R4 GATEWAY Binary Vector (R4pGWB)

1. R4pGWB vector consists of 4 kind of resistance marker seriese
   R4pGWB4xx (Pnos:NPTII), kanamycin resistance
   R4pGWB5xx (Pnos:HPT), hygromycin resistance
   R4pGWB6xx (Pnos:bar), BASTA resistance
   R4pGWB7xx (Pnos:GPT), tunicamycin resistance
2. C-terminal tags are automatically fused subsequent to the LR reaction.
3. Please read the GATEWAY instruction manual (Invitrogen) for procedures,   etc.

   Please note that our R4pGWB seriese are available only for basic research.   If you plan to use these R4pGWBs for commercial research, please contact Invitrogen about licencing of GATEWAY^{TM}.
   Please do not distribute to other researchers without permission.

Tsuyoshi Nakagawa
Department of Molecular and Functional Genomics
Interdisciplinary Center for Research
Organization of Research
Shimane University
Matsue 690, Japan
tnakagaw@life.shimane-u.ac.jp
(July 26, 2002)

The R4pGWB Series

"n" indicate selection markers, 4 (401 etc, kanamycin resistance)、5 (501 etc, hygromycin resistance)、6 (601 etc, BASTA resistance), 7 (701 etc, tunicamycin resistance).

(1) R4pGWBn01 : [--R4-CmR-*ccd*B-R2 --]

(4) R4pGWBn04 : [--R4- CmR-*ccd*B-R2–sGFP--]

(7) R4pGWBn07 : [--R4- CmR-*ccd*B-R2-6xHis--]

(10) R4pGWBn10 : [--R4- CmR-*ccd*B-R2-FLAG--]

(13) R4pGWBn13 : [--R4- CmR-*ccd*B-R2-3xHA--]

(16) R4pGWBn16 : [--R4- CmR-*ccd*B-R2-4xMyc--]

(19) R4pGWBn19 : [--R4- CmR-*ccd*B-R2-10xMyc--]

(22) R4pGWBn22 : [--R4- CmR-*ccd*B-R2-GST--]

(25) R4pGWBn25 : [--R4- CmR-*ccd*B-R2-T7--]

(28) R4pGWBn28 : [--R4- CmR-*ccd*B-R2-TAP--]

(33) R4pGWBn33 : [--R4- CmR-*ccd*B-R2-GUS--]

(35) R4pGWBn35 : [--R4- CmR-*ccd*B-R2-LUC--]

(40) R4pGWBn40 : [--R4- CmR-*ccd*B-R2-EYFP--]

(43) R4pGWBn43 : [--R4- CmR-*ccd*B-R2-ECFP--]

(50) R4pGWBn50 : [--R4- CmR-*ccd*B-R2-G3GFP--]

(53) R4pGWBn53 : [--R4- CmR-*ccd*B-R2-mRFP--]

(59) R4pGWBn59 : [--R4- CmR-*ccd*B-R2-tagRFP--]

Notes:

1. DNA sequences enclosed in [ ] are introduced into the *Hin*dIII-*Sac*I site of the binary vector pPZP221 [Hajdukiewicz et al. (1994) Plant Mol. Biol. **25**: 989-994; see next page].　R = the *att*R site.

Construction of pGWBs

The binary vector used as the starting plasmid for the construction of R4pGWBs was pPZP221 [Hajdukiewicz et al. (1994) Plant Mol. Biol. **25**: 989-994]. The aacC1 selection marker of pPZP221 was replaced with Pnos:NPTII:Tnos (R4pGWB4xx, kanamycin resistance), Pnos:HPT:Tnos (R4pGWB5xx, hygromycin resistance), Pnos:bar:Tnos (R4pGWB6xx, BASTA resistance), or Pnos:GPT:Tnos (R4pGWB7xx, tunicamycin resistance). The *Hin*dIII-*Sac*I region in pPZP221 was replaced by the GATEWAY cassette and tag in the R4pGWB as shown below.

```
RB-(HindIII)-GATEWAY-(SacI)-Tnos-[Tnos-marker-Pnos]-LB
```

The DNA fragment diagrammed above is present between LB and RB of binary vector. Note that the marker unit is in reverse orientation.

Antibiotics can be used for selection in bacteria are as below.
R4pGWBs: spectinomycin, chloramphenicol
recombinant R4pGWB (after LR reaction) :   spectinomycin.

The princple behind the GATEWAY technology and the R4pGWB vector series could be generalized as follow.
(1) Subsequent to the LR reaction with promoter and CDS Entry Clones and a R4pGWB vector, the *att*R4-(CmR, *ccd*B)-*att*R2 cassette in the R4pGWB will be replaced by the *att*B4-(Promoter)-*att*B1-(CDS)-*att*B2.
(2) The tag will be fused automatically in-frame to CDS, if the primer was designed according to the GATEWAY system (see instruction manual by Invitrogen). Note that the peptide encoded by the *att*B and linker region will be included in your construct, linking the tag to *Your Clone*.
(3) N indicates the nucleotide (and X the amino acid) that shold be included in *Your Clone*.

Important Note
(1) Linearize R4pGWBs by digestion *Swa*I or *Sal*I (*Sal*I is not available for R4pGWB6xx) before performing the LR reaction. The efficiency of the LR reaction depends on the topology of the plasmids in the following order (see instruction manual of Invitrogen).

Either or both plasmids linear > both plasmids relaxed >> both plasmids supercoiled

The *Swa*I linker was introduced into *Not*I site (of GATEWAY cassette just downstream of *att*R1) in all ImpGWBs.   This *Xho*I site is unique.

```
                         SwaI
      RB-(HindIII)-GATEWAY-(SacI)-Tnos-[Tnos-marker-Pnos]
```

(2) Select with spectinomycin (100 ug/ml) after the LR reaction.

(3) Use the *Escherichia coli* strain DH5alpha for cloning the product of the LR (BP) reactions, since negative selection with *ccd*B does not work in the *E.coli* which has F' episome (see instruction manual of Invitrogen).

(4) Sometimes, *E.coli* with both the Entry Clone and recombinant R4pGWB are obtained subsequent to the cloning.   I recommend that only the *E. coli* colonies with only the recombinant R4pGWB be used.

R4pGWBn01 : [--R4-ccdB-R2 --]

(AAGCTT) GTGGATCCCCCATC
<u>ACA ACT TTG TAT AGA AAA --(CmR,*ccd*B)--TTC TTG TAC AAA GTG GTT</u> GAT AAC AGC
GCT <u>TA(G</u> AGCTC)

(1) The *Hin*dIII (AAGCTT) and *Sac*I (GAGCTC) sites are shown
(2) The *att*R4-CmR-*ccd*B-*att*R2 cassette is underlined.

After the LR reaction, the following sequence will be obtained.

(AAGCTT) GTGGATCCCCCATC
<u>ACA ACT TTG TAT AGA AAA GTT G</u>NN (promoter) N<u>CA AGT TTG TAC AAA AAA GCA GGC</u>
T<u>NN</u> (CDS) N<u>AC CCA GCT TTC TTG TAC AAA GTG GT</u>T GAT AAC AGC GCT <u>TA(G</u> AGCTC)

(1) The *att*B4 and *att*B2 sequences are underlined.
(2) The in-frame termination codon TAG (double-underlined) is present downstream of
   *att*B2.

R4pGWBn04 : [--R1-ccdB-R2–sGFP--]

(AAGCTT) GTGGATCCCCCATC
<u>ACA ACT TTG TAT AGA AAA</u> –-(CmR, *ccd*B)-- <u>TTC TTG TAC AAA GTG GTG ATC</u> <u>ATG</u>
of GFP –-(GAGCTC)

(1) The initiation codon ATG (double-underlined) is from the GFP coding sequence.

After the LR reaction, the following sequence will be obtained.

(AAGCTT) GTGGATCCCCCATC
<u>ACA ACT TTG TAT AGA AAA GTT GNN</u> (promoter) <u>NCA AGT TTG TAC AAA AAA GCA GGC</u>
<u>TNN</u> (CDS) <u>NAC CCA GCT TTC TTG TAC AAA GTG GTG</u> ATC <u>ATG</u> of sGFP –-(GAGCTC)
         (X) (P) (A) (F) (L) (Y) (K) (V) (V) (I) M

(1) In the case of fusion with CDS and sGFP, the peptide sequence made up of the amino acid in ( ) will link CDS and GFP.
(2) N indicates the nucleotide (and X the amino acid) that should be included in Your Clone, if C-terminal GUS fusion is to be made.
(3) The sequences from the termination codon of *GFP* to *Sac*I site (see below) was generated by the follwing steps: (i) *Not*I digest, (ii) fill-in, and (iii) *Sac*I linker (8-mer) ligation.
   <u>TAA</u>AGCGGCCC(GAGCTC)

R4pGWBn07, n10, n13, n16, n19, n22, n25, n28, n33, n35, n40, n43, n50, n53, n59
(no promoter or 35S promoter, C-tag)

```
(AAGCTT) GTGGATCCCCCATC
ACA ACT TTG TAT AGA AAA --(CmR,ccdB)--TTC TTG TAC AAA GTG GTT GAT AAC AGC
tag GCT TA(G AGCTC)
```

(1) The *Hin*dIII (AAGCTT) and *Sac*I (GAGCTC) sites are shown.
(2) The *att*R4-CmR-*ccd*B-*att*R2 cassette is underlined.
(3) The sequence of each tag is indicated in the "Tags" section.

After LR reaction, the following sequence will be obtained.


```
(AAGCTT) GTGGATCCCCCATC
ACA ACT TTG TAT AGA AAA GTT GNN (promoter) NCA AGT TTG TAC AAA AAA GCA GGC
TNN (CDS) NAC CCA GCT TTC TTG TAC AAA GTG GTT GAT AAC AGC tag GCT TA(G AGCTC)
         M - X   P   A   F   L   Y   K   V   V   D   N   S  tag  A   *
```

(1) Include an initiation codon ATG in *Your Clone*.   Protein fusion with the C-terminal
    tag will be linked by the peptide sequence derived from the *att*B2 and linker region
    as shown in the above sequence.
(2) Translation will stop at the termination codon TAG (double-underlined) after the tag.

Tags

Hyphenations in the amino acid sequences are used for proper spacing.

R4pGWBn04   (sGFP)   720bp

The sequence of sGFP (ATG to TAA) is as described by Chiu et al., (1996) Current Biol., **6**: 325-330 .

Note that the C-terminal K is converted to I in pGWB6 (35S promoter, N-GFP).   (see the section on pGWB6)

```
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
 M— V- S— K— G— E— E— L— F— T— G— V— V— P— I— L— V— E— L— D- G— D- V-
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
 N- G- H- K- F- S- V- S- G- E- G- E- G- D- A- T- Y- G- K- L- T- L- K-
TTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCACCTACGGCGTG
 F- I- C- T- T- G- K- L- P- V- P- W- P- T- L- V- T- T- F- T- Y- G- V-
CAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC
 Q- C- F- S- R- Y- P- D- H- M- K- Q- H- D- F- F- K- S- A- M- P- E- G-
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTC
 Y- V- Q- E- R- T- I- F- F- K- D- D- G- N- Y- K- T- R- A- E- V- K- F-
GAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
 E- G- D- T- L- V- N- R- I- E- L- K- G- I- D- F- K- E- D- G- N- I- L-
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC
 G- H- K- L- E- Y- N- Y- N- S- H- N- V- Y- I- M- A- D- K- Q- K- N- G-
ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
 I- K- V- N- F- K- I- R- H- N- I- E- D- G- S- V- Q- L- A- D- H- Y- Q-
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCC
 Q- N- T- P- I- G- D- G- P- V- L- L- P- D- N- H- Y- L- S- T- Q- S- A-
CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
 L- S- K- D- P- N- E- K- R- D- H- M- V- L- L- E- F- V- T- A- A- G- I
ACTCACGGCATGGACGAGCTGTACAAGTAA
 T- H- G- M- D- E- L- Y- K- *
```

R4pGWBn07 (6xHis) 18bp

```
CATCATCATCATCATCAT
 H- H- H- H- H- H-
```

R4pGWBn10 (FLAG) 24bp

```
GACTACAAGGATGACGATGACAAG
 D- Y- K- D- D- D- D- K-
```

R4pGWBn13 (3xHA) 120bp

Please consult   http://pingu.salk.edu/users/forsburg/ for further information

```
GGGTTAATTAACATCTTTTACCCATACGATGTTCCTGACTATGCGGGCTATCCCTATGACGTCCCGGAC
 G- L- I- N- I- F- Y- P- Y- D- V- P- D- Y- A- G- Y- P- Y- D- V- P- D-
TATGCAGGATCCTATCCATATGACGTTCCAGATTACGCTGCTCAGTGCAGC
 Y- A- G- S- Y- P- Y- D- V- P- D- Y- A- A- Q- C- S-
```

(1) The three HA units are underlined.

(2) The 54th nucleotide is a G in the reported sequence, but it is a C in our vectors. This does not affect the identity of the amino acid, however [CCG (P) to CCC (P)].

R4pGWBn16 (4xMyc) 171bp

Please consult   http://pingu.salk.edu/users/forsburg/ for further information

```
GGGTTAATTAACGGTGAACAAAAGCTAATCTCCGAGGAAGACTTGAACGGTGAACAAAAATTAATCTCA
 G- L- I- N- G- E- Q- K- L- I- S- E- E- D- L- N- G- E- Q- K- L- I- S-
GAAGAAGACTTGAACGGACTCGACGGTGAACAAAAGTTGATTTCTGAAGAAGATTTGAACGGTGAACAA
 E- E- D- L- N- G- L- D- G- E- Q- K- L- I- S- E- E- D- L- N- G- E- Q-
AAGCTAATCTCCGAGGAAGACTTGAACGGTAGC
 K- L- I- S- E- E- D- L- N- G —S
```

The four Myc units aare underlined.

R4pGWBn19 (10xMyc)   402bp

```
GGGTTAATTAACGGTGAACAAAAGCTAATCTCCGAGGAAGACTTGAACGGTGAACAAAAATTAATCTCA
 G- L- I- N- G- E- Q- K- L- I- S- E- E- D- L- N- G- E- Q- K- L- I- S-
GAAGAAGACTTGAACGGACTCGACGGTGAACAAAAGTTGATTTCTGAAGAAGATTTGAACGGTGAACAA
 E- E- D- L- N- G- L- D- G- E- Q- K- L- I- S- E- E- D- L- N- G- E- Q-
AAGCTAATCTCCGAGGAAGACTTGAACGGTGAACAAAAATTAATCTCAGAAGAAGACTTGAACGGACTC
 K- L- I- S- E- E- D- L- N- G- E- Q- K- L- I- S- E- E- D- L- N- G- L-
GACGGTGAACAAAAGTTGATTTCTGAAGAAGATTTGAACGGTGAACAAAAGCTAATCTCCGAGGAAGAC
 D- G- E- Q- K- L- I- S- E- E- D- L- N- G- E- Q- K- L- I- S- E- E- D-
TTGAACGGTGAACAAAAATTAATCTCAGAAGAAGACTTGAACGGACTCGACGGTGAACAAAAGTTGATT
 L- N- G- E- Q- K- L- I- S- E- E- D- L- N- G- L- D- G- E- Q- K- L- I-
TCTGAAGAAGATTTGAACGGTGAACAAAAGCTAATCTCCGAGGAAGACTTGAACGGT
 S- E- E- D- L- N- G- E- Q- K- L- I- S- E- E- D- L- N- G
```

(1) The ten Myc units are underlined.

R4pGWBn22 (GST)   678bp

The full-length *GST* sequence was derived from pGEX2T

```
ATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATAT
 M- S- P- I- L- G- Y- W- K- I- K- G- L- V- Q- P- T- R- L- L- L- E- Y-
CTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTT
 L- E- E- K- Y- E- E- H- L- Y- E- R- D- E- G- D- K- W- R- N- K- K- F-
GAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATG
 E- L- G- L- E- F- P- N- L- P- Y- Y- I- D- G- D- V- K- L- T- Q- S- M-
GCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATT
 A- I- I- R- Y- I- A- D- K- H- N- M- L- G- G- C- P- K- E- R- A- E- I-
TCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTT
 S- M- L- E- G- A- V- L- D- I- R- Y- G- V- S- R- I- A- Y- S- K- D- F-
GAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGT
 E- T- L- K- V- D- F- L- S- K- L- P- E- M- L- K- M- F- E- D- R- L- C-
CATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTT
 H- K- T- Y- L- N- G- D- H- V- T- H- P- D- F- M- L- Y- D- A- L- D- V-
GTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAA
 V- L- Y- M- D- P- M- C- L- D- A- F- P- K- L- V- C- F- K- K- R- I- E-
GCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAA
 A- I- P- Q- I- D- K- Y- L- K- S- S- K- Y- I- A- W- P- L- Q- G- W- Q-
GCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGTTCCGCGTGGATCC
 A- T- F- G- G- G- D- H- P- P- K- S- D- L- V- P- R- G- S-
```


R4pGWBn25   (T7)   33bp

```
ATGGCTAGCATGACTGGTGGACAGCAAATGGGT
 M- A- S- M- T- G- G- Q- Q- M- G
```

R4pGWBn28 **(TAP)** 546bp

```
AAGAGAAGATGGAAAAAGAATTTCATAGCCGTCTCAGCAGCCAACCGCTTTAAGAAAATCTCATCCTCC
K- R- R- W- K- K- N- F- I- A- V- S- A- A- N- R- F- K- K- I- S- S- S-
GGGGCACTTGATTATGATATTCCAACTACTGCTAGCGAGAATTTGTATTTTCAGGGTGAGCTCAAAACC
G- A- L- D- Y- D- I- P- T- T- A- S- E- N- L- Y- F- Q- G- E- L- K- T-
GCGGCTCTTGCGCAACACGATGAAGCCGTGGACAACAAATTCAACAAAGAACAACAAAACGCGTTCTAT
A- A- L- A- Q- H- D- E- A- V- D- N- K- F- N- K- E- Q- Q- N- A- F- Y
GAGATCTTACATTTACCTAACTTAAACGAAGAACAACGAAACGCCTTCATCCAAAGTTTAAAAGATGAC
E- I- L- H- L- P- N- L- N- E- E- Q- R- N- A- F- I- Q- S- L- K- D- D-
CCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCTAAAAAGCTAAATGATGCTCAGGCGCCGAAAGTAGAC
P- S- Q- S- A- N- L- L- A- E- A- K- K- L- N- D- A- Q- A- P- K- V- D-
AACAAATTCAACAAAGAACAACAAAACGCGTTCTATGAGATCTTACATTTACCTAACTTAAACGAAGAA
N- K- F- N- K- E- Q- Q- N- A- F- Y- E- I- L- H- L- P- N- L- N- E- E-
CAACGAAACGCCTTCATCCAAAGTTTAAAAGATGACCCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCT
Q- R- N- A- F- I- Q- S- L- K- D- D- P- S- Q- S- A- N- L- L- A- E- A-
AAAAAGCTAAATGGTGCTCAGGCGCCGAAAGTAGACGCGAATTCCGCGGGGAAGTCAACCTGA
K- K- L- N- G- A- Q- A- P- K- V- D- A- N- S- A- G- K- S- T- *
```

(1) The CBP sequence is underlined.
(2) TEV site is underlined in bold.
(3) The protein A sequence is double-underlined.

R4pGWBn33 (GUS)   1812bp

ATGTTACGTCCTGTAGAAACCCCAACCCGTGAAATCAAAAAACTCGACGGCCTGTGGGCATTCAGTCTG
 M L R P V E T P T R E I K K L D G L W A F S L

GATCGCGAAAACTGTGGAATTGATCAGCGTTGGTGGGAAAGCGCGTTACAAGAAAGCCGGGCAATTGCT
 D R E N C G I D Q R W W E S A L Q E S R A I A

GTGCCAGGCAGTTTTAACGATCAGTTCGCCGATGCAGATATTCGTAATTATGCGGGCAACGTCTGGTAT
 V P G S F N D Q F A D A D I R N Y A G N V W Y

CAGCGCGAAGTCTTTATACCGAAAGGTTGGGCAGGCCAGCGTATCGTGCTGCGTTTCGATGCGGTCACT
 Q R E V F I P K G W A G Q R I V L R F D A V T

CATTACGGCAAAGTGTGGGTCAATAATCAGGAAGTGATGGAGCATCAGGGCGGCTATACGCCATTTGAA
 H Y G K V W V N N Q E V M E H Q G G Y T P F E

GCCGATGTCACGCCGTATGTTATTGCCGGGAAAAGTGTACGTATCACCGTTTGTGTGAACAACGAACTG
 A D V T P Y V I A G K S V R I T V C V N N E L

AACTGGCAGACTATCCCGCCGGGAATGGTGATTACCGACGAAAACGGCAAGAAAAAGCAGTCTTACTTC
 N W Q T I P P G M V I T D E N G K K K Q S Y F

CATGATTTCTTTAACTATGCCGGAATCCATCGCAGCGTAATGCTCTACACCACGCCGAACACCTGGGTG
 H D F F N Y A G I H R S V M L Y T T P N T W V

GACGATATCACCGTGGTGACGCATGTCGCGCAAGACTGTAACCACGCGTCTGTTGACTGGCAGGTGGTG
 D D I T V V T H V A Q D C N H A S V D W Q V V

GCCAATGGTGATGTCAGCGTTGAACTGCGTGATGCGGATCAACAGGTGGTTGCAACTGGACAAGGCACT
 A N G D V S V E L R D A D Q Q V V A T G Q G T

AGCGGGACTTTGCAAGTGGTGAATCCGCACCTCTGGCAACCGGGTGAAGGTTATCTCTATGAACTGTGC
 S G T L Q V V N P H L W Q P G E G Y L Y E L C

GTCACAGCCAAAAGCCAGACAGAGTGTGATATCTACCCGCTTCGCGTCGGCATCCGGTCAGTGGCAGTG
 V T A K S Q T E C D I Y P L R V G I R S V A V

AAGGGCGAACAGTTCCTGATTAACCACAAACCGTTCTACTTTACTGGCTTTGGTCGTCATGAAGATGCG
 K G E Q F L I N H K P F Y F T G F G R H E D A

GACTTGCGTGGCAAAGGATTCGATAACGTGCTGATGGTGCACGACCACGCATTAATGGACTGGATTGGG
 D L R G K G F D N V L M V H D H A L M D W I G

GCCAACTCCTACCGTACCTCGCATTACCCTTACGCTGAAGAGATGCTCGACTGGGCAGATGAACATGGC
 A N S Y R T S H Y P Y A E E M L D W A D E H G

ATCGTGGTGATTGATGAAACTGCTGCTGTCGGCTTTAACCTCTCTTTAGGCATTGGTTTCGAAGCGGGC
 I V V I D E T A A V G F N L S L G I G F E A G

AACAAGCCGAAAGAACTGTACAGCGAAGAGGCAGTCAACGGGGAAACTCAGCAAGCGCACTTACAGGCG
 N K P K E L Y S E E A V N G E T Q Q A H L Q A

ATTAAAGAGCTGATAGCGCGTGACAAAAACCACCCAAGCGTGGTGATGTGGAGTATTGCCAACGAACCG
 I K E L I A R D K N H P S V V M W S I A N E P

GATACCCGTCCGCAAGGTGCACGGGAATATTTCGCGCCACTGGCGGAAGCAACGCGTAAACTCGACCCG
 D T R P Q G A R E Y F A P L A E A T R K L D P

ACGCGTCCGATCACCTGCGTCAATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCTCTTT

```
 T  R  P  I  T  C  V  N  V  M  F  C  D  A  H  T  D  T  I  S  D  L  F
GATGTGCTGTGCCTGAACCGTTATTACGGATGGTATGTCCAAAGCGGCGATTTGGAAACGGCAGAGAAG
 D  V  L  C  L  N  R  Y  Y  G  W  Y  V  Q  S  G  D  L  E  T  A  E  K
GTACTGGAAAAAGAACTTCTGGCCTGGCAGGAGAAACTGCATCAGCCGATTATCATCACCGAATACGGC
 V  L  E  K  E  L  L  A  W  Q  E  K  L  H  Q  P  I  I  I  T  E  Y  G
GTGGATACGTTAGCCGGGCTGCACTCAATGTACACCGACATGTGGAGTGAAGAGTATCAGTGTGCATGG
 V  D  T  L  A  G  L  H  S  M  Y  T  D  M  W  S  E  E  Y  Q  C  A  W
CTGGATATGTATCACCGCGTCTTTGATCGCGTCAGCGCCGTCGTCGGTGAACAGGTATGGAATTTCGCC
 L  D  M  Y  H  R  V  F  D  R  V  S  A  V  V  G  E  Q  V  W  N  F  A
GATTTTGCGACCTCGCAAGGCATATTGCGCGTTGGCGGTAACAAGAAAGGGATCTTCACTCGCGACCGC
 D  F  A  T  S  Q  G  I  L  R  V  G  G  N  K  K  G  I  F  T  R  D  R
AAACCGAAGTCGGCGGCTTTTCTGCTGCAAAAACGCTGGACTGGCATGAACTTCGGTGAAAAACCGCAG
 K  P  K  S  A  A  F  L  L  Q  K  R  W  T  G  M  N  F  G  E  K  P  Q
CAGGGAGGCAAACAATGA
 Q  G  G  K  Q  *
```

R4pGWBn35 (LUC)   1653bp

```
ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCTGGA
 M  E  D  A  K  N  I  K  K  G  P  A  P  F  Y  P  L  E  D  G  T  A  G

GAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGCACAT
 E  Q  L  H  K  A  M  K  R  Y  A  L  V  P  G  T  I  A  F  T  D  A  H

ATCGAGGTGGACATCACTTACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGA
 I  E  V  D  I  T  Y  A  E  Y  F  E  M  S  V  R  L  A  E  A  M  K  R

TATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTG
 Y  G  L  N  T  N  H  R  I  V  V  C  S  E  N  S  L  Q  F  F  M  P  V

TTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTC
 L  G  A  L  F  I  G  V  A  V  A  P  A  N  D  I  Y  N  E  R  E  L  L

AACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAAC
 N  S  M  G  I  S  Q  P  T  V  V  F  V  S  K  K  G  L  Q  K  I  L  N

GTGCAAAAAAAGCTCCCAATCATCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTT
 V  Q  K  K  L  P  I  I  Q  K  I  I  I  M  D  S  K  T  D  Y  Q  G  F

CAGTCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGCCAGAG
 Q  S  M  Y  T  F  V  T  S  H  L  P  P  G  F  N  E  Y  D  F  V  P  E

TCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATCTACTGGTCTGCCTAAAGGT
 S  F  D  R  D  K  T  I  A  L  I  M  N  S  S  G  S  T  G  L  P  K  G

GTCGCTCTGCCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAA
 V  A  L  P  H  R  T  A  C  V  R  F  S  H  A  R  D  P  I  F  G  N  Q

ATCATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTC
 I  I  P  D  T  A  I  L  S  V  V  P  F  H  H  G  F  G  M  F  T  T  L

GGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAGGAGC
 G  Y  L  I  C  G  F  R  V  V  L  M  Y  R  F  E  E  E  L  F  L  R  S

CTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTCCTTCTTCGCCAAAAGCACT
 L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T  L  F  S  F  F  A  K  S  T

CTGATTGACAAATACGATTTATCTAATTTACACGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAA
 L  I  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P  L  S  K  E

GTCGGGGAAGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAAGGATATGGGCTCACTGAGACT
 V  G  E  A  V  A  K  R  F  H  L  P  G  I  R  Q  G  Y  G  L  T  E  T

ACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTT
 T  S  A  I  L  I  T  P  E  G  D  D  K  P  G  A  V  G  K  V  V  P  F

TTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAACTGTGT
 F  E  A  K  V  V  D  L  D  T  G  K  T  L  G  V  N  Q  R  G  E  L  C

GTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAG
 V  R  G  P  M  I  M  S  G  Y  V  N  N  P  E  A  T  N  A  L  I  D  K

GATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGC
 D  G  W  L  H  S  G  D  I  A  Y  W  D  E  D  E  H  F  F  I  V  D  R

CTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCCAA
```

```
  L   K   S   L   I   K   Y   K   G   Y   Q   V   A   P   A   E   L   E   S   I   L   L   Q
CACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCC
  H   P   N   I   F   D   A   G   V   A   G   L   P   D   D   D   A   G   E   L   P   A   A
GTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTA
  V   V   V   L   E   H   G   K   T   M   T   E   K   E   I   V   D   Y   V   A   S   Q   V
ACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAA
  T   T   A   K   K   L   R   G   G   V   V   F   V   D   E   V   P   K   G   L   T   G   K
CTCGACGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAA
  L   D   A   R   K   I   R   E   I   L   I   K   A   K   K   G   G   K   I   A   V   *
```

R4pGWBn40 (EYFP) 717bp

```
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
 M   V   S   K   G   E   E   L   F   T   G   V   V   P   I   L   V   E   L   D   G   D   V
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
 N   G   H   K   F   S   V   S   G   E   G   E   G   D   A   T   Y   G   K   L   T   L   K
TTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCCTG
 F   I   C   T   T   G   K   L   P   V   P   W   P   T   L   V   T   T   F   G   Y   G   L
CAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC
 Q   C   F   A   R   Y   P   D   H   M   K   Q   H   D   F   F   K   S   A   M   P   E   G
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTC
 Y   V   Q   E   R   T   I   F   F   K   D   D   G   N   Y   K   T   R   A   E   V   K   F
GAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
 E   G   D   T   L   V   N   R   I   E   L   K   G   I   D   F   K   E   D   G   N   I   L
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC
 G   H   K   L   E   Y   N   Y   N   S   H   N   V   Y   I   M   A   D   K   Q   K   N   G
ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
 I   K   V   N   F   K   I   R   H   N   I   E   D   G   S   V   Q   L   A   D   H   Y   Q
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCGCC
 Q   N   T   P   I   G   D   G   P   V   L   L   P   D   N   H   Y   L   S   Y   Q   S   A
CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
 L   S   K   D   P   N   E   K   R   D   H   M   V   L   L   E   F   V   T   A   A   G   I
ACTCTCGGCATGGACGAGCTGTACAAG
 T   L   G   M   D   E   L   Y   K
```

R4pGWBn43 (ECFP) 717bp

```
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
 M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
 N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K
TTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTGGGGCGTG
 F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  W  G  V
CAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC
 Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTC
 Y  V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F
GAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
 E  G  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L
GGGCACAAGCTGGAGTACAACTACATCAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGC
 G  H  K  L  E  Y  N  Y  I  S  H  N  V  Y  I  T  A  D  K  Q  K  N  G
ATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
 I  K  A  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCC
 Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  T  Q  S  A
CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
 L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I
ACTCTCGGCATGGACGAGCTGTACAAG
 T  L  G  M  D  E  L  Y  K
```

R4pGWBn50 (G3GFP) 717bp

ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAAT
M S K G E E L F T G V V P I L V E L D G D V N

GGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTT
G H K F S V S G E G E G D A T Y G K L T L K F

ATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTTACTACTTTCGCTTACGGTGTGCAG
I C T T G K L P V P W P T L V T T F A Y G V Q

TGCTTCTCAAGATACCCAGATCATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATAC
C F S R Y P D H M K R H D F F K S A M P E G Y

GTGCAGGAGAGGACCATCTTCTTCAAGGACGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAG
V Q E R T I F F K D D G N Y K T R A E V K F E

GGAGACACCCTCGTCAACAGGATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGC
G D T L V N R I E L K G I D F K E D G N I L G

CACAAGTTGGAATACAACTTCAACTCCCACAACGTGTACATCATGGCAGACAAACAAAAGAATGGAATC
H K L E Y N F N S H N V Y I M A D K Q K N G I

AAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAA
K V N F K I R H N I E D G S V Q L A D H Y Q Q

AATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTT
N T P I G D G P V L L P D N H Y L S T Q S A L

TCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACA
S K D P N E K R D H M V L L E F V T A A G I T

CATGGCATGGATGAACTATACAAATAA
H G M D E L Y K *

Termination codon (TAA) is included in pGWBn50 and pGWBn51 (C-G3GFP).   Termination codon is not included in pGWBn52 (N-G3GFP).

R4pGWBn53 (mRFP)   678bp

```
ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAAC
 M   A   S   S   E   D   V   I   K   E   F   M   R   F   K   V   R   M   E   G   S   V   N
GGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTG
 G   H   E   F   E   I   E   G   E   G   E   G   R   P   Y   E   G   T   Q   T   A   K   L
AAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCCAGTACGGCTCC
 K   V   T   K   G   G   P   L   P   F   A   W   D   I   L   S   P   Q   F   Q   Y   G   S
AAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAG
 K   A   Y   V   K   H   P   A   D   I   P   D   Y   L   K   L   S   F   P   E   G   F   K
TGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGAC
 W   E   R   V   M   N   F   E   D   G   G   V   V   T   V   T   Q   D   S   S   L   Q   D
GGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAG
 G   E   F   I   Y   K   V   K   L   R   G   T   N   F   P   S   D   G   P   V   M   Q   K
AAGACCATGGGCTGGGAGGCCTCCACCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATC
 K   T   M   G   W   E   A   S   T   E   R   M   Y   P   E   D   G   A   L   K   G   E   I
AAGATGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCACCTACATGGCCAAG
 K   M   R   L   K   L   K   D   G   G   H   Y   D   A   E   V   K   T   T   Y   M   A   K
AAGCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACATCACCTCCCACAACGAGGAC
 K   P   V   Q   L   P   G   A   Y   K   T   D   I   K   L   D   I   T   S   H   N   E   D
TACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGCCACTCCACCGGCGCCTAA
 Y   T   I   V   E   Q   Y   E   R   A   E   G   R   H   S   T   G   A   *
```

Termination codon (TAA) is included in pGWBn53 and pGWBn54 (C-mRFP).   Termination codon is not included in pGWBn55 (N-mRFP).

```
R4pGWBn59 (tagRFP)  714bp
ATGGTGTCTAAGGGCGAAGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCGTG
M  V  S  K  G  E  E  L  I  K  E  N  M  H  M  K  L  Y  M  E  G  T  V
AACAACCACCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACCCAGACCATGAGA
N  N  H  H  F  K  C  T  S  E  G  E  G  K  P  Y  E  G  T  Q  T  M  R
ATCAAGGTGGTCGAGGGCGGCCCTCTCCCCTTCGCCTTCGACATCCTGGCTACCAGCTTCATGTACGGC
I  K  V  V  E  G  G  P  L  P  F  A  F  D  I  L  A  T  S  F  M  Y  G
AGCAGAACCTTCATCAACCACACCCAGGGCATCCCCGACTTCTTTAAGCAGTCCTTCCCTGAGGGCTTC
S  R  T  F  I  N  H  T  Q  G  I  P  D  F  F  K  Q  S  F  P  E  G  F
ACATGGGAGAGAGTCACCACATACGAAGACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCTCCAG
T  W  E  R  V  T  T  Y  E  D  G  G  V  L  T  A  T  Q  D  T  S  L  Q
GACGGCTGCCTCATCTACAACGTCAAGATCAGAGGGGTGAACTTCCCATCCAACGGCCCTGTGATGCAG
D  G  C  L  I  Y  N  V  K  I  R  G  V  N  F  P  S  N  G  P  V  M  Q
AAGAAAACACTCGGCTGGGAGGCCAACACCGAGATGCTGTACCCCGCTGACGGCGGCCTGGAAGGCAGA
K  K  T  L  G  W  E  A  N  T  E  M  L  Y  P  A  D  G  G  L  E  G  R
AGCGACATGGCCCTGAAGCTCGTGGGCGGGGGCCACCTGATCTGCAACTTCAAGACCACATACAGATCC
S  D  M  A  L  K  L  V  G  G  G  H  L  I  C  N  F  K  T  T  Y  R  S
AAGAAACCCGCTAAGAACCTCAAGATGCCCGGCGTCTACTATGTGGACCACAGACTGGAAAGAATCAAG
K  K  P  A  K  N  L  K  M  P  G  V  Y  Y  V  D  H  R  L  E  R  I  K
GAGGCCGACAAAGAGACCTACGTCGAGCAGCACGAGGTGGCTGTGGCCAGATACTGCGACCTCCCTAGC
E  A  D  K  E  T  Y  V  E  Q  H  E  V  A  V  A  R  Y  C  D  L  P  S
AAACTGGGGCACAAACTTAATTGA
K  L  G  H  K  L  N  *
```

Termination codon (TGA) is included in pGWBn59 and pGWBn60 (C-tagRFP).   Termination codon is not included in pGWBn61 (N-tagRFP).

Acknowledgements