaggagaaa actcaccgag gcagttccat  28680
aggatggcaa gatcctggta tcggtctgcg attccgactc gtccaacatc aatacaacct  28740
attaatttcc cctcgtcaaa aataaggtta tcaagtgaga aatcaccatg agtgacgact  28800
gaatccggtg agaatggcaa aagtttatgc atttctttcc agacttgttc aacaggccag  28860
ccattacgct cgtcatcaaa atcactcgca tcaaccaaac cgttattcat tcgtgattgc  28920
gcctgagcga gtcgaaatac gcgatcgctg ttaaaaggac aattacaaac aggaatcgaa  28980
tgcaaccggc gcaggaacac tgccagcgca tcaacaatat tttcacctga atcaggatat  29040
tcttctaata cctggaatgc tgttttccct gggatcgcag tggtgagtaa ccatgcatca  29100
tcaggagtac ggataaaatg cttgatggtc ggaagaggca taaattccgt cagccagttt  29160
agtctgacca tctcatctgt aacatcattg gcaacgctac ctttgccatg tttcagaaac  29220
aactctggcg catcgggctt cccatacaat cggtagattg tcgcacctga ttgcccgaca  29280
```

-continued

```
ttatcgcgag cccatttata cccatataaa tcagcatcca tgttggaatt taatcgcggc  29340
cttgagcaag acgtttcccg ttgaatatgg ctcataacag aacttattat ttccttcctc  29400
ttttctacag tatttaaaga taccccaaga agctaattat aacaagacga actccaattc  29460
actgttcctt gcattctaaa accttaaata ccagaaaaca gctttttcaa agttgttttc  29520
aaagttggcg tataacatag tatcgacgga gccgattttg aaaccgcggt gatcacaggc  29580
agcaacgctc tgtcatcgtt acaatcaaca tgctaccctc cgcgagatca tccgtgtttc  29640
aaacccggca gcttagttgc cgttcttccg aatagcatcg gtaacatgag caaagtctgc  29700
cgccttacaa cggctctccc gctgacgccg tcccggactg atgggctgcc tgtatcgagt  29760
ggtgattttg tgccgagctg ccggtcgggg agctgttggc tggctgg                29807

SEQ ID NO: 40          moltype = DNA  length = 29807
FEATURE                Location/Qualifiers
source                 1..29807
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 40
tggcaggata tattgtggtg taaacataac gaattgctct tcagcaagga gaatgaggag  60
aatgtagcag aaggcatgtt gttgtgactc cgaggggttg cctcaaactc tatcttataa  120
ccggcgtgga ggcatggagg caggggtatt ttggtcattt taatagatag tggaaaatga  180
cgtggaattt acttaaagac gaagtctttg cgacaagggg gggcccacgc cgaatttaat  240
attaccggcg tggcccccc ttatcgcgag tgctttagca cgagcggtcc agatttaaag  300
tagaaaattt cccgcccact agggttaaag gtgttcacac tataaaagca tatacgatgt  360
gatggtattt gatggagcgt atattgtatc aggtatttcc gttggatacg aattattcgt  420
acgaccctcg gtatactaat gaggagtgtt gatacaacca taaaatgata attacaccca  480
taaattgata attatcacac ccataaattg atattgcctc ttcatggtct aaacttcagt  540
aagtttacga cattttcatg aagaggtcat ttcatgacaa gcttacctac gccctcagca  600
attacgactt cggcgttaga gtatgcattc catactcccg ctcgcaatct aacgaaatct  660
cgcggaaaaa atattcatcg ttatgtcagt gtaaagcgct agcaaggagt gagacttttc  720
aacaaagggt aatatcggga aacctcctcg gattccattg cccagctatc tgtcacttca  780
tcaaaaggac agtagaaaag gaaggtggca cctacaaatg ccatcattgc gataaaggaa  840
aggctatcgt tcaagatgcc tctgccgaca gtggtcccaa agatggaccc ccacccacga  900
ggagcatcgt ggaaaaagaa gacgttccaa ccacgtcttc aaagcaagtg gattgatgtg  960
atatctccac tgacgtaagg gatgacgcac aatcccacta tccttcgcaa gacccttcct  1020
ctatataagg aagttcattt catttggaga ggacacgctg acggccgcgt atttttacaa  1080
caattaccaa caacaacaaa caacaaacaa cattacaatt actatttaca attacagcgg  1140
ccgccccctt caccaaaaaa atggctgcag ccaagggcga ggagctgttc accggggtgg  1200
tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg  1260
agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gctgctctgc accaccggca  1320
agctgcccgt gccctggccc accctcgtga ccaccctggg ctacggcgtg cagtgcttcg  1380
cccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg cccgaaggct  1440
acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg  1500
tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcatc gacttcaagg  1560
aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac aacgtctata  1620
tcaccgccga caagcagaag aacggcatca aggccaactt caagatccgc cacaacatcg  1680
aggacggcgg cgtgcagctc gccgaccact accagcagaa cacccccatc ggcgacggcc  1740
ccgtgctgct gcccgacaac cactacctga gctaccagtc caaactgttc aaagacccca  1800
acgagaagcg cgatcacatg gtcttgttag aatttttgac tgctgctggt attaccgagg  1860
gtatgaatga attgtacaaa ggctccggac cgaagaagaa gaggaaggtt taaatatgaa  1920
gatgaagatg aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt  1980
tttcttggct tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat  2040
ctcattataa tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc  2100
ttctccagca tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat  2160
aagcgctata caggagagtg tccgtgactg atgcaaggtt atcggaactg taatgattat  2220
atgttttcta attttgaagt ggttcttttg atgagttttg ccagttactg aaagtaactt  2280
tcttactgca gtagttttgc tgaaatactc gattcacaaa aatatcaact tatggttgtt  2340
ttgtgagata tcaatatatg gttgttttgt ggttaagttg ctgattataa ataattatta  2400
aatatcactt tatggttgca tcaacacgct acagaggagt ctgagcagga caactcgcgt  2460
agtgagagtt acatgttcgt tgggttcttc cgacacggac ctgagttggc caacgtccca  2520
cctgaggtct gtgccccggt gatgagaagt gtgcatctcg ttcttgcagc tcgtcagtac  2580
tttcagaatc atggcgtgca tggtagaatg acccttataa cggacttcga catggcaatc  2640
gctaggtaaa tgaatggaaa aagtggcacc gagtcggtgc tttttttccc ggggcgcgcc  2700
gatatcgagc tctcccggcg cgccgatatc gagctcagtg tttgatcgcc ggcggtaccg  2760
agtgtacttc aagtcagtgg gaaatcaata aaatgattat tttatgaata tatttcattg  2820
tgcaagtaga tagaaattac atatgttaca taacacacga aataaacaaa aaaagacaat  2880
ccaaaaacaa acaccccaaa aaaaataatc actttagata aactcgtatg aggagaggca  2940
cgttcagtga ctcgacgatt cccgagcaaa aaaagtctcc ccgtcacaca tgtagtgggt  3000
gacgcaatta tctttaaagt aatccttctg ttgacttgtc attgataaca tccagtcttc  3060
gtcaggattg caaagaatta tagaagggat cccacctttt attttcttct tttttccata  3120
tttagggttg acagtgaaat cagactggca acctattaat tgcttccaca atgggacgaa  3180
cttgaagggg atgtcgtcga tgatattata ggtggcgtgt tcatcgtagt tggtgaaatc  3240
gatggtaccg ttccaatagt tgtgtcgtcc gagacttcta gcccaggtgg tctttccggt  3300
acgagttggt ccgcagatgt agaggctggg gtgtcggatt ccattccttc cattgtcctt  3360
gttaaatcgg ccatccattc aaggtcagat tgagcttgtt ggtatgagac aggatgtatg  3420
taagtataag cgtctatgct tacatggtat agatgggttt ccctccagga gtgtagatct  3480
tcgtggcagc gaagatctga ttctgtgaag ggcgacacat acggttcagg ttgtggaggg  3540
aataatttgt tggctgaata ttccagccat tgaagctttg ttgcccattc atgagggaat  3600
```

-continued

```
tcttccttga tcatgtcaag atattcctcc ttagacgttg cagtctggat aatagttctc  3660
catcgtgcgt cagatttgcg aggagaaacc ttatgatctc ggaaatctcc tctggtttta  3720
atatctccgt cctttgatat gtaatcaagg acttgtttag agtttctagc tggctggata  3780
ttagggtgat ttccttcaaa atcgaaaaaa gaaggatccc taatacaagg ttttttatca  3840
agctggagga gagcatgata gtgggtagtg ccatcttgat gaagctcaga agcaacacca  3900
aggaagaaaa taagaaaagg tgtgagtttc tcccagagaa actggaataa atcatctctt  3960
tgagatgagc acttgggata ggtaaggaaa acatatttag attggagtct gaagttctta  4020
ctagcagaag gcatgttgtt gtgactccga ggggttgcct caaactctat cttataaccg  4080
gcgtggaggc atggaggcag gggtattttg gtcattttaa tagatagtgg aaaatgacgt  4140
ggaatttact taaagacgaa gtctttgcga caaggggggg cccacgccga atttaatatt  4200
accggcgtgg cccccccttta tcgcgagtgc tttagcacga gcggtccaga tttaaagtag  4260
aaaatttccc gcccactagg gttaaaggtg ttcacactat aaaagcatat acgatgtgat  4320
ggtatttgat ggagcgtata ttgtatcagg tatttccgtt ggatacgaat tattcgtacg  4380
accctcatag tttaaactat cagtgggtag cttaatggta gacactttat tagttagatc  4440
cgctgttatt ctgcacgatt gcgcttgccg gtggaaattt aaggacgagt gtagtgcact  4500
cacccttagc cttcagcgtg acaagattag ggtagcaccg gaaacagcga ttttacggac  4560
agatagtaca gtctgagact cagtcactgg tccctaaacg ggagccggta cgtcggcggt  4620
tcgctagcaa ggagtggcag acatactgtc ccacaaatga agatggaatc tgtaaaagaa  4680
aacgcgtgaa ataatgcgtc tgacaaaggt taggtcggct gcctttaatc aataccaaag  4740
tggtccctac cacgatggaa aaactgtgca gtcggtttgg ctttttctga cgaacaaata  4800
agattcgtgg ccgacaggtg ggggtccacc atgtgaaggc atcttcagac tccaataatg  4860
gagcaatgac gtaagggctt acgaaataag taagggtagt ttgggaaatg tccactcacc  4920
cgtcagtcta taaatactta gcccctccct cattgttaag ggagcaaaat ctcagagaga  4980
tagtcctaga gagagaaaga gagcaagtag cctagaagta gtcaaggcgg cgaagtattc  5040
aggcacgtgg ccaggaagaa gaaaagccaa gacgacgaaa acaggtaaga gctaagctta  5100
atgggaaagc gtactgctga tggttcagag tttgaaagtc caaagaagaa aagaaaagtt  5160
ggtagtggta tgactgataa acggaaagat ggttccggca agcttcttta ctgcagcttt  5220
tgcgggaaat ctcaacatga agttaggaaa cttattgcag ggccttctgt ttatatctgt  5280
gacgaatgtg ttgatctctg caatgatatt attcgagaag agatcaaaga agtggctcct  5340
cacagagaga ggagtgcgtt gccgacacca cacgagatta gaaatcatct tgatgattat  5400
gtcattggtc aagaacaagc aaagaaggtc cttgctgttg cagtgtacaa ccattacaag  5460
aggctcagaa acggagatac ttctaatggt gtggaacttg gcaaatcaaa catcctgctc  5520
attggaccca ctggaagtgg aaagacactc ctagcagaga cattggcaag gctcttagac  5580
gttcctttca ctatggctga tgcaactact cttactgagg ctggttatgt tggagaagat  5640
gtggagaata ttattcaaaa gttattgcag aaatgtgatt acgatgtcca gaaagctcag  5700
agagggatag tctacataga tgagatagac aagatctcta ggaagagtga taatccctcg  5760
ataacacgcg acgtttctgg cgaaggtgtt caacaagctt tgttgaagct cattgaaggc  5820
accgtagcgg cagttccacc acaaggagga agaaagcatc ctcagcaaga gtttctccaa  5880
gtggatacct caaaaatctt atttatatgt ggtggtgcct tcgctgggct tgacaaagtt  5940
atctcacacc gtgtagaaac tggtagcggg attgggtttg gtgcaacggt gaaggcaaaa  6000
tctgataagg cttctgaagg agagcttctg gctcaagttg aaccagagga tcttatcaag  6060
tttggactaa ttcctgagtt catcggtaga cttcctgtcg ttgctacgtt gaacgagtta  6120
agtgaggaag ccctgattca aatcttgaaa gagccaaaga atgctctgac aaagcaatat  6180
caggctcttt tcaatcttga gggagttgac ttggaatttc gtgatgaagc gctggatgct  6240
atagctaaga aagcaatggc gaggaaaacc ggagctcgtg gtctaagatc gattgttgaa  6300
gccgctttat tggatacaat gtatgatttg ccctccatgg aagatgttga gaaagtggtc  6360
atcgacgagt ctgtgattga tggacagtca aagccattgc tcatatatgg taaacctgaa  6420
gcacagcagg catcaggaga atgaatatga agatgaagat gaaatatttg gtgtgtcaaa  6480
taaaaaggtt gtgtgcttaa gtttgtgttt ttttcttggc ttgttgtgtt atgaatttgt  6540
ggctttttct aatattaaat gaatgtaaca tctcattata atgaataaac aaatgtttct  6600
ataatccatt gtgaatgttt tgttggatct cttctccagc atataactac tgtatgtgct  6660
atggtatgga ctatggaata tgattaaaga taagcgctat acaggagtgg cagacatact  6720
gtcccacaaa tgaagatgga atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa  6780
ggttaggtcg gctgccttta atcaatacca aagtggtccc taccacgatg gaaaaactgt  6840
gcagtcggtt tggctttttc tgacgaacaa ataagattcg tggccgacag gtgggggtcc  6900
accatgtgaa ggcatcttca gactccaata atggagcaat gacgtaaggg cttacgaaat  6960
aagtaagggt agtttgggaa atgtccactc acccgtcagt ctataaatac ttagcccctc  7020
cctcattgtt aagggagcaa aatctcagag agatagtcct agagagagaa agagagcaag  7080
tagcctagaa gtagtcaagg cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc  7140
caagacgacg aaaacaggta agagctaagc ttaatgggaa agcgtactgc tgatggttca  7200
gagtttgaaa gtccaaagaa gaaaagaaaa gttggtagtg gtatgagtta ttctggagaa  7260
cgggataatt ttgcacctca catggcatta gttccaatgg tgattgaaca aacaagtaga  7320
ggggagcgaa gctttgacat ttactctcgt ctattaaaag aaagggtcat cttcttaacc  7380
gggcaagttg aagatcatat ggcaaatctg attgtagctc agatgctttt tcttgaagca  7440
gagaatcctg agaaagatat ctatctctac attaactctc ctggtggtgt aataactgct  7500
ggaatgtcta tatatgatac aatgcagttt atcaaacctg atgtttcaac aatatgtatg  7560
ggtcaggctg cgtcaatggg agcttttctt ttgacggctg gcgcaaaagg aaagcggttc  7620
tgcctaccga actcaagggt gatgatccat caaccacttg gaggatacca aggtcaagct  7680
accgatatag agattcatgc tagagaaatt ttgaaggtga agggtagaat gaatgaactg  7740
atggcccttc acactggaca atcgctcgag cagattgagc gtgacacaga gagagacagg  7800
tttttatccg caccagaagc tgttgagtat ggtttggttg attctattct tactcataga  7860
aactgaatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  7920
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  7980
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  8040
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  8100
tatgattaaa gataagcgct acagaggagc atcttcattc ttaagatatg aagataatct  8160
```

-continued

```
tcaaaaggcc cctgggaatc tgaaagaaga gaagcaggcc catttatatg ggaaagaaca  8220
atagtatttc ttatataggc ccatttaagt tgaaaacaat cttcaaaagt cccacatcgc  8280
ttagataaga aaacgaagct gagtttatat acagctagag tcgaagtagt gataatgtca  8340
ttactactgc aaagtagctg ataacatccc actatccttc gcaagaccct tcctctactt  8400
tactgctgaa taagtagata actacgcttt tttttttgca aaattttcca gatcgatttc  8460
ttcttcctct gttcttcggc gttcaatttc tggggttttc tcttcgtttt ctgtaactga  8520
aacctaaaat ttgacctaaa aaaaatctca aataatatga ttcagtggtt ttgtactttt  8580
cagttagttg agttttgcag ttccgatgag ataaaccaat acgctaggta tactaatgag  8640
gagtggcaga catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa  8700
taatgcgtct gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc  8760
acgatggaaa aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc  8820
cgacaggtgg gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg  8880
taagggctta cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat  8940
aaatacttag cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag  9000
agagaaagag agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc  9060
caggaagaag aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tggcgacatc  9120
ccttccaacg ccatcagcta taactacttc ggcactagag tatgctttcc atactccggc  9180
tcgcaatttg acgaagtcta gggggaagaa cattcatcga tatgttagtg tcaagatgtc  9240
taaaagaatt actgtagaat caacactaga atgcgatgct tgctatcatt ttgacttcga  9300
gccttctatc gtgagatttt gtgctcagcc tattcgcttt ctgtattatt taaatgggca  9360
gtcacactcc tatgtcccag attttcttgt gcaattcgac acaaacgagt ttgtcttgta  9420
tgaagttaaa agcgcttatg cgaaaaacaa acctgatttc gatgttgagt gggaagcaaa  9480
ggtgaaggcc gcaacagagc tgggtttgga acttgaactt gttgaggaga gcgatataag  9540
agataccgtt gttttgaaca atcttaagag gatgcataga tatgcgagta aggacgagct  9600
taataacgta cacaattccc tacttaagat cataaagtat aacggagctc aatccgcccg  9660
atgtttgggt gaacagctag ggttgaaagg tcgtacggta cttcctattc tgtgtgatct  9720
tctttcaaga tgcttgctcg acaccagact tgataagccc ctctctctgg aaagtcgttt  9780
tgaactcgcc tcatatggag gatcaggatc tggcaaaaga accgcagatg gttctgagtt  9840
cgaaagccca aagaaaaaga gaaaggtggg ttccggttct ggaggagcca agaagggttt  9900
ttctagcttt cacaggaagg cagtttcctc acaggacact ttagagagta tagaactagt  9960
ttccagtgcg aattgcctcg aatcagtcac ttatcaagat atatctgcat ttcctgaaac  10020
cattgcagtg gagatcaact tcagattgtc tattctccgg tttctcgcaa gaaaatgtga  10080
gactatcgtt gctaaatcta ttgagcccca tagagttgag cttcaacaga attattctcg  10140
gaagattcca tctgcaataa caatatacag atggtggttg gcctttcgga agtcagatta  10200
caacccgatt agtctggctc ctaatatcaa agacagaggg aacagagaaa ctaaagtatc  10260
tactgtggtt gactccatca tggaacaagc cgttgaaagg gttatctctg gtcgtaaagt  10320
caatgtttct agcgcgtaca aacgagtccg gaggaaagtt agacaatata atcttactca  10380
tggaacaaag tacacatatc caaaatacga gagcgtacga aagcgcgtaa agaagaaaac  10440
accgtttgag ttattagctg ctggtaaggg agagagggtc gccaagcgtg aatttaggag  10500
aatgggaaag aagattctga cgtcaagtgt tctcgagagg gtggagatag atcacactgt  10560
tgtggatctg ttcgctgttc atgaggaata taggattcct ctcgggaggc catggctgac  10620
tcagttagtt gattgttaca gcaaagcagt cattggtttc taccttggct ttgaacctcc  10680
tagttatgtc tccgtatcac tagctttaaa gaatgccatt caacgcaaag atgatttgat  10740
ttcctcttac gagtcaatag aaaatgaatg gttatgttat ggtataccag atctccttgt  10800
tacagacaat ggaaaagaat ttttgtccaa agcttttgat caagcttgtg agagcttatt  10860
aattaacgtg catcagaata aggtagaaac cccagataac aagcctcatg ttgaaagaaa  10920
ctacgggaca ataaatactt ccttattgga cgatttacct ggtaaaagtt tctctcagta  10980
cttgcagagg gaaggatatg actctgtggg agaagctaca ttgactctca atgagatcag  11040
agaaatctac ttgatttggc ttgtcgacat atatcacaag aaaccgaatc aaagaggaac  11100
taactgtcca aatgttgctt ggaaaaaagg ttgccaagaa tgggagccag aagaattctc  11160
cgggtcaaaa gatgagcttg attttaaatt cgcaattgtt gattataaac aacttacaaa  11220
agtaggaatc actgtttaca aggaactaag ctactcaaac gatcgtttgg cagagtatag  11280
aggcaaaaaa ggaaaccaca aagttcagtt caagtacaat ccggaatgta tggctgtgat  11340
atgggttctt gatgaagata tgaacgaata tttcacagtt aatgctatcg actatgaata  11400
cgcatctagg gtgtcactgt ggcaacataa atacaacatg aagtatcaag cagagctaaa  11460
ttctgcggaa tatgacgaag ataaagagat tgatgcagaa attaaaatcg aggaaattgc  11520
tgatagatct atagtgaaaa ccaataaaat tagagctcgt cgtcgtggtg ctaggcatca  11580
agagaatagt gcacgagcta aatcaatctc taatgcaaat cctgcaagta ttcagaagca  11640
tgaagacgaa atagtcagtg ctgataatga tgattgggat atcgattacg tgtgagctta  11700
tatgaagatg aagatgaaat atttggtgtg tcaaataaaa aggttgtgtg cttaagtttg  11760
tgtttttttc ttggcttgtt gtgttatgaa tttgtggctt tttctaatat taaatgaatg  11820
taacatctca ttataatgaa taaacaaatg tttctataat ccattgtgaa tgttttgttg  11880
gatctcttct ccagcatata actactgtat gtgctatggt atggactatg gaatatgatt  11940
aaagataagc gctagcaagg agtggcagac atactgtccc acaaatgaag atggaatctg  12000
taaaagaaaa cgcgtgaaat aatgcgtctg acaaaggtta ggtcggctgc ctttaatcaa  12060
taccaaagtg gtccctacca cgatggaaaa actgtgcagt cggtttggct ttttctgacg  12120
aacaaataag attcgtggcc gacaggtggg ggtccaccat gtgaaggcat cttcagactc  12180
caataatgga gcaatgacgt aagggcttac gaaataagta agggtagttt gggaaatgtc  12240
cactcacccg tcagtctata aatacttagc ccctccctca ttgttaaggg agcaaaatct  12300
cagagagata gtcctagaga gagaaagaga gcaagtagcc tagaagtagt caaggcggcg  12360
aagtattcag gcacgtggcc aggaagaaga aaagccaaga cgacgaaaac aggtaagagc  12420
taagcttaat gggaaagcgt actgctgatg gttcagagtt tgaaagtcca aagaagaaaa  12480
gaaaagttgg tagtggtatg agtgaaacac gagaggctag gataagccgg gccaaacggg  12540
cttttgtttc tacaccgtct gtacgcaaga tcctgagcta catggataga tgtagggact  12600
tatcagatct cgaatcagag ccaacttgta tgatggttta tggtgcctct ggtgtgggaa  12660
aaaccacggt catcaagaag tatttgaatc aaaacagaag agaatctgag gcgggtggtg  12720
```

-continued

```
atattatccc agtgttacac attgagctcc ctgataatgc aaagcctgtt gatgccgcac  12780
gggagctgtt ggttgagatg ggggatccac tggccctcta tgaaactgat ttagcaagac  12840
ttactaagag attgactgag ctgatccctg ctgttggggt gaagcttata atcattgatg  12900
aatttcagca tcttgtggaa gagagatcca atagagtgtt aactcaagtt ggaaactggc  12960
ttaagatgat cttaaacaaa accaaatgcc ctattgtaat atttggaatg ccttactcaa  13020
aggttgttct tcaagccaat tcgcagcttc acggaaggtt ttcaatacaa gtcgaactac  13080
gtcccttctc ttaccaagga ggaagaggag ttttcaagac attcttggag tatcttgata  13140
aagctttacc tttcgagaag caagcaggct tagcaaatga aagtcttcag aaaaaactgt  13200
atgctttctc tcagggcaac atgagatcct taaggaatct catatatcag gcaagtattg  13260
aagctattga caaccaacat gaaaccatta cagaggagga ttttgtgttt gcttcaaaat  13320
tgacaagcgg agacaaacca aattcatgga aaaatccttt tgaagaaggt gttgaagtga  13380
cagaagacat gcttcgtcca ccacccaaag atattggatg ggaagattac ctaaggcatt  13440
ctactccccg tgtatccaag ccgggcagaa acaagaattt cttcgaatga gcttatatga  13500
agatgaagat gaaatatttg gtgtgtcaaa taaaaaggtt gtgtgcttaa gtttgtgttt  13560
ttttcttggc ttgttgtgtt atgaatttgt ggctttttct aatattaaat gaatgtaaca  13620
tctcattata atgaataaac aaatgtttct ataatccatt gtgaatgttt tgttggatct  13680
cttctccagc atataactac tgtatgtgct atggtatgga ctatggaata tgattaaaga  13740
taagcgctat acaggagtgg cagacatact gtcccacaaa tgaagatgga atctgtaaaa  13800
gaaaacgcgt gaaataatgc gtctgacaaa ggttaggtcg gctgccttta atcaatacca  13860
aagtggtccc taccacgatg gaaaaactgt gcagtcggtt tggctttttc tgacgaacaa  13920
ataagattcg tggccgacag gtgggggtcc accatgtgaa ggcatcttca gactccaata  13980
atggagcaat gacgtaaggg cttacgaaat aagtaagggt agtttgggaa atgtccactc  14040
acccgtcagt ctataaatac ttagcccctc cctcattgtt aagggagcaa aatctcagag  14100
agatagtcct agagagagaa agagagcaag tagcctagaa gtagtcaagg cggcgaagta  14160
ttcaggcacg tggccaggaa gaagaaaagc caagacgacg aaaacaggta agagctaagc  14220
ttaatgggaa agcgtactgc tgatggttca gagtttgaaa gtccaaagaa gaaaagaaaa  14280
gttggtagtg gtatgtttct tcaaagacct aaaccgtata gcgacgaaag tctggaatct  14340
ttcttcatcc gtgttgccaa caagaatggt tatggagatg ttcacaggtt tttggaagcc  14400
acaaagcgat tccttcaaga cattgatcac aatggatacc aaacatttcc cactgatata  14460
acgcgtatca atccttattc agccaaaaac tcttcatcag caagaacagc gagtttcttg  14520
aagctagcac agctcacctt caacgaacca ccagaattgc tagggcttgc aataaacaga  14580
accaacatga agtactctcc ttctactagt gcagtggtaa gaggagcaga ggttttccca  14640
agatccctcc tgcgtactca ttccattcca tgttgccctc tctgcttgag ggaaaatggg  14700
tatgctagtt acctttggca ttttcaagga tacgagtatt gtcactctca taatgtccct  14760
ctaattacca cttgttcttg tggtaaagaa ttcgattaca gagtatctgg gttaaaagga  14820
atttgttgta agtgcaaaga gccaataact ttgacatcga gagagaatgg ccatgaagca  14880
gcttgtactg ttagtaattg gcttgccggt cacgagtcaa aaccattgcc taatttaccc  14940
aaatcttata gatggggact cgtccattgg tggatgggaa tcaaggactc tgaatttgat  15000
catttcagct ttgtgcagtt tttctcaaac tggcctcggt cgtttcatag tatcatagaa  15060
gatgaagttg aatttaatct tgaacatgct gttgtttcta catctgagct aaggctcaaa  15120
gatcttcttg ggagactgtt ctttggtagc atcaggcttc ctgagagaaa tctgcaacat  15180
aacattattt tgggtgagct gttatgttat ctagaaaacc gtttgtggca agataaaggc  15240
ttgatagcca atttaaagat gaatgctctt gaggcaacag tgatgcttaa ctgctccttg  15300
gatcagattg cttcaatggt ggagcaaaga attttaaagc ctaaccgaaa atcaaagcca  15360
aattcaccgc ttgatgtgac ggattatttg tttcactttg gtgacatttt ctgcttatgg  15420
ttagctgagt ttcagtctga tgaattcaat cgaagctttt atgtttctag gtggtgagct  15480
tatatgaaga tgaagatgaa atatttggtg tgtcaaataa aaaggttgtg tgcttaagtt  15540
tgtgtttttt tcttggcttg ttgtgttatg aatttgtggc tttttctaat attaaatgaa  15600
tgtaacatct cattataatg aataaacaaa tgtttctata atccattgtg aatgttttgt  15660
tggatctctt ctccagcata taactactgt atgtgctatg gtatggacta tggaatatga  15720
ttaaagataa gcgctacaga ggagtctgag caggacaact cgcgtagtga gagttacatg  15780
ttcgttgggt tcttccgaca cggacctgag ttggccaacg tcccacctga ggtctgtgcc  15840
ccggtgatga gaagtgtgca tctcgttctt gcagctcgtc agtactttca gaatcatggc  15900
gtgcatggta gaatgaccct tataacggac ttcgacatgg caatcgctag gtaaatgaat  15960
gaggagtggc agacatactg tcccacaaat gaagatggaa tctgtaaaag aaaacgcgtg  16020
aaataatgcg tctgacaaag gttaggtcgg ctgcctttaa tcaataccaa agtggtccct  16080
accacgatgg aaaaactgtg cagtcggttt ggctttttct gacgaacaaa taagattcgt  16140
ggccgacagg tgggggtcca ccatgtgaag gcatcttcag actccaataa tggagcaatg  16200
acgtaagggc ttacgaaata agtaagggta gtttgggaaa tgtccactca cccgtcagtc  16260
tataaatact tagcccctcc ctcattgtta agggagcaaa atctcagaga gatagtccta  16320
gagagagaaa gagagcaagt agcctagaag tagtcaaggc ggcgaagtat tcaggcacgt  16380
ggccaggaag aagaaaagcc aagacgacga aaacaggtaa gagctaagct taatgggaaa  16440
gcgtactgct gatggttcag agtttgaaag tccaaagaag aaaagaaaag ttggtagtgg  16500
tgtgaaatgg tactacaaaa caataacatt tctacctgag ctctgcaaca atgaatcact  16560
agcagcgaaa tgtctgcgtg tacttcatgg attcaactac caatatgaaa caagaaacat  16620
tggtgttagc ttccctcttt ggtgtgacgc aactgtcgga aagaagatat cctttgtgtc  16680
caagaacaag attgagcttg acttgcttct caaacaacat tactttgttc aaatggagca  16740
gcttcagtat ttccacatct caaatactgt gctggttcca gaagattgca cttatgtcag  16800
tttcaggaga tgtcaaagta ttgataagct tactgctgct ggtttagcaa gaaagatccg  16860
tcgattggag aagagagctt tgtcaagagg agaacagttt gatccctcaa gttttgctca  16920
gaaagaacac accgccattg ctcattatca ttctctcggt gaatcttcaa agcaaacaaa  16980
taggaatttt cggttaaata taaggatgtt gagtgaacaa ccaagagaag gaaattccat  17040
cttctcttct tatgggctat caaattctga gaattctttt cagcctgttc cgctaatctg  17100
agcttatatg aagatgaaga tgaaatattt ggtgtgtcaa ataaaaaggt tgtgtgctta  17160
agtttgtgtt tttttcttgg cttgttgtgt tatgaatttg tggctttttc taatattaaa  17220
tgaatgtaac atctcattat aatgaataaa caaatgtttc tataatccat tgtgaatgtt  17280
```

-continued

```
ttgttggatc tcttctccag catataacta ctgtatgtgc tatggtatgg actatggaat  17340
atgattaaag ataagcgcta gcaaggagtg gcagacatac tgtcccacaa atgaagatgg  17400
aatctgtaaa agaaaacgcg tgaaataatg cgtctgacaa aggttaggtc ggctgccttt  17460
aatcaatacc aaagtggtcc ctaccacgat ggaaaaactg tgcagtcggt ttggcttttt  17520
ctgacgaaca aataagattc gtggccgaca ggtgggggtc caccatgtga aggcatcttc  17580
agactccaat aatggagcaa tgacgtaagg gcttacgaaa taagtaaggg tagtttggga  17640
aatgtccact cacccgtcag tctataaata cttagcccct ccctcattgt taagggagca  17700
aaatctcaga gagatagtcc tagagagaga aagagagcaa gtagcctaga agtagtcaag  17760
gcggcgaagt attcaggcac gtggccagga agaagaaaag ccaagacgac gaaaacaggt  17820
aagagctaag cttaatggga aagcgtactg ctgatggttc agagtttgaa agtccaaaga  17880
agaaaagaaa agttggtagt ggtatgaaac ttccaacaaa tctagcttac gagaggtcga  17940
ttgatccttc agatgtttgt ttctttgtgg tctggccaga tgatcggaaa acaccgttga  18000
catacaattc acgtaccctc ctagggcaga tggaggctgc atctctggcg tatgatgtaa  18060
gtggtcaacc gatcaagagt gccacagcag aagcattagc tcaaggaaat cctcatcaag  18120
ttgacttttg tcatgttcct tatggagcat ctcacatcga gtgcagcttc tctgtttctt  18180
tttcatcaga gcttcgccag ccttacaaat gtaacagcag taaagtgaag caaactctcg  18240
ttcagctcgt agagttgtat gaaactaaga ttggatggac tgaactagcg acgagatatt  18300
tgatgaacat atgcaatggt aaatggctct ggaaaaatac aaggaaggct tattgttgga  18360
atatagtcct gactccatgg ccttggaacg gagagaaggt tggctttgaa gacattagga  18420
caaactacac ttcaagacaa gatttcaaga acaacaagaa ttggtctgcc attgttgaga  18480
tgattaagac agctttcagt tctactgatg gcttagcgat cttcgaggtt cgtgcgactc  18540
ttcacttacc aaccaatgct atggttagac cttcccaggt attcacagag aaagaatcag  18600
ggtcgaaatc aaaaagcaag actcagaatt ctcgagtgtt tcaatcaaca accatagatg  18660
gtgaaagatc tcccattctt ggtgcattca aaactggtgc agctatagct actattgacg  18720
actggtatcc tgaagctaca gaaccattgc gtgttggaag atttggtgtc catagagaag  18780
atgtgacttg ctatagacat ccttccacag ggaaagattt tttctccata cttcaacaag  18840
ctgaacatta cattgaagta ctttcagcaa acaaaacccc agcacaagaa accattaatg  18900
atatgcattt tcttatggct aatctcatca aaggtggcat gtttcagcac aaaggagact  18960
gagcttatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  19020
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  19080
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  19140
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  19200
tatgattaaa gataagcgct atacaggagt ggcagacata ctgtcccaca aatgaagatg  19260
gaatctgtaa aagaaaacgc gtgaaataat gcgtctgaca aaggttaggt cggctgcctt  19320
taatcaatac caaagtggtc cctaccacga tggaaaaact gtgcagtcgg tttggctttt  19380
tctgacgaac aaataagatt cgtggccgac aggtgggggt ccaccatgtg aaggcatctt  19440
cagactccaa taatggagca atgacgtaag ggcttacgaa ataagtaagg gtagtttggg  19500
aaatgtccac tcacccgtca gtctataaat acttagcccc tccctcattg ttaagggagc  19560
aaaatctcag agagatagtc ctagagagag aaagagagca agtagcctag aagtagtcaa  19620
ggcggcgaag tattcaggca cgtggccagg aagaagaaaa gccaagacga cgaaaacagg  19680
taagagctaa gcttaatggg aaagcgtact gctgatggtt cagagtttga aagtccaaag  19740
aagaaaagaa aagttggtag tggtatgcag actcttaaag aactaattgc ttctaatccg  19800
gatgacctaa caacggagct caaaagggca tttcgccccc tgactcctca cattgccatt  19860
gacggtaacg agcttgacgc tctcaccatt cttgtgaatc tcacagacaa aactgatgat  19920
caaaaggatc ttctcgatag agctaaatgt aaacagaaac tcagagatga gaaatggtgg  19980
gcttcttgca tcaattgcgt taattataga cagtctcaca atcctaaatt ccctgatatc  20040
aggtctgaag gagtgattcg cactcaagct ttgggcgaac ttccttcatt tctattaagc  20100
tccagtaaaa taccaccgta tcattggtca tattctcatg attcaaaata tgttaacaag  20160
tcagcctttt taacaaatga attctgttgg gacggagaga taagctgctt gggggaactc  20220
ctcaaagatg cggatcatcc actttggaac acactaaaga agcttggctg ctcacagaag  20280
acatgcaaag ctatggcgaa acaattggct gatatcacat taacaacgat taatgtaaca  20340
ctcgctccga attacttgac tcagatttct cttccggatt ctgatacaag ttacatttct  20400
ctctcgcctg tcgcaagtct atcaatgcaa agtcatttcc accaacgttt gcaagacgaa  20460
aacagacatt ctgctataac aagattctcc cgaacaacta acatgggtgt gactgcaatg  20520
acttgtggtg gagctttcag gatgttaaag tctggtgcaa agttctccag tccacctcat  20580
catagattaa actcgaaacg ctcttggttg accagtgagc atgttcaatc attgaagcag  20640
tatcaacgtc tgaacaaatc attgatccct gagaatagta ggattgctct taggcgcaag  20700
tacaagattg aacttcaaaa tatggttcga tcatggtttg ccatgcaaga tcacactctc  20760
gattcaaata tccttattca acatctcaat cacgatttga gttatctcgg ggcgaccaaa  20820
agatttgctt atgatccagc tatgacaaaa ttgttcactg agctgctaaa gagggagctg  20880
agtaattcta tcaacaacgg agaacaacat accaatggca gttttctggt cctcccaaac  20940
atacgggttt gtggagcaac tgcgttatct tccccagtaa ccgttggcat ccctagctta  21000
accgccttct ttggttttgt tcacgccttt gagaggaaca tcaacaggac cacttcttcc  21060
tttcgtgtcg agagctttgc aatatgtgtc catcagcttc atgtggagaa gcgtggtttg  21120
acagctgagt ttgtagagaa gggagatggg acaatttcag caccagcaac cagagatgat  21180
tggcaatgtg atgttgtgtt ttctctaatt ttgaatacga acttcgctca acacattgat  21240
caggacacac tggtcacgag tttacctaag cggctggcaa ggggttctgc taagatcgca  21300
atagatgatt tcaagcatat aaactcattc tccactcttg aaactgcaat tgagtctctt  21360
cccatcgaag ctggaagatg gcttagtctg tatgcacaga gtaataataa tctttcggac  21420
ttgttagctg ctatgaccga agaccatcaa cttatggcta gttgtgttgg ctaccacctt  21480
ctagaagaac caaaagataa gccaaatagc cttagaggat acaaacatgc catagctgaa  21540
tgtattatag gattaattaa ttcaattaca ttttcctcgg aaaccgatcc taatactata  21600
ttctggtcct taaagaacta ccaaaactat cttgtagttc agcccagatc cataaatgat  21660
gaaacaacag acaagagcag cctgtgagct tatatgaaga tgaagatgaa atatttggtg  21720
tgtcaaataa aaaggttgtg tgcttaagtt tgtgtttttt tcttggcttg ttgtgttatg  21780
aatttgtggc tttttctaat attaaatgaa tgtaacatct cattataatg aataaacaaa  21840
```

-continued

```
tgtttctata atccattgtg aatgttttgt tggatctctt ctccagcata taactactgt  21900
atgtgctatg gtatggacta tggaatatga ttaaagataa gcgctacaga ggagtctgag  21960
caggacaact cgcgtagtga gagttacatg ttcgttgggt tcttccgaca cggacctgag  22020
ttggccaacg tcccacctga ggtctgtgcc ccggtgatga gaagtgtgca tctcgttctt  22080
gcagctcgtc agtactttca gaatcatggc gtgcatggta gaatgaccct tataacggac  22140
ttcgacatgg caatcgctag gtagcttaat gaataggcgt atcacgaggc cctttcgtct  22200
tcaagaatta gcttggcaat tcccgatcta gtaacataga tgacaccgcg cgcgataatt  22260
tatcctagtt tgcgcgctat attttgtttt ctatcgcgta ttaaatgtat aattgcggga  22320
ctctaatcat aaaaacccat ctcataaata acgtcatgca ttacatgtta attattacat  22380
gcttaacgta attcaacaga aattatatga taatcgtcgc aagaccggca acaggattca  22440
atcttaagaa actttattgc caaatgtttg aacgatcggg atcgaagatc cattccaagc  22500
ctacaggaac aggtggtggc ggccctcggt gcgctcgtac tgctccacga tggtgtagtc  22560
ctcgttgtgg gaggtgatgt ccagcttggc gtccacgtag tagtagccgg gcagctgcac  22620
gggcttcttg gccatgtaga tggacttgaa ctccaccagg tagtggccgc cgtccttcag  22680
cttcagggcc ttgtgtgtct cgcccttcag cacgccgtcg cgggggtaca ggcgctcggt  22740
ggaggcctcc cagcccatcg tcttcttctg catcacgggg ccgtcggagg ggaagttcac  22800
gccgatgaac ttcaccttgt agatgaagca gccgtcctgc agggaggagt cctgggtcac  22860
ggtcgccacg ccgccgtcct cgaagttcat cacgcgctcc cacttgaagc cctcggggaa  22920
ggacagcttc ttgtagtcgg ggatgtcggc ggggtgcttc acgtacacct tggagccgta  22980
ctggaactgg ggggacagga tgtcccaggc gaagggcagg gggccgccct tggtcacctt  23040
cagcttcacg gtgttgtggc cctcgtaggg gcggccctcg ccctcgccct cgatctcgaa  23100
ctcgtggccg ttcacggtgc cctccatgcg caccttgaag cgcatgaact cggtgatgac  23160
gttctcggag gaggccattt cgcctcgatc gagttgagag tgaatatgag actctaattg  23220
gataccgagg ggaatttatg gaacgtcagt ggagcatttt tgacaagaaa tatttgctag  23280
ctgatagtga ccttaggcga cttttgaacg cgcaataatg gtttctgacg tatgtgctta  23340
gctcattaaa ctccagaaac ccgcggctca gtggctcctt caacgttgcg gttctgtcag  23400
ttccaaacgt aaaacggctt gtcccgcgtc atcggcgggg gtcataccgt gactccctta  23460
attctcatgt atggttcgct atactgaaga gcaattctga caggatatat tggcgggtaa  23520
acctaagaga aaagagcgtt tattagaata atcggatatt taaaagggcg tgaaaaggtt  23580
tatccgttcg tccatttgta tgtgcatgcc aaccacaggg ttcccctcgg gatcaaagta  23640
ctttgatcca acccctccgc tgctatagtg cagtcggctt ctgacgttca gtgcagccgt  23700
catctgaaaa cgacatgtcg cacaagtcct aagttacgcg acaggctgcc gccctgccct  23760
tttcctggcg ttttcttgtc gcgtgtttta gtcgcataaa gtagaatact tgcgactaga  23820
accggagaca ttacgccatg aacaagagcg ccgccgctgg cctgctgggc tatgcccgcg  23880
tcagcaccga cgaccaggac ttgaccaacc aacgggccga actgcacgcg gccggctgca  23940
ccaagctgtt ttccgagaag atcaccggca ccaggcgcga ccgcccggag ctggccagga  24000
tgcttgacca cctacgccct ggcgacgttg tgacagtgac caggctagac cgcctggccc  24060
gcagcacccg cgacctactg gacattgccg agcgcatcca ggaggccggc gcgggcctgc  24120
gtagcctggc agagccgtgg gccgacacca ccacgccggc cggccgcatg gtgttgaccg  24180
tgttcgccgg cattgccgag ttcgagcgtt ccctaatcat cgaccgcacc cggagcgggc  24240
gcgaggccgc caaggcccga ggcgtgaagt ttggcccccg ccctaccctc accccggcac  24300
agatcgcgca cgcccgcgag ctgatcgacc aggaaggccg caccgtgaaa gaggcggctg  24360
cactgcttgg cgtgcatcgc tcgaccctgt accgcgcact tgagcgcagc gaggaagtga  24420
cgcccaccga ggccaggcgg cgcggtgcct tccgtgagga cgcattgacc gaggccgacg  24480
ccctggcggc cgccgagaat gaacgccaag aggaacaagc atgaaaccgc accaggacgg  24540
ccaggacgaa ccgtttttca ttaccgaaga gatcgaggcg gagatgatcg cggccgggta  24600
cgtgttcgag ccgcccgcgc acctctcaac cgtgcggctg catgaaatcc tggccggttt  24660
gtctgatgcc aagctggcgg cctggccggc cagcttggcc gctgaagaaa ccgagcgccg  24720
ccgtctaaaa aggtgatgtg tatttgagta aaacagcttg cgtcatgcgg tcgctgcgta  24780
tatgatgcga tgagtaaata aacaaatacg caaggggaac gcatgaaggt tatcgctgta  24840
cttaaccaga aaggcgggtc aggcaagacg accatcgcaa cccatctagc ccgcgccctg  24900
caactcgccg gggccgatgt tctgttagtc gattccgatc cccagggcag tgcccgcgat  24960
tgggcggccg tgcgggaaga tcaaccgcta accgttgtcg gcatcgaccg cccgacgatt  25020
gaccgcgacg tgaaggccat cggccggcgc gacttcgtag tgatcgacgg agcgccccag  25080
gcggcggact tggctgtgtc cgcgatcaag gcagccgact tcgtgctgat tccggtgcag  25140
ccaagccctt acgacatatg ggccaccgcc gacctggtgg agctggttaa gcagcgcatt  25200
gaggtcacgg atggaaggct acaagcggcc tttgtcgtgt cgcgggcgat caaaggcacg  25260
cgcatcggcg gtgaggttgc cgaggcgctg gccgggtacg agctgcccat tcttgagtcc  25320
cgtatcacgc agcgcgtgag ctacccaggc actgccgccg ccggcacaac cgttcttgaa  25380
tcagaacccg agggcgacgc tgcccgcgag gtccaggcgc tggccgctga aattaaatca  25440
aaactcattt gagttaatga ggtaaagaga aaatgagcaa aagcacaaac acgctaagtg  25500
ccggccgtcc gagcgcacgc agcagcaagg ctgcaacgtt ggccagcctg gcagacacgc  25560
cagccatgaa gcgggtcaac tttcagttgc cggcggagga tcacaccaag ctgaagatgt  25620
acgcggtacg ccaaggcaag accattaccg agctgctatc tgaatacatc gcgcagctac  25680
cagagtaaat gagcaaatga ataaatgagt agatgaattt tagcggctaa aggaggcggc  25740
atggaaaatc aagaacaacc aggcaccgac gccgtggaat gccccatgtg tggaggaacg  25800
ggcggttggc caggcgtaag cggctgggtt gcctgccggc cctgcaatgg cactggaacc  25860
cccaagcccg aggaatcggc gtgagcggtc gcaaaccatc cggcccggta caaatcggcg  25920
cggcgctggg tgatgacctg gtggagaagt tgaaggccgc gcaggccgcc cagcggcaac  25980
gcatcgaggc agaagcacgc cccggtgaat cgtggcaagc ggccgctgat cgaatccgca  26040
aagaatcccg gcaaccgccg gcagccggtg cgccgtcgat taggaagccg cccaagggcg  26100
acgagcaacc agatttttc gttccgatgc tctatgacgt gggcacccgc gatagtcgca  26160
gcatcatgga cgtggccgtt ttccgtctgt cgaagcgtga ccgacgagct ggcgaggtga  26220
tccgctacga gcttccagac gggcacgtag aggtttccgc agggccggcc ggcatggcca  26280
gtgtgtggga ttacgacctg gtactgatgg cggtttccca tctaaccgaa tccatgaacc  26340
gataccggga agggaaggga gacaagcccg gccgcgtgtt ccgtccacac gttgcggacg  26400
```

-continued

```
tactcaagtt ctgccggcga gccgatggcg gaaagcagaa agacgacctg gtagaaacct  26460
gcattcggtt aaacaccacg cacgttgcca tgcagcgtac gaagaaggcc aagaacggcc  26520
gcctggtgac ggtatccgag ggtgaagcct tgattagccg ctacaagatc gtaaagagcg  26580
aaaccgggcg gccggagtac atcgagatcg agctagctga ttggatgtac cgcgagatca  26640
cagaaggcaa gaacccggac gtgctgacgg ttcaccccga ttactttttg atcgatcccg  26700
gcatcggccg ttttctctac cgcctggcac gccgcgccgc aggcaaggca gaagccagat  26760
ggttgttcaa gacgatctac gaacgcagtg gcagcgccgg agagttcaag aagttctgtt  26820
tcaccgtgcg caagctgatc gggtcaaatg acctgccgga gtacgatttg aaggaggagg  26880
cggggcaggc tggcccgatc ctagtcatgc gctaccgcaa cctgatcgag ggcgaagcat  26940
ccgccggttc ctaatgtacg gagcagatgc tagggcaaat tgccctagca ggggaaaaag  27000
gtcgaaaagg actctttcct gtggatagca cgtacattgg gaacccaaag ccgtacattg  27060
ggaaccggaa cccgtacatt gggaacccaa agccgtacat tgggaaccgg tcacacatgt  27120
aagtgactga tataaaagag aaaaaaggcg atttttccgc ctaaaactct ttaaaactta  27180
ttaaaactct taaaacccgc ctggcctgtg cataactgtc tggccagcgc acagccgaag  27240
tgctgcaaaa agcgcctacc cttcggtcgc tgcgctccct acgccccgcc gcttcgcgtc  27300
ggcctatcgc ggccgctggc cgctcaaaaa tggctggcct acggccaggc aatctaccag  27360
ggcgcggaca agccgcgccg tcgccactcg accgccggcg cccacatcaa ggcaccctgc  27420
ctcgcgcgtt tcggtgatga cggtgaaaac ctctgacaca tgcagctccc ggtgacggtc  27480
acagcttgtc tgtaagcgga tgccgggagc agacaagccc gtcagggcgc gtcagcgggt  27540
gttggcgggt gtcggggcgc agccatgacc cagtcacgta gcgatagcgg agtgtatact  27600
ggcttaacta tgcggcatca gagcagattg tactgagagt gcaccatatg cggtgtgaaa  27660
taccgcacag atgcgtaagg agaaaatacc gcatcaggcg ctcatccgct tcctcgctca  27720
ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg  27780
taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc  27840
agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc  27900
cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac  27960
tataaagata ccaggcgttt ccccctggaa gctccctcgt gcgctctcct gttccgaccc  28020
tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata  28080
gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc  28140
acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca  28200
acccggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag  28260
cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta  28320
gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg  28380
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggtttttt gtttgcaagc  28440
agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt  28500
ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgcat tctaggtgat  28560
tagaaaaact catcgagcat caaatgaaac tgcaatttat tcatatcagg attatcaata  28620
ccatattttt gaaaaagccg tttctgtaat gaaggagaaa actcaccgag gcagttccat  28680
aggatggcaa gatcctggta tcggtctgcg attccgactc gtccaacatc aatacaacct  28740
attaatttcc cctcgtcaaa aataaggtta tcaagtgaga aatcaccatg agtgacgact  28800
gaatccggtg agaatggcaa aagtttatgc atttctttcc agacttgttc aacaggccag  28860
ccattacgct cgtcatcaaa atcactcgca tcaaccaaac cgttattcat tcgtgattgc  28920
gcctgagcga gtcgaaatac gcgatcgctg ttaaaaggac aattacaaac aggaatcgaa  28980
tgcaaccggc gcaggaacac tgccagcgca tcaacaatat tttcacctga atcaggatat  29040
tcttctaata cctggaatgc tgttttccct gggatcgcag tggtgagtaa ccatgcatca  29100
tcaggagtac ggataaaatg cttgatggtc ggaagaggca taaattccgt cagccagttt  29160
agtctgacca tctcatctgt aacatcattg gcaacgctac ctttgccatg tttcagaaac  29220
aactctggcg catcgggctt cccatacaat cggtagattg tcgcacctga ttgcccgaca  29280
ttatcgcgag cccatttata cccatataaa tcagcatcca tgttggaatt taatcgcggc  29340
cttgagcaag acgtttcccg ttgaatatgg ctcataacag aacttattat ttccttcctc  29400
ttttctacag tatttaaaga taccccaaga agctaattat aacaagacga actccaattc  29460
actgttcctt gcattctaaa accttaaata ccagaaaaca gctttttcaa agttgttttc  29520
aaagttggcg tataacatag tatcgacgga gccgattttg aaaccgcggt gatcacaggc  29580
agcaacgctc tgtcatcgtt acaatcaaca tgctaccctc cgcgagatca tccgtgtttc  29640
aaacccggca gcttagttgc cgttcttccg aatagcatcg gtaacatgag caaagtctgc  29700
cgccttacaa cggctctccc gctgacgccg tcccggactg atgggctgcc tgtatcgagt  29760
ggtgattttg tgccgagctg ccggtcgggg agctgttggc tggctgg                29807
```

```
SEQ ID NO: 41          moltype = DNA  length = 29999
FEATURE                Location/Qualifiers
source                 1..29999
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 41
tggcaggata tattgtggtg taaacataac gaattgctct tcagcaagga gaatgaggag  60
aatgtagcag aaggcatgtt gttgtgactc cgaggggttg cctcaaactc tatcttataa  120
ccggcgtgga ggcatggagg caggggtatt ttggtcattt taatagatag tggaaaatga  180
cgtggaattt acttaaagac gaagtctttg cgacaagggg gggcccacgc cgaatttaat  240
attaccggcg tggcccccc ttatcgcgag tgctttagca cgagcggtcc agatttaaag  300
tagaaaattt cccgcccact agggttaaag gtgttcacac tataaaagca tatacgatgt  360
gatggtattt gatggagcgt atattgtatc aggtatttcc gttggatacg aattattcgt  420
acgaccctcg gtatactaat gaggagtgtc gctgcaacca tagtttgaca taatgcaacc  480
ataccttgac ataattaaag tcatactctg acataaagcc ccttattaaa aatgatgcta  540
aaaataaatt ccattataat caattattta agaattaatt aaaccatagt gtgacatagt  600
ggctttttat ttgtatagta tgtgtactca aataataaaa atttacttat gtatatacgc  660
aacttaagaa agccatctcc aaataaaaat gtatttcgct agcaaggagt gagacttttc  720
```

-continued

```
aacaaagggt aatatcggga aacctcctcg gattccattg cccagctatc tgtcacttca  780
tcaaaaggac agtagaaaag gaaggtggca cctacaaatg ccatcattgc gataaaggaa  840
aggctatcgt tcaagatgcc tctgccgaca gtggtcccaa agatggaccc ccacccacga  900
ggagcatcgt ggaaaaagaa gacgttccaa ccacgtcttc aaagcaagtg gattgatgtg  960
atatctccac tgacgtaagg gatgacgcac aatcccacta tccttcgcaa gacccttcct  1020
ctatataagg aagttcattt catttggaga ggacacgctg acggccgcgt atttttacaa  1080
caattaccaa caacaacaaa caacaaacaa cattacaatt actatttaca attacagcgg  1140
ccgccccctt caccaaaaaa atggctgcag ccaagggcga ggagctgttc accggggtgg  1200
tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg  1260
agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gctgctctgc accaccggca  1320
agctgcccgt gccctggccc accctcgtga ccaccctggg ctacggcgtg cagtgcttcg  1380
cccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg cccgaaggct  1440
acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg  1500
tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcatc gacttcaagg  1560
aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac aacgtctata  1620
tcaccgccga caagcagaag aacggcatca aggccaactt caagatccgc cacaacatcg  1680
aggacggcgg cgtgcagctc gccgaccact accagcagaa cacccccatc ggcgacggcc  1740
ccgtgctgct gcccgacaac cactacctga gctaccagtc caaactgttc aaagacccca  1800
acgagaagcg cgatcacatg gtcttgttag aatttttgac tgctgctggt attaccgagg  1860
gtatgaatga attgtacaaa ggctccggac cgaagaagaa gaggaaggtt taaatatgaa  1920
gatgaagatg aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt  1980
tttcttggct tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat  2040
ctcattataa tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc  2100
ttctccagca tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat  2160
aagcgctata caggagccgt ataggcagct gaagataaaa acaatcgatt cgatgtttaa  2220
gttaatgtaa ttctgccgaa aaggcagtga gtagtaattg aaaaaaccag tcttatatcc  2280
gggctggttt ttttgttttt agcgttataa cttttatttc actatgtcac tgtatggttg  2340
cattttatgt cagtgaatgg ttgcaggttt tttgtaagtt attgatatgc ctatccagtg  2400
tatgtcacgc tttgcttgca gcgacacgct acagaggagt ctgagcagga caactcgcgt  2460
agtgagagtt acatgttcgt tgggttcttc cgacacggac ctgagttggc caacgtccca  2520
cctgaggtct gtgccccggt gatgagaagt gtgcatctcg ttcttgcagc tcgtcagtac  2580
tttcagaatc atggcgtgca tggtagaatg acccttataa cggacttcga catggcaatc  2640
gctaggtaaa tgaatggaaa aagtggcacc gagtcggtgc tttttttccc ggggcgcgcc  2700
gatatcgagc tctcccggcg cgccgatatc gagctcagtg tttgatcgcc ggcggtaccg  2760
agtgtacttc aagtcagtgg gaaatcaata aaatgattat tttatgaata tatttcattg  2820
tgcaagtaga tagaaattac atatgttaca taacacacga aataaacaaa aaaagacaat  2880
ccaaaaacaa acaccccaaa aaaaataatc actttagata aactcgtatg aggagaggca  2940
cgttcagtga ctcgacgatt cccgagcaaa aaaagtctcc ccgtcacaca tgtagtgggt  3000
gacgcaatta tctttaaagt aatccttctg ttgacttgtc attgataaca tccagtcttc  3060
gtcaggattg caaagaatta tagaagggat cccacctttt attttcttct tttttccata  3120
tttagggttg acagtgaaat cagactggca acctattaat tgcttccaca atgggacgaa  3180
cttgaagggg atgtcgtcga tgatattata ggtggcgtgt tcatcgtagt tggtgaaatc  3240
gatggtaccg ttccaatagt tgtgtcgtcc gagacttcta gcccaggtgg tctttccggt  3300
acgagttggt ccgcagatgt agaggctggg gtgtcggatt ccattccttc cattgtcctt  3360
gttaaatcgg ccatccattc aaggtcagat tgagcttgtt ggtatgagac aggatgtatg  3420
taagtataag cgtctatgct tacatggtat agatgggttt ccctccagga gtgtagatct  3480
tcgtggcagc gaagatctga ttctgtgaag ggcgacacat acggttcagg ttgtggaggg  3540
aataatttgt tggctgaata ttccagccat tgaagctttg ttgcccattc atgagggaat  3600
tcttccttga tcatgtcaag atattcctcc ttagacgttg cagtctggat aatagttctc  3660
catcgtgcgt cagatttgcg aggagaaacc ttatgatctc ggaaatctcc tctggtttta  3720
atatctccgt cctttgatat gtaatcaagg acttgtttag agtttctagc tggctggata  3780
ttagggtgat ttccttcaaa atcgaaaaaa gaaggatccc taatacaagg ttttttatca  3840
agctggagga gagcatgata gtgggtagtg ccatcttgat gaagctcaga agcaacacca  3900
aggaagaaaa taagaaaagg tgtgagtttc tcccagagaa actggaataa atcatctctt  3960
tgagatgagc acttgggata ggtaaggaaa acatatttag attggagtct gaagttctta  4020
ctagcagaag gcatgttgtt gtgactccga ggggttgcct caaactctat cttataaccg  4080
gcgtggaggc atggaggcag gggtattttg gtcattttaa tagatagtgg aaaatgacgt  4140
ggaatttact taaagacgaa gtctttgcga caaggggggg cccacgccga atttaatatt  4200
accggcgtgg cccccccttа tcgcgagtgc tttagcacga gcggtccaga tttaaagtag  4260
aaaatttccc gcccactagg gttaaaggtg ttcacactat aaaagcatat acgatgtgat  4320
ggtatttgat ggagcgtata ttgtatcagg tatttccgtt ggatacgaat tattcgtacg  4380
accctcatag tttaaactat cagtgggtag cttaatggta gacactttat tagttagatc  4440
cgctgttatt ctgcacgatt gcgcttgccg gtggaaattt aaggacgagt gtagtgcact  4500
cacccttagc cttcagcgtg acaagattag ggtagcaccg gaaacagcga ttttacggac  4560
agatagtaca gtctgagact cagtcactgg tccctaaacg ggagccggta cgtcggcggt  4620
tcgctagcaa ggagtggcag acatactgtc ccacaaatga agatggaatc tgtaaaagaa  4680
aacgcgtgaa ataatgcgtc tgacaaaggt taggtcggct gcctttaatc aataccaaag  4740
tggtccctac cacgatggaa aaactgtgca gtcggtttgg ctttttctga cgaacaaata  4800
agattcgtgg ccgacaggtg ggggtccacc atgtgaaggc atcttcagac tccaataatg  4860
gagcaatgac gtaagggctt acgaaataag taagggtagt ttgggaaatg tccactcacc  4920
cgtcagtcta taaatactta gcccctccct cattgttaag ggagcaaaat ctcagagaga  4980
tagtcctaga gagagaaaga gagcaagtag cctagaagta gtcaaggcgg cgaagtattc  5040
aggcacgtgg ccaggaagaa gaaaagccaa gacgacgaaa acaggtaaga gctaagctta  5100
atgggaaagc gtactgctga tggttcagag tttgaaagtc caaagaagaa aagaaaagtt  5160
ggtagtggta tgactgataa acggaaagat ggttccggca agcttcttta ctgcagcttt  5220
tgcgggaaat ctcaacatga agttaggaaa cttattgcag ggccttctgt ttatatctgt  5280
```

-continued

```
gacgaatgtg ttgatctctg caatgatatt attcgagaag agatcaaaga agtggctcct  5340
cacagagaga ggagtgcgtt gccgacacca cacgagatta gaaatcatct tgatgattat  5400
gtcattggtc aagaacaagc aaagaaggtc cttgctgttg cagtgtacaa ccattacaag  5460
aggctcagaa acggagatac ttctaatggt gtggaacttg gcaaatcaaa catcctgctc  5520
attggaccca ctggaagtgg aaagacactc ctagcagaga cattggcaag gctcttagac  5580
gttcctttca ctatggctga tgcaactact cttactgagg ctggttatgt tggagaagat  5640
gtggagaata ttattcaaaa gttattgcag aaatgtgatt acgatgtcca gaaagctcag  5700
agagggatag tctacataga tgagatagac aagatctcta ggaagagtga taatccctcg  5760
ataacacgcg acgtttctgg cgaaggtgtt caacaagctt tgttgaagct cattgaaggc  5820
accgtagcgg cagttccacc acaaggagga agaaagcatc ctcagcaaga gtttctccaa  5880
gtggatacct caaaaatctt atttatatgt ggtggtgcct tcgctgggct tgacaaagtt  5940
atctcacacc gtgtagaaac tggtagcggg attgggtttg gtgcaacggt gaaggcaaaa  6000
tctgataagg cttctgaagg agagcttctg gctcaagttg aaccagagga tcttatcaag  6060
tttggactaa ttcctgagtt catcggtaga cttcctgtcg ttgctacgtt gaacgagtta  6120
agtgaggaag ccctgattca aatcttgaaa gagccaaaga atgctctgac aaagcaatat  6180
caggctcttt tcaatcttga gggagttgac ttggaatttc gtgatgaagc gctggatgct  6240
atagctaaga aagcaatggc gaggaaaacc ggagctcgtg gtctaagatc gattgttgaa  6300
gccgctttat tggatacaat gtatgatttg ccctccatgg aagatgttga gaaagtggtc  6360
atcgacgagt ctgtgattga tggacagtca aagccattgc tcatatatgg taaacctgaa  6420
gcacagcagg catcaggaga atgaatatga agatgaagat gaaatatttg gtgtgtcaaa  6480
taaaaaggtt gtgtgcttaa gtttgtgttt ttttcttggc ttgttgtgtt atgaatttgt  6540
ggctttttct aatattaaat gaatgtaaca tctcattata atgaataaac aaatgtttct  6600
ataatccatt gtgaatgttt tgttggatct cttctccagc atataactac tgtatgtgct  6660
atggtatgga ctatggaata tgattaaaga taagcgctat acaggagtgg cagacatact  6720
gtcccacaaa tgaagatgga atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa  6780
ggttaggtcg gctgccttta atcaatacca aagtggtccc taccacgatg gaaaaactgt  6840
gcagtcggtt tggctttttc tgacgaacaa ataagattcg tggccgacag gtgggggtcc  6900
accatgtgaa ggcatcttca gactccaata atggagcaat gacgtaaggg cttacgaaat  6960
aagtaagggt agtttgggaa atgtccactc acccgtcagt ctataaatac ttagcccctc  7020
cctcattgtt aagggagcaa aatctcagag agatagtcct agagagagaa agagagcaag  7080
tagcctagaa gtagtcaagg cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc  7140
caagacgacg aaaacaggta agagctaagc ttaatgggaa agcgtactgc tgatggttca  7200
gagtttgaaa gtccaaagaa gaaaagaaaa gttggtagtg gtatgagtta ttctggagaa  7260
cgggataatt ttgcacctca catggcatta gttccaatgg tgattgaaca aacaagtaga  7320
ggggagcgaa gctttgacat ttactctcgt ctattaaaag aaagggtcat cttcttaacc  7380
gggcaagttg aagatcatat ggcaaatctg attgtagctc agatgctttt tcttgaagca  7440
gagaatcctg agaaagatat ctatctctac attaactctc ctggtggtgt aataactgct  7500
ggaatgtcta tatatgatac aatgcagttt atcaaacctg atgtttcaac aatatgtatg  7560
ggtcaggctg cgtcaatggg agcttttctt ttgacggctg gcgcaaaagg aaagcggttc  7620
tgcctaccga actcaagggt gatgatccat caaccacttg gaggatacca aggtcaagct  7680
accgatatag agattcatgc tagagaaatt ttgaaggtga agggtagaat gaatgaactg  7740
atggcccttc acactggaca atcgctcgag cagattgagc gtgacacaga gagagacagg  7800
tttttatccg caccagaagc tgttgagtat ggtttggttg attctattct tactcataga  7860
aactgaatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  7920
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  7980
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  8040
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  8100
tatgattaaa gataagcgct acagaggagc atcttcattc ttaagatatg aagataatct  8160
tcaaaaggcc cctgggaatc tgaaagaaga gaagcaggcc catttatatg ggaaagaaca  8220
atagtatttc ttatataggc ccatttaagt tgaaaacaat cttcaaaagt cccacatcgc  8280
ttagataaga aaacgaagct gagtttatat acagctagag tcgaagtagt gataatggtg  8340
acctgccgta taggcagctg aagatttcct ctatataagg aagttcattt catttggtaa  8400
ttctgccgaa aaggcagtga gtagtgcttt tttttttgca aaattttcca gatcgatttc  8460
ttcttcctct gttcttcggc gttcaatttc tggggttttc tcttcgtttt ctgtaactga  8520
aacctaaaat ttgacctaaa aaaaatctca aataatatga ttcagtggtt ttgtactttt  8580
cagttagttg agttttgcag ttccgatgag ataaaccaat acgctaggta tactaatgag  8640
gagtggcaga catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa  8700
taatgcgtct gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc  8760
acgatggaaa aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc  8820
cgacaggtgg gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg  8880
taagggctta cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat  8940
aaatacttag cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag  9000
agagaaagag agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc  9060
caggaagaag aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tgtacattag  9120
gaacttgaga aagccttctc caaataagaa tgtttttaag ttcgcatcta ccaaggtcag  9180
ctctgtggtc atgtgtgaat ccagccttga gtttgacgct tgctttcatc atgaatataa  9240
cgacttgatt gagagttttg gttctcagcc agaaggtttc aaatacgaat tcatgggtaa  9300
atctctccct tacactccag atgctttgat ctcctacacg gacaaaactc aaaagtatca  9360
cgagtataaa ccatattcaa aaatcgcatc acctctattt agagcagagt ttgctgcaaa  9420
gagggctgca agtcttaagt tgggtattga tttggtactc gtcacagaca gacagattag  9480
agtaaaccca atactcaaca atctgaagct ccttcataga tactcgggcg tttatggtat  9540
atctggtatc cagaaggaac tacttagctt tattcacaaa agtggcgtta tcaaactcaa  9600
tgatatctca tcacaggttg ggatccccat tggagaaact cggtcatttc tttttggact  9660
gatgcataag ggacttgtaa aggctgacct aggttgtgat gacctgacaa acaaccctac  9720
tctttgggct actcctggaa gtggctcagg caaacgtact gctgacggaa gtgagttcga  9780
gtccccaaag aaaaaaagaa aagttggtag tggatctggt ggaatgactg atttcttcaa  9840
```

-continued

```
tgagtttgat gaatcgcttg taccactgaa accacaaact ccaacgcagt atgttaagct  9900
ggacgatgct aatttaattc agagagattt agatacattc tcagatactt tcaagaacca  9960
agccctccaa agatataagc taatttctac aattgataaa aagttaagcc gcggatggac 10020
tcaaagaaat ctcgacccga tcctcgacga actcttcaaa ggcggtgatg tagtgaggcc 10080
aaactggaga actgtggcaa gatggaggaa gaaatatata gaaagcaatg gtgacatcgc 10140
ttctcttgct gataaaaacc acaaaatggg aaacagaaca aatcggataa aaggagatga 10200
caaatttttt gacaaggcct tggaacgttt ccttgacgcg aaaagaccta caatagcaac 10260
agcttatcaa tattacaagg acttgatagt gattgaaaac gagtcaattg ttgaagggaa 10320
gatccccatc atctcttata atgcatttaa caaacgaatt aaagccattc ctccttatgc 10380
tgttgcagtc gctaggcatg ggaaattcaa agctgatcaa tggtttgcat attgtgcagc 10440
tcacgttcca ccaacgcgta tacttgaacg agtggagatt gatcatactc cgttggatct 10500
aatattattg gatgatgagc ttctgattcc cattggccgg ccatatctta cacttctcat 10560
agatgtgttt tcaggttgcg ttctgggatt ccatttgtca tacaagagcc cttcttatgt 10620
ctctgcagca aaagcaatca ctcacgcaat aaaacccaaa agcttagatg cgcttaatat 10680
tgagcttcag aatgactggc cctgtttcgg gaagtttgag aatttagtag ttgacaatgg 10740
agctgaattt tggtcaaaaa atttggaaca tgcttgtcaa tctgctggta taaatattca 10800
atacaatccg gtgcgtaagc catggctcaa gccttttata gagaggttct ttggggtgat 10860
gaatgaatac ttcttgccgg aactcccagg aaaaacattt tctaacattc ttgagaaaga 10920
agaatacaaa ccggagaagg atgcaataat gaggttcagc acctttgttg aggagttcca 10980
ccgttggatt gccgacgtgt atcatcaaga tagtaatagc cgtgaaacgc gtattcctat 11040
taagcgatgg caacaaggat tcgatgctta tcctccttta acaatgaacg aagaagagga 11100
gacacggttt tctatgctca tgagaatctc tgactctagg acgctaacca ggaatgggtt 11160
taagtaccaa gaattgatgt atgattctac cgccttagca gattaccgta agcattaccc 11220
tcagacaaaa gagactgtga agaagctaat aaaggtcgat ccagatgata tctccaagat 11280
ttacgtttat ctggaagagt tggaaagtta cttggaggtt ccttgcacag atccaactgg 11340
atacactgat ggtttatcca tatatgagca taaaaccatt aagaagatca accgagaagt 11400
tattagagaa tcaaaagata gccttgggtt ggcgaaagcg agaatggcta ttcatgagag 11460
agtgaaacaa gagcaagaag tattcatcga aagtaagacc aaagctaaga ttactgcagt 11520
caagaagcag gctcagatcg cggatgttag taatacaggc acatcaacaa tcaaggtgtc 11580
cgaggagtct gcggctcctg tccagaaaca catctcaaat gataacagtg atgattggga 11640
tgatgatctg gaagcttttg aatgagctta tatgaagatg aagatgaaat atttggtgtg 11700
tcaaataaaa aggttgtgtg cttaagtttg tgtttttttc ttggcttgtt gtgttatgaa 11760
tttgtggctt tttctaatat taaatgaatg taacatctca ttataatgaa taaacaaatg 11820
tttctataat ccattgtgaa tgttttgttg gatctcttct ccagcatata actactgtat 11880
gtgctatggt atggactatg gaatatgatt aaagataagc gctagcaagg agtggcagac 11940
atactgtccc acaaatgaag atggaatctg taaaagaaaa cgcgtgaaat aatgcgtctg 12000
acaaaggtta ggtcggctgc ctttaatcaa taccaaagtg gtccctacca cgatggaaaa 12060
actgtgcagt cggtttggct ttttctgacg aacaaataag attcgtggcc gacaggtggg 12120
ggtccaccat gtgaaggcat cttcagactc caataatgga gcaatgacgt aagggcttac 12180
gaaataagta agggtagttt gggaaatgtc cactcacccg tcagtctata aatacttagc 12240
ccctccctca ttgttaaggg agcaaaatct cagagagata gtcctagaga gagaaagaga 12300
gcaagtagcc tagaagtagt caaggcggcg aagtattcag gcacgtggcc aggaagaaga 12360
aaagccaaga cgacgaaaac aggtaagagc taagcttaat gggaaatgct cttacagaga 12420
ttcaaataga gaagttacgg aatttcagtg attgcatcgt gatgcatcct cagataaaga 12480
ccattttcaa tgattttgat gaactaaggc ttaatagaaa gtttcagagt gatcagcaat 12540
gtatgcttct tataggagat actggtgttg gcaaatcaca caccatcaac cattacaaga 12600
aaagagtcct cgcaacacaa aactattcaa gaaacacaat gccggtcctt gtttctcgca 12660
tatcacgtgg gaaaggtttg gatgctactt tggtgcagat gctggcggat ctagagctgt 12720
ttggttcttc tcagatcaag aagcggggct acaaaacaga ccttaccaaa aaattagttg 12780
agagtttgat caaagctcaa gtggagctgt tgattatcaa tgaattccaa gagcttattg 12840
agtttaaaag tgttcaagaa agacaacaga ttgcaaatgg tctaaagttt attagtgaag 12900
aagccaaggt tcccattgtt ttggttggaa tgccatgggc agcaaagata gctgaggaac 12960
cacaatgggc ttctcgactc gtaagaaaac gaaagctaga atatttttca ctcaagaatg 13020
acagcaagta ctttaggcag tatttgatgg gtcttgccaa aaagatgcct ttcgacgttc 13080
ctcctaaatt agaatcgaag aacactacga ttgcgctatt tgcagcttgt agaggggaaa 13140
acagagcttt gaagcacctt cttttagaag cactgaagct cgccctttca tgcaatgagt 13200
atcttgagaa caagcatttc atcacagctt atgacaaatt cgatttcttc aacgacaagg 13260
agaaactgaa atccaaaaac cctttcaagc aagatataaa agatattgaa atttatgaag 13320
ttattaaaaa ttcttcctac aatccaaatg cacttgatcc agaagatatg ttgactgatc 13380
gagtatttgc tattgtgaaa ggaagcggaa agaggactgc agatggatca gaatttgagt 13440
ccccaaagaa gaaaaggaaa gtgtgagctt atatgaagat gaagatgaaa tatttggtgt 13500
gtcaaataaa aaggttgtgt gcttaagttt gtgtttttt cttggcttgt tgtgttatga 13560
atttgtggct ttttctaata ttaaatgaat gtaacatctc attataatga ataaacaaat 13620
gtttctataa tccattgtga atgttttgtt ggatctcttc tccagcatat aactactgta 13680
tgtgctatgg tatggactat ggaatatgat taaagataag cgctatacag gagtggcaga 13740
catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa taatgcgtct 13800
gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc acgatggaaa 13860
aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc cgacaggtgg 13920
gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg taagggctta 13980
cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat aaatacttag 14040
cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag agagaaagag 14100
agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc caggaagaag 14160
aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tgggaaagcg tactgctgat 14220
ggttcagagt ttgaaagtcc aaagaagaaa agaaaagttg gtagtggtat ggcttttctt 14280
ttctcaccaa aagcaagagc gttttccgat gaatcgcttg aaagttattt acttcgagtt 14340
gtctcagaaa acttcttcga ttcatatgaa ggactatctt tggccatcag agaagaactt 14400
```

-continued

```
catgagttgg acttcgaggc tcatggagca tttcccgtag atttaaagag acttaatgtt  14460
taccatgcaa aacacaactc acactttagg atgagagctc ttggcctatt agaaactctt  14520
ttagatcttc ctcgttatga gcttcaaaag ctagcactcc tgaagtcaga cattaaattc  14580
aattcttccg ttgccctcta caacaatggt gttgatattc ctttgagatt cattagacat  14640
catgctgagg aagctgttga ctccattccg gtttgttctc aatgcttggc tgaagaagct  14700
tatataaagc aatcttggca tatcaaatgg gttaatgctt gcaccaaaca tcaatgtgct  14760
ttgcttcaca attgtcctga atgttatgca ccgatcaatt acattgagaa tgaatcaatc  14820
actcattgct cttgtggttt tgaattgagc tgcgcctcta caagtcctgt gaacacactc  14880
tcgattgaac atttgaacaa actcctggac aaaggagaac ggaatgactc aaatccattg  14940
tttaacaaca tgacgttgac agaaaggttt gctgctttat tatggtatca agagaggtat  15000
tcacagactg ataatttctg tttgaatgat gctgtgaact acttcagtaa gtggccagca  15060
gtgtttaata ctgagctaga tgagctctca aagaatgccg agatgaagct gattgatctc  15120
ttcaacaaga cagagtttaa gtttatcttt ggagatgcaa ttcttgcttg tccaagtact  15180
cagaagcaga gtgaatctca tttcatctac cgggcattac ttgattacct tgtcacttta  15240
gtggagagca atcccaaaac aaagaagcca aacgcagctg atttgcttgt ttctgtcctg  15300
gaagcagcaa ctttgttggg gacgagtgta gaacaagttt ataggctcta tcagaatgga  15360
atactgcaaa ccgcctttcg gcacaagatg aatcagagaa taaaccctta caagggtgct  15420
ttctttctcc gtcatgtgat agagtataaa accagctttg gtaacgataa agctagaatg  15480
tatctctcag cttggtgagc ttatatgaag atgaagatga aatatttggt gtgtcaaata  15540
aaaaggttgt gtgcttaagt ttgtgttttt ttcttggctt gttgtgttat gaatttgtgg  15600
ctttttctaa tattaaatga atgtaacatc tcattataat gaataaacaa atgtttctat  15660
aatccattgt gaatgttttg ttggatctct tctccagcat ataactactg tatgtgctat  15720
ggtatggact atggaatatg attaaagata agcgctacag aggagtctga gcaggacaac  15780
tcgcgtagtg agagttacat gttcgttggg ttcttccgac acggacctga gttggccaac  15840
gtcccacctg aggtctgtgc cccggtgatg agaagtgtgc atctcgttct tgcagctcgt  15900
cagtactttc agaatcatgg cgtgcatggt agaatgaccc ttataacgga cttcgacatg  15960
gcaatcgcta ggtaaatgaa tgaggagtgg cagacatact gtcccacaaa tgaagatgga  16020
atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa ggttaggtcg gctgccttta  16080
atcaatacca aagtggtccc taccacgatg gaaaaactgt gcagtcggtt tggctttttc  16140
tgacgaacaa ataagattcg tggccgacag gtgggggtcc accatgtgaa ggcatcttca  16200
gactccaata atggagcaat gacgtaaggg cttacgaaat aagtaagggt agtttgggaa  16260
atgtccactc acccgtcagt ctataaatac ttagcccctc cctcattgtt aagggagcaa  16320
aatctcagag agatagtcct agagagagaa agagagcaag tagcctagaa gtagtcaagg  16380
cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc caagacgacg aaaacaggta  16440
agagctaagc ttaatgggaa agcgtactgc tgatggttca gagtttgaaa gtccaaagaa  16500
gaaaagaaaa gttggtagtg gtatgcagag atattatttt accgttcatt ttcttcctaa  16560
acaagctaat ttggcccttc tcaccggaag atgcatatct atcatgcatg gatttatctt  16620
gaagcataac attgaaggaa tgggagtaac atttcccgca tggtctgatt caagcattgg  16680
aaacgagatt gcttttgttt acactgataa agaaatccta aacacattga aagatcaggc  16740
ttactttgtg gatatgcaag actgtggttt cttcaaggtt tctcaggtgt tggctgttcc  16800
tgattcttgc gaggaagtca ggttcatccg caatcaagcc gtcgcaaaga ttttcactgg  16860
cgaatcgagg aggagactta aacggttaca aaagagagcc ttagcaagag gagaagattt  16920
caatccaaag aagatagaag ctccaagaga aatagacatt tttcaccgtg tagcaatgac  16980
aagcaaaagt tctcaagaag actacattct gcacatccag aagcaagatg ttgattgtca  17040
agcagagcct tatttttcaa actatggtct tgcgtcaaat gagaagttca aaggaactgt  17100
accagacttg agtccgtcca ttgatagaaa ttgagcttat atgaagatga agatgaaata  17160
tttggtgtgt caaataaaaa ggttgtgtgc ttaagtttgt gtttttttct tggcttgttg  17220
tgttatgaat ttgtggcttt ttctaatatt aaatgaatgt aacatctcat tataatgaat  17280
aaacaaatgt ttctataatc cattgtgaat gttttgttgg atctcttctc cagcatataa  17340
ctactgtatg tgctatggta tggactatgg aatatgatta aagataagcg ctagcaagga  17400
gtggcagaca tactgtccca caaatgaaga tggaatctgt aaaagaaaac gcgtgaaata  17460
atgcgtctga caaaggttag gtcggctgcc tttaatcaat accaaagtgg tccctaccac  17520
gatggaaaaa ctgtgcagtc ggtttggctt tttctgacga acaaataaga ttcgtggccg  17580
acaggtgggg gtccaccatg tgaaggcatc ttcagactcc aataatggag caatgacgta  17640
agggcttacg aaataagtaa gggtagtttg ggaaatgtcc actcacccgt cagtctataa  17700
atacttagcc cctccctcat tgttaaggga gcaaaatctc agagagatag tcctagagag  17760
agaaagagag caagtagcct agaagtagtc aaggcggcga agtattcagg cacgtggcca  17820
ggaagaagaa aagccaagac gacgaaaaca ggtaagagct aagcttaatg ggaaagcgta  17880
ctgctgatgg ttcagagttt gaaagtccaa agaagaaaag aaaagttggt agtggtatgg  17940
agctttgcaa tatactcaag tatgatcgaa gtctttaccc tgggaaagcg gttttcttct  18000
acaaaacggc ggacagcgat tttgttcctt tggaagctga cattaataag ataagaggac  18060
caaaatcagg ctttacagag gctttcactc ctcaattctc tccgaaaaat atctctccac  18120
aagatctaac tcacaacaac attcttacac tcgaagagtg ttatgttcca ccaaatgtcg  18180
aacatatttt ctgccggttt tctctgcgtg tccaagctaa ttcgttagtc ccctctggtt  18240
gttccgatcc tgaggttttt agccttctca aagagcttgc agaaaccttc aaagaatgtg  18300
gtggttacaa ggagctggca gtacgctatt gtcggaacat tttgattgga acatggcttt  18360
ggagaaatca gaataccgga aacacacaga ttgagatcaa gacttcaaaa ggatcttgct  18420
acttgattga taatacaaga aaactcgctt gggaaagcaa atgggcttca gacgacttga  18480
aagttttgga agaattaagc aatgaaatcg aatctgcttt aacagatcca aatgtgtttt  18540
ggtctgcaga tataaccgcc aagatcgaag catctttctg tcaagagata tatccttcac  18600
agatattgaa tgataaagtg aagcaaggag aagcgtcaaa acagtttgtg aaagctaaat  18660
gtgctgatgg aaggtatgct gtgagtttca acagtgttaa gatcggtgct gctttgcaaa  18720
gtattgatga ttggtgggat gaagatgcta gtaaaagact tagagttcat gaatttggtg  18780
cagacaagga aattggcgta gcgaggaggc ctcccgattc cgaacaaaac ttttattcaa  18840
tcttcaagaa cactgaatgg tatctttccg ccttaaagaa ttgcatcact aacaagaatg  18900
agaagattga cccagcaata tactatttgt tcagcgtgtt gatcaaaggt ggaatgtttc  18960
```

-continued

```
agaaaaaggc agaggcaaag aaagcatgag cttatatgaa gatgaagatg aaatatttgg  19020
tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt tttcttggct tgttgtgtta  19080
tgaatttgtg gctttttcta atattaaatg aatgtaacat ctcattataa tgaataaaca  19140
aatgtttcta taatccattg tgaatgtttt gttggatctc ttctccagca tataactact  19200
gtatgtgcta tggtatggac tatggaatat gattaaagat aagcgctata caggagtggc  19260
agacatactg tcccacaaat gaagatggaa tctgtaaaag aaaacgcgtg aaataatgcg  19320
tctgacaaag gttaggtcgg ctgcctttaa tcaataccaa agtggtccct accacgatgg  19380
aaaaactgtg cagtcggttt ggctttttct gacgaacaaa taagattcgt ggccgacagg  19440
tgggggtcca ccatgtgaag gcatcttcag actccaataa tggagcaatg acgtaagggc  19500
ttacgaaata agtaagggta gtttgggaaa tgtccactca cccgtcagtc tataaatact  19560
tagcccctcc ctcattgtta agggagcaaa atctcagaga gatagtccta gagagagaaa  19620
gagagcaagt agcctagaag tagtcaaggc ggcgaagtat tcaggcacgt ggccaggaag  19680
aagaaaagcc aagacgacga aaacaggtaa gagctaagct taatgggaaa gcgtactgct  19740
gatggttcag agtttgaaag tccaaagaag aaaagaaaag ttggtagtgg tatgcacctg  19800
aaagagctgt tagagataac agatacaact gagagagata gatctctccg acgtgccttc  19860
tcgccgtata ccgccatgat cgacattaca ggatcagaag cagtcgctct aattatccta  19920
ctcaatctca cttatagaaa gaatcaggta gatgacctgt tggataagaa actcgcgaaa  19980
caagccctca agagtgagga tcatatcaac aaatgcatca aggagattgc ttggtttcat  20040
actcacaacc ttaaatatcc agacataagg gtgagtaagc aaaatcttgc agtagaacca  20100
ccaacattgc attcttatgt gctttcttct gctaattatc ccaaagctta tggttggtcc  20160
cataattcag ctaaagtgaa ttttgctaag ttgtttgtct catacttcaa atggcaaaac  20220
caagtgagct ggcttgccca agtcttggca accaactctg ataactggaa aagtgcattt  20280
acttcccttg gtttatcggt gaaagccttc aaatctttat gcgttacagt taagaattct  20340
ttacctgaag aggcaatccc ggactcagtg gatagatata gcaggcagat tagaatgcct  20400
tatcatgatg gttatcttgc tgtcactcct gtgatatctc acgttgttca aagcaaaata  20460
caacaagctg ctattgacaa aagagctagg ttttctaatg ttgagtttac cagaccagct  20520
gctgtttcaa tgctggcggc aagtctaggt ggtgtgataa atgtacttaa ttatcctcca  20580
tacattcgat caaagtacca tggactatcg aattcaaggg cttttaagtt aaacaacgga  20640
cagactgtct tcaacgtgga agccctcctg aagcctgagc ttataaaagc tcttgaagga  20700
attatattct ctaacaatgc attggcatta aaacagagaa ggcagcagaa agtaaaaaac  20760
atcaaagaac tgagaaatac cctccttgag tggttttccc ctgttttcga atggagattg  20820
gatgcaattg agaatggtta tgatctagaa caattggaat cagcgagcga acggttagag  20880
tacaagattc tgagtttgcc ggataatgaa ctaccttctt tgactatacc cttgtttaga  20940
ttacttaatg agatgttggg aggagtttca atgactcaga ggtatgcatt ccatccaaag  21000
ttgatgtctc cattaaaggc tgctctccaa tggttgctcg ttaatttaac tgatcaaaag  21060
catgttttaa ttgaagaaga tgatgagcat tacagatacc ttcacctgtc agggattagg  21120
gtttttgatg cccaagcatt gtctaatcct tactgctctg gtattccttc gcttacggcg  21180
gtttggggaa tgatccacag ttatcaacgc aaactaaacg aagctttggg cacgaacgta  21240
cggttcacaa gcttttcatg gttcatccgt aattactctg ccgttgctgg taaaaagcta  21300
ccagaactta gccttcaggg tgcacaacag tctagattga aaagacctgg gataatagac  21360
gggaagtatt gtgatttggt gtttgacttg attattcata tcgatggcta tgaagacgat  21420
ttgcaggctg ttgattcaaa gcctgatatc ctcaaggctc atttcccaag caatttcgcc  21480
ggcggcgtga tgcatcaacc cgagttgaat tctaatatta attggtgttg tctttactcc  21540
aacgaaaacc agctttttga gaagctccgt cggcttcctc tttctggatg ttgggttatg  21600
cccacagaac ataagattca agatcttgat gaactgttgt tgctgttgaa ctcggatagt  21660
aaactaagtc caagcatgat gggatacatg cttctcacag agccaatggc tcgtgtagga  21720
tcgttagaga ggcttcactg ttatgcagaa cctgctattg gtgttgtcaa atacgaggca  21780
gcaacttctg ttcgattaaa aggtatcgga aactatttca attccgcgtt ttggatgctc  21840
gatgctcaag aaaagttcat gctcatgaag aaagtgtgag cttatatgaa gatgaagatg  21900
aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt tttcttggct  21960
tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat ctcattataa  22020
tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc ttctccagca  22080
tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat aagcgctaca  22140
gaggagtctg agcaggacaa ctcgcgtagt gagagttaca tgttcgttgg gttcttccga  22200
cacggacctg agttggccaa cgtcccacct gaggtctgtg ccccggtgat gagaagtgtg  22260
catctcgttc ttgcagctcg tcagtacttt cagaatcatg gcgtgcatgg tagaatgacc  22320
cttataacgg acttcgacat ggcaatcgct aggtagctta atgaataggc gtatcacgag  22380
gccctttcgt cttcaagaat tagcttggca attcccgatc tagtaacata gatgacaccg  22440
cgcgcgataa tttatcctag tttgcgcgct atattttgtt ttctatcgcg tattaaatgt  22500
ataattgcgg gactctaatc ataaaaaccc atctcataaa taacgtcatg cattacatgt  22560
taattattac atgcttaacg taattcaaca gaaattatat gataatcgtc gcaagaccgg  22620
caacaggatt caatcttaag aaactttatt gccaaatgtt tgaacgatcg ggatcgaaga  22680
tccattccaa gcctacagga acaggtggtg gcggccctcg gtgcgctcgt actgctccac  22740
gatggtgtag tcctcgttgt gggaggtgat gtccagcttg gcgtccacgt agtagtagcc  22800
gggcagctgc acgggcttct tggccatgta gatggacttg aactccacca ggtagtggcc  22860
gccgtccttc agcttcaggg ccttgtgtgt ctcgcccttc agcacgccgt cgcgggggta  22920
caggcgctcg gtggaggcct cccagcccat cgtcttcttc tgcatcacgg ggccgtcgga  22980
ggggaagttc acgccgatga acttcacctt gtagatgaag cagccgtcct gcagggagga  23040
gtcctgggtc acggtcgcca cgccgccgtc ctcgaagttc atcacgcgct cccacttgaa  23100
gccctcgggg aaggacagct tcttgtagtc ggggatgtcg gcggggtgct tcacgtacac  23160
cttggagccg tactggaact gggggggacag gatgtcccag gcgaagggca gggggccgcc  23220
cttggtcacc ttcagcttca cggtgttgtg gccctcgtag gggcggccct cgccctcgcc  23280
ctcgatctcg aactcgtggc cgttcacggt gccctccatg cgcaccttga agcgcatgaa  23340
ctcggtgatg acgttctcgg aggaggccat ttcgcctcga tcgagttgag agtgaatatg  23400
agactctaat tggataccga ggggaattta tggaacgtca gtggagcatt tttgacaaga  23460
aatatttgct agctgatagt gaccttaggc gacttttgaa cgcgcaataa tggtttctga  23520
```

-continued

```
cgtatgtgct tagctcatta aactccagaa acccgcggct cagtggctcc ttcaacgttg  23580
cggttctgtc agttccaaac gtaaaacggc ttgtcccgcg tcatcggcgg gggtcatacc  23640
gtgactccct taattctcat gtatggttcg ctatactgaa gagcaattct gacaggatat  23700
attggcgggt aaacctaaga gaaaagagcg tttattagaa taatcggata tttaaaaggg  23760
cgtgaaaagg tttatccgtt cgtccatttg tatgtgcatg ccaaccacag ggttcccctc  23820
gggatcaaag tactttgatc caacccctcc gctgctatag tgcagtcggc ttctgacgtt  23880
cagtgcagcc gtcatctgaa aacgacatgt cgcacaagtc ctaagttacg cgacaggctg  23940
ccgccctgcc cttttcctgg cgttttcttg tcgcgtgttt tagtcgcata aagtagaata  24000
cttgcgacta gaaccggaga cattacgcca tgaacaagag cgccgccgct ggcctgctgg  24060
gctatgcccg cgtcagcacc gacgaccagg acttgaccaa ccaacgggcc gaactgcacg  24120
cggccggctg caccaagctg ttttccgaga agatcaccgg caccaggcgc gaccgcccgg  24180
agctggccag gatgcttgac cacctacgcc ctggcgacgt tgtgacagtg accaggctag  24240
accgcctggc ccgcagcacc cgcgacctac tggacattgc cgagcgcatc caggaggccg  24300
gcgcgggcct gcgtagcctg gcagagccgt gggccgacac caccacgccg gccggccgca  24360
tggtgttgac cgtgttcgcc ggcattgccg agttcgagcg ttccctaatc atcgaccgca  24420
cccggagcgg gcgcgaggcc gccaaggccc gaggcgtgaa gtttggcccc cgccctaccc  24480
tcaccccggc acagatcgcg cacgcccgcg agctgatcga ccaggaaggc cgcaccgtga  24540
aagaggcggc tgcactgctt ggcgtgcatc gctcgaccct gtaccgcgca cttgagcgca  24600
gcgaggaagt gacgcccacc gaggccaggc ggcgcggtgc cttccgtgag gacgcattga  24660
ccgaggccga cgccctggcg gccgccgaga atgaacgcca agaggaacaa gcatgaaacc  24720
gcaccaggac ggccaggacg aaccgttttt cattaccgaa gagatcgagg cggagatgat  24780
cgcggccggg tacgtgttcg agccgcccgc gcacctctca accgtgcggc tgcatgaaat  24840
cctggccggt ttgtctgatg ccaagctggc ggcctggccg gccagcttgg ccgctgaaga  24900
aaccgagcgc cgccgtctaa aaaggtgatg tgtatttgag taaaacagct tgcgtcatgc  24960
ggtcgctgcg tatatgatgc gatgagtaaa taaacaaata cgcaagggga acgcatgaag  25020
gttatcgctg tacttaacca gaaaggcggg tcaggcaaga cgaccatcgc aacccatcta  25080
gcccgcgccc tgcaactcgc cggggccgat gttctgttag tcgattccga tccccagggc  25140
agtgcccgcg attgggcggc cgtgcgggaa gatcaaccgc taaccgttgt cggcatcgac  25200
cgcccgacga ttgaccgcga cgtgaaggcc atcggccggc gcgacttcgt agtgatcgac  25260
ggagcgcccc aggcggcgga cttggctgtg tccgcgatca aggcagccga cttcgtgctg  25320
attccggtgc agccaagccc ttacgacata tgggccaccg ccgacctggt ggagctggtt  25380
aagcagcgca ttgaggtcac ggatggaagg ctacaagcgg cctttgtcgt gtcgcgggcg  25440
atcaaaggca cgcgcatcgg cggtgaggtt gccgaggcgc tggccgggta cgagctgccc  25500
attcttgagt cccgtatcac gcagcgcgtg agctacccag gcactgccgc cgccggcaca  25560
accgttcttg aatcagaacc cgagggcgac gctgcccgcg aggtccaggc gctggccgct  25620
gaaattaaat caaaactcat ttgagttaat gaggtaaaga gaaaatgagc aaaagcacaa  25680
acacgctaag tgccggccgt ccgagcgcac gcagcagcaa ggctgcaacg ttggccagcc  25740
tggcagacac gccagccatg aagcgggtca actttcagtt gccggcggag gatcacacca  25800
agctgaagat gtacgcggta cgccaaggca agaccattac cgagctgcta tctgaataca  25860
tcgcgcagct accagagtaa atgagcaaat gaataaatga gtagatgaat tttagcggct  25920
aaaggaggcg gcatggaaaa tcaagaacaa ccaggcaccg acgccgtgga atgccccatg  25980
tgtggaggaa cgggcggttg gccaggcgta agcggctggg ttgcctgccg gccctgcaat  26040
ggcactggaa cccccaagcc cgaggaatcg gcgtgagcgg tcgcaaacca tccggcccgg  26100
tacaaatcgg cgcggcgctg ggtgatgacc tggtggagaa gttgaaggcc gcgcaggccg  26160
cccagcggca acgcatcgag gcagaagcac gccccggtga atcgtggcaa gcggccgctg  26220
atcgaatccg caaagaatcc cggcaaccgc cggcagccgg tgcgccgtcg attaggaagc  26280
cgcccaaggg cgacgagcaa ccagattttt tcgttccgat gctctatgac gtgggcaccc  26340
gcgatagtcg cagcatcatg gacgtggccg ttttccgtct gtcgaagcgt gaccgacgag  26400
ctggcgaggt gatccgctac gagcttccag acgggcacgt agaggtttcc gcagggccgg  26460
ccggcatggc cagtgtgtgg gattacgacc tggtactgat ggcggtttcc catctaaccg  26520
aatccatgaa ccgataccgg gaagggaagg gagacaagcc cggccgcgtg ttccgtccac  26580
acgttgcgga cgtactcaag ttctgccggc gagccgatgg cggaaagcag aaagacgacc  26640
tggtagaaac ctgcattcgg ttaaacacca cgcacgttgc catgcagcgt acgaagaagg  26700
ccaagaacgg ccgcctggtg acggtatccg agggtgaagc cttgattagc cgctacaaga  26760
tcgtaaagag cgaaaccggg cggccggagt acatcgagat cgagctagct gattggatgt  26820
accgcgagat cacagaaggc aagaacccgg acgtgctgac ggttcacccc gattactttt  26880
tgatcgatcc cggcatcggc cgttttctct accgcctggc acgccgcgcc gcaggcaagg  26940
cagaagccag atggttgttc aagacgatct acgaacgcag tggcagcgcc ggagagttca  27000
agaagttctg tttcaccgtg cgcaagctga tcgggtcaaa tgacctgccg gagtacgatt  27060
tgaaggagga ggcggggcag gctggcccga tcctagtcat gcgctaccgc aacctgatcg  27120
agggcgaagc atccgccggt tcctaatgta cggagcagat gctagggcaa attgccctag  27180
caggggaaaa aggtcgaaaa ggactctttc ctgtggatag cacgtacatt gggaacccaa  27240
agccgtacat tgggaaccgg aacccgtaca ttgggaaccc aaagccgtac attgggaacc  27300
ggtcacacat gtaagtgact gatataaaag agaaaaaagg cgatttttcc gcctaaaact  27360
ctttaaaact tattaaaact cttaaaaccc gcctggcctg tgcataactg tctggccagc  27420
gcacagccga agtgctgcaa aaagcgccta cccttcggtc gctgcgctcc ctacgccccg  27480
ccgcttcgcg tcggcctatc gcggccgctg gccgctcaaa aatggctggc ctacggccag  27540
gcaatctacc agggcgcgga caagccgcgc cgtcgccact cgaccgccgg cgcccacatc  27600
aaggcaccct gcctcgcgcg tttcggtgat gacggtgaaa acctctgaca catgcagctc  27660
ccggtgacgg tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc  27720
gcgtcagcgg gtgttggcgg gtgtcggggc gcagccatga cccagtcacg tagcgatagc  27780
ggagtgtata ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaccata  27840
tgcggtgtga aataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgctcatccg  27900
cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc  27960
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt  28020
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc  28080
```

-continued

```
ataggctccg ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa  28140
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc  28200
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg  28260
cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc  28320
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc  28380
gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca  28440
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact  28500
acggctacac tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg  28560
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt  28620
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct  28680
tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatgc  28740
attctaggtg attagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca  28800
ggattatcaa taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg  28860
aggcagttcc ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca  28920
tcaatacaac ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca  28980
tgagtgacga ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt  29040
tcaacaggcc agccattacg ctcgtcatca aaatcactcg catcaaccaa accgttattc  29100
attcgtgatt gcgcctgagc gagtcgaaat acgcgatcgc tgttaaaagg acaattacaa  29160
acaggaatcg aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct  29220
gaatcaggat attcttctaa tacctggaat gctgttttcc ctgggatcgc agtggtgagt  29280
aaccatgcat catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc  29340
gtcagccagt ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca  29400
tgtttcagaa acaactctgg cgcatcgggc ttcccataca atcggtagat tgtcgcacct  29460
gattgcccga cattatcgcg agcccattta tacccatata aatcagcatc catgttggaa  29520
tttaatcgcg gccttgagca agacgtttcc cgttgaatat ggctcataac agaacttatt  29580
atttccttcc tcttttctac agtatttaaa gataccccaa gaagctaatt ataacaagac  29640
gaactccaat tcactgttcc ttgcattcta aaaccttaaa taccagaaaa cagctttttc  29700
aaagttgttt tcaaagttgg cgtataacat agtatcgacg gagccgattt tgaaaccgcg  29760
gtgatcacag gcagcaacgc tctgtcatcg ttacaatcaa catgctaccc tccgcgagat  29820
catccgtgtt tcaaacccgg cagcttagtt gccgttcttc cgaatagcat cggtaacatg  29880
agcaaagtct gccgccttac aacggctctc ccgctgacgc cgtcccggac tgatgggctg  29940
cctgtatcga gtggtgattt tgtgccgagc tgccggtcgg ggagctgttg gctggctgg   29999
```

SEQ ID NO: 42 moltype = DNA length = 29999
FEATURE Location/Qualifiers
source 1..29999
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 42

```
tggcaggata tattgtggtg taaacataac gaattgctct tcagcaagga gaatgaggag  60
aatgtagcag aaggcatgtt gttgtgactc cgaggggttg cctcaaactc tatcttataa  120
ccggcgtgga ggcatggagg caggggtatt ttggtcattt taatagatag tggaaaatga  180
cgtggaattt acttaaagac gaagtctttg cgacaagggg gggcccacgc cgaatttaat  240
attaccggcg tggccccccc ttatcgcgag tgctttagca cgagcggtcc agatttaaag  300
tagaaaattt cccgcccact agggttaaag gtgttcacac tataaaagca tatacgatgt  360
gatggtattt gatggagcgt atattgtatc aggtatttcc gttggatacg aattattcgt  420
acgaccctcg gtatactaat gaggagtgtc gctgcaacca tagtttgaca taatgcaacc  480
ataccttgac ataattaaag tcatactctg acataaagcc ccttattaaa aatgatgcta  540
aaaataaatt ccattataat caattattta agaattaatt aaaccatagt gtgacatagt  600
ggctttttat ttgtatagta tgtgtactca aataataaaa atttacttat gtatatacgc  660
aacttaagaa agccatctcc aaataaaaat gtatttcgct agcaaggagt gagacttttc  720
aacaaagggt aatatcggga aacctcctcg gattccattg cccagctatc tgtcacttca  780
tcaaaaggac agtagaaaag gaaggtggca cctacaaatg ccatcattgc gataaaggaa  840
aggctatcgt tcaagatgcc tctgccgaca gtggtcccaa agatggaccc ccacccacga  900
ggagcatcgt ggaaaaagaa gacgttccaa ccacgtcttc aaagcaagtg gattgatgtg  960
atatctccac tgacgtaagg gatgacgcac aatcccacta tccttcgcaa gacccttcct  1020
ctatataagg aagttcattt catttggaga ggacacgctg acggccgcgt atttttacaa  1080
caattaccaa caacaacaaa caacaaacaa cattacaatt actatttaca attacagcgg  1140
ccgccccctt caccaaaaaa atggctgcag ccaagggcga ggagctgttc accggggtgg  1200
tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg  1260
agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gctgctctgc accaccggca  1320
agctgcccgt gccctggccc accctcgtga ccaccctggg ctacggcgtg cagtgcttcg  1380
cccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg cccgaaggct  1440
acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg  1500
tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcatc gacttcaagg  1560
aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac aacgtctata  1620
tcaccgccga caagcagaag aacggcatca aggccaactt caagatccgc cacaacatcg  1680
aggacggcgg cgtgcagctc gccgaccact accagcagaa cacccccatc ggcgacggcc  1740
ccgtgctgct gcccgacaac cactacctga gctaccagtc caaactgttc aaagacccca  1800
acgagaagcg cgatcacatg gtcttgttag aatttttgac tgctgctggt attaccgagg  1860
gtatgaatga attgtacaaa ggctccggac cgaagaagaa gaggaaggtt taaatatgaa  1920
gatgaagatg aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt  1980
tttcttggct tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat  2040
ctcattataa tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc  2100
ttctccagca tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat  2160
aagcgctata caggagccgt ataggcagct gaagataaaa acaatcgatt cgatgtttaa  2220
```

-continued

```
gttaatgtaa ttctgccgaa aaggcagtga gtagtaattg aaaaaaccag tcttatatcc  2280
gggctggttt ttttgttttt agcgttataa cttttatttc actatgtcac tgtatggttg  2340
cattttatgt cagtgaatgg ttgcaggttt tttgtaagtt attgatatgc ctatccagtg  2400
tatgtcacgc tttgcttgca gcgacacgct acagaggagt ctgagcagga caactcgcgt  2460
agtgagagtt acatgttcgt tgggttcttc cgacacggac ctgagttggc caacgtccca  2520
cctgaggtct gtgccccggt gatgagaagt gtgcatctcg ttcttgcagc tcgtcagtac  2580
tttcagaatc atggcgtgca tggtagaatg acccttataa cggacttcga catggcaatc  2640
gctaggtaaa tgaatggaaa aagtggcacc gagtcggtgc tttttttccc ggggcgcgcc  2700
gatatcgagc tctcccggcg cgccgatatc gagctcagtg tttgatcgcc ggcggtaccg  2760
agtgtacttc aagtcagtgg gaaatcaata aaatgattat tttatgaata tatttcattg  2820
tgcaagtaga tagaaattac atatgttaca taacacacga aataaacaaa aaaagacaat  2880
ccaaaaacaa acaccccaaa aaaaataatc actttagata aactcgtatg aggagaggca  2940
cgttcagtga ctcgacgatt cccgagcaaa aaaagtctcc ccgtcacaca tgtagtgggt  3000
gacgcaatta tctttaaagt aatccttctg ttgacttgtc attgataaca tccagtcttc  3060
gtcaggattg caaagaatta tagaagggat cccacctttt attttcttct tttttccata  3120
tttagggttg acagtgaaat cagactggca acctattaat tgcttccaca atgggacgaa  3180
cttgaagggg atgtcgtcga tgatattata ggtggcgtgt tcatcgtagt tggtgaaatc  3240
gatggtaccg ttccaatagt tgtgtcgtcc gagacttcta gcccaggtgg tctttccggt  3300
acgagttggt ccgcagatgt agaggctggg gtgtcggatt ccattccttc cattgtcctt  3360
gttaaatcgg ccatccattc aaggtcagat tgagcttgtt ggtatgagac aggatgtatg  3420
taagtataag cgtctatgct tacatggtat agatgggttt ccctccagga gtgtagatct  3480
tcgtggcagc gaagatctga ttctgtgaag ggcgacacat acggttcagg ttgtggaggg  3540
aataatttgt tggctgaata ttccagccat tgaagctttg ttgcccattc atgagggaat  3600
tcttccttga tcatgtcaag atattcctcc ttagacgttg cagtctggat aatagttctc  3660
catcgtgcgt cagatttgcg aggagaaacc ttatgatctc ggaaatctcc tctggtttta  3720
atatctccgt cctttgatat gtaatcaagg acttgtttag agtttctagc tggctggata  3780
ttagggtgat ttccttcaaa atcgaaaaaa gaaggatccc taatacaagg ttttttatca  3840
agctggagga gagcatgata gtgggtagtg ccatcttgat gaagctcaga agcaacacca  3900
aggaagaaaa taagaaaagg tgtgagtttc tcccagagaa actggaataa atcatctctt  3960
tgagatgagc acttgggata ggtaaggaaa acatatttag attggagtct gaagttctta  4020
ctagcagaag gcatgttgtt gtgactccga ggggttgcct caaactctat cttataaccg  4080
gcgtggaggc atggaggcag gggtattttg gtcattttaa tagatagtgg aaaatgacgt  4140
ggaatttact taaagacgaa gtctttgcga caaggggggg cccacgccga atttaatatt  4200
accggcgtgg cccccccttta tcgcgagtgc tttagcacga gcggtccaga tttaaagtag  4260
aaaatttccc gcccactagg gttaaaggtg ttcacactat aaaagcatat acgatgtgat  4320
ggtatttgat ggagcgtata ttgtatcagg tatttccgtt ggatacgaat tattcgtacg  4380
accctcatag tttaaactat cagtgggtag cttaatggta gacactttat tagttagatc  4440
cgctgttatt ctgcacgatt gcgcttgccg gtggaaattt aaggacgagt gtagtgcact  4500
cacccttagc cttcagcgtg acaagattag ggtagcaccg gaaacagcga ttttacggac  4560
agatagtaca gtctgagact cagtcactgg tccctaaacg ggagccggta cgtcggcggt  4620
tcgctagcaa ggagtggcag acatactgtc ccacaaatga agatggaatc tgtaaaagaa  4680
aacgcgtgaa ataatgcgtc tgacaaaggt taggtcggct gcctttaatc aataccaaag  4740
tggtccctac cacgatggaa aaactgtgca gtcggtttgg ctttttctga cgaacaaata  4800
agattcgtgg ccgacaggtg ggggtccacc atgtgaaggc atcttcagac tccaataatg  4860
gagcaatgac gtaagggctt acgaaataag taagggtagt ttgggaaatg tccactcacc  4920
cgtcagtcta taaatactta gcccctccct cattgttaag ggagcaaaat ctcagagaga  4980
tagtcctaga gagagaaaga gagcaagtag cctagaagta gtcaaggcgg cgaagtattc  5040
aggcacgtgg ccaggaagaa gaaaagccaa gacgacgaaa acaggtaaga gctaagctta  5100
atgggaaagc gtactgctga tggttcagag tttgaaagtc caaagaagaa aagaaaagtt  5160
ggtagtggta tgactgataa acggaaagat ggttccggca agcttcttta ctgcagcttt  5220
tgcgggaaat ctcaacatga agttaggaaa cttattgcag ggccttctgt ttatatctgt  5280
gacgaatgtg ttgatctctg caatgatatt attcgagaag agatcaaaga agtggctcct  5340
cacagagaga ggagtgcgtt gccgacacca cacgagatta gaaatcatct tgatgattat  5400
gtcattggtc aagaacaagc aaagaaggtc cttgctgttg cagtgtacaa ccattacaag  5460
aggctcagaa acggagatac ttctaatggt gtggaacttg gcaaatcaaa catcctgctc  5520
attggaccca ctggaagtgg aaagacactc ctagcagaga cattggcaag gctcttagac  5580
gttcctttca ctatggctga tgcaactact cttactgagg ctggttatgt tggagaagat  5640
gtggagaata ttattcaaaa gttattgcag aaatgtgatt acgatgtcca gaaagctcag  5700
agagggatag tctacataga tgagatagac aagatctcta ggaagagtga taatccctcg  5760
ataacacgcg acgtttctgg cgaaggtgtt caacaagctt tgttgaagct cattgaaggc  5820
accgtagcgg cagttccacc acaaggagga agaaagcatc ctcagcaaga gtttctccaa  5880
gtggatacct caaaaatctt atttatatgt ggtggtgcct tcgctgggct tgacaaagtt  5940
atctcacacc gtgtagaaac tggtagcggg attgggtttg gtgcaacggt gaaggcaaaa  6000
tctgataagg cttctgaagg agagcttctg gctcaagttg aaccagagga tcttatcaag  6060
tttggactaa ttcctgagtt catcggtaga cttcctgtcg ttgctacgtt gaacgagtta  6120
agtgaggaag ccctgattca aatcttgaaa gagccaaaga atgctctgac aaagcaatat  6180
caggctcttt tcaatcttga gggagttgac ttggaatttc gtgatgaagc gctggatgct  6240
atagctaaga aagcaatggc gaggaaaacc ggagctcgtg gtctaagatc gattgttgaa  6300
gccgctttat tggatacaat gtatgatttg ccctccatgg aagatgttga gaaagtggtc  6360
atcgacgagt ctgtgattga tggacagtca aagccattgc tcatatatgg taaacctgaa  6420
gcacagcagg catcaggaga atgaatatga agatgaagat gaaatatttg gtgtgtcaaa  6480
taaaaaggtt gtgtgcttaa gtttgtgttt ttttcttggc ttgttgtgtt atgaatttgt  6540
ggctttttct aatattaaat gaatgtaaca tctcattata atgaataaac aaatgtttct  6600
ataatccatt gtgaatgttt tgttggatct cttctccagc atataactac tgtatgtgct  6660
atggtatgga ctatggaata tgattaaaga taagcgctat acaggagtgg cagacatact  6720
gtcccacaaa tgaagatgga atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa  6780
```

-continued

```
ggttaggtcg gctgccttta atcaatacca aagtggtccc taccacgatg gaaaaactgt  6840
gcagtcggtt tggctttttc tgacgaacaa ataagattcg tggccgacag gtgggggtcc  6900
accatgtgaa ggcatcttca gactccaata atggagcaat gacgtaaggg cttacgaaat  6960
aagtaagggt agtttgggaa atgtccactc acccgtcagt ctataaatac ttagcccctc  7020
cctcattgtt aagggagcaa aatctcagag agatagtcct agagagagaa agagagcaag  7080
tagcctagaa gtagtcaagg cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc  7140
caagacgacg aaaacaggta agagctaagc ttaatgggaa agcgtactgc tgatggttca  7200
gagtttgaaa gtccaaagaa gaaaagaaaa gttggtagtg gtatgagtta ttctggagaa  7260
cgggataatt ttgcacctca catggcatta gttccaatgg tgattgaaca aacaagtaga  7320
ggggagcgaa gctttgacat ttactctcgt ctattaaaag aaagggtcat cttcttaacc  7380
gggcaagttg aagatcatat ggcaaatctg attgtagctc agatgctttt tcttgaagca  7440
gagaatcctg agaaagatat ctatctctac attaactctc ctggtggtgt aataactgct  7500
ggaatgtcta tatatgatac aatgcagttt atcaaacctg atgtttcaac aatatgtatg  7560
ggtcaggctg cgtcaatggg agcttttctt ttgacggctg gcgcaaaagg aaagcggttc  7620
tgcctaccga actcaagggt gatgatccat caaccacttg gaggatacca aggtcaagct  7680
accgatatag agattcatgc tagagaaatt ttgaaggtga agggtagaat gaatgaactg  7740
atggcccttc acactggaca atcgctcgag cagattgagc gtgacacaga gagagacagg  7800
tttttatccg caccagaagc tgttgagtat ggtttggttg attctattct tactcataga  7860
aactgaatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  7920
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  7980
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  8040
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  8100
tatgattaaa gataagcgct acagaggagc atcttcattc ttaagatatg aagataatct  8160
tcaaaaggcc cctgggaatc tgaaagaaga gaagcaggcc catttatatg ggaaagaaca  8220
atagtatttc ttatataggc ccatttaagt tgaaaacaat cttcaaaagt cccacatcgc  8280
ttagataaga aaacgaagct gagtttatat acagctagag tcgaagtagt gataatggtg  8340
acctgccgta taggcagctg aagatactga cgtaagggat gacgcacaat cccactataa  8400
ttctgccgaa aaggcagtga gtagtgcttt tttttttgca aaattttcca gatcgatttc  8460
ttcttcctct gttcttcggc gttcaatttc tggggttttc tcttcgtttt ctgtaactga  8520
aacctaaaat ttgacctaaa aaaaatctca aataatatga ttcagtggtt ttgtactttt  8580
cagttagttg agttttgcag ttccgatgag ataaaccaat acgctaggta tactaatgag  8640
gagtggcaga catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa  8700
taatgcgtct gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc  8760
acgatggaaa aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc  8820
cgacaggtgg gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg  8880
taagggctta cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat  8940
aaatacttag cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag  9000
agagaaagag agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc  9060
caggaagaag aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tgtacattag  9120
gaacttgaga aagccttctc caaataagaa tgtttttaag ttcgcatcta ccaaggtcag  9180
ctctgtggtc atgtgtgaat ccagccttga gtttgacgct tgctttcatc atgaatataa  9240
cgacttgatt gagagttttg gttctcagcc agaaggtttc aaatacgaat tcatgggtaa  9300
atctctccct tacactccag atgctttgat ctcctacacg gacaaaactc aaaagtatca  9360
cgagtataaa ccatattcaa aaatcgcatc acctctattt agagcagagt ttgctgcaaa  9420
gagggctgca agtcttaagt tgggtattga tttggtactc gtcacagaca gacagattag  9480
agtaaaccca atactcaaca atctgaagct ccttcataga tactcgggcg tttatggtat  9540
atctggtatc cagaaggaac tacttagctt tattcacaaa agtggcgtta tcaaactcaa  9600
tgatatctca tcacaggttg ggatccccat tggagaaact cggtcatttc tttttggact  9660
gatgcataag ggacttgtaa aggctgacct aggttgtgat gacctgacaa acaaccctac  9720
tctttgggct actcctggaa gtggctcagg caaacgtact gctgacggaa gtgagttcga  9780
gtccccaaag aaaaaaagaa aagttggtag tggatctggt ggaatgactg atttcttcaa  9840
tgagtttgat gaatcgcttg taccactgaa accacaaact ccaacgcagt atgttaagct  9900
ggacgatgct aatttaattc agagagattt agatacattc tcagatactt tcaagaacca  9960
agccctccaa agatataagc taatttctac aattgataaa aagttaagcc gcggatggac  10020
tcaaagaaat ctcgacccga tcctcgacga actcttcaaa ggcggtgatg tagtgaggcc  10080
aaactggaga actgtggcaa gatggaggaa gaaatatata gaaagcaatg gtgacatcgc  10140
ttctcttgct gataaaaacc acaaaatggg aaacagaaca aatcggataa aaggagatga  10200
caaatttttt gacaaggcct tggaacgttt ccttgacgcg aaaagaccta caatagcaac  10260
agcttatcaa tattacaagg acttgatagt gattgaaaac gagtcaattg ttgaagggaa  10320
gatccccatc atctcttata atgcatttaa caaacgaatt aaagccattc ctccttatgc  10380
tgttgcagtc gctaggcatg ggaaattcaa agctgatcaa tggtttgcat attgtgcagc  10440
tcacgttcca ccaacgcgta tacttgaacg agtggagatt gatcatactc cgttggatct  10500
aatattattg gatgatgagc ttctgattcc cattggccgg ccatatctta cacttctcat  10560
agatgtgttt tcaggttgcg ttctgggatt ccatttgtca tacaagagcc cttcttatgt  10620
ctctgcagca aaagcaatca ctcacgcaat aaaacccaaa agcttagatg cgcttaatat  10680
tgagcttcag aatgactggc cctgtttcgg gaagtttgag aatttagtag ttgacaatgg  10740
agctgaattt tggtcaaaaa atttggaaca tgcttgtcaa tctgctggta taaatattca  10800
atacaatccg gtgcgtaagc catggctcaa gccttttata gagaggttct ttggggtgat  10860
gaatgaatac ttcttgccgg aactcccagg aaaaacattt tctaacattc ttgagaaaga  10920
agaatacaaa ccggagaagg atgcaataat gaggttcagc acctttgttg aggagttcca  10980
ccgttggatt gccgacgtgt atcatcaaga tagtaatagc cgtgaaacgc gtattcctat  11040
taagcgatgg caacaaggat tcgatgctta tcctccttta acaatgaacg aagaagagga  11100
gacacggttt tctatgctca tgagaatctc tgactctagg acgctaacca ggaatgggtt  11160
taagtaccaa gaattgatgt atgattctac cgccttagca gattaccgta agcattaccc  11220
tcagacaaaa gagactgtga agaagctaat aaaggtcgat ccagatgata tctccaagat  11280
ttacgtttat ctggaagagt tggaaagtta cttggaggtt ccttgcacag atccaactgg  11340
```

-continued

```
atacactgat ggtttatcca tatatgagca taaaaccatt aagaagatca accgagaagt  11400
tattagagaa tcaaaagata gccttgggtt ggcgaaagcg agaatggcta ttcatgagag  11460
agtgaaacaa gagcaagaag tattcatcga aagtaagacc aaagctaaga ttactgcagt  11520
caagaagcag gctcagatcg cggatgttag taatacaggc acatcaacaa tcaaggtgtc  11580
cgaggagtct gcggctcctg tccagaaaca catctcaaat gataacagtg atgattggga  11640
tgatgatctg gaagcttttg aatgagctta tatgaagatg aagatgaaat atttggtgtg  11700
tcaaataaaa aggttgtgtg cttaagtttg tgtttttttc ttggcttgtt gtgttatgaa  11760
tttgtggctt tttctaatat taaatgaatg taacatctca ttataatgaa taaacaaatg  11820
tttctataat ccattgtgaa tgttttgttg gatctcttct ccagcatata actactgtat  11880
gtgctatggt atggactatg gaatatgatt aaagataagc gctagcaagg agtggcagac  11940
atactgtccc acaaatgaag atggaatctg taaaagaaaa cgcgtgaaat aatgcgtctg  12000
acaaaggtta ggtcggctgc ctttaatcaa taccaaagtg gtccctacca cgatggaaaa  12060
actgtgcagt cggtttggct ttttctgacg aacaaataag attcgtggcc gacaggtggg  12120
ggtccaccat gtgaaggcat cttcagactc caataatgga gcaatgacgt aagggcttac  12180
gaaataagta agggtagttt gggaaatgtc cactcacccg tcagtctata aatacttagc  12240
ccctccctca ttgttaaggg agcaaaatct cagagagata gtcctagaga gagaaagaga  12300
gcaagtagcc tagaagtagt caaggcggcg aagtattcag gcacgtggcc aggaagaaga  12360
aaagccaaga cgacgaaaac aggtaagagc taagcttaat gggaaatgct cttacagaga  12420
ttcaaataga gaagttacgg aatttcagtg attgcatcgt gatgcatcct cagataaaga  12480
ccattttcaa tgattttgat gaactaaggc ttaatagaaa gtttcagagt gatcagcaat  12540
gtatgcttct tataggagat actggtgttg gcaaatcaca caccatcaac cattacaaga  12600
aaagagtcct cgcaacacaa aactattcaa gaaacacaat gccggtcctt gtttctcgca  12660
tatcacgtgg gaaaggtttg gatgctactt tggtgcagat gctggcggat ctagagctgt  12720
ttggttcttc tcagatcaag aagcggggct acaaaacaga ccttaccaaa aaattagttg  12780
agagtttgat caaagctcaa gtggagctgt tgattatcaa tgaattccaa gagcttattg  12840
agtttaaaag tgttcaagaa agacaacaga ttgcaaatgg tctaaagttt attagtgaag  12900
aagccaaggt tcccattgtt ttggttggaa tgccatgggc agcaaagata gctgaggaac  12960
cacaatgggc ttctcgactc gtaagaaaac gaaagctaga atatttttca ctcaagaatg  13020
acagcaagta ctttaggcag tatttgatgg gtcttgccaa aaagatgcct ttcgacgttc  13080
ctcctaaatt agaatcgaag aacactacga ttgcgctatt tgcagcttgt agaggggaaa  13140
acagagcttt gaagcacctt cttttagaag cactgaagct cgccctttca tgcaatgagt  13200
atcttgagaa caagcatttc atcacagctt atgacaaatt cgatttcttc aacgacaagg  13260
agaaactgaa atccaaaaac cctttcaagc aagatataaa agatattgaa atttatgaag  13320
ttattaaaaa ttcttcctac aatccaaatg cacttgatcc agaagatatg ttgactgatc  13380
gagtatttgc tattgtgaaa ggaagcggaa agaggactgc agatggatca gaatttgagt  13440
ccccaaagaa gaaaaggaaa gtgtgagctt atatgaagat gaagatgaaa tatttggtgt  13500
gtcaaataaa aaggttgtgt gcttaagttt gtgtttttt cttggcttgt tgtgttatga  13560
atttgtggct ttttctaata ttaaatgaat gtaacatctc attataatga ataaacaaat  13620
gtttctataa tccattgtga atgttttgtt ggatctcttc tccagcatat aactactgta  13680
tgtgctatgg tatggactat ggaatatgat taaagataag cgctatacag gagtggcaga  13740
catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa taatgcgtct  13800
gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc acgatggaaa  13860
aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc cgacaggtgg  13920
gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg taagggctta  13980
cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat aaatacttag  14040
cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag agagaaagag  14100
agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc caggaagaag  14160
aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tgggaaagcg tactgctgat  14220
ggttcagagt ttgaaagtcc aaagaagaaa agaaaagttg gtagtggtat ggcttttctt  14280
ttctcaccaa aagcaagagc gttttccgat gaatcgcttg aaagttattt acttcgagtt  14340
gtctcagaaa acttcttcga ttcatatgaa ggactatctt tggccatcag agaagaactt  14400
catgagttgg acttcgaggc tcatggagca tttcccgtag atttaaagag acttaatgtt  14460
taccatgcaa aacacaactc acactttagg atgagagctc ttggcctatt agaaactctt  14520
ttagatcttc ctcgttatga gcttcaaaag ctagcactcc tgaagtcaga cattaaattc  14580
aattcttccg ttgccctcta caacaatggt gttgatattc ctttgagatt cattagacat  14640
catgctgagg aagctgttga ctccattccg gtttgttctc aatgcttggc tgaagaagct  14700
tatataaagc aatcttggca tatcaaatgg gttaatgctt gcaccaaaca tcaatgtgct  14760
ttgcttcaca attgtcctga atgttatgca ccgatcaatt acattgagaa tgaatcaatc  14820
actcattgct cttgtggttt tgaattgagc tgcgcctcta caagtcctgt gaacacactc  14880
tcgattgaac atttgaacaa actcctggac aaaggagaac ggaatgactc aaatccattg  14940
tttaacaaca tgacgttgac agaaaggttt gctgctttat tatggtatca agagaggtat  15000
tcacagactg ataatttctg tttgaatgat gctgtgaact acttcagtaa gtggccagca  15060
gtgtttaata ctgagctaga tgagctctca aagaatgccg agatgaagct gattgatctc  15120
ttcaacaaga cagagtttaa gtttatcttt ggagatgcaa ttcttgcttg tccaagtact  15180
cagaagcaga gtgaatctca tttcatctac cgggcattac ttgattacct tgtcacttta  15240
gtggagagca atcccaaaac aaagaagcca aacgcagctg atttgcttgt ttctgtcctg  15300
gaagcagcaa ctttgttggg gacgagtgta gaacaagttt ataggctcta tcagaatgga  15360
atactgcaaa ccgcctttcg gcacaagatg aatcagagaa taaaccctta caagggtgct  15420
ttctttctcc gtcatgtgat agagtataaa accagctttg gtaacgataa agctagaatg  15480
tatctctcag cttggtgagc ttatatgaag atgaagatga aatatttggt gtgtcaaata  15540
aaaaggttgt gtgcttaagt ttgtgttttt ttcttggctt gttgtgttat gaatttgtgg  15600
ctttttctaa tattaaatga atgtaacatc tcattataat gaataaacaa atgtttctat  15660
aatccattgt gaatgttttg ttggatctct tctccagcat ataactactg tatgtgctat  15720
ggtatggact atggaatatg attaaagata agcgctacag aggagtctga gcaggacaac  15780
tcgcgtagtg agagttacat gttcgttggg ttcttccgac acggacctga gttggccaac  15840
gtcccacctg aggtctgtgc cccggtgatg agaagtgtgc atctcgttct tgcagctcgt  15900
```

-continued

```
cagtactttc agaatcatgg cgtgcatggt agaatgaccc ttataacgga cttcgacatg  15960
gcaatcgcta ggtaaatgaa tgaggagtgg cagacatact gtcccacaaa tgaagatgga  16020
atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa ggttaggtcg gctgccttta  16080
atcaatacca aagtggtccc taccacgatg gaaaaactgt gcagtcggtt tggctttttc  16140
tgacgaacaa ataagattcg tggccgacag gtgggggtcc accatgtgaa ggcatcttca  16200
gactccaata atggagcaat gacgtaaggg cttacgaaat aagtaagggt agtttgggaa  16260
atgtccactc acccgtcagt ctataaatac ttagcccctc cctcattgtt aagggagcaa  16320
aatctcagag agatagtcct agagagagaa agagagcaag tagcctagaa gtagtcaagg  16380
cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc caagacgacg aaaacaggta  16440
agagctaagc ttaatgggaa agcgtactgc tgatggttca gagtttgaaa gtccaaagaa  16500
gaaaagaaaa gttggtagtg gtatgcagag atattatttt accgttcatt ttcttcctaa  16560
acaagctaat ttggcccttc tcaccggaag atgcatatct atcatgcatg gatttatctt  16620
gaagcataac attgaaggaa tgggagtaac atttcccgca tggtctgatt caagcattgg  16680
aaacgagatt gcttttgttt acactgataa agaaatccta aacacattga aagatcaggc  16740
ttactttgtg gatatgcaag actgtggttt cttcaaggtt tctcaggtgt tggctgttcc  16800
tgattcttgc gaggaagtca ggttcatccg caatcaagcc gtcgcaaaga ttttcactgg  16860
cgaatcgagg aggagactta aacggttaca aaagagagcc ttagcaagag gagaagattt  16920
caatccaaag aagatagaag ctccaagaga aatagacatt tttcaccgtg tagcaatgac  16980
aagcaaaagt tctcaagaag actacattct gcacatccag aagcaagatg ttgattgtca  17040
agcagagcct tatttttcaa actatggtct tgcgtcaaat gagaagttca aaggaactgt  17100
accagacttg agtccgtcca ttgatagaaa ttgagcttat atgaagatga agatgaaata  17160
tttggtgtgt caaataaaaa ggttgtgtgc ttaagtttgt gtttttttct tggcttgttg  17220
tgttatgaat ttgtggcttt ttctaatatt aaatgaatgt aacatctcat tataatgaat  17280
aaacaaatgt ttctataatc cattgtgaat gttttgttgg atctcttctc cagcatataa  17340
ctactgtatg tgctatggta tggactatgg aatatgatta aagataagcg ctagcaagga  17400
gtggcagaca tactgtccca caaatgaaga tggaatctgt aaaagaaaac gcgtgaaata  17460
atgcgtctga caaaggttag gtcggctgcc tttaatcaat accaaagtgg tccctaccac  17520
gatggaaaaa ctgtgcagtc ggtttggctt tttctgacga acaaataaga ttcgtggccg  17580
acaggtgggg gtccaccatg tgaaggcatc ttcagactcc aataatggag caatgacgta  17640
agggcttacg aaataagtaa gggtagtttg ggaaatgtcc actcacccgt cagtctataa  17700
atacttagcc cctccctcat tgttaaggga gcaaaatctc agagagatag tcctagagag  17760
agaaagagag caagtagcct agaagtagtc aaggcggcga agtattcagg cacgtggcca  17820
ggaagaagaa aagccaagac gacgaaaaca ggtaagagct aagcttaatg ggaaagcgta  17880
ctgctgatgg ttcagagttt gaaagtccaa agaagaaaag aaaagttggt agtggtatgg  17940
agctttgcaa tatactcaag tatgatcgaa gtctttaccc tgggaaagcg gttttcttct  18000
acaaaacggc ggacagcgat tttgttcctt tggaagctga cattaataag ataagaggac  18060
caaaatcagg ctttacagag gctttcactc ctcaattctc tccgaaaaat atctctccac  18120
aagatctaac tcacaacaac attcttacac tcgaagagtg ttatgttcca ccaaatgtcg  18180
aacatatttt ctgccggttt tctctgcgtg tccaagctaa ttcgttagtc ccctctggtt  18240
gttccgatcc tgaggttttt agccttctca aagagcttgc agaaaccttc aaagaatgtg  18300
gtggttacaa ggagctggca gtacgctatt gtcggaacat tttgattgga acatggcttt  18360
ggagaaatca gaataccgga aacacacaga ttgagatcaa gacttcaaaa ggatcttgct  18420
acttgattga taatacaaga aaactcgctt gggaaagcaa atgggcttca gacgacttga  18480
aagttttgga agaattaagc aatgaaatcg aatctgcttt aacagatcca aatgtgtttt  18540
ggtctgcaga tataaccgcc aagatcgaag catctttctg tcaagagata tatccttcac  18600
agatattgaa tgataaagtg aagcaaggag aagcgtcaaa acagtttgtg aaagctaaat  18660
gtgctgatgg aaggtatgct gtgagtttca acagtgttaa gatcggtgct gctttgcaaa  18720
gtattgatga ttggtgggat gaagatgcta gtaaaagact tagagttcat gaatttggtg  18780
cagacaagga aattggcgta gcgaggaggc ctcccgattc cgaacaaaac ttttattcaa  18840
tcttcaagaa cactgaatgg tatctttccg ccttaaagaa ttgcatcact aacaagaatg  18900
agaagattga cccagcaata tactatttgt tcagcgtgtt gatcaaaggt ggaatgtttc  18960
agaaaaaggc agaggcaaag aaagcatgag cttatatgaa gatgaagatg aaatatttgg  19020
tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt tttcttggct tgttgtgtta  19080
tgaatttgtg gctttttcta atattaaatg aatgtaacat ctcattataa tgaataaaca  19140
aatgtttcta taatccattg tgaatgtttt gttggatctc ttctccagca tataactact  19200
gtatgtgcta tggtatggac tatggaatat gattaaagat aagcgctata caggagtggc  19260
agacatactg tcccacaaat gaagatggaa tctgtaaaag aaaacgcgtg aaataatgcg  19320
tctgacaaag gttaggtcgg ctgcctttaa tcaataccaa agtggtccct accacgatgg  19380
aaaaactgtg cagtcggttt ggctttttct gacgaacaaa taagattcgt ggccgacagg  19440
tgggggtcca ccatgtgaag gcatcttcag actccaataa tggagcaatg acgtaagggc  19500
ttacgaaata agtaagggta gtttgggaaa tgtccactca cccgtcagtc tataaatact  19560
tagcccctcc ctcattgtta agggagcaaa atctcagaga gatagtccta gagagagaaa  19620
gagagcaagt agcctagaag tagtcaaggc ggcgaagtat tcaggcacgt ggccaggaag  19680
aagaaaagcc aagacgacga aaacaggtaa gagctaagct taatgggaaa gcgtactgct  19740
gatggttcag agtttgaaag tccaaagaag aaaagaaaag ttggtagtgg tatgcacctg  19800
aaagagctgt tagagataac agatacaact gagagagata gatctctccg acgtgccttc  19860
tcgccgtata ccgccatgat cgacattaca ggatcagaag cagtcgctct aattatccta  19920
ctcaatctca cttatagaaa gaatcaggta gatgacctgt tggataagaa actcgcgaaa  19980
caagccctca agagtgagga tcatatcaac aaatgcatca aggagattgc ttggtttcat  20040
actcacaacc ttaaatatcc agacataagg gtgagtaagc aaaatcttgc agtagaacca  20100
ccaacattgc attcttatgt gctttcttct gctaattatc ccaaagctta tggttggtcc  20160
cataattcag ctaaagtgaa ttttgctaag ttgtttgtct catacttcaa atggcaaaac  20220
caagtgagct ggcttgccca agtcttggca accaactctg ataactggaa aagtgcattt  20280
acttcccttg gtttatcggt gaaagccttc aaatctttat gcgttacagt taagaattct  20340
ttacctgaag aggcaatccc ggactcagtg gatagatata gcaggcagat tagaatgcct  20400
tatcatgatg gttatcttgc tgtcactcct gtgatatctc acgttgttca aagcaaaata  20460
```

-continued

```
caacaagctg ctattgacaa aagagctagg ttttctaatg ttgagtttac cagaccagct  20520
gctgtttcaa tgctggcggc aagtctaggt ggtgtgataa atgtacttaa ttatcctcca  20580
tacattcgat caaagtacca tggactatcg aattcaaggg cttttaagtt aaacaacgga  20640
cagactgtct tcaacgtgga agccctcctg aagcctgagc ttataaaagc tcttgaagga  20700
attatattct ctaacaatgc attggcatta aaacagagaa ggcagcagaa agtaaaaaac  20760
atcaaagaac tgagaaatac cctccttgag tggttttccc ctgttttcga atggagattg  20820
gatgcaattg agaatggtta tgatctagaa caattggaat cagcgagcga acggttagag  20880
tacaagattc tgagtttgcc ggataatgaa ctaccttctt tgactatacc cttgtttaga  20940
ttacttaatg agatgttggg aggagtttca atgactcaga ggtatgcatt ccatccaaag  21000
ttgatgtctc cattaaaggc tgctctccaa tggttgctcg ttaatttaac tgatcaaaag  21060
catgttttaa ttgaagaaga tgatgagcat tacagatacc ttcacctgtc agggattagg  21120
gtttttgatg cccaagcatt gtctaatcct tactgctctg gtattccttc gcttacggcg  21180
gtttggggaa tgatccacag ttatcaacgc aaactaaacg aagctttggg cacgaacgta  21240
cggttcacaa gcttttcatg gttcatccgt aattactctg ccgttgctgg taaaaagcta  21300
ccagaactta gccttcaggg tgcacaacag tctagattga aaagacctgg gataatagac  21360
gggaagtatt gtgatttggt gtttgacttg attattcata tcgatggcta tgaagacgat  21420
ttgcaggctg ttgattcaaa gcctgatatc ctcaaggctc atttcccaag caatttcgcc  21480
ggcggcgtga tgcatcaacc cgagttgaat tctaatatta attggtgttg tctttactcc  21540
aacgaaaacc agctttttga gaagctccgt cggcttcctc tttctggatg ttgggttatg  21600
cccacagaac ataagattca agatcttgat gaactgttgt tgctgttgaa ctcggatagt  21660
aaactaagtc caagcatgat gggatacatg cttctcacag agccaatggc tcgtgtagga  21720
tcgttagaga ggcttcactg ttatgcagaa cctgctattg gtgttgtcaa atacgaggca  21780
gcaacttctg ttcgattaaa aggtatcgga aactatttca attccgcgtt ttggatgctc  21840
gatgctcaag aaaagttcat gctcatgaag aaagtgtgag cttatatgaa gatgaagatg  21900
aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt tttcttggct  21960
tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat ctcattataa  22020
tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc ttctccagca  22080
tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat aagcgctaca  22140
gaggagtctg agcaggacaa ctcgcgtagt gagagttaca tgttcgttgg gttcttccga  22200
cacggacctg agttggccaa cgtcccacct gaggtctgtg ccccggtgat gagaagtgtg  22260
catctcgttc ttgcagctcg tcagtacttt cagaatcatg gcgtgcatgg tagaatgacc  22320
cttataacgg acttcgacat ggcaatcgct aggtagctta atgaataggc gtatcacgag  22380
gccctttcgt cttcaagaat tagcttggca attcccgatc tagtaacata gatgacaccg  22440
cgcgcgataa tttatcctag tttgcgcgct atattttgtt ttctatcgcg tattaaatgt  22500
ataattgcgg gactctaatc ataaaaaccc atctcataaa taacgtcatg cattacatgt  22560
taattattac atgcttaacg taattcaaca gaaattatat gataatcgtc gcaagaccgg  22620
caacaggatt caatcttaag aaactttatt gccaaatgtt tgaacgatcg ggatcgaaga  22680
tccattccaa gcctacagga acaggtggtg gcggccctcg gtgcgctcgt actgctccac  22740
gatggtgtag tcctcgttgt gggaggtgat gtccagcttg gcgtccacgt agtagtagcc  22800
gggcagctgc acgggcttct tggccatgta gatggacttg aactccacca ggtagtggcc  22860
gccgtccttc agcttcaggg ccttgtgtgt ctcgcccttc agcacgccgt cgcgggggta  22920
caggcgctcg gtggaggcct cccagcccat cgtcttcttc tgcatcacgg ggccgtcgga  22980
ggggaagttc acgccgatga acttcacctt gtagatgaag cagccgtcct gcagggagga  23040
gtcctgggtc acggtcgcca cgccgcgctc ctcgaagttc atcacgcgct cccacttgaa  23100
gccctcgggg aaggacagct tcttgtagtc ggggatgtcg gcggggtgct tcacgtacac  23160
cttggagccg tactggaact gggggggacag gatgtcccag gcgaagggca gggggccgcc  23220
cttggtcacc ttcagcttca cggtgttgtg gccctcgtag gggcggccct cgccctcgcc  23280
ctcgatctcg aactcgtggc cgttcacggt gccctccatg cgcaccttga agcgcatgaa  23340
ctcggtgatg acgttctcgg aggaggccat ttcgcctcga tcgagttgag agtgaatatg  23400
agactctaat tggataccga ggggaattta tggaacgtca gtggagcatt tttgacaaga  23460
aatatttgct agctgatagt gaccttaggc gacttttgaa cgcgcaataa tggtttctga  23520
cgtatgtgct tagctcatta aactccagaa acccgcggct cagtggctcc ttcaacgttg  23580
cggttctgtc agttccaaac gtaaaacggc ttgtcccgcg tcatcggcgg gggtcatacc  23640
gtgactccct taattctcat gtatggttcg ctatactgaa gagcaattct gacaggatat  23700
attggcgggt aaacctaaga gaaaagagcg tttattagaa taatcggata tttaaaaggg  23760
cgtgaaaagg tttatccgtt cgtccatttg tatgtgcatg ccaaccacag ggttcccctc  23820
gggatcaaag tactttgatc caacccctcc gctgctatag tgcagtcggc ttctgacgtt  23880
cagtgcagcc gtcatctgaa aacgacatgt cgcacaagtc ctaagttacg cgacaggctg  23940
ccgccctgcc cttttcctgg cgttttcttg tcgcgtgttt tagtcgcata aagtagaata  24000
cttgcgacta gaaccggaga cattacgcca tgaacaagag cgccgccgct ggcctgctgg  24060
gctatgcccg cgtcagcacc gacgaccagg acttgaccaa ccaacgggcc gaactgcacg  24120
cggccggctg caccaagctg ttttccgaga agatcaccgg caccaggcgc gaccgcccgg  24180
agctggccag gatgcttgac cacctacgcc ctggcgacgt tgtgacagtg accaggctag  24240
accgcctggc ccgcagcacc cgcgacctac tggacattgc cgagcgcatc caggaggccg  24300
gcgcgggcct gcgtagcctg gcagagccgt gggccgacac caccacgccg gccggccgca  24360
tggtgttgac cgtgttcgcc ggcattgccg agttcgagcg ttccctaatc atcgaccgca  24420
cccggagcgg gcgcgaggcc gccaaggccc gaggcgtgaa gtttggcccc cgccctaccc  24480
tcaccccggc acagatcgcg cacgcccgcg agctgatcga ccaggaaggc cgcaccgtga  24540
aagaggcggc tgcactgctt ggcgtgcatc gctcgaccct gtaccgcgca cttgagcgca  24600
gcgaggaagt gacgcccacc gaggccaggc ggcgcggtgc cttccgtgag gacgcattga  24660
ccgaggccga cgccctggcg gccgccgaga atgaacgcca agaggaacaa gcatgaaacc  24720
gcaccaggac ggccaggacg aaccgttttt cattaccgaa gagatcgagg cggagatgat  24780
cgcggccggg tacgtgttcg agccgcccgc gcacctctca accgtgcggc tgcatgaaat  24840
cctggccggt ttgtctgatg ccaagctggc ggcctggccg gccagcttgg ccgctgaaga  24900
aaccgagcgc cgccgtctaa aaaggtgatg tgtatttgag taaaacagct tgcgtcatgc  24960
ggtcgctgcg tatatgatgc gatgagtaaa taaacaaata cgcaagggga acgcatgaag  25020
```

-continued

```
gttatcgctg tacttaacca gaaaggcggg tcaggcaaga cgaccatcgc aacccatcta  25080
gcccgcgccc tgcaactcgc cggggccgat gttctgttag tcgattccga tccccagggc  25140
agtgcccgcg attgggcggc cgtgcgggaa gatcaaccgc taaccgttgt cggcatcgac  25200
cgcccgacga ttgaccgcga cgtgaaggcc atcggccggc gcgacttcgt agtgatcgac  25260
ggagcgcccc aggcggcgga cttggctgtg tccgcgatca aggcagccga cttcgtgctg  25320
attccggtgc agccaagccc ttacgacata tgggccaccg ccgacctggt ggagctggtt  25380
aagcagcgca ttgaggtcac ggatggaagg ctacaagcgg cctttgtcgt gtcgcgggcg  25440
atcaaaggca cgcgcatcgg cggtgaggtt gccgaggcgc tggccgggta cgagctgccc  25500
attcttgagt cccgtatcac gcagcgcgtg agctacccag gcactgccgc cgccggcaca  25560
accgttcttg aatcagaacc cgagggcgac gctgcccgcg aggtccaggc gctggccgct  25620
gaaattaaat caaaactcat ttgagttaat gaggtaaaga gaaaatgagc aaaagcacaa  25680
acacgctaag tgccggccgt ccgagcgcac gcagcagcaa ggctgcaacg ttggccagcc  25740
tggcagacac gccagccatg aagcgggtca actttcagtt gccggcggag gatcacacca  25800
agctgaagat gtacgcggta cgccaaggca agaccattac cgagctgcta tctgaataca  25860
tcgcgcagct accagagtaa atgagcaaat gaataaatga gtagatgaat tttagcggct  25920
aaaggaggcg gcatggaaaa tcaagaacaa ccaggcaccg acgccgtgga atgccccatg  25980
tgtggaggaa cgggcggttg gccaggcgta agcggctggg ttgcctgccg gccctgcaat  26040
ggcactggaa cccccaagcc cgaggaatcg gcgtgagcgg tcgcaaacca tccggcccgg  26100
tacaaatcgg cgcggcgctg ggtgatgacc tggtggagaa gttgaaggcc gcgcaggccg  26160
cccagcggca acgcatcgag gcagaagcac gccccggtga atcgtggcaa gcggccgctg  26220
atcgaatccg caaagaatcc cggcaaccgc cggcagccgg tgcgccgtcg attaggaagc  26280
cgcccaaggg cgacgagcaa ccagattttt tcgttccgat gctctatgac gtgggcaccc  26340
gcgatagtcg cagcatcatg gacgtggccg ttttccgtct gtcgaagcgt gaccgacgag  26400
ctggcgaggt gatccgctac gagcttccag acgggcacgt agaggtttcc gcagggccgg  26460
ccggcatggc cagtgtgtgg gattacgacc tggtactgat ggcggtttcc catctaaccg  26520
aatccatgaa ccgataccgg gaagggaagg gagacaagcc cggccgcgtg ttccgtccac  26580
acgttgcgga cgtactcaag ttctgccggc gagccgatgg cggaaagcag aaagacgacc  26640
tggtagaaac ctgcattcgg ttaaacacca cgcacgttgc catgcagcgt acgaagaagg  26700
ccaagaacgg ccgcctggtg acggtatccg agggtgaagc cttgattagc cgctacaaga  26760
tcgtaaagag cgaaaccggg cggccggagt acatcgagat cgagctagct gattggatgt  26820
accgcgagat cacagaaggc aagaacccgg acgtgctgac ggttcacccc gattactttt  26880
tgatcgatcc cggcatcggc cgttttctct accgcctggc acgccgcgcc gcaggcaagg  26940
cagaagccag atggttgttc aagacgatct acgaacgcag tggcagcgcc ggagagttca  27000
agaagttctg tttcaccgtg cgcaagctga tcgggtcaaa tgacctgccg gagtacgatt  27060
tgaaggagga ggcggggcag gctggcccga tcctagtcat gcgctaccgc aacctgatcg  27120
agggcgaagc atccgccggt tcctaatgta cggagcagat gctagggcaa attgccctag  27180
caggggaaaa aggtcgaaaa ggactctttc ctgtggatag cacgtacatt gggaacccaa  27240
agccgtacat tgggaaccgg aacccgtaca ttgggaaccc aaagccgtac attgggaacc  27300
ggtcacacat gtaagtgact gatataaaag agaaaaaagg cgatttttcc gcctaaaact  27360
ctttaaaact tattaaaact cttaaaaccc gcctggcctg tgcataactg tctggccagc  27420
gcacagccga agtgctgcaa aaagcgccta cccttcggtc gctgcgctcc ctacgccccg  27480
ccgcttcgcg tcggcctatc gcggccgctg gccgctcaaa aatggctggc ctacggccag  27540
gcaatctacc agggcgcgga caagccgcgc cgtcgccact cgaccgccgg cgcccacatc  27600
aaggcaccct gcctcgcgcg tttcggtgat gacggtgaaa acctctgaca catgcagctc  27660
ccggtgacgg tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc  27720
gcgtcagcgg gtgttggcgg gtgtcggggc gcagccatga cccagtcacg tagcgatagc  27780
ggagtgtata ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaccata  27840
tgcggtgtga aataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgctcatccg  27900
cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc  27960
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt  28020
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc  28080
ataggctccg ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa  28140
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc  28200
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg  28260
cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc  28320
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc  28380
gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca  28440
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact  28500
acggctacac tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg  28560
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt  28620
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct  28680
tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatgc  28740
attctaggtg attagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca  28800
ggattatcaa taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg  28860
aggcagttcc ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca  28920
tcaatacaac ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca  28980
tgagtgacga ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt  29040
tcaacaggcc agccattacg ctcgtcatca aaatcactcg catcaaccaa accgttattc  29100
attcgtgatt gcgcctgagc gagtcgaaat acgcgatcgc tgttaaaagg acaattacaa  29160
acaggaatcg aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct  29220
gaatcaggat attcttctaa tacctggaat gctgttttcc ctgggatcgc agtggtgagt  29280
aaccatgcat catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc  29340
gtcagccagt ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca  29400
tgtttcagaa acaactctgg cgcatcgggc ttcccataca atcggtagat tgtcgcacct  29460
gattgcccga cattatcgcg agcccattta tacccatata aatcagcatc catgttggaa  29520
tttaatcgcg gccttgagca agacgtttcc cgttgaatat ggctcataac agaacttatt  29580
```

-continued

```
atttccttcc tcttttctac agtatttaaa gataccccaa gaagctaatt ataacaagac  29640
gaactccaat tcactgttcc ttgcattcta aaaccttaaa taccagaaaa cagctttttc  29700
aaagttgttt tcaaagttgg cgtataacat agtatcgacg gagccgattt tgaaaccgcg  29760
gtgatcacag gcagcaacgc tctgtcatcg ttacaatcaa catgctaccc tccgcgagat  29820
catccgtgtt tcaaacccgg cagcttagtt gccgttcttc cgaatagcat cggtaacatg  29880
agcaaagtct gccgccttac aacggctctc ccgctgacgc cgtcccggac tgatgggctg  29940
cctgtatcga gtggtgattt tgtgccgagc tgccggtcgg ggagctgttg gctggctgg   29999

SEQ ID NO: 43          moltype = DNA  length = 29999
FEATURE                Location/Qualifiers
source                 1..29999
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 43
tggcaggata tattgtggtg taaacataac gaattgctct tcagcaagga gaatgaggag  60
aatgtagcag aaggcatgtt gttgtgactc cgaggggttg cctcaaactc tatcttataa  120
ccggcgtgga ggcatggagg caggggtatt ttggtcattt taatagatag tggaaaatga  180
cgtggaattt acttaaagac gaagtctttg cgacaagggg gggcccacgc cgaatttaat  240
attaccggcg tggcccccc ttatcgcgag tgctttagca cgagcggtcc agatttaaag  300
tagaaaattt cccgcccact agggttaaag gtgttcacac tataaaagca tatacgatgt  360
gatggtattt gatggagcgt atattgtatc aggtatttcc gttggatacg aattattcgt  420
acgaccctcg gtatactaat gaggagtgtc gctgcaacca tagtttgaca taatgcaacc  480
ataccttgac ataattaaag tcatactctg acataaagcc ccttattaaa aatgatgcta  540
aaaataaatt ccattataat caattattta agaattaatt aaaccatagt gtgacatagt  600
ggctttttat ttgtatagta tgtgtactca aataataaaa atttacttat gtatatacgc  660
aacttaagaa agccatctcc aaataaaaat gtatttcgct agcaaggagt gagacttttc  720
aacaaagggt aatatcggga aacctcctcg gattccattg cccagctatc tgtcacttca  780
tcaaaaggac agtagaaaag gaaggtggca cctacaaatg ccatcattgc gataaaggaa  840
aggctatcgt tcaagatgcc tctgccgaca gtggtcccaa agatggaccc ccacccacga  900
ggagcatcgt ggaaaaagaa gacgttccaa ccacgtcttc aaagcaagtg gattgatgtg  960
atatctccac tgacgtaagg gatgacgcac aatcccacta tccttcgcaa gacccttcct  1020
ctatataagg aagttcattt catttggaga ggacacgctg acggccgcgt atttttacaa  1080
caattaccaa caacaacaaa caacaaacaa cattacaatt actatttaca attacagcgg  1140
ccgccccctt caccaaaaaa atggctgcag ccaagggcga ggagctgttc accggggtgg  1200
tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg  1260
agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gctgctctgc accaccggca  1320
agctgcccgt gccctggccc accctcgtga ccaccctggg ctacggcgtg cagtgcttcg  1380
cccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg cccgaaggct  1440
acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg  1500
tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcatc gacttcaagg  1560
aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac aacgtctata  1620
tcaccgccga caagcagaag aacggcatca aggccaactt caagatccgc cacaacatcg  1680
aggacggcgg cgtgcagctc gccgaccact accagcagaa cacccccatc ggcgacggcc  1740
ccgtgctgct gcccgacaac cactacctga gctaccagtc caaactgttc aaagacccca  1800
acgagaagcg cgatcacatg gtcttgttag aatttttgac tgctgctggt attaccgagg  1860
gtatgaatga attgtacaaa ggctccggac cgaagaagaa gaggaaggtt taaatatgaa  1920
gatgaagatg aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt  1980
tttcttggct tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat  2040
ctcattataa tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc  2100
ttctccagca tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat  2160
aagcgctata caggagccgt ataggcagct gaagataaaa acaatcgatt cgatgtttaa  2220
gttaatgtaa ttctgccgaa aaggcagtga gtagtaattg aaaaaaccag tcttatatcc  2280
gggctggttt ttttgttttt agcgttataa cttttatttc actatgtcac tgtatggttg  2340
cattttatgt cagtgaatgg ttgcaggttt tttgtaagtt attgatatgc ctatccagtg  2400
tatgtcacgc tttgcttgca gcgacacgct acagaggagt ctgagcagga caactcgcgt  2460
agtgagagtt acatgttcgt tgggttcttc cgacacggac ctgagttggc caacgtccca  2520
cctgaggtct gtgccccggt gatgagaagt gtgcatctcg ttcttgcagc tcgtcagtac  2580
tttcagaatc atggcgtgca tggtagaatg acccttataa cggacttcga catggcaatc  2640
gctaggtaaa tgaatggaaa aagtggcacc gagtcggtgc tttttttccc ggggcgcgcc  2700
gatatcgagc tctcccggcg cgccgatatc gagctcagtg tttgatcgcc ggcggtaccg  2760
agtgtacttc aagtcagtgg gaaatcaata aaatgattat tttatgaata tatttcattg  2820
tgcaagtaga tagaaattac atatgttaca taacacacga aataaacaaa aaaagacaat  2880
ccaaaaacaa acaccccaaa aaaaataatc actttagata aactcgtatg aggagaggca  2940
cgttcagtga ctcgacgatt cccgagcaaa aaaagtctcc ccgtcacaca tgtagtgggt  3000
gacgcaatta tctttaaagt aatccttctg ttgacttgtc attgataaca tccagtcttc  3060
gtcaggattg caaagaatta tagaagggat cccacctttt attttcttct tttttccata  3120
tttagggttg acagtgaaat cagactggca acctattaat tgcttccaca atgggacgaa  3180
cttgaagggg atgtcgtcga tgatattata ggtggcgtgt tcatcgtagt tggtgaaatc  3240
gatggtaccg ttccaatagt tgtgtcgtcc gagacttcta gcccaggtgg tctttccggt  3300
acgagttggt ccgcagatgt agaggctggg gtgtcggatt ccattccttc cattgtcctt  3360
gttaaatcgg ccatccattc aaggtcagat tgagcttgtt ggtatgagac aggatgtatg  3420
taagtataag cgtctatgct tacatggtat agatgggttt ccctccagga gtgtagatct  3480
tcgtggcagc gaagatctga ttctgtgaag ggcgacacat acggttcagg ttgtggaggg  3540
aataatttgt tggctgaata ttccagccat tgaagctttg ttgcccattc atgagggaat  3600
tcttccttga tcatgtcaag atattcctcc ttagacgttg cagtctggat aatagttctc  3660
catcgtgcgt cagatttgcg aggagaaacc ttatgatctc ggaaatctcc tctggtttta  3720
```

-continued

```
atatctccgt cctttgatat gtaatcaagg acttgtttag agtttctagc tggctggata  3780
ttagggtgat ttccttcaaa atcgaaaaaa gaaggatccc taatacaagg ttttttatca  3840
agctggagga gagcatgata gtgggtagtg ccatcttgat gaagctcaga agcaacacca  3900
aggaagaaaa taagaaaagg tgtgagtttc tcccagagaa actggaataa atcatctctt  3960
tgagatgagc acttgggata ggtaaggaaa acatatttag attggagtct gaagttctta  4020
ctagcagaag gcatgttgtt gtgactccga ggggttgcct caaactctat cttataaccg  4080
gcgtggaggc atggaggcag gggtattttg gtcattttaa tagatagtgg aaaatgacgt  4140
ggaatttact taaagacgaa gtctttgcga caaggggggg cccacgccga atttaatatt  4200
accggcgtgg cccccccttta tcgcgagtgc tttagcacga gcggtccaga tttaaagtag  4260
aaaatttccc gcccactagg gttaaaggtg ttcacactat aaaagcatat acgatgtgat  4320
ggtatttgat ggagcgtata ttgtatcagg tatttccgtt ggatacgaat tattcgtacg  4380
accctcatag tttaaactat cagtgggtag cttaatggta gacactttat tagttagatc  4440
cgctgttatt ctgcacgatt gcgcttgccg gtggaaattt aaggacgagt gtagtgcact  4500
cacccttagc cttcagcgtg acaagattag ggtagcaccg gaaacagcga ttttacggac  4560
agatagtaca gtctgagact cagtcactgg tccctaaacg ggagccggta cgtcggcggt  4620
tcgctagcaa ggagtggcag acatactgtc ccacaaatga agatggaatc tgtaaaagaa  4680
aacgcgtgaa ataatgcgtc tgacaaaggt taggtcggct gcctttaatc aataccaaag  4740
tggtccctac cacgatggaa aaactgtgca gtcggtttgg ctttttctga cgaacaaata  4800
agattcgtgg ccgacaggtg ggggtccacc atgtgaaggc atcttcagac tccaataatg  4860
gagcaatgac gtaagggctt acgaaataag taagggtagt ttgggaaatg tccactcacc  4920
cgtcagtcta taaatactta gcccctccct cattgttaag ggagcaaaat ctcagagaga  4980
tagtcctaga gagagaaaga gagcaagtag cctagaagta gtcaaggcgg cgaagtattc  5040
aggcacgtgg ccaggaagaa gaaaagccaa gacgacgaaa acaggtaaga gctaagctta  5100
atgggaaagc gtactgctga tggttcagag tttgaaagtc caaagaagaa aagaaaagtt  5160
ggtagtggta tgactgataa acggaaagat ggttccggca agcttcttta ctgcagcttt  5220
tgcgggaaat ctcaacatga agttaggaaa cttattgcag ggccttctgt ttatatctgt  5280
gacgaatgtg ttgatctctg caatgatatt attcgagaag agatcaaaga agtggctcct  5340
cacagagaga ggagtgcgtt gccgacacca cacgagatta gaaatcatct tgatgattat  5400
gtcattggtc aagaacaagc aaagaaggtc cttgctgttg cagtgtacaa ccattacaag  5460
aggctcagaa acggagatac ttctaatggt gtggaacttg gcaaatcaaa catcctgctc  5520
attggaccca ctggaagtgg aaagacactc ctagcagaga cattggcaag gctcttagac  5580
gttcctttca ctatggctga tgcaactact cttactgagg ctggttatgt tggagaagat  5640
gtggagaata ttattcaaaa gttattgcag aaatgtgatt acgatgtcca gaaagctcag  5700
agagggatag tctacataga tgagatagac aagatctcta ggaagagtga taatccctcg  5760
ataacacgcg acgtttctgg cgaaggtgtt caacaagctt tgttgaagct cattgaaggc  5820
accgtagcgg cagttccacc acaaggagga agaaagcatc ctcagcaaga gtttctccaa  5880
gtggatacct caaaaatctt atttatatgt ggtggtgcct tcgctgggct tgacaaagtt  5940
atctcacacc gtgtagaaac tggtagcggg attgggtttg gtgcaacggt gaaggcaaaa  6000
tctgataagg cttctgaagg agagcttctg gctcaagttg aaccagagga tcttatcaag  6060
tttggactaa ttcctgagtt catcggtaga cttcctgtcg ttgctacgtt gaacgagtta  6120
agtgaggaag ccctgattca aatcttgaaa gagccaaaga atgctctgac aaagcaatat  6180
caggctcttt tcaatcttga gggagttgac ttggaatttc gtgatgaagc gctggatgct  6240
atagctaaga aagcaatggc gaggaaaacc ggagctcgtg gtctaagatc gattgttgaa  6300
gccgctttat tggatacaat gtatgatttg ccctccatgg aagatgttga gaaagtggtc  6360
atcgacgagt ctgtgattga tggacagtca aagccattgc tcatatatgg taaacctgaa  6420
gcacagcagg catcaggaga atgaatatga agatgaagat gaaatatttg gtgtgtcaaa  6480
taaaaaggtt gtgtgcttaa gtttgtgttt ttttcttggc ttgttgtgtt atgaatttgt  6540
ggctttttct aatattaaat gaatgtaaca tctcattata atgaataaac aaatgtttct  6600
ataatccatt gtgaatgttt tgttggatct cttctccagc atataactac tgtatgtgct  6660
atggtatgga ctatggaata tgattaaaga taagcgctat acaggagtgg cagacatact  6720
gtcccacaaa tgaagatgga atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa  6780
ggttaggtcg gctgccttta atcaatacca aagtggtccc taccacgatg gaaaaactgt  6840
gcagtcggtt tggctttttc tgacgaacaa ataagattcg tggccgacag gtgggggtcc  6900
accatgtgaa ggcatcttca gactccaata atggagcaat gacgtaaggg cttacgaaat  6960
aagtaagggt agtttgggaa atgtccactc acccgtcagt ctataaatac ttagcccctc  7020
cctcattgtt aagggagcaa aatctcagag agatagtcct agagagagaa agagagcaag  7080
tagcctagaa gtagtcaagg cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc  7140
caagacgacg aaaacaggta agagctaagc ttaatgggaa agcgtactgc tgatggttca  7200
gagtttgaaa gtccaaagaa gaaaagaaaa gttggtagtg gtatgagtta ttctggagaa  7260
cgggataatt ttgcacctca catggcatta gttccaatgg tgattgaaca aacaagtaga  7320
ggggagcgaa gctttgacat ttactctcgt ctattaaaag aaagggtcat cttcttaacc  7380
gggcaagttg aagatcatat ggcaaatctg attgtagctc agatgctttt tcttgaagca  7440
gagaatcctg agaaagatat ctatctctac attaactctc ctggtggtgt aataactgct  7500
ggaatgtcta tatatgatac aatgcagttt atcaaacctg atgtttcaac aatatgtatg  7560
ggtcaggctg cgtcaatggg agcttttctt ttgacggctg gcgcaaaagg aaagcggttc  7620
tgcctaccga actcaagggt gatgatccat caaccacttg gaggatacca aggtcaagct  7680
accgatatag agattcatgc tagagaaatt ttgaaggtga agggtagaat gaatgaactg  7740
atggcccttc acactggaca atcgctcgag cagattgagc gtgacacaga gagagacagg  7800
tttttatccg caccagaagc tgttgagtat ggtttggttg attctattct tactcataga  7860
aactgaatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  7920
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  7980
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  8040
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  8100
tatgattaaa gataagcgct acagaggagc atcttcattc ttaagatatg aagataatct  8160
tcaaaaggcc cctgggaatc tgaaagaaga gaagcaggcc catttatatg ggaaagaaca  8220
atagtatttc ttatataggc ccatttaagt tgaaaacaat cttcaaaagt cccacatcgc  8280
```

-continued

```
ttagataaga aaacgaagct gagtttatat acagctagag tcgaagtagt gataatggtg  8340
acctgccgta taggcagctg aagatatccc actatccttc gcaagaccct tcctctataa  8400
ttctgccgaa aaggcagtga gtagtgcttt ttttttgca aaattttcca gatcgatttc  8460
ttcttcctct gttcttcggc gttcaatttc tggggttttc tcttcgtttt ctgtaactga  8520
aacctaaaat ttgacctaaa aaaaatctca aataatatga ttcagtggtt ttgtactttt  8580
cagttagttg agttttgcag ttccgatgag ataaaccaat acgctaggta tactaatgag  8640
gagtggcaga catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa  8700
taatgcgtct gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc  8760
acgatggaaa aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc  8820
cgacaggtgg gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg  8880
taagggctta cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat  8940
aaatacttag cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag  9000
agagaaagag agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc  9060
caggaagaag aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tgtacattag  9120
gaacttgaga aagccttctc caaataagaa tgtttttaag ttcgcatcta ccaaggtcag  9180
ctctgtggtc atgtgtgaat ccagccttga gtttgacgct tgctttcatc atgaatataa  9240
cgacttgatt gagagttttg gttctcagcc agaaggtttc aaatacgaat tcatgggtaa  9300
atctctccct tacactccag atgctttgat ctcctacacg gacaaaactc aaaagtatca  9360
cgagtataaa ccatattcaa aaatcgcatc acctctattt agagcagagt ttgctgcaaa  9420
gagggctgca agtcttaagt tgggtattga tttggtactc gtcacagaca gacagattag  9480
agtaaaccca atactcaaca atctgaagct ccttcataga tactcgggcg tttatggtat  9540
atctggtatc cagaaggaac tacttagctt tattcacaaa agtggcgtta tcaaactcaa  9600
tgatatctca tcacaggttg ggatccccat tggagaaact cggtcatttc tttttggact  9660
gatgcataag ggacttgtaa aggctgacct aggttgtgat gacctgacaa acaaccctac  9720
tctttgggct actcctggaa gtggctcagg caaacgtact gctgacggaa gtgagttcga  9780
gtccccaaag aaaaaaagaa aagttggtag tggatctggt ggaatgactg atttcttcaa  9840
tgagtttgat gaatcgcttg taccactgaa accacaaact ccaacgcagt atgttaagct  9900
ggacgatgct aatttaattc agagagattt agatacattc tcagatactt tcaagaacca  9960
agccctccaa agatataagc taatttctac aattgataaa aagttaagcc gcggatggac  10020
tcaaagaaat ctcgacccga tcctcgacga actcttcaaa ggcggtgatg tagtgaggcc  10080
aaactggaga actgtggcaa gatggaggaa gaaatatata gaaagcaatg gtgacatcgc  10140
ttctcttgct gataaaaacc acaaaatggg aaacagaaca aatcggataa aaggagatga  10200
caaatttttt gacaaggcct tggaacgttt ccttgacgcg aaaagaccta caatagcaac  10260
agcttatcaa tattacaagg acttgatagt gattgaaaac gagtcaattg ttgaagggaa  10320
gatccccatc atctcttata atgcatttaa caaacgaatt aaagccattc ctccttatgc  10380
tgttgcagtc gctaggcatg ggaaattcaa agctgatcaa tggtttgcat attgtgcagc  10440
tcacgttcca ccaacgcgta tacttgaacg agtggagatt gatcatactc cgttggatct  10500
aatattattg gatgatgagc ttctgattcc cattggccgg ccatatctta cacttctcat  10560
agatgtgttt tcaggttgcg ttctgggatt ccatttgtca tacaagagcc cttcttatgt  10620
ctctgcagca aaagcaatca ctcacgcaat aaaacccaaa agcttagatg cgcttaatat  10680
tgagcttcag aatgactggc cctgtttcgg gaagtttgag aatttagtag ttgacaatgg  10740
agctgaattt tggtcaaaaa atttggaaca tgcttgtcaa tctgctggta taaatattca  10800
atacaatccg gtgcgtaagc catggctcaa gccttttata gagaggttct ttggggtgat  10860
gaatgaatac ttcttgccgg aactcccagg aaaaacattt tctaacattc ttgagaaaga  10920
agaatacaaa ccggagaagg atgcaataat gaggttcagc acctttgttg aggagttcca  10980
ccgttggatt gccgacgtgt atcatcaaga tagtaatagc cgtgaaacgc gtattcctat  11040
taagcgatgg caacaaggat tcgatgctta tcctccttta acaatgaacg aagaagagga  11100
gacacggttt tctatgctca tgagaatctc tgactctagg acgctaacca ggaatgggtt  11160
taagtaccaa gaattgatgt atgattctac cgccttagca gattaccgta agcattaccc  11220
tcagacaaaa gagactgtga agaagctaat aaaggtcgat ccagatgata tctccaagat  11280
ttacgtttat ctggaagagt tggaaagtta cttggaggtt ccttgcacag atccaactgg  11340
atacactgat ggtttatcca tatatgagca taaaaccatt aagaagatca accgagaagt  11400
tattagagaa tcaaaagata gccttgggtt ggcgaaagcg agaatggcta ttcatgagag  11460
agtgaaacaa gagcaagaag tattcatcga aagtaagacc aaagctaaga ttactgcagt  11520
caagaagcag gctcagatcg cggatgttag taatacaggc acatcaacaa tcaaggtgtc  11580
cgaggagtct gcggctcctg tccagaaaca catctcaaat gataacagtg atgattggga  11640
tgatgatctg gaagcttttg aatgagctta tatgaagatg aagatgaaat atttggtgtg  11700
tcaaataaaa aggttgtgtg cttaagtttg tgtttttttc ttggcttgtt gtgttatgaa  11760
tttgtggctt tttctaatat taaatgaatg taacatctca ttataatgaa taaacaaatg  11820
tttctataat ccattgtgaa tgttttgttg gatctcttct ccagcatata actactgtat  11880
gtgctatggt atggactatg gaatatgatt aaagataagc gctagcaagg agtggcagac  11940
atactgtccc acaaatgaag atggaatctg taaaagaaaa cgcgtgaaat aatgcgtctg  12000
acaaaggtta ggtcggctgc ctttaatcaa taccaaagtg gtccctacca cgatggaaaa  12060
actgtgcagt cggtttggct ttttctgacg aacaaataag attcgtggcc gacaggtggg  12120
ggtccaccat gtgaaggcat cttcagactc caataatgga gcaatgacgt aagggcttac  12180
gaaataagta agggtagttt gggaaatgtc cactcacccg tcagtctata aatacttagc  12240
ccctccctca ttgttaaggg agcaaaatct cagagagata gtcctagaga gagaaagaga  12300
gcaagtagcc tagaagtagt caaggcggcg aagtattcag gcacgtggcc aggaagaaga  12360
aaagccaaga cgacgaaaac aggtaagagc taagcttaat gggaaatgct cttacagaga  12420
ttcaaataga gaagttacgg aatttcagtg attgcatcgt gatgcatcct cagataaaga  12480
ccattttcaa tgattttgat gaactaaggc ttaatagaaa gtttcagagt gatcagcaat  12540
gtatgcttct tataggagat actggtgttg gcaaatcaca caccatcaac cattacaaga  12600
aaagagtcct cgcaacacaa aactattcaa gaaacacaat gccggtcctt gtttctcgca  12660
tatcacgtgg gaaaggtttg gatgctactt tggtgcagat gctggcggat ctagagctgt  12720
ttggttcttc tcagatcaag aagcggggct acaaaacaga ccttaccaaa aaattagttg  12780
agagtttgat caaagctcaa gtggagctgt tgattatcaa tgaattccaa gagcttattg  12840
```

-continued

```
agtttaaaag tgttcaagaa agacaacaga ttgcaaatgg tctaaagttt attagtgaag  12900
aagccaaggt tcccattgtt ttggttggaa tgccatgggc agcaaagata gctgaggaac  12960
cacaatgggc ttctcgactc gtaagaaaac gaaagctaga atatttttca ctcaagaatg  13020
acagcaagta ctttaggcag tatttgatgg gtcttgccaa aaagatgcct ttcgacgttc  13080
ctcctaaatt agaatcgaag aacactacga ttgcgctatt tgcagcttgt agaggggaaa  13140
acagagcttt gaagcacctt cttttagaag cactgaagct cgccctttca tgcaatgagt  13200
atcttgagaa caagcatttc atcacagctt atgacaaatt cgatttcttc aacgacaagg  13260
agaaactgaa atccaaaaac cctttcaagc aagatataaa agatattgaa atttatgaag  13320
ttattaaaaa ttcttcctac aatccaaatg cacttgatcc agaagatatg ttgactgatc  13380
gagtatttgc tattgtgaaa ggaagcggaa agaggactgc agatggatca gaatttgagt  13440
ccccaaagaa gaaaaggaaa gtgtgagctt atatgaagat gaagatgaaa tatttggtgt  13500
gtcaaataaa aaggttgtgt gcttaagttt gtgtttttt cttggcttgt tgtgttatga  13560
atttgtggct ttttctaata ttaaatgaat gtaacatctc attataatga ataaacaaat  13620
gtttctataa tccattgtga atgttttgtt ggatctcttc tccagcatat aactactgta  13680
tgtgctatgg tatggactat ggaatatgat taaagataag cgctatacag gagtggcaga  13740
catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa taatgcgtct  13800
gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc acgatggaaa  13860
aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc cgacaggtgg  13920
gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg taagggctta  13980
cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat aaatacttag  14040
cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag agagaaagag  14100
agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc caggaagaag  14160
aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tgggaaagcg tactgctgat  14220
ggttcagagt ttgaaagtcc aaagaagaaa agaaaagttg gtagtggtat ggcttttctt  14280
ttctcaccaa aagcaagagc gttttccgat gaatcgcttg aaagttattt acttcgagtt  14340
gtctcagaaa acttcttcga ttcatatgaa ggactatctt tggccatcag agaagaactt  14400
catgagttgg acttcgaggc tcatggagca tttcccgtag atttaaagag acttaatgtt  14460
taccatgcaa aacacaactc acactttagg atgagagctc ttggcctatt agaaactctt  14520
ttagatcttc ctcgttatga gcttcaaaag ctagcactcc tgaagtcaga cattaaattc  14580
aattcttccg ttgccctcta caacaatggt gttgatattc ctttgagatt cattagacat  14640
catgctgagg aagctgttga ctccattccg gtttgttctc aatgcttggc tgaagaagct  14700
tatataaagc aatcttggca tatcaaatgg gttaatgctt gcaccaaaca tcaatgtgct  14760
ttgcttcaca attgtcctga atgttatgca ccgatcaatt acattgagaa tgaatcaatc  14820
actcattgct cttgtggttt tgaattgagc tgcgcctcta caagtcctgt gaacacactc  14880
tcgattgaac atttgaacaa actcctggac aaaggagaac ggaatgactc aaatccattg  14940
tttaacaaca tgacgttgac agaaaggttt gctgctttat tatggtatca agagaggtat  15000
tcacagactg ataatttctg tttgaatgat gctgtgaact acttcagtaa gtggccagca  15060
gtgtttaata ctgagctaga tgagctctca aagaatgccg agatgaagct gattgatctc  15120
ttcaacaaga cagagtttaa gtttatcttt ggagatgcaa ttcttgcttg tccaagtact  15180
cagaagcaga gtgaatctca tttcatctac cgggcattac ttgattacct tgtcacttta  15240
gtggagagca atcccaaaac aaagaagcca aacgcagctg atttgcttgt ttctgtcctg  15300
gaagcagcaa ctttgttggg gacgagtgta gaacaagttt ataggctcta tcagaatgga  15360
atactgcaaa ccgcctttcg gcacaagatg aatcagagaa taaaccctta caagggtgct  15420
ttctttctcc gtcatgtgat agagtataaa accagctttg gtaacgataa agctagaatg  15480
tatctctcag cttggtgagc ttatatgaag atgaagatga aatatttggt gtgtcaaata  15540
aaaaggttgt gtgcttaagt ttgtgttttt ttcttggctt gttgtgttat gaatttgtgg  15600
ctttttctaa tattaaatga atgtaacatc tcattataat gaataaacaa atgtttctat  15660
aatccattgt gaatgttttg ttggatctct tctccagcat ataactactg tatgtgctat  15720
ggtatggact atggaatatg attaaagata agcgctacag aggagtctga gcaggacaac  15780
tcgcgtagtg agagttacat gttcgttggg ttcttccgac acggacctga gttggccaac  15840
gtcccacctg aggtctgtgc cccggtgatg agaagtgtgc atctcgttct tgcagctcgt  15900
cagtactttc agaatcatgg cgtgcatggt agaatgaccc ttataacgga cttcgacatg  15960
gcaatcgcta ggtaaatgaa tgaggagtgg cagacatact gtcccacaaa tgaagatgga  16020
atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa ggttaggtcg gctgccttta  16080
atcaatacca aagtggtccc taccacgatg gaaaaactgt gcagtcggtt tggctttttc  16140
tgacgaacaa ataagattcg tggccgacag gtgggggtcc accatgtgaa ggcatcttca  16200
gactccaata atggagcaat gacgtaaggg cttacgaaat aagtaagggt agtttgggaa  16260
atgtccactc acccgtcagt ctataaatac ttagcccctc cctcattgtt aagggagcaa  16320
aatctcagag agatagtcct agagagagaa agagagcaag tagcctagaa gtagtcaagg  16380
cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc caagacgacg aaaacaggta  16440
agagctaagc ttaatgggaa agcgtactgc tgatggttca gagtttgaaa gtccaaagaa  16500
gaaaagaaaa gttggtagtg gtatgcagag atattatttt accgttcatt ttcttcctaa  16560
acaagctaat ttggcccttc tcaccggaag atgcatatct atcatgcatg gatttatctt  16620
gaagcataac attgaaggaa tgggagtaac atttcccgca tggtctgatt caagcattgg  16680
aaacgagatt gcttttgttt acactgataa agaaatccta aacacattga aagatcaggc  16740
ttactttgtg gatatgcaag actgtggttt cttcaaggtt tctcaggtgt tggctgttcc  16800
tgattcttgc gaggaagtca ggttcatccg caatcaagcc gtcgcaaaga ttttcactgg  16860
cgaatcgagg aggagactta aacggttaca aaagagagcc ttagcaagag gagaagattt  16920
caatccaaag aagatagaag ctccaagaga aatagacatt tttcaccgtg tagcaatgac  16980
aagcaaaagt tctcaagaag actacattct gcacatccag aagcaagatg ttgattgtca  17040
agcagagcct tatttttcaa actatggtct tgcgtcaaat gagaagttca aaggaactgt  17100
accagacttg agtccgtcca ttgatagaaa ttgagcttat atgaagatga agatgaaata  17160
tttggtgtgt caaataaaaa ggttgtgtgc ttaagtttgt gtttttttct tggcttgttg  17220
tgttatgaat ttgtggcttt ttctaatatt aaatgaatgt aacatctcat tataatgaat  17280
aaacaaatgt ttctataatc cattgtgaat gttttgttgg atctcttctc cagcatataa  17340
ctactgtatg tgctatggta tggactatgg aatatgatta aagataagcg ctagcaagga  17400
```

-continued

```
gtggcagaca tactgtccca caaatgaaga tggaatctgt aaaagaaaac gcgtgaaata  17460
atgcgtctga caaaggttag gtcggctgcc tttaatcaat accaaagtgg tccctaccac  17520
gatggaaaaa ctgtgcagtc ggtttggctt tttctgacga acaaataaga ttcgtggccg  17580
acaggtgggg gtccaccatg tgaaggcatc ttcagactcc aataatggag caatgacgta  17640
agggcttacg aaataagtaa gggtagtttg ggaaatgtcc actcacccgt cagtctataa  17700
atacttagcc cctccctcat tgttaaggga gcaaaatctc agagagatag tcctagagag  17760
agaaagagag caagtagcct agaagtagtc aaggcggcga agtattcagg cacgtggcca  17820
ggaagaagaa aagccaagac gacgaaaaca ggtaagagct aagcttaatg ggaaagcgta  17880
ctgctgatgg ttcagagttt gaaagtccaa agaagaaaag aaaagttggt agtggtatgg  17940
agctttgcaa tatactcaag tatgatcgaa gtctttaccc tgggaaagcg gttttcttct  18000
acaaaacggc ggacagcgat tttgttcctt tggaagctga cattaataag ataagaggac  18060
caaaatcagg ctttacagag gctttcactc ctcaattctc tccgaaaaat atctctccac  18120
aagatctaac tcacaacaac attcttacac tcgaagagtg ttatgttcca ccaaatgtcg  18180
aacatatttt ctgccggttt tctctgcgtg tccaagctaa ttcgttagtc ccctctggtt  18240
gttccgatcc tgaggttttt agccttctca aagagcttgc agaaaccttc aaagaatgtg  18300
gtggttacaa ggagctggca gtacgctatt gtcggaacat tttgattgga acatggcttt  18360
ggagaaatca gaataccgga aacacacaga ttgagatcaa gacttcaaaa ggatcttgct  18420
acttgattga taatacaaga aaactcgctt gggaaagcaa atgggcttca gacgacttga  18480
aagttttgga agaattaagc aatgaaatcg aatctgcttt aacagatcca aatgtgtttt  18540
ggtctgcaga tataaccgcc aagatcgaag catctttctg tcaagagata tatccttcac  18600
agatattgaa tgataaagtg aagcaaggag aagcgtcaaa acagtttgtg aaagctaaat  18660
gtgctgatgg aaggtatgct gtgagtttca acagtgttaa gatcggtgct gctttgcaaa  18720
gtattgatga ttggtgggat gaagatgcta gtaaaagact tagagttcat gaatttggtg  18780
cagacaagga aattggcgta gcgaggaggc ctcccgattc cgaacaaaac ttttattcaa  18840
tcttcaagaa cactgaatgg tatctttccg ccttaaagaa ttgcatcact aacaagaatg  18900
agaagattga cccagcaata tactatttgt tcagcgtgtt gatcaaaggt ggaatgtttc  18960
agaaaaaggc agaggcaaag aaagcatgag cttatatgaa gatgaagatg aaatatttgg  19020
tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt tttcttggct tgttgtgtta  19080
tgaatttgtg gctttttcta atattaaatg aatgtaacat ctcattataa tgaataaaca  19140
aatgtttcta taatccattg tgaatgtttt gttggatctc ttctccagca tataactact  19200
gtatgtgcta tggtatggac tatggaatat gattaaagat aagcgctata caggagtggc  19260
agacatactg tcccacaaat gaagatggaa tctgtaaaag aaaacgcgtg aaataatgcg  19320
tctgacaaag gttaggtcgg ctgcctttaa tcaataccaa agtggtccct accacgatgg  19380
aaaaactgtg cagtcggttt ggctttttct gacgaacaaa taagattcgt ggccgacagg  19440
tgggggtcca ccatgtgaag gcatcttcag actccaataa tggagcaatg acgtaagggc  19500
ttacgaaata agtaagggta gtttgggaaa tgtccactca cccgtcagtc tataaatact  19560
tagcccctcc ctcattgtta agggagcaaa atctcagaga gatagtccta gagagagaaa  19620
gagagcaagt agcctagaag tagtcaaggc ggcgaagtat tcaggcacgt ggccaggaag  19680
aagaaaagcc aagacgacga aaacaggtaa gagctaagct taatgggaaa gcgtactgct  19740
gatggttcag agtttgaaag tccaaagaag aaaagaaaag ttggtagtgg tatgcacctg  19800
aaagagctgt tagagataac agatacaact gagagagata gatctctccg acgtgccttc  19860
tcgccgtata ccgccatgat cgacattaca ggatcagaag cagtcgctct aattatccta  19920
ctcaatctca cttatagaaa gaatcaggta gatgacctgt tggataagaa actcgcgaaa  19980
caagccctca agagtgagga tcatatcaac aaatgcatca aggagattgc ttggtttcat  20040
actcacaacc ttaaatatcc agacataagg gtgagtaagc aaaatcttgc agtagaacca  20100
ccaacattgc attcttatgt gctttcttct gctaattatc ccaaagctta tggttggtcc  20160
cataattcag ctaaagtgaa ttttgctaag ttgtttgtct catacttcaa atggcaaaac  20220
caagtgagct ggcttgccca agtcttggca accaactctg ataactggaa aagtgcattt  20280
acttcccttg gtttatcggt gaaagccttc aaatctttat gcgttacagt taagaattct  20340
ttacctgaag aggcaatccc ggactcagtg gatagatata gcaggcagat tagaatgcct  20400
tatcatgatg gttatcttgc tgtcactcct gtgatatctc acgttgttca aagcaaaata  20460
caacaagctg ctattgacaa aagagctagg ttttctaatg ttgagtttac cagaccagct  20520
gctgtttcaa tgctggcggc aagtctaggt ggtgtgataa atgtacttaa ttatcctcca  20580
tacattcgat caaagtacca tggactatcg aattcaaggg cttttaagtt aaacaacgga  20640
cagactgtct tcaacgtgga agccctcctg aagcctgagc ttataaaagc tcttgaagga  20700
attatattct ctaacaatgc attggcatta aaacagagaa ggcagcagaa agtaaaaaac  20760
atcaaagaac tgagaaatac cctccttgag tggttttccc ctgttttcga atggagattg  20820
gatgcaattg agaatggtta tgatctagaa caattggaat cagcgagcga acggttagag  20880
tacaagattc tgagtttgcc ggataatgaa ctaccttctt tgactatacc cttgtttaga  20940
ttacttaatg agatgttggg aggagtttca atgactcaga ggtatgcatt ccatccaaag  21000
ttgatgtctc cattaaaggc tgctctccaa tggttgctcg ttaatttaac tgatcaaaag  21060
catgttttaa ttgaagaaga tgatgagcat tacagatacc ttcacctgtc agggattagg  21120
gtttttgatg cccaagcatt gtctaatcct tactgctctg gtattccttc gcttacggcg  21180
gtttggggaa tgatccacag ttatcaacgc aaactaaacg aagctttggg cacgaacgta  21240
cggttcacaa gcttttcatg gttcatccgt aattactctg ccgttgctgg taaaaagcta  21300
ccagaactta gccttcaggg tgcacaacag tctagattga aaagacctgg gataatagac  21360
gggaagtatt gtgatttggt gtttgacttg attattcata tcgatggcta tgaagacgat  21420
ttgcaggctg ttgattcaaa gcctgatatc ctcaaggctc atttcccaag caatttcgcc  21480
ggcggcgtga tgcatcaacc cgagttgaat tctaatatta attggtgttg tctttactcc  21540
aacgaaaacc agctttttga gaagctccgt cggcttcctc tttctggatg ttgggttatg  21600
cccacagaac ataagattca agatcttgat gaactgttgt tgctgttgaa ctcggatagt  21660
aaactaagtc caagcatgat gggatacatg cttctcacag agccaatggc tcgtgtagga  21720
tcgttagaga ggcttcactg ttatgcagaa cctgctattg gtgttgtcaa atacgaggca  21780
gcaacttctg ttcgattaaa aggtatcgga aactatttca attccgcgtt ttggatgctc  21840
gatgctcaag aaaagttcat gctcatgaag aaagtgtgag cttatatgaa gatgaagatg  21900
aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt tttcttggct  21960
```

-continued

```
tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat ctcattataa  22020
tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc ttctccagca  22080
tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat aagcgctaca  22140
gaggagtctg agcaggacaa ctcgcgtagt gagagttaca tgttcgttgg gttcttccga  22200
cacggacctg agttggccaa cgtcccacct gaggtctgtg ccccggtgat gagaagtgtg  22260
catctcgttc ttgcagctcg tcagtacttt cagaatcatg gcgtgcatgg tagaatgacc  22320
cttataacgg acttcgacat ggcaatcgct aggtagctta atgaataggc gtatcacgag  22380
gccctttcgt cttcaagaat tagcttggca attcccgatc tagtaacata gatgacaccg  22440
cgcgcgataa tttatcctag tttgcgcgct atattttgtt ttctatcgcg tattaaatgt  22500
ataattgcgg gactctaatc ataaaaaccc atctcataaa taacgtcatg cattacatgt  22560
taattattac atgcttaacg taattcaaca gaaattatat gataatcgtc gcaagaccgg  22620
caacaggatt caatcttaag aaactttatt gccaaatgtt tgaacgatcg ggatcgaaga  22680
tccattccaa gcctacagga acaggtggtg gcggccctcg gtgcgctcgt actgctccac  22740
gatggtgtag tcctcgttgt gggaggtgat gtccagcttg gcgtccacgt agtagtagcc  22800
gggcagctgc acgggcttct tggccatgta gatggacttg aactccacca ggtagtggcc  22860
gccgtccttc agcttcaggg ccttgtgtgt ctcgcccttc agcacgccgt cgcgggggta  22920
caggcgctcg gtggaggcct cccagcccat cgtcttcttc tgcatcacgg ggccgtcgga  22980
ggggaagttc acgccgatga acttcacctt gtagatgaag cagccgtcct gcagggagga  23040
gtcctgggtc acggtcgcca cgccgccgtc ctcgaagttc atcacgcgct cccacttgaa  23100
gccctcgggg aaggacagct tcttgtagtc ggggatgtcg gcggggtgct tcacgtacac  23160
cttggagccg tactggaact ggggggacag gatgtcccag gcgaagggca gggggccgcc  23220
cttggtcacc ttcagcttca cggtgttgtg gccctcgtag gggcggccct cgccctcgcc  23280
ctcgatctcg aactcgtggc cgttcacggt gccctccatg cgcaccttga agcgcatgaa  23340
ctcggtgatg acgttctcgg aggaggccat ttcgcctcga tcgagttgag agtgaatatg  23400
agactctaat tggataccga ggggaattta tggaacgtca gtggagcatt tttgacaaga  23460
aatatttgct agctgatagt gaccttaggc gacttttgaa cgcgcaataa tggtttctga  23520
cgtatgtgct tagctcatta aactccagaa acccgcggct cagtggctcc ttcaacgttg  23580
cggttctgtc agttccaaac gtaaaacggc ttgtcccgcg tcatcggcgg gggtcatacc  23640
gtgactccct taattctcat gtatggttcg ctatactgaa gagcaattct gacaggatat  23700
attggcgggt aaacctaaga gaaaagagcg tttattagaa taatcggata tttaaaaggg  23760
cgtgaaaagg tttatccgtt cgtccatttg tatgtgcatg ccaaccacag ggttcccctc  23820
gggatcaaag tactttgatc caacccctcc gctgctatag tgcagtcggc ttctgacgtt  23880
cagtgcagcc gtcatctgaa aacgacatgt cgcacaagtc ctaagttacg cgacaggctg  23940
ccgccctgcc cttttcctgg cgttttcttg tcgcgtgttt tagtcgcata aagtagaata  24000
cttgcgacta gaaccggaga cattacgcca tgaacaagag cgccgccgct ggcctgctgg  24060
gctatgcccg cgtcagcacc gacgaccagg acttgaccaa ccaacgggcc gaactgcacg  24120
cggccggctg caccaagctg ttttccgaga agatcaccgg caccaggcgc gaccgcccgg  24180
agctggccag gatgcttgac cacctacgcc ctggcgacgt tgtgacagtg accaggctag  24240
accgcctggc ccgcagcacc cgcgacctac tggacattgc cgagcgcatc caggaggccg  24300
gcgcgggcct gcgtagcctg gcagagccgt gggccgacac caccacgccg gccggccgca  24360
tggtgttgac cgtgttcgcc ggcattgccg agttcgagcg ttccctaatc atcgaccgca  24420
cccggagcgg gcgcgaggcc gccaaggccc gaggcgtgaa gtttggcccc cgccctaccc  24480
tcaccccggc acagatcgcg cacgcccgcg agctgatcga ccaggaaggc cgcaccgtga  24540
aagaggcggc tgcactgctt ggcgtgcatc gctcgaccct gtaccgcgca cttgagcgca  24600
gcgaggaagt gacgcccacc gaggccaggc ggcgcggtgc cttccgtgag gacgcattga  24660
ccgaggccga cgccctggcg gccgccgaga atgaacgcca agaggaacaa gcatgaaacc  24720
gcaccaggac ggccaggacg aaccgttttt cattaccgaa gagatcgagg cggagatgat  24780
cgcggccggg tacgtgttcg agccgcccgc gcacctctca accgtgcggc tgcatgaaat  24840
cctggccggt ttgtctgatg ccaagctggc ggcctggccg gccagcttgg ccgctgaaga  24900
aaccgagcgc cgccgtctaa aaaggtgatg tgtatttgag taaaacagct tgcgtcatgc  24960
ggtcgctgcg tatatgatgc gatgagtaaa taaacaaata cgcaagggga acgcatgaag  25020
gttatcgctg tacttaacca gaaaggcggg tcaggcaaga cgaccatcgc aacccatcta  25080
gcccgcgccc tgcaactcgc cggggccgat gttctgttag tcgattccga tccccagggc  25140
agtgcccgcg attgggcggc cgtgcgggaa gatcaaccgc taaccgttgt cggcatcgac  25200
cgcccgacga ttgaccgcga cgtgaaggcc atcggccggc gcgacttcgt agtgatcgac  25260
ggagcgcccc aggcggcgga cttggctgtg tccgcgatca aggcagccga cttcgtgctg  25320
attccggtgc agccaagccc ttacgacata tgggccaccg ccgacctggt ggagctggtt  25380
aagcagcgca ttgaggtcac ggatggaagg ctacaagcgg cctttgtcgt gtcgcgggcg  25440
atcaaaggca cgcgcatcgg cggtgaggtt gccgaggcgc tggccgggta cgagctgccc  25500
attcttgagt cccgtatcac gcagcgcgtg agctacccag gcactgccgc cgccggcaca  25560
accgttcttg aatcagaacc cgagggcgac gctgcccgcg aggtccaggc gctggccgct  25620
gaaattaaat caaaactcat ttgagttaat gaggtaaaga gaaaatgagc aaaagcacaa  25680
acacgctaag tgccggccgt ccgagcgcac gcagcagcaa ggctgcaacg ttggccagcc  25740
tggcagacac gccagccatg aagcgggtca actttcagtt gccggcggag gatcacacca  25800
agctgaagat gtacgcggta cgccaaggca agaccattac cgagctgcta tctgaataca  25860
tcgcgcagct accagagtaa atgagcaaat gaataaatga gtagatgaat tttagcggct  25920
aaaggaggcg gcatggaaaa tcaagaacaa ccaggcaccg acgccgtgga atgccccatg  25980
tgtggaggaa cgggcggttg gccaggcgta agcggctggg ttgcctgccg gccctgcaat  26040
ggcactggaa cccccaagcc cgaggaatcg gcgtgagcgg tcgcaaacca tccggcccgg  26100
tacaaatcgg cgcggcgctg ggtgatgacc tggtggagaa gttgaaggcc gcgcaggccg  26160
cccagcggca acgcatcgag gcagaagcac gccccggtga atcgtggcaa gcggccgctg  26220
atcgaatccg caaagaatcc cggcaaccgc cggcagccgg tgcgccgtcg attaggaagc  26280
cgcccaaggg cgacgagcaa ccagattttt tcgttccgat gctctatgac gtgggcaccc  26340
gcgatagtcg cagcatcatg gacgtggccg ttttccgtct gtcgaagcgt gaccgacgag  26400
ctggcgaggt gatccgctac gagcttccag acgggcacgt agaggtttcc gcagggccgg  26460
ccggcatggc cagtgtgtgg gattacgacc tggtactgat ggcggtttcc catctaaccg  26520
```

-continued

```
aatccatgaa ccgataccgg gaagggaagg gagacaagcc cggccgcgtg ttccgtccac  26580
acgttgcgga cgtactcaag ttctgccggc gagccgatgg cggaaagcag aaagacgacc  26640
tggtagaaac ctgcattcgg ttaaacacca cgcacgttgc catgcagcgt acgaagaagg  26700
ccaagaacgg ccgcctggtg acggtatccg agggtgaagc cttgattagc cgctacaaga  26760
tcgtaaagag cgaaaccggg cggccggagt acatcgagat cgagctagct gattggatgt  26820
accgcgagat cacagaaggc aagaacccgg acgtgctgac ggttcacccc gattactttt  26880
tgatcgatcc cggcatcggc cgttttctct accgcctggc acgccgcgcc gcaggcaagg  26940
cagaagccag atggttgttc aagacgatct acgaacgcag tggcagcgcc ggagagttca  27000
agaagttctg tttcaccgtg cgcaagctga tcgggtcaaa tgacctgccg gagtacgatt  27060
tgaaggagga ggcggggcag gctggcccga tcctagtcat gcgctaccgc aacctgatcg  27120
agggcgaagc atccgccggt tcctaatgta cggagcagat gctagggcaa attgccctag  27180
caggggaaaa aggtcgaaaa ggactctttc ctgtggatag cacgtacatt gggaacccaa  27240
agccgtacat tgggaaccgg aacccgtaca ttgggaaccc aaagccgtac attgggaacc  27300
ggtcacacat gtaagtgact gatataaaag agaaaaaagg cgatttttcc gcctaaaact  27360
ctttaaaact tattaaaact cttaaaaccc gcctggcctg tgcataactg tctggccagc  27420
gcacagccga agtgctgcaa aaagcgccta cccttcggtc gctgcgctcc ctacgccccg  27480
ccgcttcgcg tcggcctatc gcggccgctg gccgctcaaa aatggctggc ctacggccag  27540
gcaatctacc agggcgcgga caagccgcgc cgtcgccact cgaccgccgg cgcccacatc  27600
aaggcaccct gcctcgcgcg tttcggtgat gacggtgaaa acctctgaca catgcagctc  27660
ccggtgacgg tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc  27720
gcgtcagcgg gtgttggcgg gtgtcggggc gcagccatga cccagtcacg tagcgatagc  27780
ggagtgtata ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaccata  27840
tgcggtgtga aataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgctcatccg  27900
cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc  27960
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt  28020
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc  28080
ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa  28140
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc  28200
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg  28260
cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc  28320
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc  28380
gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca  28440
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact  28500
acggctacac tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg  28560
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt  28620
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct  28680
tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatgc  28740
attctaggtg attagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca  28800
ggattatcaa taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg  28860
aggcagttcc ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca  28920
tcaatacaac ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca  28980
tgagtgacga ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt  29040
tcaacaggcc agccattacg ctcgtcatca aaatcactcg catcaaccaa accgttattc  29100
attcgtgatt gcgcctgagc gagtcgaaat acgcgatcgc tgttaaaagg acaattacaa  29160
acaggaatcg aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct  29220
gaatcaggat attcttctaa tacctggaat gctgttttcc ctgggatcgc agtggtgagt  29280
aaccatgcat catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc  29340
gtcagccagt ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca  29400
tgtttcagaa acaactctgg cgcatcgggc ttcccataca atcggtagat tgtcgcacct  29460
gattgcccga cattatcgcg agcccattta tacccatata aatcagcatc catgttggaa  29520
tttaatcgcg gccttgagca agacgtttcc cgttgaatat ggctcataac agaacttatt  29580
atttccttcc tcttttctac agtatttaaa gataccccaa gaagctaatt ataacaagac  29640
gaactccaat tcactgttcc ttgcattcta aaaccttaaa taccagaaaa cagctttttc  29700
aaagttgttt tcaaagttgg cgtataacat agtatcgacg gagccgattt tgaaaccgcg  29760
gtgatcacag gcagcaacgc tctgtcatcg ttacaatcaa catgctaccc tccgcgagat  29820
catccgtgtt tcaaacccgg cagcttagtt gccgttcttc cgaatagcat cggtaacatg  29880
agcaaagtct gccgccttac aacggctctc ccgctgacgc cgtcccggac tgatgggctg  29940
cctgtatcga gtggtgattt tgtgccgagc tgccggtcgg ggagctgttg gctggctgg   29999
```

```
SEQ ID NO: 44          moltype = DNA  length = 28739
FEATURE                Location/Qualifiers
source                 1..28739
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 44
tggcaggata tattgtggtg taaacataac gaattgctct tcagcaagga gaatgaggag  60
aatgtagcag aaggcatgtt gttgtgactc cgaggggttg cctcaaactc tatcttataa  120
ccggcgtgga ggcatggagg caggggtatt ttggtcattt taatagatag tggaaaatga  180
cgtggaattt acttaaagac gaagtctttg cgacaagggg gggcccacgc cgaatttaat  240
attaccggcg tggcccccc ttatcgcgag tgctttagca cgagcggtcc agatttaaag  300
tagaaaattt cccgcccact agggttaaag gtgttcacac tataaaagca tatacgatgt  360
gatggtattt gatggagcgt atattgtatc aggtatttcc gttggatacg aattattcgt  420
acgaccctcg gtatactaat gaggagtgtc gctgcaacca tagtttgaca taatgcaacc  480
ataccttgac ataattaaag tcatactctg acataaagcc ccttattaaa aatgatgcta  540
aaaataaatt ccattataat caattattta agaattaatt aaaccatagt gtgacatagt  600
ggctttttat ttgtatagta tgtgtactca aataataaaa atttacttat gtatatacgc  660
```

-continued

```
aacttaagaa agccatctcc aaataaaaat gtatttctac ctgttccatg gccaaccgct  720
agcaaggagg tggagatcct tcctcctaga tccaaaatac ggtaccgtca acaatctcac  780
tcgagagtcg acgtatgtgc aaaactaagc attccgaagc cattgttagc cgtatgaata  840
gggaaactaa acccagtgat aagacctgat gttttcgctt ctttaattac atttggagat  900
gcccgctata caggagccgt ataggcagct gaagataaaa acaatcgatt cgatgtttaa  960
gttaatgtaa ttctgccgaa aaggcagtga gtagtaattg aaaaaaccag tcttatatcc 1020
gggctggttt ttttgttttt agcgttataa cttttatttc actatgtcac tgtatggttg 1080
cattttatgt cagtgaatgg ttgcaggttt tttgtaagtt attgatatgc ctatccagtg 1140
tatgtcacgc tttgcttgca gcgacacgct acagaggagt ctgagcagga caactcgcgt 1200
agtgagagtt acatgttcgt tgggttcttc cgacacggac ctgagttggc caacgtccca 1260
cctgaggtct gtgccccggt gatgagaagt gtgcatctcg ttcttgcagc tcgtcagtac 1320
tttcagaatc atggcgtgca tggtagaatg acccttataa cggacttcga catggcaatc 1380
gctaggtaaa tgaatggaaa aagtggcacc gagtcggtgc tttttttccc ggggcgcgcc 1440
gatatcgagc tctcccggcg cgccgatatc gagctcagtg tttgatcgcc ggcggtaccg 1500
agtgtacttc aagtcagtgg gaaatcaata aaatgattat tttatgaata tatttcattg 1560
tgcaagtaga tagaaattac atatgttaca taacacacga aataaacaaa aaaagacaat 1620
ccaaaaacaa acaccccaaa aaaaataatc actttagata aactcgtatg aggagaggca 1680
cgttcagtga ctcgacgatt cccgagcaaa aaaagtctcc ccgtcacaca tgtagtgggt 1740
gacgcaatta tctttaaagt aatccttctg ttgacttgtc attgataaca tccagtcttc 1800
gtcaggattg caaagaatta tagaagggat cccacctttt attttcttct tttttccata 1860
tttagggttg acagtgaaat cagactggca acctattaat tgcttccaca atgggacgaa 1920
cttgaagggg atgtcgtcga tgatattata ggtggcgtgt tcatcgtagt tggtgaaatc 1980
gatggtaccg ttccaatagt tgtgtcgtcc gagacttcta gcccaggtgg tctttccggt 2040
acgagttggt ccgcagatgt agaggctggg gtgtcggatt ccattccttc cattgtcctt 2100
gttaaatcgg ccatccattc aaggtcagat tgagcttgtt ggtatgagac aggatgtatg 2160
taagtataag cgtctatgct tacatggtat agatgggttt ccctccagga gtgtagatct 2220
tcgtggcagc gaagatctga ttctgtgaag ggcgacacat acggttcagg ttgtggaggg 2280
aataatttgt tggctgaata ttccagccat tgaagctttg ttgcccattc atgagggaat 2340
tcttccttga tcatgtcaag atattcctcc ttagacgttg cagtctggat aatagttctc 2400
catcgtgcgt cagatttgcg aggagaaacc ttatgatctc ggaaatctcc tctggtttta 2460
atatctccgt cctttgatat gtaatcaagg acttgtttag agtttctagc tggctggata 2520
ttagggtgat ttccttcaaa atcgaaaaaa gaaggatccc taatacaagg ttttttatca 2580
agctggagga gagcatgata gtgggtagtg ccatcttgat gaagctcaga agcaacacca 2640
aggaagaaaa taagaaaagg tgtgagtttc tcccagagaa actggaataa atcatctctt 2700
tgagatgagc acttgggata ggtaaggaaa acatatttag attggagtct gaagttctta 2760
ctagcagaag gcatgttgtt gtgactccga ggggttgcct caaactctat cttataaccg 2820
gcgtggaggc atggaggcag gggtattttg gtcattttaa tagatagtgg aaaatgacgt 2880
ggaatttact taaagacgaa gtctttgcga caaggggggg cccacgccga atttaatatt 2940
accggcgtgg cccccccta tcgcgagtgc tttagcacga gcggtccaga tttaaagtag 3000
aaaatttccc gcccactagg gttaaaggtg ttcacactat aaaagcatat acgatgtgat 3060
ggtatttgat ggagcgtata ttgtatcagg tatttccgtt ggatacgaat tattcgtacg 3120
accctcatag tttaaactat cagtgggtag cttaatggta gacactttat tagttagatc 3180
cgctgttatt ctgcacgatt gcgcttgccg gtggaaattt aaggacgagt gtagtgcact 3240
cacccttagc cttcagcgtg acaagattag ggtagcaccg gaaacagcga ttttacggac 3300
agatagtaca gtctgagact cagtcactgg tccctaaacg ggagccggta cgtcggcggt 3360
tcgctagcaa ggagtggcag acatactgtc ccacaaatga agatggaatc tgtaaaagaa 3420
aacgcgtgaa ataatgcgtc tgacaaaggt taggtcggct gcctttaatc aataccaaag 3480
tggtccctac cacgatggaa aaactgtgca gtcggtttgg ctttttctga cgaacaaata 3540
agattcgtgg ccgacaggtg ggggtccacc atgtgaaggc atcttcagac tccaataatg 3600
gagcaatgac gtaagggctt acgaaataag taagggtagt ttgggaaatg tccactcacc 3660
cgtcagtcta taaatactta gcccctccct cattgttaag ggagcaaaat ctcagagaga 3720
tagtcctaga gagagaaaga gagcaagtag cctagaagta gtcaaggcgg cgaagtattc 3780
aggcacgtgg ccaggaagaa gaaaagccaa gacgacgaaa acaggtaaga gctaagctta 3840
atgggaaagc gtactgctga tggttcagag tttgaaagtc caaagaagaa aagaaaagtt 3900
ggtagtggta tgactgataa acggaaagat ggttccggca agcttcttta ctgcagcttt 3960
tgcgggaaat ctcaacatga agttaggaaa cttattgcag ggccttctgt ttatatctgt 4020
gacgaatgtg ttgatctctg caatgatatt attcgagaag agatcaaaga agtggctcct 4080
cacagagaga ggagtgcgtt gccgacacca cacgagatta gaaatcatct tgatgattat 4140
gtcattggtc aagaacaagc aaagaaggtc cttgctgttg cagtgtacaa ccattacaag 4200
aggctcagaa acggagatac ttctaatggt gtggaacttg gcaaatcaaa catcctgctc 4260
attggaccca ctggaagtgg aaagacactc ctagcagaga cattggcaag gctcttagac 4320
gttcctttca ctatggctga tgcaactact cttactgagg ctggttatgt tggagaagat 4380
gtggagaata ttattcaaaa gttattgcag aaatgtgatt acgatgtcca gaaagctcag 4440
agagggatag tctacataga tgagatagac aagatctcta ggaagagtga taatccctcg 4500
ataacacgcg acgtttctgg cgaaggtgtt caacaagctt tgttgaagct cattgaaggc 4560
accgtagcgg cagttccacc acaaggagga agaaagcatc ctcagcaaga gtttctccaa 4620
gtggatacct caaaaatctt atttatatgt ggtggtgcct tcgctgggct tgacaaagtt 4680
atctcacacc gtgtagaaac tggtagcggg attgggtttg gtgcaacggt gaaggcaaaa 4740
tctgataagg cttctgaagg agagcttctg gctcaagttg aaccagagga tcttatcaag 4800
tttggactaa ttcctgagtt catcggtaga cttcctgtcg ttgctacgtt gaacgagtta 4860
agtgaggaag ccctgattca aatcttgaaa gagccaaaga atgctctgac aaagcaatat 4920
caggctcttt tcaatcttga gggagttgac ttggaatttc gtgatgaagc gctggatgct 4980
atagctaaga aagcaatggc gaggaaaacc ggagctcgtg gtctaagatc gattgttgaa 5040
gccgctttat tggatacaat gtatgatttg ccctccatgg aagatgttga gaaagtggtc 5100
atcgacgagt ctgtgattga tggacagtca aagccattgc tcatatatgg taaacctgaa 5160
gcacagcagg catcaggaga atgaatatga agatgaagat gaaatatttg gtgtgtcaaa 5220
```

-continued

```
taaaaaggtt gtgtgcttaa gtttgtgttt ttttcttggc ttgttgtgtt atgaatttgt  5280
ggctttttct aatattaaat gaatgtaaca tctcattata atgaataaac aaatgtttct  5340
ataatccatt gtgaatgttt tgttggatct cttctccagc atataactac tgtatgtgct  5400
atggtatgga ctatggaata tgattaaaga taagcgctat acaggagtgg cagacatact  5460
gtcccacaaa tgaagatgga atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa  5520
ggttaggtcg gctgccttta atcaatacca aagtggtccc taccacgatg gaaaaactgt  5580
gcagtcggtt tggctttttc tgacgaacaa ataagattcg tggccgacag gtgggggtcc  5640
accatgtgaa ggcatcttca gactccaata atggagcaat gacgtaaggg cttacgaaat  5700
aagtaagggt agtttgggaa atgtccactc acccgtcagt ctataaatac ttagcccctc  5760
cctcattgtt aagggagcaa aatctcagag agatagtcct agagagagaa agagagcaag  5820
tagcctagaa gtagtcaagg cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc  5880
caagacgacg aaaacaggta agagctaagc ttaatgggaa agcgtactgc tgatggttca  5940
gagtttgaaa gtccaaagaa gaaaagaaaa gttggtagtg gtatgagtta ttctggagaa  6000
cgggataatt ttgcacctca catggcatta gttccaatgg tgattgaaca aacaagtaga  6060
ggggagcgaa gctttgacat ttactctcgt ctattaaaag aaagggtcat cttcttaacc  6120
gggcaagttg aagatcatat ggcaaatctg attgtagctc agatgctttt tcttgaagca  6180
gagaatcctg agaaagatat ctatctctac attaactctc ctggtggtgt aataactgct  6240
ggaatgtcta tatatgatac aatgcagttt atcaaacctg atgtttcaac aatatgtatg  6300
ggtcaggctg cgtcaatggg agcttttctt ttgacggctg gcgcaaaagg aaagcggttc  6360
tgcctaccga actcaagggt gatgatccat caaccacttg gaggatacca aggtcaagct  6420
accgatatag agattcatgc tagagaaatt ttgaaggtga agggtagaat gaatgaactg  6480
atggcccttc acactggaca atcgctcgag cagattgagc gtgacacaga gagagacagg  6540
tttttatccg caccagaagc tgttgagtat ggtttggttg attctattct tactcataga  6600
aactgaatat gaagatgaag atgaaatatt tggtgtgtca aataaaaagg ttgtgtgctt  6660
aagtttgtgt ttttttcttg gcttgttgtg ttatgaattt gtggcttttt ctaatattaa  6720
atgaatgtaa catctcatta taatgaataa acaaatgttt ctataatcca ttgtgaatgt  6780
tttgttggat ctcttctcca gcatataact actgtatgtg ctatggtatg gactatggaa  6840
tatgattaaa gataagcgct acagaggagc atcttcattc ttaagatatg aagataatct  6900
tcaaaaggcc cctgggaatc tgaaagaaga gaagcaggcc catttatatg ggaaagaaca  6960
atagtatttc ttatataggc ccatttaagt tgaaaacaat cttcaaaagt cccacatcgc  7020
ttagataaga aaacgaagct gagtttatat acagctagag tcgaagtagt gataatggtg  7080
acctgccgta taggcagctg aagatactga cgtaagggat gacgcacaat cccactataa  7140
ttctgccgaa aaggcagtga gtagtgcttt tttttttgca aaattttcca gatcgatttc  7200
ttcttcctct gttcttcggc gttcaatttc tggggttttc tcttcgtttt ctgtaactga  7260
aacctaaaat ttgacctaaa aaaaatctca aataatatga ttcagtggtt ttgtactttt  7320
cagttagttg agttttgcag ttccgatgag ataaaccaat acgctaggta tactaatgag  7380
gagtggcaga catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa  7440
taatgcgtct gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc  7500
acgatggaaa aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc  7560
cgacaggtgg gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg  7620
taagggctta cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat  7680
aaatacttag cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag  7740
agagaaagag agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc  7800
caggaagaag aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tgtacattag  7860
gaacttgaga aagccttctc caaataagaa tgtttttaag ttcgcatcta ccaaggtcag  7920
ctctgtggtc atgtgtgaat ccagccttga gtttgacgct tgctttcatc atgaatataa  7980
cgacttgatt gagagttttg gttctcagcc agaaggtttc aaatacgaat tcatgggtaa  8040
atctctccct tacactccag atgctttgat ctcctacacg gacaaaactc aaaagtatca  8100
cgagtataaa ccatattcaa aaatcgcatc acctctattt agagcagagt ttgctgcaaa  8160
gagggctgca agtcttaagt tgggtattga tttggtactc gtcacagaca gacagattag  8220
agtaaaccca atactcaaca atctgaagct ccttcataga tactcgggcg tttatggtat  8280
atctggtatc cagaaggaac tacttagctt tattcacaaa agtggcgtta tcaaactcaa  8340
tgatatctca tcacaggttg ggatccccat tggagaaact cggtcatttc tttttggact  8400
gatgcataag ggacttgtaa aggctgacct aggttgtgat gacctgacaa acaaccctac  8460
tctttgggct actcctggaa gtggctcagg caaacgtact gctgacggaa gtgagttcga  8520
gtccccaaag aaaaaaagaa aagttggtag tggatctggt ggaatgactg atttcttcaa  8580
tgagtttgat gaatcgcttg taccactgaa accacaaact ccaacgcagt atgttaagct  8640
ggacgatgct aatttaattc agagagattt agatacattc tcagatactt tcaagaacca  8700
agccctccaa agatataagc taatttctac aattgataaa aagttaagcc gcggatggac  8760
tcaaagaaat ctcgacccga tcctcgacga actcttcaaa ggcggtgatg tagtgaggcc  8820
aaactggaga actgtggcaa gatggaggaa gaaatatata gaaagcaatg gtgacatcgc  8880
ttctcttgct gataaaaacc acaaaatggg aaacagaaca aatcggataa aaggagatga  8940
caaatttttt gacaaggcct tggaacgttt ccttgacgcg aaaagaccta caatagcaac  9000
agcttatcaa tattacaagg acttgatagt gattgaaaac gagtcaattg ttgaagggaa  9060
gatccccatc atctcttata atgcatttaa caaacgaatt aaagccattc ctccttatgc  9120
tgttgcagtc gctaggcatg ggaaattcaa agctgatcaa tggtttgcat attgtgcagc  9180
tcacgttcca ccaacgcgta tacttgaacg agtggagatt gatcatactc cgttggatct  9240
aatattattg gatgatgagc ttctgattcc cattggccgg ccatatctta cacttctcat  9300
agatgtgttt tcaggttgcg ttctgggatt ccatttgtca tacaagagcc cttcttatgt  9360
ctctgcagca aaagcaatca ctcacgcaat aaaacccaaa agcttagatg cgcttaatat  9420
tgagcttcag aatgactggc cctgtttcgg gaagtttgag aatttagtag ttgacaatgg  9480
agctgaattt tggtcaaaaa atttggaaca tgcttgtcaa tctgctggta taaatattca  9540
atacaatccg gtgcgtaagc catggctcaa gccttttata gagaggttct ttggggtgat  9600
gaatgaatac ttcttgccgg aactcccagg aaaaacattt tctaacattc ttgagaaaga  9660
agaatacaaa ccggagaagg atgcaataat gaggttcagc acctttgttg aggagttcca  9720
ccgttggatt gccgacgtgt atcatcaaga tagtaatagc cgtgaaacgc gtattcctat  9780
```

-continued

```
taagcgatgg caacaaggat tcgatgctta tcctccttta acaatgaacg aagaagagga  9840
gacacggttt tctatgctca tgagaatctc tgactctagg acgctaacca ggaatgggtt  9900
taagtaccaa gaattgatgt atgattctac cgccttagca gattaccgta agcattaccc  9960
tcagacaaaa gagactgtga agaagctaat aaaggtcgat ccagatgata tctccaagat 10020
ttacgtttat ctggaagagt tggaaagtta cttggaggtt ccttgcacag atccaactgg 10080
atacactgat ggtttatcca tatatgagca taaaaccatt aagaagatca accgagaagt 10140
tattagagaa tcaaaagata gccttgggtt ggcgaaagcg agaatggcta ttcatgagag 10200
agtgaaacaa gagcaagaag tattcatcga aagtaagacc aaagctaaga ttactgcagt 10260
caagaagcag gctcagatcg cggatgttag taatacaggc acatcaacaa tcaaggtgtc 10320
cgaggagtct gcggctcctg tccagaaaca catctcaaat gataacagtg atgattggga 10380
tgatgatctg gaagcttttg aatgagctta tatgaagatg aagatgaaat atttggtgtg 10440
tcaaataaaa aggttgtgtg cttaagtttg tgtttttttc ttggcttgtt gtgttatgaa 10500
tttgtggctt tttctaatat taaatgaatg taacatctca ttataatgaa taaacaaatg 10560
tttctataat ccattgtgaa tgttttgttg gatctcttct ccagcatata actactgtat 10620
gtgctatggt atggactatg gaatatgatt aaagataagc gctagcaagg agtggcagac 10680
atactgtccc acaaatgaag atggaatctg taaaagaaaa cgcgtgaaat aatgcgtctg 10740
acaaaggtta ggtcggctgc ctttaatcaa taccaaagtg gtccctacca cgatggaaaa 10800
actgtgcagt cggtttggct ttttctgacg aacaaataag attcgtggcc gacaggtggg 10860
ggtccaccat gtgaaggcat cttcagactc caataatgga gcaatgacgt aagggcttac 10920
gaaataagta agggtagttt gggaaatgtc cactcacccg tcagtctata aatacttagc 10980
ccctccctca ttgttaaggg agcaaaatct cagagagata gtcctagaga gagaaagaga 11040
gcaagtagcc tagaagtagt caaggcggcg aagtattcag gcacgtggcc aggaagaaga 11100
aaagccaaga cgacgaaaac aggtaagagc taagcttaat gggaaatgct cttacagaga 11160
ttcaaataga gaagttacgg aatttcagtg attgcatcgt gatgcatcct cagataaaga 11220
ccattttcaa tgattttgat gaactaaggc ttaatagaaa gtttcagagt gatcagcaat 11280
gtatgcttct tataggagat actggtgttg gcaaatcaca caccatcaac cattacaaga 11340
aaagagtcct cgcaacacaa aactattcaa gaaacacaat gccggtcctt gtttctcgca 11400
tatcacgtgg gaaaggtttg gatgctactt tggtgcagat gctggcggat ctagagctgt 11460
ttggttcttc tcagatcaag aagcggggct acaaaacaga ccttaccaaa aaattagttg 11520
agagtttgat caaagctcaa gtggagctgt tgattatcaa tgaattccaa gagcttattg 11580
agtttaaaag tgttcaagaa agacaacaga ttgcaaatgg tctaaagttt attagtgaag 11640
aagccaaggt tcccattgtt ttggttggaa tgccatgggc agcaaagata gctgaggaac 11700
cacaatgggc ttctcgactc gtaagaaaac gaaagctaga atatttttca ctcaagaatg 11760
acagcaagta ctttaggcag tatttgatgg gtcttgccaa aaagatgcct ttcgacgttc 11820
ctcctaaatt agaatcgaag aacactacga ttgcgctatt tgcagcttgt agaggggaaa 11880
acagagcttt gaagcacctt cttttagaag cactgaagct cgccctttca tgcaatgagt 11940
atcttgagaa caagcatttc atcacagctt atgacaaatt cgatttcttc aacgacaagg 12000
agaaactgaa atccaaaaac cctttcaagc aagatataaa agatattgaa atttatgaag 12060
ttattaaaaa ttcttcctac aatccaaatg cacttgatcc agaagatatg ttgactgatc 12120
gagtatttgc tattgtgaaa ggaagcggaa agaggactgc agatggatca gaatttgagt 12180
ccccaaagaa gaaaaggaaa gtgtgagctt atatgaagat gaagatgaaa tatttggtgt 12240
gtcaaataaa aaggttgtgt gcttaagttt gtgtttttttt cttggcttgt tgtgttatga 12300
atttgtggct ttttctaata ttaaatgaat gtaacatctc attataatga ataaacaaat 12360
gtttctataa tccattgtga atgttttgtt ggatctcttc tccagcatat aactactgta 12420
tgtgctatgg tatggactat ggaatatgat taaagataag cgctatacag gagtggcaga 12480
catactgtcc cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa taatgcgtct 12540
gacaaaggtt aggtcggctg cctttaatca ataccaaagt ggtccctacc acgatggaaa 12600
aactgtgcag tcggtttggc tttttctgac gaacaaataa gattcgtggc cgacaggtgg 12660
gggtccacca tgtgaaggca tcttcagact ccaataatgg agcaatgacg taagggctta 12720
cgaaataagt aagggtagtt tgggaaatgt ccactcaccc gtcagtctat aaatacttag 12780
cccctccctc attgttaagg gagcaaaatc tcagagagat agtcctagag agagaaagag 12840
agcaagtagc ctagaagtag tcaaggcggc gaagtattca ggcacgtggc caggaagaag 12900
aaaagccaag acgacgaaaa caggtaagag ctaagcttaa tgggaaagcg tactgctgat 12960
ggttcagagt ttgaaagtcc aaagaagaaa agaaaagttg gtagtggtat ggcttttctt 13020
ttctcaccaa aagcaagagc gttttccgat gaatcgcttg aaagttattt acttcgagtt 13080
gtctcagaaa acttcttcga ttcatatgaa ggactatctt tggccatcag agaagaactt 13140
catgagttgg acttcgaggc tcatggagca tttcccgtag atttaaagag acttaatgtt 13200
taccatgcaa aacacaactc acactttagg atgagagctc ttggcctatt agaaactctt 13260
ttagatcttc ctcgttatga gcttcaaaag ctagcactcc tgaagtcaga cattaaattc 13320
aattcttccg ttgccctcta caacaatggt gttgatattc ctttgagatt cattagacat 13380
catgctgagg aagctgttga ctccattccg gtttgttctc aatgcttggc tgaagaagct 13440
tatataaagc aatcttggca tatcaaatgg gttaatgctt gcaccaaaca tcaatgtgct 13500
ttgcttcaca attgtcctga atgttatgca ccgatcaatt acattgagaa tgaatcaatc 13560
actcattgct cttgtggttt tgaattgagc tgcgcctcta caagtcctgt gaacacactc 13620
tcgattgaac atttgaacaa actcctggac aaaggagaac ggaatgactc aaatccattg 13680
tttaacaaca tgacgttgac agaaaggttt gctgctttat tatggtatca agagaggtat 13740
tcacagactg ataatttctg tttgaatgat gctgtgaact acttcagtaa gtggccagca 13800
gtgtttaata ctgagctaga tgagctctca aagaatgccg agatgaagct gattgatctc 13860
ttcaacaaga cagagtttaa gtttatcttt ggagatgcaa ttcttgcttg tccaagtact 13920
cagaagcaga gtgaatctca tttcatctac cgggcattac ttgattacct tgtcacttta 13980
gtggagagca atcccaaaac aaagaagcca aacgcagctg atttgcttgt ttctgtcctg 14040
gaagcagcaa ctttgttggg gacgagtgta gaacaagttt ataggctcta tcagaatgga 14100
atactgcaaa ccgcctttcg gcacaagatg aatcagagaa taaaccctta caagggtgct 14160
ttctttctcc gtcatgtgat agagtataaa accagctttg gtaacgataa agctagaatg 14220
tatctctcag cttggtgagc ttatatgaag atgaagatga aatatttggt gtgtcaaata 14280
aaaaggttgt gtgcttaagt ttgtgttttt ttcttggctt gttgtgttat gaatttgtgg 14340
```

-continued

```
ctttttctaa tattaaatga atgtaacatc tcattataat gaataaacaa atgtttctat  14400
aatccattgt gaatgttttg ttggatctct tctccagcat ataactactg tatgtgctat  14460
ggtatggact atggaatatg attaaagata agcgctacag aggagtctga gcaggacaac  14520
tcgcgtagtg agagttacat gttcgttggg ttcttccgac acggacctga gttggccaac  14580
gtcccacctg aggtctgtgc cccggtgatg agaagtgtgc atctcgttct tgcagctcgt  14640
cagtactttc agaatcatgg cgtgcatggt agaatgaccc ttataacgga cttcgacatg  14700
gcaatcgcta ggtaaatgaa tgaggagtgg cagacatact gtcccacaaa tgaagatgga  14760
atctgtaaaa gaaaacgcgt gaaataatgc gtctgacaaa ggttaggtcg gctgccttta  14820
atcaatacca aagtggtccc taccacgatg gaaaaactgt gcagtcggtt tggctttttc  14880
tgacgaacaa ataagattcg tggccgacag gtgggggtcc accatgtgaa ggcatcttca  14940
gactccaata atggagcaat gacgtaaggg cttacgaaat aagtaagggt agtttgggaa  15000
atgtccactc acccgtcagt ctataaatac ttagcccctc cctcattgtt aagggagcaa  15060
aatctcagag agatagtcct agagagagaa agagagcaag tagcctagaa gtagtcaagg  15120
cggcgaagta ttcaggcacg tggccaggaa gaagaaaagc caagacgacg aaaacaggta  15180
agagctaagc ttaatgggaa agcgtactgc tgatggttca gagtttgaaa gtccaaagaa  15240
gaaaagaaaa gttggtagtg gtatgcagag atattatttt accgttcatt ttcttcctaa  15300
acaagctaat ttggcccttc tcaccggaag atgcatatct atcatgcatg gatttatctt  15360
gaagcataac attgaaggaa tgggagtaac atttcccgca tggtctgatt caagcattgg  15420
aaacgagatt gcttttgttt acactgataa agaaatccta aacacattga aagatcaggc  15480
ttactttgtg gatatgcaag actgtggttt cttcaaggtt tctcaggtgt tggctgttcc  15540
tgattcttgc gaggaagtca ggttcatccg caatcaagcc gtcgcaaaga ttttcactgg  15600
cgaatcgagg aggagactta aacggttaca aaagagagcc ttagcaagag gagaagattt  15660
caatccaaag aagatagaag ctccaagaga aatagacatt tttcaccgtg tagcaatgac  15720
aagcaaaagt tctcaagaag actacattct gcacatccag aagcaagatg ttgattgtca  15780
agcagagcct tatttttcaa actatggtct tgcgtcaaat gagaagttca aaggaactgt  15840
accagacttg agtccgtcca ttgatagaaa ttgagcttat atgaagatga agatgaaata  15900
tttggtgtgt caaataaaaa ggttgtgtgc ttaagtttgt gtttttttct tggcttgttg  15960
tgttatgaat ttgtggcttt ttctaatatt aaatgaatgt aacatctcat tataatgaat  16020
aaacaaatgt ttctataatc cattgtgaat gttttgttgg atctcttctc cagcatataa  16080
ctactgtatg tgctatggta tggactatgg aatatgatta aagataagcg ctagcaagga  16140
gtggcagaca tactgtccca caaatgaaga tggaatctgt aaaagaaaac gcgtgaaata  16200
atgcgtctga caaaggttag gtcggctgcc tttaatcaat accaaagtgg tccctaccac  16260
gatggaaaaa ctgtgcagtc ggtttggctt tttctgacga acaaataaga ttcgtggccg  16320
acaggtgggg gtccaccatg tgaaggcatc ttcagactcc aataatggag caatgacgta  16380
agggcttacg aaataagtaa gggtagtttg ggaaatgtcc actcacccgt cagtctataa  16440
atacttagcc cctccctcat tgttaaggga gcaaaatctc agagagatag tcctagagag  16500
agaaagagag caagtagcct agaagtagtc aaggcggcga agtattcagg cacgtggcca  16560
ggaagaagaa aagccaagac gacgaaaaca ggtaagagct aagcttaatg ggaaagcgta  16620
ctgctgatgg ttcagagttt gaaagtccaa agaagaaaag aaaagttggt agtggtatgg  16680
agctttgcaa tatactcaag tatgatcgaa gtctttaccc tgggaaagcg gttttcttct  16740
acaaaacggc ggacagcgat tttgttcctt tggaagctga cattaataag ataagaggac  16800
caaaatcagg ctttacagag gctttcactc ctcaattctc tccgaaaaat atctctccac  16860
aagatctaac tcacaacaac attcttacac tcgaagagtg ttatgttcca ccaaatgtcg  16920
aacatatttt ctgccggttt tctctgcgtg tccaagctaa ttcgttagtc ccctctggtt  16980
gttccgatcc tgaggttttt agccttctca aagagcttgc agaaaccttc aaagaatgtg  17040
gtggttacaa ggagctggca gtacgctatt gtcggaacat tttgattgga acatggcttt  17100
ggagaaatca gaataccgga aacacacaga ttgagatcaa gacttcaaaa ggatcttgct  17160
acttgattga taatacaaga aaactcgctt gggaaagcaa atgggcttca gacgacttga  17220
aagttttgga agaattaagc aatgaaatcg aatctgcttt aacagatcca aatgtgtttt  17280
ggtctgcaga tataaccgcc aagatcgaag catctttctg tcaagagata tatccttcac  17340
agatattgaa tgataaagtg aagcaaggag aagcgtcaaa acagtttgtg aaagctaaat  17400
gtgctgatgg aaggtatgct gtgagtttca acagtgttaa gatcggtgct gctttgcaaa  17460
gtattgatga ttggtgggat gaagatgcta gtaaaagact tagagttcat gaatttggtg  17520
cagacaagga aattggcgta gcgaggaggc ctcccgattc cgaacaaaac ttttattcaa  17580
tcttcaagaa cactgaatgg tatctttccg ccttaaagaa ttgcatcact aacaagaatg  17640
agaagattga cccagcaata tactatttgt tcagcgtgtt gatcaaaggt ggaatgtttc  17700
agaaaaaggc agaggcaaag aaagcatgag cttatatgaa gatgaagatg aaatatttgg  17760
tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt tttcttggct tgttgtgtta  17820
tgaatttgtg gctttttcta atattaaatg aatgtaacat ctcattataa tgaataaaca  17880
aatgtttcta taatccattg tgaatgtttt gttggatctc ttctccagca tataactact  17940
gtatgtgcta tggtatggac tatggaatat gattaaagat aagcgctata caggagtggc  18000
agacatactg tcccacaaat gaagatggaa tctgtaaaag aaaacgcgtg aaataatgcg  18060
tctgacaaag gttaggtcgg ctgcctttaa tcaataccaa agtggtccct accacgatgg  18120
aaaaactgtg cagtcggttt ggctttttct gacgaacaaa taagattcgt ggccgacagg  18180
tgggggtcca ccatgtgaag gcatcttcag actccaataa tggagcaatg acgtaagggc  18240
ttacgaaata agtaagggta gtttgggaaa tgtccactca cccgtcagtc tataaatact  18300
tagcccctcc ctcattgtta agggagcaaa atctcagaga gatagtccta gagagagaaa  18360
gagagcaagt agcctagaag tagtcaaggc ggcgaagtat tcaggcacgt ggccaggaag  18420
aagaaaagcc aagacgacga aaacaggtaa gagctaagct taatgggaaa gcgtactgct  18480
gatggttcag agtttgaaag tccaaagaag aaaagaaaag ttggtagtgg tatgcacctg  18540
aaagagctgt tagagataac agatacaact gagagagata gatctctccg acgtgccttc  18600
tcgccgtata ccgccatgat cgacattaca ggatcagaag cagtcgctct aattatccta  18660
ctcaatctca cttatagaaa gaatcaggta gatgacctgt tggataagaa actcgcgaaa  18720
caagccctca agagtgagga tcatatcaac aaatgcatca aggagattgc ttggtttcat  18780
actcacaacc ttaaatatcc agacataagg gtgagtaagc aaaatcttgc agtagaacca  18840
ccaacattgc attcttatgt gctttcttct gctaattatc ccaaagctta tggttggtcc  18900
```

-continued

```
cataattcag ctaaagtgaa ttttgctaag ttgtttgtct catacttcaa atggcaaaac  18960
caagtgagct ggcttgccca agtcttggca accaactctg ataactggaa aagtgcattt  19020
acttcccttg gtttatcggt gaaagccttc aaatctttat gcgttacagt taagaattct  19080
ttacctgaag aggcaatccc ggactcagtg gatagatata gcaggcagat tagaatgcct  19140
tatcatgatg gttatcttgc tgtcactcct gtgatatctc acgttgttca aagcaaaata  19200
caacaagctg ctattgacaa aagagctagg ttttctaatg ttgagtttac cagaccagct  19260
gctgtttcaa tgctggcggc aagtctaggt ggtgtgataa atgtacttaa ttatcctcca  19320
tacattcgat caaagtacca tggactatcg aattcaaggg cttttaagtt aaacaacgga  19380
cagactgtct tcaacgtgga agccctcctg aagcctgagc ttataaaagc tcttgaagga  19440
attatattct ctaacaatgc attggcatta aaacagagaa ggcagcagaa agtaaaaaac  19500
atcaaagaac tgagaaatac cctccttgag tggttttccc ctgttttcga atggagattg  19560
gatgcaattg agaatggtta tgatctagaa caattggaat cagcgagcga acggttagag  19620
tacaagattc tgagtttgcc ggataatgaa ctaccttctt tgactatacc cttgtttaga  19680
ttacttaatg agatgttggg aggagtttca atgactcaga ggtatgcatt ccatccaaag  19740
ttgatgtctc cattaaaggc tgctctccaa tggttgctcg ttaatttaac tgatcaaaag  19800
catgttttaa ttgaagaaga tgatgagcat tacagatacc ttcacctgtc agggattagg  19860
gtttttgatg cccaagcatt gtctaatcct tactgctctg gtattccttc gcttacggcg  19920
gtttggggaa tgatccacag ttatcaacgc aaactaaacg aagctttggg cacgaacgta  19980
cggttcacaa gcttttcatg gttcatccgt aattactctg ccgttgctgg taaaaagcta  20040
ccagaactta gccttcaggg tgcacaacag tctagattga aaagacctgg gataatagac  20100
gggaagtatt gtgatttggt gtttgacttg attattcata tcgatggcta tgaagacgat  20160
ttgcaggctg ttgattcaaa gcctgatatc ctcaaggctc atttcccaag caatttcgcc  20220
ggcggcgtga tgcatcaacc cgagttgaat tctaatatta attggtgttg tctttactcc  20280
aacgaaaacc agctttttga gaagctccgt cggcttcctc tttctggatg ttgggttatg  20340
cccacagaac ataagattca agatcttgat gaactgttgt tgctgttgaa ctcggatagt  20400
aaactaagtc caagcatgat gggatacatg cttctcacag agccaatggc tcgtgtagga  20460
tcgttagaga ggcttcactg ttatgcagaa cctgctattg gtgttgtcaa atacgaggca  20520
gcaacttctg ttcgattaaa aggtatcgga aactatttca attccgcgtt ttggatgctc  20580
gatgctcaag aaaagttcat gctcatgaag aaagtgtgag cttatatgaa gatgaagatg  20640
aaatatttgg tgtgtcaaat aaaaaggttg tgtgcttaag tttgtgtttt tttcttggct  20700
tgttgtgtta tgaatttgtg gctttttcta atattaaatg aatgtaacat ctcattataa  20760
tgaataaaca aatgtttcta taatccattg tgaatgtttt gttggatctc ttctccagca  20820
tataactact gtatgtgcta tggtatggac tatggaatat gattaaagat aagcgctaca  20880
gaggagtctg agcaggacaa ctcgcgtagt gagagttaca tgttcgttgg gttcttccga  20940
cacggacctg agttggccaa cgtcccacct gaggtctgtg ccccggtgat gagaagtgtg  21000
catctcgttc ttgcagctcg tcagtacttt cagaatcatg gcgtgcatgg tagaatgacc  21060
cttataacgg acttcgacat ggcaatcgct aggtagctta atgaataggc gtatcacgag  21120
gccctttcgt cttcaagaat tagcttggca attcccgatc tagtaacata gatgacaccg  21180
cgcgcgataa tttatcctag tttgcgcgct atattttgtt ttctatcgcg tattaaatgt  21240
ataattgcgg gactctaatc ataaaaaccc atctcataaa taacgtcatg cattacatgt  21300
taattattac atgcttaacg taattcaaca gaaattatat gataatcgtc gcaagaccgg  21360
caacaggatt caatcttaag aaactttatt gccaaatgtt tgaacgatcg ggatcgaaga  21420
tccattccaa gcctacagga acaggtggtg gcggccctcg gtgcgctcgt actgctccac  21480
gatggtgtag tcctcgttgt gggaggtgat gtccagcttg gcgtccacgt agtagtagcc  21540
gggcagctgc acgggcttct tggccatgta gatggacttg aactccacca ggtagtggcc  21600
gccgtccttc agcttcaggg ccttgtgtgt ctcgcccttc agcacgccgt cgcgggggta  21660
caggcgctcg gtggaggcct cccagcccat cgtcttcttc tgcatcacgg ggccgtcgga  21720
ggggaagttc acgccgatga acttcacctt gtagatgaag cagccgtcct gcagggagga  21780
gtcctgggtc acggtcgcca cgccgccgtc ctcgaagttc atcacgcgct cccacttgaa  21840
gccctcgggg aaggacagct tcttgtagtc ggggatgtcg gcggggtgct tcacgtacac  21900
cttggagccg tactggaact gggggggacag gatgtcccag gcgaagggca gggggccgcc  21960
cttggtcacc ttcagcttca cggtgttgtg gccctcgtag gggcggccct cgccctcgcc  22020
ctcgatctcg aactcgtggc cgttcacggt gccctccatg cgcaccttga agcgcatgaa  22080
ctcggtgatg acgttctcgg aggaggccat ttcgcctcga tcgagttgag agtgaatatg  22140
agactctaat tggataccga ggggaattta tggaacgtca gtggagcatt tttgacaaga  22200
aatatttgct agctgatagt gaccttaggc gacttttgaa cgcgcaataa tggtttctga  22260
cgtatgtgct tagctcatta aactccagaa acccgcggct cagtggctcc ttcaacgttg  22320
cggttctgtc agttccaaac gtaaaacggc ttgtcccgcg tcatcggcgg gggtcatacc  22380
gtgactccct taattctcat gtatggttcg ctatactgaa gagcaattct gacaggatat  22440
attggcgggt aaacctaaga gaaaagagcg tttattagaa taatcggata tttaaaaggg  22500
cgtgaaaagg tttatccgtt cgtccatttg tatgtgcatg ccaaccacag ggttcccctc  22560
gggatcaaag tactttgatc caacccctcc gctgctatag tgcagtcggc ttctgacgtt  22620
cagtgcagcc gtcatctgaa aacgacatgt cgcacaagtc ctaagttacg cgacaggctg  22680
ccgccctgcc cttttcctgg cgttttcttg tcgcgtgttt tagtcgcata aagtagaata  22740
cttgcgacta gaaccggaga cattacgcca tgaacaagag cgccgccgct ggcctgctgg  22800
gctatgcccg cgtcagcacc gacgaccagg acttgaccaa ccaacgggcc gaactgcacg  22860
cggccggctg caccaagctg ttttccgaga agatcaccgg caccaggcgc gaccgcccgg  22920
agctggccag gatgcttgac cacctacgcc ctggcgacgt tgtgacagtg accaggctag  22980
accgcctggc ccgcagcacc cgcgacctac tggacattgc cgagcgcatc caggaggccg  23040
gcgcgggcct gcgtagcctg gcagagccgt gggccgacac caccacgccg gccggccgca  23100
tggtgttgac cgtgttcgcc ggcattgccg agttcgagcg ttccctaatc atcgaccgca  23160
cccggagcgg gcgcgaggcc gccaaggccc gaggcgtgaa gtttggcccc cgccctaccc  23220
tcaccccggc acagatcgcg cacgcccgcg agctgatcga ccaggaaggc cgcaccgtga  23280
aagaggcggc tgcactgctt ggcgtgcatc gctcgaccct gtaccgcgca cttgagcgca  23340
gcgaggaagt gacgcccacc gaggccaggc ggcgcggtgc cttccgtgag gacgcattga  23400
ccgaggccga cgccctggcg gccgccgaga atgaacgcca agaggaacaa gcatgaaacc  23460
```

-continued

```
gcaccaggac ggccaggacg aaccgttttt cattaccgaa gagatcgagg cggagatgat  23520
cgcggccggg tacgtgttcg agccgcccgc gcacctctca accgtgcggc tgcatgaaat  23580
cctggccggt ttgtctgatg ccaagctggc ggcctggccg gccagcttgg ccgctgaaga  23640
aaccgagcgc cgccgtctaa aaaggtgatg tgtatttgag taaaacagct tgcgtcatgc  23700
ggtcgctgcg tatatgatgc gatgagtaaa taaacaaata cgcaaggggga acgcatgaag  23760
gttatcgctg tacttaacca gaaaggcggg tcaggcaaga cgaccatcgc aacccatcta  23820
gcccgcgccc tgcaactcgc cggggccgat gttctgttag tcgattccga tccccagggc  23880
agtgcccgcg attgggcggc cgtgcgggaa gatcaaccgc taaccgttgt cggcatcgac  23940
cgcccgacga ttgaccgcga cgtgaaggcc atcggccggc gcgacttcgt agtgatcgac  24000
ggagcgcccc aggcggcgga cttggctgtg tccgcgatca aggcagccga cttcgtgctg  24060
attccggtgc agccaagccc ttacgacata tgggccaccg ccgacctggt ggagctggtt  24120
aagcagcgca ttgaggtcac ggatggaagg ctacaagcgg cctttgtcgt gtcgcgggcg  24180
atcaaaggca cgcgcatcgg cggtgaggtt gccgaggcgc tggccgggta cgagctgccc  24240
attcttgagt cccgtatcac gcagcgcgtg agctacccag gcactgccgc cgccggcaca  24300
accgttcttg aatcagaacc cgagggcgac gctgcccgcg aggtccaggc gctggccgct  24360
gaaattaaat caaaactcat ttgagttaat gaggtaaaga gaaaatgagc aaaagcacaa  24420
acacgctaag tgccggccgt ccgagcgcac gcagcagcaa ggctgcaacg ttggccagcc  24480
tggcagacac gccagccatg aagcgggtca actttcagtt gccggcggag gatcacacca  24540
agctgaagat gtacgcggta cgccaaggca agaccattac cgagctgcta tctgaataca  24600
tcgcgcagct accagagtaa atgagcaaat gaataaatga gtagatgaat tttagcggct  24660
aaaggaggcg gcatggaaaa tcaagaacaa ccaggcaccg acgccgtgga atgccccatg  24720
tgtggaggaa cgggcggttg gccaggcgta agcggctggg ttgcctgccg gccctgcaat  24780
ggcactggaa cccccaagcc cgaggaatcg gcgtgagcgg tcgcaaacca tccggcccgg  24840
tacaaatcgg cgcggcgctg ggtgatgacc tggtggagaa gttgaaggcc gcgcaggccg  24900
cccagcggca acgcatcgag gcagaagcac gccccggtga atcgtggcaa gcggccgctg  24960
atcgaatccg caaagaatcc cggcaaccgc cggcagccgg tgcgccgtcg attaggaagc  25020
cgcccaaggg cgacgagcaa ccagattttt tcgttccgat gctctatgac gtgggcaccc  25080
gcgatagtcg cagcatcatg gacgtggccg ttttccgtct gtcgaagcgt gaccgacgag  25140
ctggcgaggt gatccgctac gagcttccag acgggcacgt agaggtttcc gcagggccgg  25200
ccggcatggc cagtgtgtgg gattacgacc tggtactgat ggcggtttcc catctaaccg  25260
aatccatgaa ccgataccgg gaagggaagg gagacaagcc cggccgcgtg ttccgtccac  25320
acgttgcgga cgtactcaag ttctgccggc gagccgatgg cggaaagcag aaagacgacc  25380
tggtagaaac ctgcattcgg ttaaacacca cgcacgttgc catgcagcgt acgaagaagg  25440
ccaagaacgg ccgcctggtg acggtatccg agggtgaagc cttgattagc cgctacaaga  25500
tcgtaaagag cgaaaccggg cggccggagt acatcgagat cgagctagct gattggatgt  25560
accgcgagat cacagaaggc aagaacccgg acgtgctgac ggttcacccc gattactttt  25620
tgatcgatcc cggcatcggc cgttttctct accgcctggc acgccgcgcc gcaggcaagg  25680
cagaagccag atggttgttc aagacgatct acgaacgcag tggcagcgcc ggagagttca  25740
agaagttctg tttcaccgtg cgcaagctga tcgggtcaaa tgacctgccg gagtacgatt  25800
tgaaggagga ggcggggcag gctggcccga tcctagtcat gcgctaccgc aacctgatcg  25860
agggcgaagc atccgccggt tcctaatgta cggagcagat gctagggcaa attgccctag  25920
caggggaaaa aggtcgaaaa ggactctttc ctgtggatag cacgtacatt gggaacccaa  25980
agccgtacat tgggaaccgg aacccgtaca ttgggaaccc aaagccgtac attgggaacc  26040
ggtcacacat gtaagtgact gatataaaag agaaaaaagg cgatttttcc gcctaaaact  26100
ctttaaaact tattaaaact cttaaaaccc gcctggcctg tgcataactg tctggccagc  26160
gcacagccga agtgctgcaa aaagcgccta cccttcggtc gctgcgctcc ctacgccccg  26220
ccgcttcgcg tcggcctatc gcggccgctg gccgctcaaa aatggctggc ctacggccag  26280
gcaatctacc agggcgcgga caagccgcgc cgtcgccact cgaccgccgg cgcccacatc  26340
aaggcaccct gcctcgcgcg tttcggtgat gacggtgaaa acctctgaca catgcagctc  26400
ccggtgacgg tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc  26460
gcgtcagcgg gtgttggcgg gtgtcggggc gcagccatga cccagtcacg tagcgatagc  26520
ggagtgtata ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaccata  26580
tgcggtgtga aataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgctcatccg  26640
cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc  26700
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt  26760
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc  26820
ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa  26880
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc  26940
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg  27000
cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc  27060
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc  27120
gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca  27180
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact  27240
acggctacac tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg  27300
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt  27360
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct  27420
tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatgc  27480
attctaggtg attagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca  27540
ggattatcaa taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg  27600
aggcagttcc ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca  27660
tcaatacaac ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca  27720
tgagtgacga ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt  27780
tcaacaggcc agccattacg ctcgtcatca aaatcactcg catcaaccaa accgttattc  27840
attcgtgatt gcgcctgagc gagtcgaaat acgcgatcgc tgttaaaagg acaattacaa  27900
acaggaatcg aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct  27960
gaatcaggat attcttctaa tacctggaat gctgttttcc ctgggatcgc agtggtgagt  28020
```

-continued

```
aaccatgcat catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc  28080
gtcagccagt ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca  28140
tgtttcagaa acaactctgg cgcatcgggc ttcccataca atcggtagat tgtcgcacct  28200
gattgcccga cattatcgcg agcccattta tacccatata aatcagcatc catgttggaa  28260
tttaatcgcg gccttgagca agacgtttcc cgttgaatat ggctcataac agaacttatt  28320
atttccttcc tcttttctac agtatttaaa gataccccaa gaagctaatt ataacaagac  28380
gaactccaat tcactgttcc ttgcattcta aaaccttaaa taccagaaaa cagctttttc  28440
aaagttgttt tcaaagttgg cgtataacat agtatcgacg gagccgattt tgaaaccgcg  28500
gtgatcacag gcagcaacgc tctgtcatcg ttacaatcaa catgctaccc tccgcgagat  28560
catccgtgtt tcaaacccgg cagcttagtt gccgttcttc cgaatagcat cggtaacatg  28620
agcaaagtct gccgccttac aacggctctc ccgctgacgc cgtcccggac tgatgggctg  28680
cctgtatcga gtggtgattt tgtgccgagc tgccggtcgg ggagctgttg gctggctgg   28739

SEQ ID NO: 45          moltype = DNA  length = 22059
FEATURE                Location/Qualifiers
source                 1..22059
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 45
tggcaggata tattgtggtg taaacataac gaattgctct tcacagagga gaatggtaga  60
cactttatta gttagatccg ctgttattct gcacgattgc gcttgccggt ggaaatttaa  120
ggacgagtgt agtgcactca cccttagcct tcagcgtgac aagattaggg tagcaccgga  180
aacagcgatt ttacggacag atagtacagt ctgagactca gtcactggtc cctaaacggg  240
agccggtacg tcggcggtat actaatgagg agtggcagac atactgtccc acaaatgaag  300
atggaatctg taaaagaaaa cgcgtgaaat aatgcgtctg acaaaggtta ggtcggctgc  360
ctttaatcaa taccaaagtg gtccctacca cgatggaaaa actgtgcagt cggtttggct  420
ttttctgacg aacaaataag attcgtggcc gacaggtggg ggtccaccat gtgaaggcat  480
cttcagactc caataatgga gcaatgacgt aagggcttac gaaataagta agggtagttt  540
gggaaatgtc cactcacccg tcagtctata aatacttagc ccctccctca ttgttaaggg  600
agcaaaatct cagagagata gtcctagaga gagaaagaga gcaagtagcc tagaagtagt  660
caaggcggcg aagtattcag gcacgtggcc aggaagaaga aaagccaaga cgacgaaaac  720
aggtaagagc taagcttaat gtacattagg aacttgagaa agccttctcc aaataagaat  780
gtttttaagt tcgcatctac caaggtcagc tctgtggtca tgtgtgaatc cagccttgag  840
tttgacgctt gctttcatca tgaatataac gacttgattg agagttttgg ttctcagcca  900
gaaggtttca aatacgaatt catgggtaaa tctctccctt acactccaga tgctttgatc  960
tcctacacgg acaaaactca aaagtatcac gagtataaac catattcaaa aatcgcatca  1020
cctctattta gagcagagtt tgctgcaaag agggctgcaa gtcttaagtt gggtattgat  1080
ttggtactcg tcacagacag acagattaga gtaaacccaa tactcaacaa tctgaagctc  1140
cttcatagat actcgggcgt ttatggtata tctggtatcc agaaggaact acttagcttt  1200
attcacaaaa gtggcgttat caaactcaat gatatctcat cacaggttgg gatccccatt  1260
ggagaaactc ggtcatttct ttttggactg atgcataagg gacttgtaaa ggctgaccta  1320
ggttgtgatg acctgacaaa caaccctact ctttgggcta ctcctggaag tggctcaggc  1380
aaacgtactg ctgacggaag tgagttcgag tccccaaaga aaaaaagaaa agttggtagt  1440
ggatctggtg gaatgactga tttcttcaat gagtttgatg aatcgcttgt accactgaaa  1500
ccacaaactc caacgcagta tgttaagctg gacgatgcta atttaattca gagagattta  1560
gatacattct cagatacttt caagaaccaa gccctccaaa gatataagct aatttctaca  1620
attgataaaa agttaagccg cggatggact caaagaaatc tcgacccgat cctcgacgaa  1680
ctcttcaaag gcggtgatgt agtgaggcca aactggagaa ctgtggcaag atggaggaag  1740
aaatatatag aaagcaatgg tgacatcgct tctcttgctg ataaaaacca caaaatggga  1800
aacagaacaa atcggataaa aggagatgac aaatttttg acaaggcctt ggaacgtttc  1860
cttgacgcga aaagacctac aatagcaaca gcttatcaat attacaagga cttgatagtg  1920
attgaaaacg agtcaattgt tgaagggaag atccccatca tctcttataa tgcatttaac  1980
aaacgaatta aagccattcc tccttatgct gttgcagtcg ctaggcatgg gaaattcaaa  2040
gctgatcaat ggtttgcata ttgtgcagct cacgttccac caacgcgtat acttgaacga  2100
gtggagattg atcatactcc gttggatcta atattattgg atgatgagct tctgattccc  2160
attggccggc catatcttac acttctcata gatgtgtttt caggttgcgt tctgggattc  2220
catttgtcat acaagagccc ttcttatgtc tctgcagcaa aagcaatcac tcacgcaata  2280
aaacccaaaa gcttagatgc gcttaatatt gagcttcaga atgactggcc ctgtttcggg  2340
aagtttgaga atttagtagt tgacaatgga gctgaatttt ggtcaaaaaa tttggaacat  2400
gcttgtcaat ctgctggtat aaatattcaa tacaatccgg tgcgtaagcc atggctcaag  2460
ccttttatag agaggttctt tggggtgatg aatgaatact tcttgccgga actcccagga  2520
aaaacatttt ctaacattct tgagaaagaa gaatacaaac cggagaagga tgcaataatg  2580
aggttcagca cctttgttga ggagttccac cgttggattg ccgacgtgta tcatcaagat  2640
agtaatagcc gtgaaacgcg tattcctatt aagcgatggc aacaaggatt cgatgcttat  2700
cctcctttaa caatgaacga agaagaggag acacggtttt ctatgctcat gagaatctct  2760
gactctagga cgctaaccag gaatgggttt aagtaccaag aattgatgta tgattctacc  2820
gccttagcag attaccgtaa gcattaccct cagacaaaag agactgtgaa gaagctaata  2880
aaggtcgatc cagatgatat ctccaagatt tacgtttatc tggaagagtt ggaaagttac  2940
ttggaggttc cttgcacaga tccaactgga tacactgatg gtttatccat atatgagcat  3000
aaaaccatta agaagatcaa ccgagaagtt attagagaat caaaagatag ccttgggttg  3060
gcgaaagcga gaatggctat tcatgagaga gtgaaacaag agcaagaagt attcatcgaa  3120
agtaagacca aagctaagat tactgcagtc aagaagcagg ctcagatcgc ggatgttagt  3180
aatacaggca catcaacaat caaggtgtcc gaggagtctg cggctcctgt ccagaaacac  3240
atctcaaatg ataacagtga tgattgggat gatgatctgg aagcttttga atgagcttat  3300
atgaagatga agatgaaata tttggtgtgt caaataaaaa ggttgtgtgc ttaagtttgt  3360
gtttttttct tggcttgttg tgttatgaat ttgtggcttt ttctaatatt aaatgaatgt  3420
```

-continued

```
aacatctcat tataatgaat aaacaaatgt ttctataatc cattgtgaat gttttgttgg  3480
atctcttctc cagcatataa ctactgtatg tgctatggta tggactatgg aatatgatta  3540
aagataagcg ctagcaagga gtggcagaca tactgtccca caaatgaaga tggaatctgt  3600
aaaagaaaac gcgtgaaata atgcgtctga caaaggttag gtcggctgcc tttaatcaat  3660
accaaagtgg tccctaccac gatggaaaaa ctgtgcagtc ggtttggctt tttctgacga  3720
acaaataaga ttcgtggccg acaggtgggg gtccaccatg tgaaggcatc ttcagactcc  3780
aataatggag caatgacgta agggcttacg aaataagtaa gggtagtttg ggaaatgtcc  3840
actcacccgt cagtctataa atacttagcc cctccctcat tgttaaggga gcaaaatctc  3900
agagagatag tcctagagag agaaagagag caagtagcct agaagtagtc aaggcggcga  3960
agtattcagg cacgtggcca ggaagaagaa aagccaagac gacgaaaaca ggtaagagct  4020
aagcttaatg ggaaatgctc ttacagagat tcaaatagag aagttacgga atttcagtga  4080
ttgcatcgtg atgcatcctc agataaagac cattttcaat gattttgatg aactaaggct  4140
taatagaaag tttcagagtg atcagcaatg tatgcttctt ataggagata ctggtgttgg  4200
caaatcacac accatcaacc attacaagaa aagagtcctc gcaacacaaa actattcaag  4260
aaacacaatg ccggtccttg tttctcgcat atcacgtggg aaaggtttgg atgctacttt  4320
ggtgcagatg ctggcggatc tagagctgtt tggttcttct cagatcaaga agcggggcta  4380
caaaacagac cttaccaaaa aattagttga gagtttgatc aaagctcaag tggagctgtt  4440
gattatcaat gaattccaag agcttattga gtttaaaagt gttcaagaaa gacaacagat  4500
tgcaaatggt ctaaagttta ttagtgaaga agccaaggtt cccattgttt tggttggaat  4560
gccatgggca gcaaagatag ctgaggaacc acaatgggct tctcgactcg taagaaaacg  4620
aaagctagaa tatttttcac tcaagaatga cagcaagtac tttaggcagt atttgatggg  4680
tcttgccaaa aagatgcctt tcgacgttcc tcctaaatta gaatcgaaga acactacgat  4740
tgcgctattt gcagcttgta gaggggaaaa cagagctttg aagcaccttc ttttagaagc  4800
actgaagctc gccctttcat gcaatgagta tcttgagaac aagcatttca tcacagctta  4860
tgacaaattc gatttcttca acgacaagga gaaactgaaa tccaaaaacc ctttcaagca  4920
agatataaaa gatattgaaa tttatgaagt tattaaaaat tcttcctaca atccaaatgc  4980
acttgatcca gaagatatgt tgactgatcg agtatttgct attgtgaaag gaagcggaaa  5040
gaggactgca gatggatcag aatttgagtc cccaaagaag aaaaggaaag tgtgagctta  5100
tatgaagatg aagatgaaat atttggtgtg tcaaataaaa aggttgtgtg cttaagtttg  5160
tgtttttttc ttggcttgtt gtgttatgaa tttgtggctt tttctaatat taaatgaatg  5220
taacatctca ttataatgaa taaacaaatg tttctataat ccattgtgaa tgttttgttg  5280
gatctcttct ccagcatata actactgtat gtgctatggt atggactatg gaatatgatt  5340
aaagataagc gctatacagg agtggcagac atactgtccc acaaatgaag atggaatctg  5400
taaaagaaaa cgcgtgaaat aatgcgtctg acaaaggtta ggtcggctgc ctttaatcaa  5460
taccaaagtg gtccctacca cgatggaaaa actgtgcagt cggtttggct ttttctgacg  5520
aacaaataag attcgtggcc gacaggtggg ggtccaccat gtgaaggcat cttcagactc  5580
caataatgga gcaatgacgt aagggcttac gaaataagta agggtagttt gggaaatgtc  5640
cactcacccg tcagtctata aatacttagc ccctccctca ttgttaaggg agcaaaatct  5700
cagagagata gtcctagaga gagaaagaga gcaagtagcc tagaagtagt caaggcggcg  5760
aagtattcag gcacgtggcc aggaagaaga aaagccaaga cgacgaaaac aggtaagagc  5820
taagcttaat gggaaagcgt actgctgatg gttcagagtt tgaaagtcca aagaagaaaa  5880
gaaaagttgg tagtggtatg gcttttcttt tctcaccaaa agcaagagcg ttttccgatg  5940
aatcgcttga aagttattta cttcgagttg tctcagaaaa cttcttcgat tcatatgaag  6000
gactatcttt ggccatcaga gaagaacttc atgagttgga cttcgaggct catggagcat  6060
ttcccgtaga tttaaagaga cttaatgttt accatgcaaa acacaactca cactttagga  6120
tgagagctct tggcctatta gaaactcttt tagatcttcc tcgttatgag cttcaaaagc  6180
tagcactcct gaagtcagac attaaattca attcttccgt tgccctctac aacaatggtg  6240
ttgatattcc tttgagattc attagacatc atgctgagga agctgttgac tccattccgg  6300
tttgttctca atgcttggct gaagaagctt atataaagca atcttggcat atcaaatggg  6360
ttaatgcttg caccaaacat caatgtgctt tgcttcacaa ttgtcctgaa tgttatgcac  6420
cgatcaatta cattgagaat gaatcaatca ctcattgctc ttgtggtttt gaattgagct  6480
gcgcctctac aagtcctgtg aacacactct cgattgaaca tttgaacaaa ctcctggaca  6540
aaggagaacg gaatgactca aatccattgt ttaacaacat gacgttgaca gaaaggtttg  6600
ctgctttatt atggtatcaa gagaggtatt cacagactga taatttctgt ttgaatgatg  6660
ctgtgaacta cttcagtaag tggccagcag tgtttaatac tgagctagat gagctctcaa  6720
agaatgccga gatgaagctg attgatctct tcaacaagac agagtttaag tttatctttg  6780
gagatgcaat tcttgcttgt ccaagtactc agaagcagag tgaatctcat ttcatctacc  6840
gggcattact tgattacctt gtcactttag tggagagcaa tcccaaaaca aagaagccaa  6900
acgcagctga tttgcttgtt tctgtcctgg aagcagcaac tttgttgggg acgagtgtag  6960
aacaagttta taggctctat cagaatggaa tactgcaaac cgcctttcgg cacaagatga  7020
atcagagaat aaacccttac aagggtgctt tctttctccg tcatgtgata gagtataaaa  7080
ccagctttgg taacgataaa gctagaatgt atctctcagc ttggtgagct tatatgaaga  7140
tgaagatgaa atatttggtg tgtcaaataa aaaggttgtg tgcttaagtt tgtgtttttt  7200
tcttggcttg ttgtgttatg aatttgtggc tttttctaat attaaatgaa tgtaacatct  7260
cattataatg aataaacaaa tgtttctata atccattgtg aatgttttgt tggatctctt  7320
ctccagcata taactactgt atgtgctatg gtatggacta tggaatatga ttaaagataa  7380
gcgctacaga ggagtctgag caggacaact cgcgtagtga gagttacatg ttcgttgggt  7440
tcttccgaca cggacctgag ttggccaacg tcccacctga ggtctgtgcc ccggtgatga  7500
gaagtgtgca tctcgttctt gcagctcgtc agtactttca gaatcatggc gtgcatggta  7560
gaatgaccct tataacggac ttcgacatgg caatcgctag gtaaatgaat gaggagtggc  7620
agacatactg tcccacaaat gaagatggaa tctgtaaaag aaaacgcgtg aaataatgcg  7680
tctgacaaag gttaggtcgg ctgcctttaa tcaataccaa agtggtccct accacgatgg  7740
aaaaactgtg cagtcggttt ggctttttct gacgaacaaa taagattcgt ggccgacagg  7800
tgggggtcca ccatgtgaag gcatcttcag actccaataa tggagcaatg acgtaagggc  7860
ttacgaaata agtaagggta gtttgggaaa tgtccactca cccgtcagtc tataaatact  7920
tagcccctcc ctcattgtta agggagcaaa atctcagaga gatagtccta gagagagaaa  7980
```

-continued

```
gagagcaagt agcctagaag tagtcaaggc ggcgaagtat tcaggcacgt ggccaggaag  8040
aagaaaagcc aagacgacga aaacaggtaa gagctaagct taatgggaaa gcgtactgct  8100
gatggttcag agtttgaaag tccaaagaag aaaagaaaag ttggtagtgg tatgcagaga  8160
tattatttta ccgttcattt tcttcctaaa caagctaatt tggcccttct caccggaaga  8220
tgcatatcta tcatgcatgg atttatcttg aagcataaca ttgaaggaat gggagtaaca  8280
tttcccgcat ggtctgattc aagcattgga aacgagattg cttttgttta cactgataaa  8340
gaaatcctaa acacattgaa agatcaggct tactttgtgg atatgcaaga ctgtggtttc  8400
ttcaaggttt ctcaggtgtt ggctgttcct gattcttgcg aggaagtcag gttcatccgc  8460
aatcaagccg tcgcaaagat tttcactggc gaatcgagga ggagacttaa acggttacaa  8520
aagagagcct tagcaagagg agaagatttc aatccaaaga agatagaagc tccaagagaa  8580
atagacattt ttcaccgtgt agcaatgaca agcaaaagtt ctcaagaaga ctacattctg  8640
cacatccaga agcaagatgt tgattgtcaa gcagagcctt atttttcaaa ctatggtctt  8700
gcgtcaaatg agaagttcaa aggaactgta ccagacttga gtccgtccat tgatagaaat  8760
tgagcttata tgaagatgaa gatgaaatat ttggtgtgtc aaataaaaag gttgtgtgct  8820
taagtttgtg tttttttctt ggcttgttgt gttatgaatt tgtggctttt tctaatatta  8880
aatgaatgta acatctcatt ataatgaata aacaaatgtt tctataatcc attgtgaatg  8940
ttttgttgga tctcttctcc agcatataac tactgtatgt gctatggtat ggactatgga  9000
atatgattaa agataagcgc tagcaaggag tggcagacat actgtcccac aaatgaagat  9060
ggaatctgta aaagaaaacg cgtgaaataa tgcgtctgac aaaggttagg tcggctgcct  9120
ttaatcaata ccaaagtggt ccctaccacg atggaaaaac tgtgcagtcg gtttggcttt  9180
ttctgacgaa caaataagat tcgtggccga caggtggggg tccaccatgt gaaggcatct  9240
tcagactcca ataatggagc aatgacgtaa gggcttacga aataagtaag ggtagtttgg  9300
gaaatgtcca ctcacccgtc agtctataaa tacttagccc ctccctcatt gttaagggag  9360
caaaatctca gagagatagt cctagagaga gaaagagagc aagtagccta gaagtagtca  9420
aggcggcgaa gtattcaggc acgtggccag gaagaagaaa agccaagacg acgaaaacag  9480
gtaagagcta agcttaatgg gaaagcgtac tgctgatggt tcagagtttg aaagtccaaa  9540
gaagaaaaga aaagttggta gtggtatgga gctttgcaat atactcaagt atgatcgaag  9600
tctttaccct gggaaagcgg ttttcttcta caaaacggcg gacagcgatt ttgttccttt  9660
ggaagctgac attaataaga taagaggacc aaaatcaggc tttacagagg ctttcactcc  9720
tcaattctct ccgaaaaata tctctccaca agatctaact cacaacaaca ttcttacact  9780
cgaagagtgt tatgttccac caaatgtcga acatattttc tgccggtttt ctctgcgtgt  9840
ccaagctaat tcgttagtcc cctctggttg ttccgatcct gaggttttta gccttctcaa  9900
agagcttgca gaaaccttca aagaatgtgg tggttacaag gagctggcag tacgctattg  9960
tcggaacatt ttgattggaa catggctttg gagaaatcag aataccggaa acacacagat  10020
tgagatcaag acttcaaaag gatcttgcta cttgattgat aatacaagaa aactcgcttg  10080
ggaaagcaaa tgggcttcag acgacttgaa agttttggaa gaattaagca atgaaatcga  10140
atctgcttta acagatccaa atgtgttttg gtctgcagat ataaccgcca agatcgaagc  10200
atctttctgt caagagatat atccttcaca gatattgaat gataaagtga agcaaggaga  10260
agcgtcaaaa cagtttgtga aagctaaatg tgctgatgga aggtatgctg tgagtttcaa  10320
cagtgttaag atcggtgctg ctttgcaaag tattgatgat tggtgggatg aagatgctag  10380
taaaagactt agagttcatg aatttggtgc agacaaggaa attggcgtag cgaggaggcc  10440
tcccgattcc gaacaaaact tttattcaat cttcaagaac actgaatggt atctttccgc  10500
cttaaagaat tgcatcacta acaagaatga gaagattgac ccagcaatat actatttgtt  10560
cagcgtgttg atcaaaggtg gaatgtttca gaaaaaggca gaggcaaaga aagcatgagc  10620
ttatatgaag atgaagatga aatatttggt gtgtcaaata aaaaggttgt gtgcttaagt  10680
ttgtgttttt ttcttggctt gttgtgttat gaatttgtgg ctttttctaa tattaaatga  10740
atgtaacatc tcattataat gaataaacaa atgtttctat aatccattgt gaatgttttg  10800
ttggatctct tctccagcat ataactactg tatgtgctat ggtatggact atggaatatg  10860
attaaagata agcgctatac aggagtggca gacatactgt cccacaaatg aagatggaat  10920
ctgtaaaaga aaacgcgtga aataatgcgt ctgacaaagg ttaggtcggc tgcctttaat  10980
caataccaaa gtggtcccta ccacgatgga aaaactgtgc agtcggtttg gctttttctg  11040
acgaacaaat aagattcgtg gccgacaggt gggggtccac catgtgaagg catcttcaga  11100
ctccaataat ggagcaatga cgtaagggct tacgaaataa gtaagggtag tttgggaaat  11160
gtccactcac ccgtcagtct ataaatactt agcccctccc tcattgttaa gggagcaaaa  11220
tctcagagag atagtcctag agagagaaag agagcaagta gcctagaagt agtcaaggcg  11280
gcgaagtatt caggcacgtg gccaggaaga agaaaagcca agacgacgaa aacaggtaag  11340
agctaagctt aatgggaaag cgtactgctg atggttcaga gtttgaaagt ccaaagaaga  11400
aaagaaaagt tggtagtggt atgcacctga aagagctgtt agagataaca gatacaactg  11460
agagagatag atctctccga cgtgccttct cgccgtatac cgccatgatc gacattacag  11520
gatcagaagc agtcgctcta attatcctac tcaatctcac ttatagaaag aatcaggtag  11580
atgacctgtt ggataagaaa ctcgcgaaac aagccctcaa gagtgaggat catatcaaca  11640
aatgcatcaa ggagattgct tggtttcata ctcacaacct taaatatcca gacataaggg  11700
tgagtaagca aaatcttgca gtagaaccac caacattgca ttcttatgtg ctttcttctg  11760
ctaattatcc caaagcttat ggttggtccc ataattcagc taaagtgaat tttgctaagt  11820
tgtttgtctc atacttcaaa tggcaaaacc aagtgagctg gcttgcccaa gtcttggcaa  11880
ccaactctga taactggaaa agtgcattta cttcccttgg tttatcggtg aaagccttca  11940
aatctttatg cgttacagtt aagaattctt tacctgaaga ggcaatcccg gactcagtgg  12000
atagatatag caggcagatt agaatgcctt atcatgatgg ttatcttgct gtcactcctg  12060
tgatatctca cgttgttcaa agcaaaatac aacaagctgc tattgacaaa agagctaggt  12120
tttctaatgt tgagtttacc agaccagctg ctgtttcaat gctggcggca agtctaggtg  12180
gtgtgataaa tgtacttaat tatcctccat acattcgatc aaagtaccat ggactatcga  12240
attcaagggc ttttaagtta aacaacggac agactgtctt caacgtggaa gccctcctga  12300
agcctgagct tataaaagct cttgaaggaa ttatattctc taacaatgca ttggcattaa  12360
aacagagaag gcagcagaaa gtaaaaaaca tcaaagaact gagaaatacc ctccttgagt  12420
ggttttcccc tgttttcgaa tggagattgg atgcaattga gaatggttat gatctagaac  12480
aattggaatc agcgagcgaa cggttagagt acaagattct gagtttgccg gataatgaac  12540
```

-continued

```
taccttcttt gactataccc ttgtttagat tacttaatga gatgttggga ggagtttcaa  12600
tgactcagag gtatgcattc catccaaagt tgatgtctcc attaaaggct gctctccaat  12660
ggttgctcgt taatttaact gatcaaaagc atgttttaat tgaagaagat gatgagcatt  12720
acagatacct tcacctgtca gggattaggg tttttgatgc ccaagcattg tctaatcctt  12780
actgctctgg tattccttcg cttacggcgg tttggggaat gatccacagt tatcaacgca  12840
aactaaacga agctttgggc acgaacgtac ggttcacaag cttttcatgg ttcatccgta  12900
attactctgc cgttgctggt aaaaagctac cagaacttag ccttcagggt gcacaacagt  12960
ctagattgaa aagacctggg ataatagacg ggaagtattg tgatttggtg tttgacttga  13020
ttattcatat cgatggctat gaagacgatt tgcaggctgt tgattcaaag cctgatatcc  13080
tcaaggctca tttcccaagc aatttcgccg gcggcgtgat gcatcaaccc gagttgaatt  13140
ctaatattaa ttggtgttgt ctttactcca acgaaaacca gctttttgag aagctccgtc  13200
ggcttcctct ttctggatgt tgggttatgc ccacagaaca taagattcaa gatcttgatg  13260
aactgttgtt gctgttgaac tcggatagta aactaagtcc aagcatgatg ggatacatgc  13320
ttctcacaga gccaatggct cgtgtaggat cgttagagag gcttcactgt tatgcagaac  13380
ctgctattgg tgttgtcaaa tacgaggcag caacttctgt tcgattaaaa ggtatcggaa  13440
actatttcaa ttccgcgttt tggatgctcg atgctcaaga aaagttcatg ctcatgaaga  13500
aagtgtgagc ttatatgaag atgaagatga aatatttggt gtgtcaaata aaaaggttgt  13560
gtgcttaagt ttgtgttttt ttcttggctt gttgtgttat gaatttgtgg ctttttctaa  13620
tattaaatga atgtaacatc tcattataat gaataaacaa atgtttctat aatccattgt  13680
gaatgttttg ttggatctct tctccagcat ataactactg tatgtgctat ggtatggact  13740
atggaatatg attaaagata agcgctacag aggagtctga gcaggacaac tcgcgtagtg  13800
agagttacat gttcgttggg ttcttccgac acggacctga gttggccaac gtcccacctg  13860
aggtctgtgc cccggtgatg agaagtgtgc atctcgttct tgcagctcgt cagtactttc  13920
agaatcatgg cgtgcatggt agaatgaccc ttataacgga cttcgacatg gcaatcgcta  13980
ggtagcttaa tgggaatgga tcttcgatcc cgatcgttca aacatttggc aataaagttt  14040
cttaagattg aatcctgttg ccggtcttgc gacgattatc atataatttc tgttgaatta  14100
cgttaagcat gtaataatta acatgtaatg catgacgtta tttatgagat gggtttttat  14160
gattagagtc ccgcaattat acatttaata cgcgatagaa aacaaaatat agcgcgcaaa  14220
ctaggataaa ttatcgcgcg cggtgtcatc tatgttacta gatcgggaat tgccaagcta  14280
attcttgaag atgaaagggc ctcgtgatac gcctattttt ataggttaat gtcatgataa  14340
taatggtttc ttagacgtca ggtggcactt ttcggggaaa tgtgcgcgga acccctattt  14400
gtttattttt ctaaatacat tcaaatatgt atccgctcat gagacaataa ccctgataaa  14460
tgcttcaata atgggaccga ctcgtcactt gtacagctcg tccatgccga gagtgatccc  14520
ggcggcggtc acgaactcca gcaggaccat gtgatcgcgc ttctcgttgg ggtctttgct  14580
cagcttggac tgggtgctca ggtagtggtt gtcgggcagc agcacggggc cgtcgccgat  14640
gggggtgttc tgctggtagt ggtcggcgag ctgcacgccg ccgtcctcga tgttgtggcg  14700
gatcttgaag ttggccttga tgccgttctt ctgcttgtcg gcggtgatat agacgttgtc  14760
gctaaagtag ttgtactcca gcttgtgccc caggatgttg ccgtcctcct tgaagtcgat  14820
gcccttcagc tcgatgcggt tcaccagggt gtcgccctcg aacttcacct cggcgcgggt  14880
cttgtagttg ccgtcgtcct tgaagaagat ggtgcgctcc tggacgtagc cttcgggcat  14940
ggcggacttg aagaagtcgt gctgcttcat gtggtcgggg tagcgggcga agcactgcac  15000
gccccaggac agggtggtca cgagggtggg ccagggcacg ggcagcttgc cggtggtgca  15060
gatgaacttc agggtcagct tgccgtaggt ggcatcgccc tcgccctcgc cggacacgct  15120
gaacttgtgg ccgtttacgt cgccgtccag ctcgaccagg atgggcacca ccccggtgaa  15180
cagctcctcg cccttgctca ccatacaaac ttacaaattt ctctgaagtt gtatcctcag  15240
tacttcaaag aaaatagctt acaccaaatt ttttcttgtt ttcacaaatg ccgaacttgg  15300
ttccttatat aggaaaactc aagggcaaaa atgacacgga aaaatataaa aggataagta  15360
gtgggggata agattccttt gtgataaggt tactttccgc ccttacattt tccaccttac  15420
atgtgtcctc tatgtctctt tcacaatcac cgaccttatc ttcttctttt cattgttgtc  15480
gtcagtgctt acgtcatcaa gattcttttc ttcgcctggt tcttcttttt caatttctac  15540
gtattcttct tcgtattctg gcagtatagg atcttgtatc tgtacattct tcatttttga  15600
acataggttg catatgtgcc gcatattgat ctgcttcttg ctgagcttac ataatacttc  15660
catagttttt cccgtaaaca ttggattctt gatgctacat cttggataat taccttctgg  15720
ctccggttcg ctaggttgaa gagcaattct gacaggatat attggcgggt aaacctaaga  15780
gaaaagagcg tttattagaa taatcggata tttaaaaggg cgtgaaaagg tttatccgtt  15840
cgtccatttg tatgtgcatg ccaaccacag ggttcccctc gggatcaaag tactttgatc  15900
caacccctcc gctgctatag tgcagtcggc ttctgacgtt cagtgcagcc gtcatctgaa  15960
aacgacatgt cgcacaagtc ctaagttacg cgacaggctg ccgccctgcc cttttcctgg  16020
cgttttcttg tcgcgtgttt tagtcgcata aagtagaata cttgcgacta gaaccggaga  16080
cattacgcca tgaacaagag cgccgccgct ggcctgctgg gctatgcccg cgtcagcacc  16140
gacgaccagg acttgaccaa ccaacgggcc gaactgcacg cggccggctg caccaagctg  16200
ttttccgaga agatcaccgg caccaggcgc gaccgcccgg agctggccag gatgcttgac  16260
cacctacgcc ctggcgacgt tgtgacagtg accaggctag accgcctggc ccgcagcacc  16320
cgcgacctac tggacattgc cgagcgcatc caggaggccg gcgcgggcct gcgtagcctg  16380
gcagagccgt gggccgacac caccacgccg gccggccgca tggtgttgac cgtgttcgcc  16440
ggcattgccg agttcgagcg ttccctaatc atcgaccgca cccggagcgg gcgcgaggcc  16500
gccaaggccc gaggcgtgaa gtttggcccc cgccctaccc tcaccccggc acagatcgcg  16560
cacgcccgcg agctgatcga ccaggaaggc cgcaccgtga aagaggcggc tgcactgctt  16620
ggcgtgcatc gctcgaccct gtaccgcgca cttgagcgca gcgaggaagt gacgcccacc  16680
gaggccaggc ggcgcggtgc cttccgtgag gacgcattga ccgaggccga cgccctggcg  16740
gccgccgaga atgaacgcca agaggaacaa gcatgaaacc gcaccaggac ggccaggacg  16800
aaccgttttt cattaccgaa gagatcgagg cggagatgat cgcggccggg tacgtgttcg  16860
agccgcccgc gcacctctca accgtgcggc tgcatgaaat cctggccggt ttgtctgatg  16920
ccaagctggc ggcctggccg gccagcttgg ccgctgaaga aaccgagcgc cgccgtctaa  16980
aaaggtgatg tgtatttgag taaaacagct tgcgtcatgc ggtcgctgcg tatatgatgc  17040
gatgagtaaa taaacaaata cgcaagggga acgcatgaag gttatcgctg tacttaacca  17100
```

-continued

```
gaaaggcggg tcaggcaaga cgaccatcgc aacccatcta gcccgcgccc tgcaactcgc  17160
cggggccgat gttctgttag tcgattccga tccccagggc agtgcccgcg attgggcggc  17220
cgtgcgggaa gatcaaccgc taaccgttgt cggcatcgac cgcccgacga ttgaccgcga  17280
cgtgaaggcc atcggccggc gcgacttcgt agtgatcgac ggagcgcccc aggcggcgga  17340
cttggctgtg tccgcgatca aggcagccga cttcgtgctg attccggtgc agccaagccc  17400
ttacgacata tgggccaccg ccgacctggt ggagctggtt aagcagcgca ttgaggtcac  17460
ggatggaagg ctacaagcgg cctttgtcgt gtcgcgggcg atcaaaggca cgcgcatcgg  17520
cggtgaggtt gccgaggcgc tggccgggta cgagctgccc attcttgagt cccgtatcac  17580
gcagcgcgtg agctacccag gcactgccgc cgccggcaca accgttcttg aatcagaacc  17640
cgagggcgac gctgcccgcg aggtccaggc gctggccgct gaaattaaat caaaactcat  17700
ttgagttaat gaggtaaaga gaaaatgagc aaaagcacaa acacgctaag tgccggccgt  17760
ccgagcgcac gcagcagcaa ggctgcaacg ttggccagcc tggcagacac gccagccatg  17820
aagcgggtca actttcagtt gccggcggag gatcacacca agctgaagat gtacgcggta  17880
cgccaaggca agaccattac cgagctgcta tctgaataca tcgcgcagct accagagtaa  17940
atgagcaaat gaataaatga gtagatgaat tttagcggct aaaggaggcg gcatggaaaa  18000
tcaagaacaa ccaggcaccg acgccgtgga atgccccatg tgtggaggaa cgggcggttg  18060
gccaggcgta agcggctggg ttgtctgccg gccctgcaat ggcactggaa cccccaagcc  18120
cgaggaatcg gcgtgagcgg tcgcaaacca tccggcccgg tacaaatcgg cgcggcgctg  18180
ggtgatgacc tggtggagaa gttgaaggcc gcgcaggccg cccagcggca acgcatcgag  18240
gcagaagcac gccccggtga atcgtggcaa gcggccgctg atcgaatccg caaagaatcc  18300
cggcaaccgc cggcagccgg tgcgccgtcg attaggaagc cgcccaaggg cgacgagcaa  18360
ccagattttt tcgttccgat gctctatgac gtgggcaccc gcgatagtcg cagcatcatg  18420
gacgtggccg ttttccgtct gtcgaagcgt gaccgacgag ctggcgaggt gatccgctac  18480
gagcttccag acgggcacgt agaggtttcc gcagggccgg ccggcatggc cagtgtgtgg  18540
gattacgacc tggtactgat ggcggtttcc catctaaccg aatccatgaa ccgataccgg  18600
gaagggaagg gagacaagcc cggccgcgtg ttccgtccac acgttgcgga cgtactcaag  18660
ttctgccggc gagccgatgg cggaaagcag aaagacgacc tggtagaaac ctgcattcgg  18720
ttaaacacca cgcacgttgc catgcagcgt acgaagaagg ccaagaacgg ccgcctggtg  18780
acggtatccg agggtgaagc cttgattagc cgctacaaga tcgtaaagag cgaaaccggg  18840
cggccggagt acatcgagat cgagctagct gattggatgt accgcgagat cacagaaggc  18900
aagaacccgg acgtgctgac ggttcacccc gattactttt tgatcgatcc cggcatcggc  18960
cgttttctct accgcctggc acgccgcgcc gcaggcaagg cagaagccag atggttgttc  19020
aagacgatct acgaacgcag tggcagcgcc ggagagttca agaagttctg tttcaccgtg  19080
cgcaagctga tcgggtcaaa tgacctgccg gagtacgatt tgaaggagga ggcggggcag  19140
gctggcccga tcctagtcat gcgctaccgc aacctgatcg agggcgaagc atccgccggt  19200
tcctaatgta cggagcagat gctagggcaa attgccctag caggggaaaa aggtcgaaaa  19260
ggactctttc ctgtggatag cacgtacatt gggaacccaa agccgtacat tgggaaccgg  19320
aacccgtaca ttgggaaccc aaagccgtac attgggaacc ggtcacacat gtaagtgact  19380
gatataaaag agaaaaaagg cgatttttcc gcctaaaact ctttaaaact tattaaaact  19440
cttaaaaccc gcctggcctg tgcataactg tctggccagc gcacagccga agtgctgcaa  19500
aaagcgccta cccttcggtc gctgcgctcc ctacgccccg ccgcttcgcg tcggcctatc  19560
gcggccgctg gccgctcaaa aatggctggc ctacggccag gcaatctacc agggcgcgga  19620
caagccgcgc cgtcgccact cgaccgccgg cgcccacatc aaggcaccct gcctcgcgcg  19680
tttcggtgat gacggtgaaa acctctgaca catgcagctc ccggtgacgg tcacagcttg  19740
tctgtaagcg gatgccggga gcagacaagc ccgtcagggc gcgtcagcgg gtgttggcgg  19800
gtgtcggggc gcagccatga cccagtcacg tagcgatagc ggagtgtata ctggcttaac  19860
tatgcggcat cagagcagat tgtactgaga gtgcaccata tgcggtgtga aataccgcac  19920
agatgcgtaa ggagaaaata ccgcatcagg cgctcatccg cttcctcgct cactgactcg  19980
ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg  20040
ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag  20100
gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg cccccctgac  20160
gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga  20220
taccaggcgt ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt  20280
accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc  20340
tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc  20400
cccgttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caacccggta  20460
agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggtat  20520
gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca  20580
gtatttggta tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct  20640
tgatccggca aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt  20700
acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgct  20760
cagtggaacg aaaactcacg ttaagggatt ttggtcatgc attctaggtg attagaaaaa  20820
ctcatcgagc atcaaatgaa actgcaattt attcatatca ggattatcaa taccatattt  20880
ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg aggcagttcc ataggatggc  20940
aagatcctgg tatcggtctg cgattccgac tcgtccaaca tcaatacaac ctattaattt  21000
cccctcgtca aaaataaggt tatcaagtga gaaatcacca tgagtgacga ctgaatccgg  21060
tgagaatggc aaaagtttat gcatttcttt ccagacttgt tcaacaggcc agccattacg  21120
ctcgtcatca aaatcactcg catcaaccaa accgttattc attcgtgatt gcgcctgagc  21180
gagtcgaaat acgcgatcgc tgttaaaagg acaattacaa acaggaatcg aatgcaaccg  21240
gcgcaggaac actgccagcg catcaacaat attttcacct gaatcaggat attcttctaa  21300
tacctggaat gctgttttcc ctgggatcgc agtggtgagt aaccatgcat catcaggagt  21360
acggataaaa tgcttgatgg tcggaagagg cataaattcc gtcagccagt ttagtctgac  21420
catctcatct gtaacatcat tggcaacgct acctttgcca tgtttcagaa acaactctgg  21480
cgcatcgggc ttcccataca atcggtagat tgtcgcacct gattgcccga cattatcgcg  21540
agcccattta tacccatata aatcagcatc catgttggaa tttaatcgcg gccttgagca  21600
agacgtttcc cgttgaatat ggctcataac agaacttatt atttccttcc tcttttctac  21660
```

-continued

```
agtatttaaa gataccccaa gaagctaatt ataacaagac gaactccaat tcactgttcc  21720
ttgcattcta aaaccttaaa taccagaaaa cagctttttc aaagttgttt tcaaagttgg  21780
cgtataacat agtatcgacg gagccgattt tgaaaccgcg gtgatcacag gcagcaacgc  21840
tctgtcatcg ttacaatcaa catgctaccc tccgcgagat catccgtgtt tcaaacccgg  21900
cagcttagtt gccgttcttc cgaatagcat cggtaacatg agcaaagtct gccgccttac  21960
aacggctctc ccgctgacgc cgtcccggac tgatgggctg cctgtatcga gtggtgattt  22020
tgtgccgagc tgccggtcgg ggagctgttg gctggctgg                         22059

SEQ ID NO: 46          moltype = DNA  length = 14022
FEATURE                Location/Qualifiers
source                 1..14022
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 46
tggcaggata tattgtggtg taaacataac ggatccggtc tcaggagaat gaggagaatg  60
tagcagaagg catgttgttg tgactccgag gggttgcctc aaactctatc ttataaccgg  120
cgtggaggca tggaggcagg ggtattttgg tcattttaat agatagtgga aaatgacgtg  180
gaatttactt aaagacgaag tctttgcgac aagggggggc ccacgccgaa tttaatatta  240
ccggcgtggc cccccttat cgcgagtgct ttagcacgag cggtccagat ttaaagtaga  300
aaatttcccg cccactaggg ttaaaggtgt tcacactata aaagcatata cgatgtgatg  360
gtatttgatg gagcgtatat tgtatcaggt atttccgttg gatacgaatt attcgtacga  420
ccctcggtat actaatgagg agtgtcgctg caaccatagt ttgacataat gcaaccatac  480
cttgacataa ttaaagtcat actctgacat aaagcccctt attaaaaatg atgctaaaaa  540
taaattccat tataatcaat tatttaagaa ttaattaaac catagtgtga catagtggct  600
ttttatttgt atagtatgtg tactcaaata ataaaaattt acttatgtat atacgcaact  660
taagaaagcc atctccaaat aaaaatgtat ttctacctgt tccatggcca accgctagca  720
aggaggtgga gatccttcct cctagatcca aaatacggta ccgtcaacaa tctcactcga  780
gagtcgacgt atgtgcaaaa ctaagcattc cgaagccatt gttagccgta tgaataggga  840
aactaaaccc agtgataaga cctgatgttt tcgcttcttt aattacattt ggagatgccc  900
gctatacagg agccgtatag gcagctgaag ataaaaacaa tcgattcgat gtttaagtta  960
atgtaattct gccgaaaagg cagtgagtag taattgaaaa aaccagtctt atatccgggc  1020
tggtttttt gtttttagcg ttataacttt tatttcacta tgtcactgta tggttgcatt  1080
ttatgtcagt gaatggttgc aggtttttg taagttattg atatgcctat ccagtgtatg  1140
tcacgctttg cttgcagcga cacgctacag aggagtctga gcaggacaac tcgcgtagtg  1200
agagttacat gttcgttggg ttcttccgac acggacctga gttggccaac gtcccacctg  1260
aggtctgtgc cccggtgatg agaagtgtgc atctcgttct tgcagctcgt cagtactttc  1320
agaatcatgg cgtgcatggt agaatgaccc ttataacgga cttcgacatg gcaatcgcta  1380
ggtaaatgaa tggaaaaagt ggcaccgagt cggtgctttt tttcccgggg cgcgccgata  1440
tcgagctctc ccggcgcgcc gatatcgagc tcagtgtttg atcgccggcg gtaccgagtg  1500
tacttcaagt cagtgggaaa tcaataaaat gattatttta tgaatatatt tcattgtgca  1560
agtagataga aattacatat gttacataac acacgaaata aacaaaaaaa gacaatccaa  1620
aaacaaacac cccaaaaaaa ataatcactt tagataaact cgtatgagga gaggcacgtt  1680
cagtgactcg acgattcccg agcaaaaaaa gtctccccgt cacacatgta gtgggtgacg  1740
caattatctt taaagtaatc cttctgttga cttgtcattg ataacatcca gtcttcgtca  1800
ggattgcaaa gaattataga agggatccca ccttttattt tcttcttttt tccatattta  1860
gggttgacag tgaaatcaga ctggcaacct attaattgct tccacaatgg gacgaacttg  1920
aaggggatgt cgtcgatgat attataggtg gcgtgttcat cgtagttggt gaaatcgatg  1980
gtaccgttcc aatagttgtg tcgtccgaga cttctagccc aggtggtctt tccggtacga  2040
gttggtccgc agatgtagag gctggggtgt cggattccat tccttccatt gtccttgtta  2100
aatcggccat ccattcaagg tcagattgag cttgttggta tgagacagga tgtatgtaag  2160
tataagcgtc tatgcttaca tggtatagat gggtttccct ccaggagtgt agatcttcgt  2220
ggcagcgaag atctgattct gtgaagggcg acacatacgg ttcaggttgt ggagggaata  2280
atttgttggc tgaatattcc agccattgaa gctttgttgc ccattcatga gggaattctt  2340
ccttgatcat gtcaagatat tcctccttag acgttgcagt ctggataata gttctccatc  2400
gtgcgtcaga tttgcgagga gaaaccttat gatctcggaa atctcctctg gttttaatat  2460
ctccgtcctt tgatatgtaa tcaaggactt gtttagagtt tctagctggc tggatattag  2520
ggtgatttcc ttcaaaatcg aaaaaagaag gatccctaat acaaggtttt ttatcaagct  2580
ggaggagagc atgatagtgg gtagtgccat cttgatgaag ctcagaagca acaccaagga  2640
agaaaataag aaaaggtgtg agtttctccc agagaaactg gaataaatca tctctttgag  2700
atgagcactt gggataggta aggaaaacat atttagattg gagtctgaag ttcttactag  2760
cagaaggcat gttgttgtga ctccgagggg ttgcctcaaa ctctatctta taaccggcgt  2820
ggaggcatgg aggcaggggt attttggtca ttttaataga tagtggaaaa tgacgtggaa  2880
tttacttaaa gacgaagtct ttgcgacaag ggggggccca cgccgaattt aatattaccg  2940
gcgtggcccc cccttatcgc gagtgcttta gcacgagcgg tccagattta aagtagaaaa  3000
tttcccgccc actagggtta aaggtgttca cactataaaa gcatatacga tgtgatggta  3060
tttgatggag cgtatattgt atcaggtatt tccgttggat acgaattatt cgtacgaccc  3120
tcatagttta aactatcagt gggtagctta atggtagaca ctttattagt tagatccgct  3180
gttattctgc acgattgcgc ttgccggtgg aaatttaagg acgagtgtag tgcactcacc  3240
cttagccttc agcgtgacaa gattagggta gcaccggaaa cagcgatttt acggacagat  3300
agtacagtct gagactcagt cactggtccc taaacgggag ccggtacgtc ggcggttcgc  3360
tagcaaggag tggcagacat actgtcccac aaatgaagat ggaatctgta aaagaaaacg  3420
cgtgaaataa tgcgtctgac aaaggttagg tcggctgcct ttaatcaata ccaaagtggt  3480
ccctaccacg atggaaaaac tgtgcagtcg gtttggcttt ttctgacgaa caaataagat  3540
tcgtggccga caggtggggg tccaccatgt gaaggcatct tcagactcca ataatggagc  3600
aatgacgtaa gggcttacga aataagtaag ggtagtttgg gaaatgtcca ctcacccgtc  3660
agtctataaa tacttagccc ctccctcatt gttaagggag caaaatctca gagagatagt  3720
```

-continued

```
cctagagaga gaaagagagc aagtagccta gaagtagtca aggcggcgaa gtattcaggc  3780
acgtggccag gaagaagaaa agccaagacg acgaaaacag gtaagagcta agcttaatgg  3840
gaaagcgtac tgctgatggt tcagagtttg aaagtccaaa gaagaaaaga aaagttggta  3900
gtggtatgac tgataaacgg aaagatggtt ccggcaagct tctttactgc agcttttgcg  3960
ggaaatctca acatgaagtt aggaaactta ttgcagggcc ttctgtttat atctgtgacg  4020
aatgtgttga tctctgcaat gatattattc gagaagagat caaagaagtg gctcctcaca  4080
gagagaggag tgcgttgccg acaccacacg agattagaaa tcatcttgat gattatgtca  4140
ttggtcaaga acaagcaaag aaggtccttg ctgttgcagt gtacaaccat tacaagaggc  4200
tcagaaacgg agatacttct aatggtgtgg aacttggcaa atcaaacatc ctgctcattg  4260
gacccactgg aagtggaaag acactcctag cagagacatt ggcaaggctc ttagacgttc  4320
ctttcactat ggctgatgca actactctta ctgaggctgg ttatgttgga gaagatgtgg  4380
agaatattat tcaaaagtta ttgcagaaat gtgattacga tgtccagaaa gctcagagag  4440
ggatagtcta catagatgag atagacaaga tctctaggaa gagtgataat ccctcgataa  4500
cacgcgacgt ttctggcgaa ggtgttcaac aagctttgtt gaagctcatt gaaggcaccg  4560
tagcggcagt tccaccacaa ggaggaagaa agcatcctca gcaagagttt ctccaagtgg  4620
atacctcaaa aatcttattt atatgtggtg gtgccttcgc tgggcttgac aaagttatct  4680
cacaccgtgt agaaactggt agcgggattg ggtttggtgc aacggtgaag gcaaaatctg  4740
ataaggcttc tgaaggagag cttctggctc aagttgaacc agaggatctt atcaagtttg  4800
gactaattcc tgagttcatc ggtagacttc ctgtcgttgc tacgttgaac gagttaagtg  4860
aggaagccct gattcaaatc ttgaaagagc caaagaatgc tctgacaaag caatatcagg  4920
ctcttttcaa tcttgaggga gttgacttgg aatttcgtga tgaagcgctg gatgctatag  4980
ctaagaaagc aatggcgagg aaaaccggag ctcgtggtct aagatcgatt gttgaagccg  5040
ctttattgga tacaatgtat gatttgccct ccatggaaga tgttgagaaa gtggtcatcg  5100
acgagtctgt gattgatgga cagtcaaagc cattgctcat atatggtaaa cctgaagcac  5160
agcaggcatc aggagaatga atatgaagat gaagatgaaa tatttggtgt gtcaaataaa  5220
aaggttgtgt gcttaagttt gtgtttttt cttggcttgt tgtgttatga atttgtggct  5280
ttttctaata ttaaatgaat gtaacatctc attataatga ataaacaaat gtttctataa  5340
tccattgtga atgttttgtt ggatctcttc tccagcatat aactactgta tgtgctatgg  5400
tatggactat ggaatatgat taaagataag cgctatacag gagtggcaga catactgtcc  5460
cacaaatgaa gatggaatct gtaaaagaaa acgcgtgaaa taatgcgtct gacaaaggtt  5520
aggtcggctg cctttaatca ataccaaagt ggtccctacc acgatggaaa aactgtgcag  5580
tcggtttggc tttttctgac gaacaaataa gattcgtggc cgacaggtgg gggtccacca  5640
tgtgaaggca tcttcagact ccaataatgg agcaatgacg taagggctta cgaaataagt  5700
aagggtagtt tgggaaatgt ccactcaccc gtcagtctat aaatacttag cccctccctc  5760
attgttaagg gagcaaaatc tcagagagat agtcctagag agagaaagag agcaagtagc  5820
ctagaagtag tcaaggcggc gaagtattca ggcacgtggc caggaagaag aaaagccaag  5880
acgacgaaaa caggtaagag ctaagcttaa tgggaaagcg tactgctgat ggttcagagt  5940
ttgaaagtcc aaagaagaaa agaaaagttg gtagtggtat gagttattct ggagaacggg  6000
ataattttgc acctcacatg gcattagttc caatggtgat tgaacaaaca agtagagggg  6060
agcgaagctt tgacatttac tctcgtctat taaaagaaag ggtcatcttc ttaaccgggc  6120
aagttgaaga tcatatggca aatctgattg tagctcagat gctttttctt gaagcagaga  6180
atcctgagaa agatatctat ctctacatta actctcctgg tggtgtaata actgctggaa  6240
tgtctatata tgatacaatg cagtttatca aacctgatgt ttcaacaata tgtatgggtc  6300
aggctgcgtc aatgggagct tttcttttga cggctggcgc aaaaggaaag cggttctgcc  6360
taccgaactc aagggtgatg atccatcaac cacttggagg ataccaaggt caagctaccg  6420
atatagagat tcatgctaga gaaattttga aggtgaaggg tagaatgaat gaactgatgg  6480
cccttcacac tggacaatcg ctcgagcaga ttgagcgtga cacagagaga gacaggtttt  6540
tatccgcacc agaagctgtt gagtatggtt tggttgattc tattcttact catagaaact  6600
gaatatgaag atgaagatga aatatttggt gtgtcaaata aaaaggttgt gtgcttaagt  6660
ttgtgttttt ttcttggctt gttgtgttat gaatttgtgg ctttttctaa tattaaatga  6720
atgtaacatc tcattataat gaataaacaa atgtttctat aatccattgt gaatgttttg  6780
ttggatctct tctccagcat ataactactg tatgtgctat ggtatggact atggaatatg  6840
attaaagata agcgctacag aggagcatct tcattcttaa gatatgaaga taatcttcaa  6900
aaggcccctg ggaatctgaa agaagagaag caggcccatt tatatgggaa agaacaatag  6960
tatttcttat ataggcccat ttaagttgaa aacaatcttc aaaagtccca catcgcttag  7020
ataagaaaac gaagctgagt ttatatacag ctagagtcga agtagtgata atggtgacct  7080
gccgtatagg cagctgaaga tactgacgta agggatgacg cacaatccca ctataattct  7140
gccgaaaagg cagtgagtag tgctttttt tttgcaaaat tttccagatc gatttcttct  7200
tcctctgttc ttcggcgttc aatttctggg gttttctctt cgttttctgt aactgaaacc  7260
taaaatttga cctaaaaaaa atctcaaata atatgattca gtggttttgt acttttcagt  7320
tagttgagtt ttgcagttcc gatgagataa accaatacgc taggtatact tgagaccgga  7380
tcctgacagg atatattggc gggtaaacct aagagaaaag agcgtttatt agaataatcg  7440
gatatttaaa agggcgtgaa aaggtttatc cgttcgtcca tttgtatgtg catgccaacc  7500
acagggttcc cctcgggatc aaagtacttt gatccaaccc ctccgctgct atagtgcagt  7560
cggcttctga cgttcagtgc agccgtcatc tgaaaacgac atgtcgcaca agtcctaagt  7620
tacgcgacag gctgccgccc tgccctttc ctggcgtttt cttgtcgcgt gttttagtcg  7680
cataaagtag aatacttgcg actagaaccg gagacattac gccatgaaca agagcgccgc  7740
cgctggcctg ctgggctatg cccgcgtcag caccgacgac caggacttga ccaaccaacg  7800
ggccgaactg cacgcggccg gctgcaccaa gctgttttcc gagaagatca ccggcaccag  7860
gcgcgaccgc ccggagctgg ccaggatgct tgaccaccta cgccctggcg acgttgtgac  7920
agtgaccagg ctagaccgcc tggcccgcag cacccgcgac ctactggaca ttgccgagcg  7980
catccaggag gccggcgcgg gcctgcgtag cctggcagag ccgtgggccg acaccaccac  8040
gccggccggc cgcatggtgt tgaccgtgtt cgccggcatt gccgagttcg agcgttccct  8100
aatcatcgac cgcacccgga gcgggcgcga ggccgccaag gcccgaggcg tgaagtttgg  8160
cccccgccct accctcaccc cggcacagat cgcgcacgcc cgcgagctga tcgaccagga  8220
aggccgcacc gtgaaagagg cggctgcact gcttggcgtg catcgctcga ccctgtaccg  8280
```

-continued

```
cgcacttgag cgcagcgagg aagtgacgcc caccgaggcc aggcggcgcg gtgccttccg  8340
tgaggacgca ttgaccgagg ccgacgccct ggcggccgcc gagaatgaac gccaagagga  8400
acaagcatga aaccgcacca ggacggccag gacgaaccgt ttttcattac cgaagagatc  8460
gaggcggaga tgatcgcggc cgggtacgtg ttcgagccgc ccgcgcacct ctcaaccgtg  8520
cggctgcatg aaatcctggc cggtttgtct gatgccaagc tggcggcctg gccggccagc  8580
ttggccgctg aagaaaccga gcgccgccgt ctaaaaaggt gatgtgtatt tgagtaaaac  8640
agcttgcgtc atgcggtcgc tgcgtatatg atgcgatgag taaataaaca aatacgcaag  8700
gggaacgcat gaaggttatc gctgtactta accagaaagg cgggtcaggc aagacgacca  8760
tcgcaaccca tctagcccgc gccctgcaac tcgccggggc cgatgttctg ttagtcgatt  8820
ccgatcccca gggcagtgcc cgcgattggg cggccgtgcg ggaagatcaa ccgctaaccg  8880
ttgtcggcat cgaccgcccg acgattgacc gcgacgtgaa ggccatcggc cggcgcgact  8940
tcgtagtgat cgacggagcg ccccaggcgg cggacttggc tgtgtccgcg atcaaggcag  9000
ccgacttcgt gctgattccg gtgcagccaa gcccttacga catatgggcc accgccgacc  9060
tggtggagct ggttaagcag cgcattgagg tcacggatgg aaggctacaa gcggcctttg  9120
tcgtgtcgcg ggcgatcaaa ggcacgcgca tcggcggtga ggttgccgag gcgctggccg  9180
ggtacgagct gcccattctt gagtcccgta tcacgcagcg cgtgagctac ccaggcactg  9240
ccgccgccgg cacaaccgtt cttgaatcag aacccgaggg cgacgctgcc cgcgaggtcc  9300
aggcgctggc cgctgaaatt aaatcaaaac tcatttgagt taatgaggta aagagaaaat  9360
gagcaaaagc acaaacacgc taagtgccgg ccgtccgagc gcacgcagca gcaaggctgc  9420
aacgttggcc agcctggcag acacgccagc catgaagcgg gtcaactttc agttgccggc  9480
ggaggatcac accaagctga agatgtacgc ggtacgccaa ggcaagacca ttaccgagct  9540
gctatctgaa tacatcgcgc agctaccaga gtaaatgagc aaatgaataa atgagtagat  9600
gaattttagc ggctaaagga ggcggcatgg aaaatcaaga acaaccaggc accgacgccg  9660
tggaatgccc catgtgtgga ggaacgggcg gttggccagg cgtaagcggc tgggttgcct  9720
gccggccctg caatggcact ggaaccccca agcccgagga atcggcgtga gcggtcgcaa  9780
accatccggc ccggtacaaa tcggcgcggc gctgggtgat gacctggtgg agaagttgaa  9840
ggccgcgcag gccgcccagc ggcaacgcat cgaggcagaa gcacgccccg gtgaatcgtg  9900
gcaagcggcc gctgatcgaa tccgcaaaga atcccggcaa ccgccggcag ccggtgcgcc  9960
gtcgattagg aagccgccca agggcgacga gcaaccagat tttttcgttc cgatgctcta  10020
tgacgtgggc acccgcgata gtcgcagcat catggacgtg gccgttttcc gtctgtcgaa  10080
gcgtgaccga cgagctggcg aggtgatccg ctacgagctt ccagacgggc acgtagaggt  10140
ttccgcaggg ccggccggca tggccagtgt gtgggattac gacctggtac tgatggcggt  10200
ttcccatcta accgaatcca tgaaccgata ccgggaaggg aagggagaca agcccggccg  10260
cgtgttccgt ccacacgttg cggacgtact caagttctgc cggcgagccg atggcggaaa  10320
gcagaaagac gacctggtag aaacctgcat tcggttaaac accacgcacg ttgccatgca  10380
gcgtacgaag aaggccaaga acggccgcct ggtgacggta tccgagggtg aagccttgat  10440
tagccgctac aagatcgtaa agagcgaaac cgggcggccg gagtacatcg agatcgagct  10500
agctgattgg atgtaccgcg agatcacaga aggcaagaac ccggacgtgc tgacggttca  10560
ccccgattac tttttgatcg atcccggcat cggccgtttt ctctaccgcc tggcacgccg  10620
cgccgcaggc aaggcagaag ccagatggtt gttcaagacg atctacgaac gcagtggcag  10680
cgccggagag ttcaagaagt tctgtttcac cgtgcgcaag ctgatcgggt caaatgacct  10740
gccggagtac gatttgaagg aggaggcggg gcaggctggc ccgatcctag tcatgcgcta  10800
ccgcaacctg atcgagggcg aagcatccgc cggttcctaa tgtacggagc agatgctagg  10860
gcaaattgcc ctagcagggg aaaaaggtcg aaaaggactc tttcctgtgg atagcacgta  10920
cattgggaac ccaaagccgt acattgggaa ccggaacccg tacattggga acccaaagcc  10980
gtacattggg aaccggtcac acatgtaagt gactgatata aaagagaaaa aaggcgattt  11040
ttccgcctaa aactctttaa aacttattaa aactcttaaa acccgcctgg cctgtgcata  11100
actgtctggc cagcgcacag ccgaagtgct gcaaaaagcg cctacccttc ggtcgctgcg  11160
ctccctacgc cccgccgctt cgcgtcggcc tatcgcggcc gctggccgct caaaaatggc  11220
tggcctacgg ccaggcaatc taccagggcg cggacaagcc gcgccgtcgc cactcgaccg  11280
ccggcgccca catcaaggca ccctgcctcg cgcgtttcgg tgatgacggt gaaaacctct  11340
gacacatgca gctcccggtg acggtcacag cttgtctgta agcggatgcc gggagcagac  11400
aagcccgtca gggcgcgtca gcgggtgttg gcgggtgtcg gggcgcagcc atgacccagt  11460
cacgtagcga tagcggagtg tatactggct taactatgcg gcatcagagc agattgtact  11520
gagagtgcac catatgcggt gtgaaatacc gcacagatgc gtaaggagaa aataccgcat  11580
caggcgctca tccgcttcct cgctcactga ctcgctgcgc tcggtcgttc ggctgcggcg  11640
agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc  11700
aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt  11760
gctggcgttt ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag  11820
tcagaggtgg cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc  11880
cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg cctttctccc  11940
ttcgggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cggtgtaggt  12000
cgttcgctcc aagctgggct gtgtgcacga accccccgtt cagcccgacc gctgcgcctt  12060
atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc cactggcagc  12120
agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa  12180
gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg ctctgctgaa  12240
gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg  12300
tagcggtggt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga  12360
agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact cacgttaagg  12420
gattttggtc atgcattcta ggtgattatt tgccgactac cttggtgatc tcgcctttca  12480
cgtagtggac aaattcttcc aactgatctg cgcgcgaggc caagcgatct tcttcttgtc  12540
caagataagc ctgtctagct tcaagtatga cgggctgata ctgggccggc aggcgctcca  12600
ttgcccagtc ggcagcgaca tccttcggcg cgattttgcc ggttactgcg ctgtaccaaa  12660
tgcgggacaa cgtaagcact acatttcgct catcaccagc ccagtcgggc ggcgagttcc  12720
atagcgttaa ggtttcattt agcgcctcaa atagatcctg ttcaggaacc ggatcaaaga  12780
gttcctccgc cgctggacct accaaggcaa cgctatgttc tcttgctttt gtcagcaaga  12840
```

-continued

```
tagccagatc aatgtcgatc gtggctggct cgaagatacc tgcaagaatg tcattgcgct  12900
gccattctcc aaattgcagt tcgcgcttag ctggataacg ccacggaatg atgtcgtcgt  12960
gcacaacaat ggtgacttct acagcgcgga gaatctcgct ctctccaggg gaagccgaag  13020
tttccaaaag gtcgttgatc aaagctcgcc gcgttgtttc atcaagcctt acggtcaccg  13080
taaccagcaa atcaatatca ctgtgtggct tcaggccgcc atccactgcg gagccgtaca  13140
aatgtacggc cagcaacgtc ggttcgagat ggcgctcgat gacgccaact acctctgata  13200
gttgagtcga tacttcggcg atcaccgctt ccctcataat gtttaacttt gttttagggc  13260
gactgccctg ctgcgtaaca tcgttgctgc tccataacat caaacatcga cccacggcgt  13320
aacgcgcttg ctgcttggat gcccgaggca tagactgtac cccaaaaaaa cagtcataac  13380
aagccatgaa aaccgccact gcgccgttac caccgctgcg ttcggtcaag gttctggacc  13440
agttgcgtga gcgcatacgc tacttgcatt acagcttacg aaccgaacag gcttatgtcc  13500
actgggttcg tgccttcatc cgtttccacg gtgtgcgtca cccggcaacc ttgggtagca  13560
gcgaagtcga ggcatttctg tcctggctgg aacagaactt attatttcct tcctcttttc  13620
tacagtattt aaagataccc caagaagcta attataacaa gacgaactcc aattcactgt  13680
tccttgcatt ctaaaacctt aaataccaga aaacagcttt ttcaaagttg ttttcaaagt  13740
tggcgtataa catagtatcg acggagccga ttttgaaacc gcggtgatca caggcagcaa  13800
cgctctgtca tcgttacaat caacatgcta ccctccgcga gatcatccgt gtttcaaacc  13860
cggcagctta gttgccgttc ttccgaatag catcggtaac atgagcaaag tctgccgcct  13920
tacaacggct ctcccgctga cgccgtcccg gactgatggg ctgcctgtat cgagtggtga  13980
ttttgtgccg agctgccggt cggggagctg ttggctggct gg                     14022
```

SEQ ID NO: 47 moltype = DNA length = 25
FEATURE Location/Qualifiers
source 1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 47
gccgacagtg gtcccaaaga tggac 25

SEQ ID NO: 48 moltype = DNA length = 18
FEATURE Location/Qualifiers
source 1..18
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 48
cttgccggtg gtgcagag 18

SEQ ID NO: 49 moltype = DNA length = 21
FEATURE Location/Qualifiers
source 1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 49
ccacccacga ggagcatcgt g 21

SEQ ID NO: 50 moltype = DNA length = 24
FEATURE Location/Qualifiers
source 1..24
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 50
ctgaagttta gaccatgaag aggc 24

SEQ ID NO: 51 moltype = DNA length = 24
FEATURE Location/Qualifiers
source 1..24
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 51
aggggcttta tgtcagagta tgac 24

SEQ ID NO: 52 moltype = DNA length = 24
FEATURE Location/Qualifiers
source 1..24
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 52
gactgctgct ggtattaccg aggg 24

SEQ ID NO: 53 moltype = DNA length = 37
FEATURE Location/Qualifiers
source 1..37
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 53
gcggatctaa ctaataaagt gtctaccatt agtatac 37

-continued

```
SEQ ID NO: 54          moltype = DNA  length = 25
FEATURE                Location/Qualifiers
source                 1..25
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 54
tccgtgactg atgcaaggtt atcgg                                          25

SEQ ID NO: 55          moltype = DNA  length = 24
FEATURE                Location/Qualifiers
source                 1..24
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 55
aaaccagtct tatatccggg ctgg                                           24

SEQ ID NO: 56          moltype = DNA  length = 37
FEATURE                Location/Qualifiers
source                 1..37
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 56
gcggatctaa ctaataaagt gtctaccatt agtatac                             37

SEQ ID NO: 57          moltype = DNA  length = 25
FEATURE                Location/Qualifiers
source                 1..25
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 57
tggcgaacag ttcatacaga gtctc                                          25

SEQ ID NO: 58          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 58
tccagctcga ccaggatgg                                                 19

SEQ ID NO: 59          moltype = DNA  length = 24
FEATURE                Location/Qualifiers
source                 1..24
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 59
tcttcgtcaa catggtggag cacg                                           24

SEQ ID NO: 60          moltype = DNA  length = 24
FEATURE                Location/Qualifiers
source                 1..24
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 60
ctgaagttta gaccatgaag aggc                                           24

SEQ ID NO: 61          moltype = DNA  length = 24
FEATURE                Location/Qualifiers
source                 1..24
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 61
aggggcttta tgtcagagta tgac                                           24

SEQ ID NO: 62          moltype = DNA  length = 18
FEATURE                Location/Qualifiers
source                 1..18
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 62
catttggaga tgttgata                                                  18

SEQ ID NO: 63          moltype = DNA  length = 18
FEATURE                Location/Qualifiers
source                 1..18
```

-continued

```
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 63
catcaacagg agagaaca                                                18

SEQ ID NO: 64           moltype = DNA  length = 18
FEATURE                 Location/Qualifiers
source                  1..18
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 64
actctagagg tgttgata                                                18

SEQ ID NO: 65           moltype = DNA  length = 18
FEATURE                 Location/Qualifiers
source                  1..18
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 65
catcaacaag aggatcca                                                18

SEQ ID NO: 66           moltype = DNA  length = 18
FEATURE                 Location/Qualifiers
source                  1..18
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 66
tcatttggag tgtcgctg                                                18

SEQ ID NO: 67           moltype = DNA  length = 18
FEATURE                 Location/Qualifiers
source                  1..18
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 67
cagcgacatg gagagaac                                                18

SEQ ID NO: 68           moltype = DNA  length = 18
FEATURE                 Location/Qualifiers
source                  1..18
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 68
actctagagg tgtcgctg                                                18

SEQ ID NO: 69           moltype = DNA  length = 18
FEATURE                 Location/Qualifiers
source                  1..18
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 69
cagcgacaag aggatcca                                                18
```

1. A system for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell, comprising:

i) an RNA-guided DNA binding complex or one or more first helper polynucleotides each comprising a sequence encoding a component of the RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combination thereof, optionally the system comprises one or more helper accessory proteins or one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP;

ii) a transposition complex or one or more second helper polynucleotides each comprising a sequence encoding a component of the transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and

iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

2. A nucleic acid composition for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell, comprising:

i) one or more first helper polynucleotides each comprising a sequence encoding a component of an RNA-guided DNA binding complex, wherein the RNA-guided DNA binding complex comprises one or more Cas proteins, a transposase, a crRNA, or any combi-

nation thereof, optionally the nucleic acid composition comprises one or more helper accessory polynucleotides each comprising a sequence encoding at least one of the one or more helper accessory proteins, wherein the one or more helper accessory proteins comprise ClpX and ClpP;

ii) one or more second helper polynucleotides each comprising a sequence encoding a component of a transposition complex, wherein the transposition complex comprises one or more transposases and one or more transposons; and

iii) a donor polynucleotide comprising a cargo sequence flanked by a first transposon end sequence (RE) on the 5' end of the cargo sequence and a second transposon end sequence (LE) on the 3' end of the cargo sequence, wherein the donor polynucleotide is comprised within a first autonomous replicon.

3. (canceled)

4. The system of claim **1**, wherein the first autonomous replicon is derived from a geminivirus.

5. (canceled)

6. The system of claim **1**, wherein the donor polynucleotide comprised within the first autonomous replicon comprises, from 5' to 3': a first long intergenic region (LIR), the RE, the cargo sequence, the LE, a short intergenic region (SIR), a sequence encoding RepA, and a second LIR.

7. (canceled)

8. The system of claim **6**, wherein:

the first and/or second LIR comprise the sequence of SEQ ID NO: 1 or a sequence having one, two, or three mismatches relative to the sequence of SEQ ID NO: 1;

the SIR comprises the sequence of SEQ ID NO: 2 or a sequence having one, two, or three mismatches relative to the sequence of SEQ ID NO: 2; and

the sequence encoding RepA comprises the sequence of SEQ ID NO: 3 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 3.

9. (canceled)

10. (canceled)

11. The system of claim **1**, wherein the LE comprises the sequence of any one of SEQ ID NOs: 4-5 or a sequence having one, two or three mismatches relative to the sequence of any one of SEQ ID NOs: 4-5 and wherein the RE comprises the sequence of any one of SEQ ID NOs: 6-7 or a sequence having one, two or three mismatches relative to the sequence of any one of SEQ ID NOs: 6-7.

12. The system of claim **1**, wherein the cargo sequence is 0.2 to 1000 kb in length, optionally 200 to 1200 bp in length.

13. (canceled)

14. (canceled)

15. (canceled)

16. The system of claim **1**, wherein the one or more Cas proteins comprise a Cas6 protein, a Cas7 protein, and a Cas8 protein.

17. The system of claim **16**, wherein:

the Cas6 protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 8-9;

the Cas7 protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 10-11; and/or

the Cas8 protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 12-13.

18. The system of claim **1**, wherein the transposase of the RNA-guided DNA binding complex comprises a TniQ protein, optionally the TniQ protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 14-15.

19. The system of claim **1**, wherein the one or more transposases of the transposition complex comprise: a TnsA transposase, a TnsB transposase, and a TnsC protein; or a TnsAB fusion protein and a TnsC protein.

20. (canceled)

21. The system of claim **19**, wherein:

the TnsAB fusion protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 16-17; and/or

the TnsC protein comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 18-19.

22. (canceled)

23. (canceled)

24. The system of claim **1**, wherein:

the ClpX comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 21; and/or

the ClpP comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 20.

25.-**28**. (canceled)

29. The system of claim **1**, wherein the cargo sequence is capable of being integrated at an integration site following binding of the RNA-guided DNA binding complex to a search target sequence on a first strand of the double stranded target sequence, wherein the integration site is about 48 to 52 base pairs downstream of the double stranded target sequence.

30.-**38**. (canceled)

39. The system of claim **1**, comprising:

at least three first helper polynucleotides each comprising a sequence encoding a Cas protein, wherein the sequence encoding the Cas protein is operably linked to a pCmY promoter and a tHSP terminator;

a first helper polynucleotide comprising a sequence encoding a transposase protein, wherein the sequence encoding the transposase protein is operably linked to a pCmY promoter and a tHSP terminator;

a first helper polynucleotide comprising a sequence encoding a crRNA, wherein the sequence encoding the crRNA is operably linked to a pU6 promoter and a tU6 terminator;

at least two second helper polynucleotides each comprising a sequence encoding a transposase, wherein the sequence encoding a transposase is operably linked to a pCmY promoter and a tHSP terminator; and

at least two helper accessory polynucleotides, wherein the sequence encoding at least one of the one or more helper accessory proteins is operably linked to a pCmY promoter and a tHSP terminator.

40. (canceled)

41. The system of claim **1**, wherein the sequence encoding the component of the RNA-guided DNA binding complex, the sequence encoding the component of the transposition

complex, the sequence encoding at least one of the one or more helper accessory proteins, or any combination thereof, is codon optimized for expression in the plant cell.

42. The system of claim **1**, wherein:

the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas6 protein, optionally the sequence encoding the Cas6 protein comprises the nucleotide sequence of any one of SEQ ID NOs: 24-25 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 24-25;

the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas7 protein, optionally the sequence encoding the Cas7 protein comprises the nucleotide sequence of any one of SEQ ID NOs: 26-27 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 26-27;

the sequence encoding the component of the RNA-guided DNA binding complex encodes a Cas8 protein, optionally the sequence encoding the Cas8 protein comprises the nucleotide sequence of any one of SEQ ID NOs: 28-29 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 28-29;

the sequence encoding the component of the RNA-guided DNA binding complex encodes a TniQ protein, optionally the sequence encoding the TniQ protein comprises the nucleotide sequence of any one of SEQ ID NOs: 30-31 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 30-31;

the sequence encoding the component of the transposition complex encodes a TnsAB fusion protein, optionally the sequence encoding the TnsAB fusion protein comprises the nucleotide sequence of any one of SEQ ID NOs: 32-33 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 32-33;

the sequence encoding the component of the transposition complex encodes a TnsC protein, optionally the sequence encoding the TnsC protein comprises the nucleotide sequence of any one of SEQ ID NOs: 34-35 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to any one of SEQ ID NOs: 34-35; and/or

the sequence encoding ClpX comprises the nucleotide sequence of SEQ ID NO: 37 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 37 and the sequence encoding ClpP comprises the nucleotide sequence of SEQ ID NO: 36 or a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to SEQ ID NO: 36.

43.-**52**. (canceled)

53. The system of claim **1**, wherein the one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide are comprised within one or more vectors, wherein the one or more vectors comprise an *Agrobacterium tumefaciens* Ti vector, and the one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide are comprised within a T-DNA region of the *Agrobacterium tumefaciens* Ti vector.

54. (canceled)

55. The system of claim **53**, wherein the T-DNA region comprising the one or more first helper polynucleotides, the one or more second helper polynucleotides, and/or the donor polynucleotide comprises the sequence of any one of SEQ ID NOs: 38-46.

56. A method for integration of a nucleic acid sequence into a double-stranded target sequence of a genome of a plant cell or a target plasmid in a plant cell, comprising: contacting the plant cell with the system of claim **1**, wherein the cargo sequence is integrated at an integration site in the genome of the plant cell or at a target site of the target plasmid upon expression of the RNA-guided DNA binding complex and the transposition complex in the plant cell.

57.-**73**. (canceled)

* * * * *