



# Sequence Alignments

Name of Variant	# of Occur	Alignment	Size (bp)	# Var bp	Sources*
CMV_enh-003	198+9	-----CGTTACATAA	304	0	Cl, Ev, Ox, In
CMV_enh-015	97+4	GACATTGATTATTGAC	380	2	Or, Si
CMV_enh-009	96	GACATTGATTATTGAC	380	0	In, Cl, Or
CMV_enh-008	47	-----CGTTACATAA	305	0	Ag, Mo
CMV_enh-005	23	GACATTGATTATTGAC	380	1	In, Ne
CMV_enh-010	22	-----CGTTACATAA	304	2	Pr, No, Or
CMV_enh-011	10	-----CGTTACATAA	304	1	Cl
CMV_enh-016	7	-----CGGTTACATAA	306	0	Qi
CMV_enh-002	4	GACATTGATTATTGAC	381	0/1i	Cl
CMV_enh-006	4	-----CGTTACATAA	304	1	No
CMV_enh-013	3	-----GCGTTACATAA	305	2	Pr, Cl
CMV_enh-012	2	-----CGTTACATAA	304	1	Cl
CMV_enh-004	2	-----CGTTACATAA	304	2	Lu
CMV_enh-007	2	GACATTGATTATTGAC	380	2	Cl
CMV_pro-007	2	-----GGAGTCCGCGT	517	0	Cl

```

CMV_enh-003      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-015      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-009      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-008      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-005      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-010      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-011      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-016      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-002      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-006      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-013      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-012      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-004      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_enh-007      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
CMV_pro-007      ACGACCCCGCCCATTTGACGTC AATAATGACGTATGTTCCCATAGTAAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
*****

```

```

CMV_enh-003      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-015      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-009      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-008      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-005      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-010      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-011      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-016      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-002      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-006      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-013      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-012      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-004      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_enh-007      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
CMV_pro-007      AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGC-CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
*****

```

```

CMV_enh-003      TTAGTCATCGCTATTACCATG
CMV_enh-015      TTAGTCATCGCTATTACCATG
CMV_enh-009      TTAGTCATCGCTATTACCATG
CMV_enh-008      TTAGTCATCGCTATTACCATG
CMV_enh-005      TTAGTCATCGCTATTACCATG
CMV_enh-010      TTAGTCATCGCTATTACCATG
CMV_enh-011      TTAGTCATCGCTATTACCATG
CMV_enh-016      TTAGTCATCGCTATTACCATG
CMV_enh-002      TTAGTCATCGCTATTACCATG
CMV_enh-006      TTAGTCATCGCTATTACCATG
CMV_enh-013      TTAGTCATCGCTATTACCATG
CMV_enh-012      TTAGTCATCGCTATTACCATG
CMV_enh-004      TTAGTCATCGCTATTACCATG
CMV_enh-007      TTAGTCATCGCTATTACCATG
CMV_pro-007      TTAGTCATCGCTATTACCATGTTGATGCGGTTTGGCAGTACATCAATGGGCGTGGATAG...
*****

```

CMV\_enh-001 is reverse complement of CMV\_enh-003. CMV\_enh-001 was merged with CMV\_enh-003  
 CMV\_enh-014 is reverse complement of CMV\_enh-015. CMV\_enh-014 was merged with CMV\_enh-015  
 CMV\_pro-007 is actually promoter plus enhancer. When split the enhancer segment is unique among the enhancers.

Key **ATGC** = change in nucleotide relative to consensus sequence – non-coding sequence

# Sequence Alignments

Name of Variant	# of Alignment Occur	Size (bp)	# Var bp	Sources*
CMV_pro-004	338+2	204	0	Cl, In, Ag, Ev
CMV_pro-009	114	204	4	Or, Si, Pr
CMV_pro-008	19	199	0	Or
CMV_pro-010	12	204	1	Ox
CMV_pro-005	10	200	4	No
CMV_pro-006	10	204	1	Cl
CMV_pro-002	4	204	1	In
CMV_pro-007	1	517	0	Cl
CMV_pro-001	1	204	7	Iv
CMV_pro-003	1	212	3/8i	Cl

```

CMV_pro-004 -----
CMV_pro-009 -----
CMV_pro-008 -----
CMV_pro-010 -----
CMV_pro-005 -----
CMV_pro-006 -----
CMV_pro-002 -----
CMV_pro-007 -----
CMV_pro-001 -----
CMV_pro-003 -----

```

```

CMV_pro-004 -----GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-009 -----GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-008 -----TGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-010 -----GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-005 -----TGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-006 -----GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-002 -----GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-007 -----TGCCAGTACATGACCTTATGGGACTTTCCTACTGGCAGTACATCTACGTATTAGTCATCGCTATTCCATGGTGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-001 -----GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA
CMV_pro-003 -----GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA

```

```

CMV_pro-004 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-009 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-008 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-010 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-005 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-006 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-002 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-007 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-001 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA
CMV_pro-003 -----CTCACGGGGATTTCCTCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCACAAAATCAACGGGACTTTCCTCAAAATGTCGTAACAAAC-----TCCGCCCCATTGACGCAA

```

```

CMV_pro-004 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-009 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-008 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-010 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-005 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-006 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-002 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-007 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-001 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT
CMV_pro-003 -----ATGGGCGGTAGGCGGTACGGTGGGAGGTTCTATATAAGCAGAGCT

```

CMV\_pro-007 is actually promoter plus enhancer. When split the promoter segment is identical to CMV\_pro-004.

Key **ATGC** = change in nucleotide relative to consensus sequence – non-coding sequence

# Sequence Alignments

Name of Variant	# of Alignment Occur	Size (bp)	# Var bp	Sources*
SV40_ori-002	1	135	1	Or
SV40_ori-001	1	136	1	Lu
SV40_ori-003	139	136	1	In,Ne,GE,OB,Lu, etc
SV40_ori-006	8	136	1	Cl
SV40_ori-004	261+12	136	0	Cl,Or,Pr,Ev,Ag, etc
SV40_ori-007	5	136	1	Pr,Ne
SV40_prom-010	8	196	1/li	Cl
SV40_prom-014	2	195	1	Ne
SV40_prom-008	10	197	0	Or,Cl,Pr
SV40_prom-013	5	197	1	Pr
SV40_prom-002	6	303	0	Si
SV40_prom-001	4	317	1	In
SV40_prom-006	1	317	0	?
SV40_prom-009	1	302	1/2d3i	Si
SV40_prom-015	3	299	1	Si
SV40_prom-005	94	330	0	Or,Cl
SV40_prom-007	77	298	1	In,Th,Cl,GE
SV40_prom-012	23	298	1/li	In,Ne
SV40_prom-003	1	358	1	Lu
SV40_prom-004	182	358	0	Cl,Ag,Ev,Pr,In,Or
SV40_prom-011	48	358	1	Pr,Lu
SR-alpha_prom	1	615	1	?
SV40_enhancer	4	230	0	Pr

72 bp repeat II (enhancer)

```

SV40_ori-002 -----TCCCGCCCC--TAACTC
SV40_ori-001 -----ATCCCGCCCC--TAACTC
SV40_ori-003 -----ATCCCGCCCC--TAACTC
SV40_ori-006 -----ATCCCGCCCC--TAACTC
SV40_ori-004 -----ATCCCGCCCC--TAACTC
SV40_ori-007 -----ATCCCGCCCC--TAACTC
SV40_prom-010 -----CAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-014 -----GCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-008 -----TGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-013 -----TGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-002 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-001 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-006 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-009 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-015 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-005 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-007 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-012 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-003 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-004 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_prom-011 GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC
SV40_enhancer GGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC GCCCCTAACTCCGCCCATCCCGCCCC--TAACTC

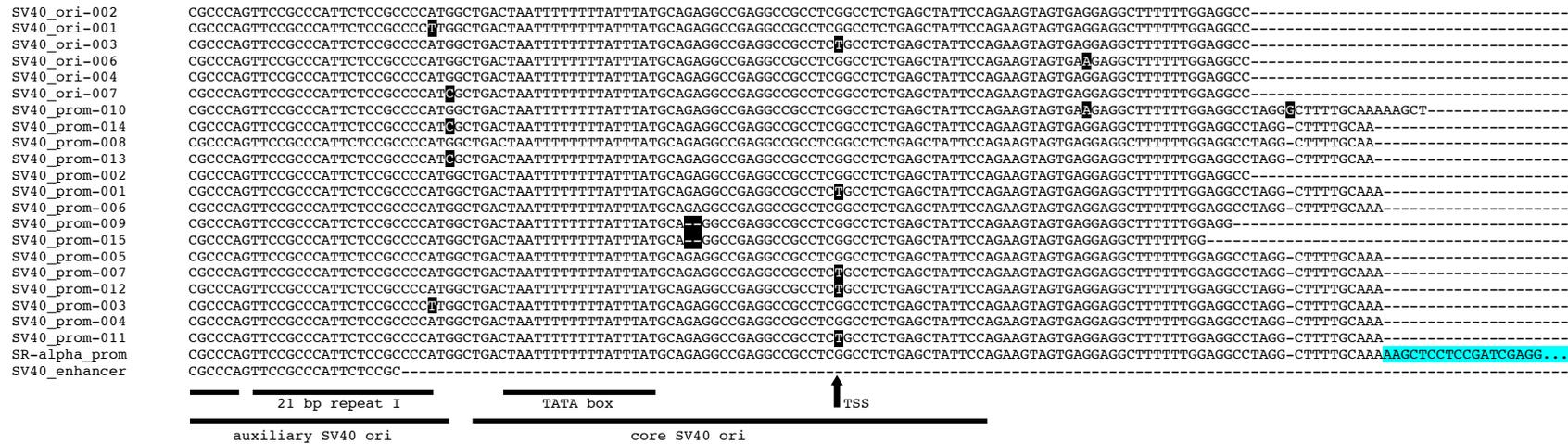
```

72 bp repeat I (enhancer)

21 bp repeat III      21 bp repeat II

auxiliary SV40 ori

## Sequence Alignments



Sequence structural elements are from Byrne et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:721-725.

Enhancer consists of 72 bp repeats

Promoter consists of 21 bp repeats and TATA box

SR-alpha promoter is a composite of the SV40 enhancer/promoter and the LTR from HTLV-1

SV40 ori-006 is reverse complement but unique

SV40 ori-005 is identical to SV40 ori-004 but reverse complement - SV40 ori-005 was merged with -004

SV40 enhancer is reverse complement

Key **ATGC** = change in nucleotide relative to consensus sequence - non-coding sequence

**ATGC** = HTLV-1 LTR



## Sequence Alignments

```

f1_ori-008      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-007      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-009      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-003      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-006      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-004      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
M13_ori-003     GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTTA
f1_ori-011      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-014      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-017      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-005      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
M13_ori-004     GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-001      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-002      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-010      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-012      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-013      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-015      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
f1_ori-016      GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAATCACATTAACGTTTACAATTT-
M13_ori-001     GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
M13_ori-002     GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
M13_ori-005     GGCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAAT---ATTAACGTTTACAATTT-
** *****

```

Domain B (AT-rich enhancer of + strand synthesis)

```

f1_ori-008      -----
f1_ori-007      -----
f1_ori-009      -----
f1_ori-003      -----
f1_ori-006      -----
f1_ori-004      -----
M13_ori-003     AATATTGGCTTATACAATCTCCTGTTTTGGGGCTTTTCTGATTATCAACCGGGG
f1_ori-011      -----
f1_ori-014      -----
f1_ori-017      -----
f1_ori-005      -----
M13_ori-004     -----
f1_ori-001      -----
f1_ori-002      -----
f1_ori-010      -----
f1_ori-012      -----
f1_ori-013      -----
f1_ori-015      -----
f1_ori-016      -----
M13_ori-001     -----
M13_ori-002     -----
M13_ori-005     -----

```

Structures are from Dotto et al. (1984) *J. Mol. Biol.* **172**:507-521 and from Baas (1985) *Biochimica et Biophysica Acta* **825**:11-139.

Yellow highlights indicate nucleotides that form bulges or loops in the indicated stem-loop structures formed when the DNA is single-stranded.  
**f1 ori variants -006, -013 and -015 are actually M13 ori as the two origins differ only by an A or T at the indicated position.**

If the sequences of variants -013 and -017 are correct, these origins are almost certainly compromised or devoid of minus strand and plus strand synthesis, respectively.

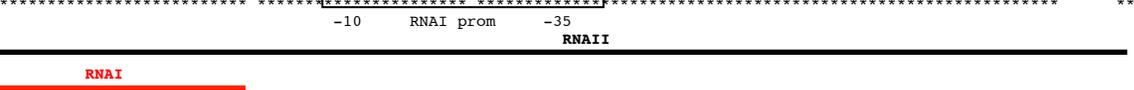
Key **ATGC** = change in nucleotide relative to consensus sequence – non-coding sequence

# Sequence Alignments

Name of Variant	# of Occur	Alignment	Size (bp)	# Var bp	Sources*
ori-014	602	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	1	No,Cl,In,Ca,Ev, etc
ori-013	524	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	0	Cl,In,Or,Pr,Ox, etc
ori-009	220	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	3	In,Qi,MBL,Ne, etc
ori-012	55	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	585	0/4d	Or,Cl,In,GE
ori-008	19	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	2	No,Qi,GE
ori-003	15	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	1	Th
ori-021	12	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	4	24050148
ori-002	8	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	588	1/1d	Lu
ori-016	8	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	4	2659436
ori-010	5	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	1	10890530
ori-011	5	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	588	0/1d	Qi,No
ori-005	4	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	2	Pr
ori-018	4	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	4	15644173
ori-001	1	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	1	Or
ori-004	1	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	2	IMAGE
ori-006	1	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	1	iGEM
ori-007	1	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	588	0/1d	Or
ori-015	1	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	2	In
ori-017	1	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	569	3	In
ori-019	1	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	583	2/6d	No
ori-020	1	TTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTCCGAAAGGTAAGTGGCTT	589	1	Cl
pBR322ori-t	1	-----	353	0	765517



Name of Variant	Alignment	Size (bp)	# Var bp	Sources*
ori-014	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	1	No,Cl,In,Ca,Ev, etc
ori-013	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	0	Cl,In,Or,Pr,Ox, etc
ori-009	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	3	In,Qi,MBL,Ne, etc
ori-012	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	585	0/4d	Or,Cl,In,GE
ori-008	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	2	No,Qi,GE
ori-003	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	1	Th
ori-021	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	4	24050148
ori-002	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	588	1/1d	Lu
ori-016	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	4	2659436
ori-010	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	1	10890530
ori-011	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	588	0/1d	Qi,No
ori-005	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	2	Pr
ori-018	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	4	15644173
ori-001	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	1	Or
ori-004	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	2	IMAGE
ori-006	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	1	iGEM
ori-007	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	588	0/1d	Or
ori-015	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	2	In
ori-017	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	569	3	In
ori-019	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	583	2/6d	No
ori-020	CAGCAGAGCGCAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTACATACCTCGCTCTGCTAACTCTGTACCAGTGGCTGC	589	1	Cl
pBR322ori-t	-----	353	0	765517











# Sequence Alignments

```

IRES-014  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-001  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-013  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-012  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-004  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-008  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-015  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-010  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-002  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-005  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-003  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-006  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES-011  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES2-001  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES2-002  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG
IRES2-003  ACCTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGA-TCTGATCTGGGCGCCTCGG

```

Stem-Loop I Stem-Loop J Stem-Loop K

```

IRES-014  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-001  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-013  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-012  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-004  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-008  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-015  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-010  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-002  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-005  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-003  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-006  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES-011  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAA-----
IRES2-001  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAAATGGCCACAACCATG
IRES2-002  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAAATGGCCACAACCATG
IRES2-003  TGCACATGCTTTACATGTGTTAGTCGAAGTTAAAAAAGCTTAGGCCCCCGAACCCAGGGGACGTGGTTTCCTTTGAAAAACACCATGATAAATGGCCACAACCATG

```

Stem-Loop K A Bucle Stem-Loop J Stem-Loop L 834

Red highlighting indicates the IRES borders were mis-annotated. The sequence actually includes the highlighted region.

IRES-009 is derived from foot and mouth disease virus (FMDV) and does not align well with other IRES variants derived from encephalomyocarditis virus (EMCV). However both EMCV and FMDV are picornaviruses and utilize Group III IRES, which initiate translation directly at the START codon without ribosome scanning. This means that the placement of the START codon is important for translation efficiency. The native START codon is highlighted in green. An alternative START codon is highlighted in cyan.

The 5' portion of IRES is dispensable for function up to stem-loop H (not shown), but interference of stem-loop H formation can occur in constructs in which the first cistron is too close. Therefore, it is preferable to include the 5' stem-loops as spacers.

IRES-002, -004, -005, -012 and -013 have wild-type number of A (A6) in the oligo(A) bulge at the K-J border. This bulge is important for interaction with eIF4G. A7 (e.g. in pIRES) has reduced translational activity compared to A6 (Bochkov and Palmenberg, 2006. Biotechniques 41:283-292). Because A6 and A7 variants have slightly different interactions with the translation machinery, they show different activities in different cell types.

The position of the START codon of the downstream ORF is also important for translational efficiency. Translation of the polypeptide from the native START codon at position 834 (native IRES numbering) provides optimal spacing. Including the first few amino acids of native IRES (IRES2 variants) allows the leader polypeptide to interact properly with the ribosome for the most efficient translation. Using the IRES START codon at position 826 reduces translational efficiency due to steric constraints on this leader peptide ribosome interaction. Moreover, placing the start codon too far downstream in a MCS reduces translation efficiency as the ribosome does not scan in Group III IRESs.

Structures are from Kaminski and Jackson (1998) *RNA* 4:626-638. Invariant stem-loops are not shown.

**IRES-007 is identical to IRES-008 but reverse complement - IRES-007 was merged with -008**

Key ATGC = change in nucleotide relative to consensus sequence - non-coding sequence

# Sequence Alignments

## Un-Annotated Promoters

Alignments do not include all plasmids carrying the un-annotated promoter

### Un-annotated AmpR promoters (colored segments = vector backbone)

Name of Plasmid Or Feature	Alignment
AmpR_prom-009	CGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-001	CGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-006	CGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-008	CGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-011	CGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-002	CGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-012	CGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCMV-Cypridina_Luc	ACACTGACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pTK-Cypridina	ACACTGACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-43.1a(+)	TGGTCATGACCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-44a(+)	TGGTCATGACCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-45b(+)	TGGTCATGACCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-51b(+)	TGGTCATGACCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-52b(+)	TGGTCATGACCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-005	-----TTTGTATTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-003	-----TTTGTATTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-010	-----TTTGTATTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-004	-----TTTGTATTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-007	-----TTTGTATTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCR4-TOPO	TCTACGGGCTGACGGCTCAGTGGAAACGAAACATGCGTTAAGGGATTGTTGTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCAS-Guide	GCCCGTGTCTCAAATCTCTGATGTACATGTCAGAAAGATAAAATAATATCATCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCMV_SPORT	AATTATGTGCTGTGTAGGGATCGCTGGTATCAAATATGTTGCCACCCCTGGCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCMV_SPORT2	AATTATGTGCTGTGTAGGGATCGCTGGTATCAAATATGTTGCCACCCCTGGCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pZL1	AATTATGTGCTGTGTAGGGATCGCTGGTATCAAATATGTTGCCACCCCTGGCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pTriEx-1.1	GC CGGTTTATACACATCTTGGGATTTGATTAAGAATGTCAGAAACCGCGGGGATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pTriEx-2	GC CGGTTTATACACATCTTGGGATTTGATTAAGAATGTCAGAAACCGCGGGGATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pQE-TriSystem_5	GC CGGTTTATACACATCTTGGGATTTGATTAAGAATGTCAGAAACCGCGGGGATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pQE-TriSystem_6	GC CGGTTTATACACATCTTGGGATTTGATTAAGAATGTCAGAAACCGCGGGGATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pDream2.1_MCS	GC CGGTTTATACACATCTTGGGATTTGATTAAGAATGTCAGAAACCGCGGGGATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pQE-TriSystem_1	----GTTTATACACATCTTGGGATTTGATTAAGAATGTCAGAAACCGCGGGGATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pQE-TriSystem_His-Str	----GTTTATACACATCTTGGGATTTGATTAAGAATGTCAGAAACCGCGGGGATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pGL4.70-hRluc	-CGCGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCTTGAGAGACGTACTAGTACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pLightSwitch_Prom	-CGCGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCTTGAGAGACGTACTAGTACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pML1.1-Nluc	-CGCGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCTTGAGAGACGTACTAGTACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pML2.1-Nluc	-CGCGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCTTGAGAGACGTACTAGTACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pGL4.10-Luc2	-CGCGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCTTGAGAGACGTACTAGTACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
Lafmid_BA	CTTATACAATCTTCTGTTTTGGGGCTTTCTGTTTATACAACCGGGTACATATGATGACATGCTACTGTTTACGATTACCGTTCATCGATTGAAAAAGGAAGAGT
pLIC-SGCl	AAGCTCTAAATCGGGGCTCCCTTAGGGTCCGATTTAGTCTTTTACGGCACCTCGACCCCAAACTGATTAGGGTATGGTTCACATTGAAAAAGGAAGAGT
pIB_V5-His-DEST	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGCTACTAAAATGAAATTAGAGCTTAAATTAATATACAGTTATTACCCATTGAAAAAGGAAGAGT
pIB_V5-His	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGCTACTAAAATGAAATTAGAGCTTAAATTAATATACAGTTATTACCCATTGAAAAAGGAAGAGT
pMIB_V5-His_A	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGCTACTAAAATGAAATTAGAGCTTAAATTAATATACAGTTATTACCCATTGAAAAAGGAAGAGT
pYC6_CT	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGCTACTAAAATGAAATTAGAGCTTAAATTAATATACAGTTATTACCCATTGAAAAAGGAAGAGT
pYES6_CT	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGCTACTAAAATGAAATTAGAGCTTAAATTAATATACAGTTATTACCCATTGAAAAAGGAAGAGT
pYC2_CT	-----AAAACGTATTATAAGTAAATGCGATGATACTAAAATGAAATTAGAGCTTAAATTAATATACAGTTATTACCCATTGAAAAAGGAAGAGT
pYES2_CT	-----AAAACGTATTATAAGTAAATGCGATGATACTAAAATGAAATTAGAGCTTAAATTAATATACAGTTATTACCCATTGAAAAAGGAAGAGT
pCR2.1-TOPO	-----AAAACGTATTATAAGTAAATGCGATGATACTAAAATGAAATTAGAGCTTAAATTAATATACAGTTATTACCCATTGAAAAAGGAAGAGT

-35 *bla* TEM prom -10 RBS

Sequences were trimmed at the maximum length shown or at the junction of an adjacent feature (producing the shorter sequences). The boxed -35 and -10 regions correspond to those of native *bla* TEM from *E. coli* Tn3.

pCAS-Guide uses the KanR (*aph*(3')-Ia) promoter from Tn5 (brown region; -35 and -10 sequences highlighted in yellow).

The violet region for pCMV-SPORT, pCMV-SPORT2 and pZL1 are part of a defunct *incA* (RepA binding sites) region, but BPROM promoter prediction algorithm suggests ATGTGT forms a viable -35 region with the intact -10.

The pink region for the pTriEx and pQE-TriSystem series are part of baculovirus recombination region (Ief2), but BPROM suggests ATGCA forms a viable -35 region with the intact -10.

The light green region for the luciferase vectors appears to be an artificial promoter (there are several synthetic sequences in these plasmids), but BPROM does not predict a -35 and -10. The blue region in Lafmid\_BA corresponds to the 3' end of M13 ori-003 and is predicted by BPROM to form -35 and -10 (yellow highlight).

The orange sequence of pLIC-SGCl is identical to a portion of fl origin. The highlighted areas correspond to -35 and -10 sites for RNAPol  $\sigma^{70}$  which normally produces an RNA template for DNA polymerase, but has weak promoter activity when the DNA is double-stranded (Higashitani et al. 1997. PNAS USA 94:2909).

The cyan region of the PIB and pMIB series, and pYC and pYES series is *S. cerevisiae* URA3 3' flanking sequence, but is predicted by BPROM to form a viable -35 and -10 (yellow highlight). The AmpR of pCR2.1-TOPO is produced as a bicistronic transcript with NeoR/KanR from the NeoR/KanR promoter.

Key **ATGC** = change in nucleotide relative to consensus sequence - non-coding sequence

# Sequence Alignments

## Un-annotated CAT promoters (colored segments = vector backbone)

Name of Plasmid Or Feature	Alignment
CAT_prom-001	TGATCGG- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTT <b>TCGAGATTTTCAGGAGCTAAGGAAGCTAAA</b>
CAT_prom-002	TGATCGG- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
CAT_prom-003	TGATCGG- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
CAT_prom-004	TGATCGG- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pSMART_HCKan (KanR)	TGATCGG- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pSMART_LCKan (KanR)	TGATCGG- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pSB1C3 (CmR)	TGATCGG- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRANGER-BTB-1 (AmpR)	----- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRANGER-BTB-2 (KanR)	----- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRANGER-BTB-3 (CmR)	----- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRANGER-BTB-5 (TpR)	----- <b>ACGTAAGAGGTTCCAACCTTCCACCATAAT</b> AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRFP-C-RS (CmR)	<b>ATATGTATCGCTCATGACTATGTTTGACAGCTTATC</b> <b>CGATAGCTAGCTAGCGCCGCTAGC</b> <b>TTTATGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA</b>
pVP13 (CmR)	----- <b>CCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAA</b>
pVP16 (CmR)	----- <b>CCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAA</b>
pVP33K (CmR)	----- <b>CCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAA</b>
pVP56K (CmR)	----- <b>CCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAA</b>
pDNR-LIB (CmR)	----- <b>GGATCTTTCAGGAGCTAAGGAAGCTAAA</b>
pDNR-Dual (CmR)	----- <b>GGATCTTTCAGGAGCTAAGGAAGCTAAA</b>

Sequences were trimmed at the maximum length shown or at the junction of an adjacent feature (producing the shorter sequences). These CAT promoters drive expression of various marker genes indicated in parentheses.

In pSMART\_LCKan, CmR is the only selectable marker. Therefore this sequences must be active. The cyan portion of pRFP-C-RS is pBR322 backbone but is predicted by BPROM to have -35 and -10 sequences (highlighted in yellow). The pVP series has two CmR genes – one behind lac UV5 promoter, and the other with the above 5' region, which is not active. pDNR-LIB/Dual have only CmR as the selectable marker. In the case of pDNR-LIB, CmR is head-to-tail with SacB and is likely produced from a bicistronic transcript with SacB. However, in pDNR-Dual, CmR and SacB are head-to-head. In this case, it seems that the SacB upstream region includes a bidirectional promoter that can drive expression of CmR. Alternatively, the loxP site between CmR and SacB in both plasmids has cryptic promoter activity.

## Un-annotated NeoR/KanR promoters (colored segments = unknown sequence)

Name of Plasmid Or Feature	Alignment
pREP4	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP56K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP68K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP65K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP81K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP33K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pSpark_III	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GAGGATCGTTTCGC
pCAMBIA5105	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GAGGATCGTTTCGC
pEer0-2	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGGCCGCAAGGATCTGATGGCCGAGGGATCAAGATCTGATCAAGAGACAGGAT---GAGGATCGTTTCGC
pSF-pA-PromMCS-Fluc	<b>GGCAGCGTACCGATCTGTTTAAACCTAGATA</b> <b>TTGATAGTCTGATCGGTCAACG</b> <b>TATAA</b> CGAGTCCTAGCTTTTGCAAACATCT-ATCAAGAGACAGGAT <b>CAGCAGGAGGCTTTCGC</b>
pSF-pA-CMVe-Rluc	<b>GGCAGCGTACCGATCTGTTTAAACCTAGATA</b> <b>TTGATAGTCTGATCGGTCAACG</b> <b>TATAA</b> CGAGTCCTAGCTTTTGCAAACATCT-ATCAAGAGACAGGAT <b>CAGCAGGAGGCTTTCGC</b>
pSF-pA-PromMCS-Rluc	<b>GGCAGCGTACCGATCTGTTTAAACCTAGATA</b> <b>TTGATAGTCTGATCGGTCAACG</b> <b>TATAA</b> CGAGTCCTAGCTTTTGCAAACATCT-ATCAAGAGACAGGAT <b>CAGCAGGAGGCTTTCGC</b>
pSF-PromMCS-Fluc	<b>GGCAGCGTACCGATCTGTTTAAACCTAGATA</b> <b>TTGATAGTCTGATCGGTCAACG</b> <b>TATAA</b> CGAGTCCTAGCTTTTGCAAACATCT-ATCAAGAGACAGGAT <b>CAGCAGGAGGCTTTCGC</b>
pSF-PromMCS-Rluc	<b>GGCAGCGTACCGATCTGTTTAAACCTAGATA</b> <b>TTGATAGTCTGATCGGTCAACG</b> <b>TATAA</b> CGAGTCCTAGCTTTTGCAAACATCT-ATCAAGAGACAGGAT <b>CAGCAGGAGGCTTTCGC</b>
pSF-pA-CMVe-Fluc	<b>GGCAGCGTACCGATCTGTTTAAACCTAGATA</b> <b>TTGATAGTCTGATCGGTCAACG</b> <b>TATAA</b> CGAGTCCTAGCTTTTGCAAACATCT-ATCAAGAGACAGGAT <b>CAGCAGGAGGCTTTCGC</b>
pSF-CMVe-Fluc	<b>GGCAGCGTACCGATCTGTTTAAACCTAGATA</b> <b>TTGATAGTCTGATCGGTCAACG</b> <b>TATAA</b> CGAGTCCTAGCTTTTGCAAACATCT-ATCAAGAGACAGGAT <b>CAGCAGGAGGCTTTCGC</b>
pSF-CMVe-Rluc	<b>GGCAGCGTACCGATCTGTTTAAACCTAGATA</b> <b>TTGATAGTCTGATCGGTCAACG</b> <b>TATAA</b> CGAGTCCTAGCTTTTGCAAACATCT-ATCAAGAGACAGGAT <b>CAGCAGGAGGCTTTCGC</b>

The top 9 promoters show identity to the native NeoR/KanR (*nptII*) promoter from *E. coli* transposon *Tn5*.

BPROM did not identify putative -35 and -10 sequences in the native promoter.

The pSF series from Oxford Genetics appear to have a related promoter, but this sequence does not appear to be natural as it generates no hits from homology searches against public databases. Highlighted regions are potential -35 and -10 sequences (BPROM). This promoter is also used to drive AmpR in pSF-Core, pSF-CMV-Fluc and pSF-CMV-Rluc.

Key **ATGC** = change in nucleotide relative to consensus sequence – non-coding sequence

# Sequence Alignments

Un-annotated KanR promoters (colored segments = plasmid backbone)

Name of Plasmid Or Feature	Alignment
pShuttle2	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pSIREN-Shuttle	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pACYC177	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pMCSG77	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pHSG298	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pHSG299	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pGreenII_0049	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pGreen_0029	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pGreen	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pKF_18k-2	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pKF_19k-2	ACAAAGCCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pOSIP-KT	TGTAACGCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pOSIP-KP	TGTAACGCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pOSIP-KL	TGTAACGCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pOSIP-KO	TGTAACGCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pOSIP-KH	TGTAACGCACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pOSIP-KC	CTACTTACACGTTGTGCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pUC5.7-Kan	GCAGCTCGGCCCGTGTCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pCONR201	GCAGCTCGGCCCGTGTCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pMCentr2	GCAGCTCGGCCCGTGTCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pMCentr3	GCAGCTCGGCCCGTGTCTCAAAATCCTGTGATGTTACATGGCACAAGATAAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pQE-80L	TTCTCGAGGTGAAGACGAAAGGGCCCTCGTATAGCCCTATTTTATAGGTAAAGTTCATGGTTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pQE-81L	TTCTCGAGGTGAAGACGAAAGGGCCCTCGTATAGCCCTATTTTATAGGTAAAGTTCATGGTTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pQE-82L	TTCTCGAGGTGAAGACGAAAGGGCCCTCGTATAGCCCTATTTTATAGGTAAAGTTCATGGTTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pET-9a	TGTAATCTTTTACGCGGGTCTGACGCTCAGTGGACGAAACTCACGTTAAGGGATTTGGTTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pET-9b	TGTAATCTTTTACGCGGGTCTGACGCTCAGTGGACGAAACTCACGTTAAGGGATTTGGTTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pET-9c	TGTAATCTTTTACGCGGGTCTGACGCTCAGTGGACGAAACTCACGTTAAGGGATTTGGTTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pET-9d	TGTAATCTTTTACGCGGGTCTGACGCTCAGTGGACGAAACTCACGTTAAGGGATTTGGTTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pGFP-B-RS	GGAAATGTGCGAAACCCATTTGTTATGTTCTTAAATACCTCAAAATATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pGFP-V-RS	GGAAATGTGCGAAACCCATTTGTTATGTTCTTAAATACCTCAAAATATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
pCMV6-Entry2	GGAAATGTGCGAAACCCATTTGTTATGTTCTTAAATACCTCAAAATATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAGGGGGTGT
AmpR_prom-009	-----CGGGAAACCCCTATTTGTTATTTTCTTAAATACATTCAAATATGATATCCGCTCATGA...*****

-35 aph(3')-Ia prom -10 RBS

These promoters are the natural *aph(3')*-Ia promoter from *E. coli* Tn5 and all drive expression of KanR (*aph(3')*-IaJ). Highlighting shows putative -35 and -10 sequences (BPROM). The pGFP and pCMV series have a hybrid promoter with AmpR promoter (cyan highlighting) from replacing AmpR with KanR during plasmid construction (TCATGA is a BspHI restriction site).

Name of Plasmid Or Feature	Alignment
pCambia0305.1	GATCACCGCGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pCambia0305.2	GATCACCGCGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pCambia03080	GATCACCGCGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pCambia03090	GATCACCGCGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pEarleyGate_100	GATCACCGCGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pEarleyGate_101	GATCACCGCGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pEarleyGate_102	GATCACCGCGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pFGC5941	GATCACCGCGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pBI121	CATAATTTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pRI_101-AN	CATAATTTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pRI_101-ON	CATAATTTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pRI_201-AN	CATAATTTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pRI_201-ON	CATAATTTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pRI_909	CATAATTTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pRI_910	CATAATTTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA
pBINPLUS	CATAATTTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGC. GAACAGTGAATTGGAGTTCGCTTCTGTTATAATFAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAA

-35 aphA-3 prom -10

These plasmids all use KanR. These promoters are the natural *aphA-3* (KanR-020) promoter from *Staphylococcus aureus*.

Key **ATGC** = change in nucleotide relative to consensus sequence - non-coding sequence

# Sequence Alignments

## Un-annotated SmR promoters

Name of Plasmid Or Feature	Alignment
pDEST20-Pc_prom	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pIDS	-----AAGTAGCGTA-----TGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pDONR223	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP100	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP101	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP222	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP221	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP212	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP211	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP202	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP201	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pZP200	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pCAMBIA1200	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pCAMBIA1105.1R	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pCAMBIA1105.1	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pCAMBIA0105.1R	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT
pCAMBIA5105	ACCACGTTGACATAAGCCTGTTTCGGTTCGTAAGCTATAATGCAAGTAGCGTAATGCGCTCAGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTTCAT



pDEST20-Pc_prom	GGC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACG-----
pIDS	GGC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pDONR223	GGC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP100	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP101	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP222	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP221	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP212	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP211	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP202	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP201	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pZP200	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pCAMBIA1200	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pCAMBIA1105.1R	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pCAMBIA1105.1	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pCAMBIA0105.1R	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC
pCAMBIA5105	GGCTTC---TTGTTATGACATGTTTTTTTGGGTTACAGTATATGCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGTTCGATGTTTATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAAACAAGTTAAACATC

inactivated att1 site

The native SmR promoter for these plasmids is derived from *E. coli* plasmid R538 and drives expression of the spectinomycin and streptomycin resistance gene (*SmR*) *aadA* (aminoglycoside 3'-adenyltransferase), which is a class 1 integron and includes two promoters to drive expression of gene cassettes that integrate at the *attI* site (mutated in these constructs) of the plasmid. The Pc promoter is embedded in the 5' end of a gene, *IntI1*, that encodes the integrase responsible for recombination between the *attC* containing integron and the *attI* site behind the promoter. *IntI1* expression is driven by a promoter, *Pint* (not shown), that partially overlaps the P2 promoter, but is inverted. With the exception of pIDS, which is missing Pc, these plasmids all have a truncated *IntI1*.

Promoter strength analysis based on Jove et al. (2010) PLoS Genet. 6:e1000793. The Pc promoter of pDEST20 and derivatives is a strong version of the promoter called PcS. It has optimal -35 to -10 spacing and two bp changes relative to the other Pc promoters in the alignment, which are weak (PcW). However, the P2 promoter of pDEST20 is probably inactive as it has only 14 bp between the -35 and -10 hexamers.

SmR expression from pIDS is driven only by P2 which has the optimal spacing (17 bp) between the -35 and -10 hexamers.

Both the PcW and P2 promoters of pDONR223 have optimal spacing.

The pZP and pCAMBIA series have sub-optimal spacing (18 bp) in both the PcS and P2 promoters.

In pDEST20 and derivatives, the Pc promoter is used to express GmR (gentamycin resistance; *aacC1*, gentamycin acetyltransferase). In pZP1XX series, Pc drives expression of CmR.

# Sequence Alignments

## Un-annotated SmR promoters

Name of Plasmid Or Feature	Alignment
pHELMSGATE-sat-aadA1	CAATTTGTACGCAAAATGTGGCTTACTCTCGGGCGCATTGACCTGTTACAG.CAATTCATTCAAGCCGACACCGCTTCGCGGGCGGGCTTAATTCAGGAGTTAAACATC
pHELMSGATE4-sat-aadA1	CAATTTGTACGCAAAATGTGGCTTACTCTCGGGCGCATTGACCTGTTACAG.CAATTCATTCAAGCCGACACCGCTTCGCGGGCGGGCTTAATTCAGGAGTTAAACATC
pHELMSGATE8-sat-aadA1	CAATTTGTACGCAAAATGTGGCTTACTCTCGGGCGCATTGACCTGTTACAG.CAATTCATTCAAGCCGACACCGCTTCGCGGGCGGGCTTAATTCAGGAGTTAAACATC
pHELMSGATE12-sat-aadA1	CAATTTGTACGCAAAATGTGGCTTACTCTCGGGCGCATTGACCTGTTACAG.CAATTCATTCAAGCCGACACCGCTTCGCGGGCGGGCTTAATTCAGGAGTTAAACATC
pSB11-sat-aadA1	CAATTTGTACGCAAAATGTGGCTTACTCTCGGGCGCATTGACCTGTTACAG.CAATTCATTCAAGCCGACACCGCTTCGCGGGCGGGCTTAATTCAGGAGTTAAACATC

Name of Plasmid Or Feature	Alignment
pHELMSGATE_5'sat1	AAGTGGCAGCAACGGATTCGCAAACTGTCACGCCTTTTGTGCCAAAAGCCGCCAGGTTTGCATCCGCTGTGCCAGGCGTTAGGCGTCAT
pHELMSGATE4_5'sat1	AAGTGGCAGCAACGGATTCGCAAACTGTCACGCCTTTTGTGCCAAAAGCCGCCAGGTTTGCATCCGCTGTGCCAGGCGTTAGGCGTCAT
pHELMSGATE8_5'sat1	AAGTGGCAGCAACGGATTCGCAAACTGTCACGCCTTTTGTGCCAAAAGCCGCCAGGTTTGCATCCGCTGTGCCAGGCGTTAGGCGTCAT
pHELMSGATE12_5'sat1	AAGTGGCAGCAACGGATTCGCAAACTGTCACGCCTTTTGTGCCAAAAGCCGCCAGGTTTGCATCCGCTGTGCCAGGCGTTAGGCGTCAT
pSB11_5'sat1	AAGTGGCAGCAACGGATTCGCAAACTGTCACGCCTTTTGTGCCAAAAGCCGCCAGGTTTGCATCCGCTGTGCCAGGCGTTAGGCGTCAT

SmR in these plasmids is derived from *E. coli* Tn7 which contains a class 2 integron including *dfrA1* (dihydrofolate reductase conferring resistance to trimethoprim), *sat* (streptothricin acetyltransferase), *aadA1* (SmR - spectinomycin and streptomycin resistance, aminoglycoside adenyltransferase), and *ybeA* (pseudouridine methyltransferase).

Class 2 integrons have a similar structure to class 1, but the sequences of the Pc promoter (-35 TTTAAT 16 bp TAAAAT -10) and the integrase (*IntI2*) are different. The Pc promoter of Tn7 drives expression of a polycistronic transcript (*dfrA1-sat-aadA1-ybeA*). However, neither the pSB11 nor the pHELMSGATE plasmids include the Pc promoter from Tn7, which is upstream of *dfrA1*. These plasmids all include *sat*, *aadA1* and *ybeA*, but pSB11 includes just a portion of the end of *dfrA1*, while the pHELMSGATE plasmids start just 5' to *sat*. Therefore, it is not clear what sequence constitutes a promoter in these constructs.

In the alignments, we show the region covering the 3' end of *sat* and intergenic region between *sat* and SmR (top alignment), and the region upstream of *sat* common to pSB11 and pHELMSGATE (bottom alignment), which has no putative -35 and -10 site. It seems likely that there is a cryptic promoter within these regions or the *sat* ORF. BPROM predicts a putative promoter at the end of the *sat* ORF (top alignment; yellow highlight). This region is ~100 bp upstream of the *aadA1* START codon.

Key **ATGC** = change in nucleotide relative to consensus sequence - non-coding sequence

# Sequence Alignments

## Un-annotated *ccdB* promoters

Name of Plasmid Or Feature	Alignment
ccdB_nat_prom	TAAAATACATAAGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pDEST26	AGATCTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pLenti4_V5-DEST	AGATCTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pT-Tex-DEST30	AGATCTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pcDNA-DEST40	AGATCTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pANT7-nHA	AGATCTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pDEST8	ACCGTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pAd_BLOCK-iT-DEST	ACCGTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pDONR201	ACCGTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pAd_FL-DEST	ACCGTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pET-53-DEST	ACCGTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pIB_V5-His-DEST	ACCGTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pMT_BioEase-DEST	ACCGTGGATCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pCMV_SPORT6ccdB	GGCCGCGCTACCCGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA
pOSIP-CH	GGTCCCTAGACGGCTTACTAAAAGCCAGATAAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTATATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACA



ccdB_nat_prom	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pDEST26	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pLenti4_V5-DEST	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pT-Tex-DEST30	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pcDNA-DEST40	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pANT7-nHA	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pDEST8	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pAd_BLOCK-iT-DEST	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pDONR201	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pAd_FL-DEST	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pET-53-DEST	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pIB_V5-His-DEST	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pMT_BioEase-DEST	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pCMV_SPORT6ccdB	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA
pOSIP-CH	GCGACAGTATCACTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGCTGCGTGCCGAACGC. AATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAA



The native *ccdB* promoter (*ccdB\_nat\_prom*) from *E. coli* F plasmid is shown in the first row. *ccdB* is transcribed as part of a multi-cistronic transcript along with *ccdA* (arrow), an inhibitor of *ccdB* protein. *ccdA* is inactivated in *ccdB* plasmids by a two bp insertion, or by complete deletion of the ORF (pCMV\_SPORT6ccdB and pOSIP-CH).

Restriction enzyme sites used for cloning (MCS in pOSIP-CH) or for construction of the plasmids are shown.

The -10 and -35 sequences are missing from pMT\_BioEase-DEST, but a second potential -35 and -10 is present in the disrupted *ccdA* ORF (BPROM). Deletion of the natural promoter may allow more efficient transcription from this cryptic promoter, bypassing transcription of the entire *ccdA* ORF.







# Sequence Alignments

```

AmpR-013      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-017      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-001      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-021      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-016      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-022      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-010      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-004      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-019      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-020      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-014      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-007      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-015      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-011      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-002      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-018      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-003      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-008      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-005      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-006      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-009      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-012      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-023      GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
bla(M)-001    GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
bla(M)-002    GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
bla(M)-003    GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
bla(M)-004    GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-007corr  GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-009corr  GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
bla(M)-001corr  GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
bla(M)-003corr  GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
bla(M)-004corr  GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
** * * * *

```

```

AmpR-013      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-017      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-001      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-021      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-016      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-022      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-010      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-004      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-019      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-020      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-014      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-007      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-015      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-011      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-002      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-018      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-003      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-008      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-005      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-006      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-009      GAGATAGGTGCCTCACTGATTAAGCATTGGTAACTG-----
AmpR-012      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-023      GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
bla(M)-001    GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
bla(M)-002    GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
bla(M)-003    GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
bla(M)-004    GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
AmpR-007corr  GAGATAGGTGCCTCACTGATTAAGCATTGGCTCAGGATTAG-----
AmpR-009corr  GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
bla(M)-001corr  GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
bla(M)-003corr  GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----
bla(M)-004corr  GAGATAGGTGCCTCACTGATTAAGCATTGGCTGTATCAACAAGTTGTACAAAAAAGCTGAACGAGAAACGTAA
*****

```

ATGC = synonymous  
ATCC = conservative/semi-conservative amino acid substitution  
ATCC = non-conservative amino acid substitution

AmpR gene is bla ( $\beta$ -lactamase) from *E. coli*  
bla(M) lacks the signal peptide  
Sequences in bold are incorrectly annotated

**AmpR-009 was identified by coordinates that were +3 from the actual start codon and is actually identical to AmpR-016 - -009 merged with -016**  
**AmpR-002 is probably identical to AmpR-004 but the sequence contains an in-frame STOP codon. This is likely a sequence error as AmpR is only bacterial selection marker for the plasmids carrying AmpR-002 - -002 merged with -004**  
**bla(M)-001 coordinates did not start at the START codon. It is identical to AmpR-001 and has been merged with it**  
**Similarly bla(M)-003 is identical to AmpR-008 and has been merged with it**



## Sequence Alignments

```

AmpR-013      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-017      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-001      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-021      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-016      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-022      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-010      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-004      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-019      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-020      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-014      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-007corr  ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHWPD-----
AmpR-015      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-011      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-002      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-018      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-003      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-008      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-005      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-006      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-009      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-012      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
AmpR-023      ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
bla(M)-002    ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW-----
bla(M)-004corr ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHWLLSTSLYKKAERET
*****;*****

```

Key

- ATGC = change in nucleotide relative to consensus sequence – synonymous
- ATGC = change in nucleotide relative to consensus sequence – conservative/semi-conservative amino acid substitution
- ATGC = change in nucleotide relative to consensus sequence – non-conservative amino acid substitution
- ATGC = predicted protein sequence following a nonsense mutation



# Sequence Alignments

```

CmR-013      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-016      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-015      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-014      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-010      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-004      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-011      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-002      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-012      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-001      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-005      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-003      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-006      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-007      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-008      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-009      TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
CmR-009corr   TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCATTTCCTTAAAGGGTTATTGAGAATATGTTTTTCGCTCAGCCAATCCCTGGGTGAGTTTACCAGTTTTGATTTAAACGTG
*****

```

```

CmR-013      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-016      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-015      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-014      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-010      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-004      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-011      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-002      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-012      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-001      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-005      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-003      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-006      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-007      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-008      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-009      GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
CmR-009corr   GCCAATATGGACAACCTTCGCCCCCGTTTTCCACCATGGGCAAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCCGTTTGTGATGGCTCCAT
*****

```

```

CmR-013      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-016      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-015      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-014      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-010      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-004      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-011      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-002      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-012      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-001      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-005      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-003      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-006      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-007      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-008      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA-----
CmR-009      GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAACTATA--GAGGATCCGGCTTACTAA
CmR-009corr   GTCCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAAACCCCGCGTGGATCCGGCTTACTAA
*****

```

CmR gene is cat (chloramphenicol acetyltransferase) from *E. coli*  
**Sequences in bold are incorrectly annotated**  
**CmR-009 missing STOP codon. Corrected sequence still unique variant**

Key  
ATGC = change in nucleotide relative to consensus sequence – synonymous  
ATGC = change in nucleotide relative to consensus sequence – conservative/semi-conservative amino acid substitution  
ATGC = change in nucleotide relative to consensus sequence – non-conservative amino acid substitution

# Sequence Alignments

Name of Variant	Alignment of Translated Protein	Size (aa)	# Var aa
CmR-013	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-016	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-015	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-014	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-010	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-004	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-011	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-002	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-012	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-001	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-005	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	1
CmR-003	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-006	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-007	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-008	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	226	2
CmR-009corr	MEKKITGYTTVDISQWHRKEHFEAFQSVAACTYNQTVQLDITAPLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY *****:*****	227	1

CmR-013	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-016	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-015	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-014	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-010	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-004	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-011	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-002	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-012	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-001	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-005	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-003	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-006	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-007	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG-----
CmR-008	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG <sup>Q</sup> GNLE-DPAY
CmR-009corr	SQDVACYGENLAYFPKGF IENMFFVSNPWSFTSFDLNVANMNDNFFAPVFTMGKYITQGDKVLMLPLAIQVHHA VCDGPHVGRMLNELQQYCEWQGGG <sup>R</sup> KRRVDPAY *****

Key ATGC = change in nucleotide relative to consensus sequence – synonymous  
ATGC = change in nucleotide relative to consensus sequence – conservative/semi-conservative amino acid substitution  
ATGC = change in nucleotide relative to consensus sequence – non-conservative amino acid substitution

# Sequence Alignments

Name of Variant	# of Occur	Alignment	Size (bp)	# Var bp	Sources*
HygR-010	35	ATG---AAAGCCGGAAGCTACCCGACAGCGTTCGAGAAATTTCTGATCGAGAAGTTCGACAGCGTTCGACGCACCTGATGCAGCTTCGGAGGGCCGAAGAAGGCGAGCTTCAGCTTC	1038	222	Pr
HygR-012	26	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1026	8	Ca
HygR-008	11	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1026	0	In, Cl, OB
HygR-005	7	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1020	0/6d	Cl
HygR-002	4	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1038	13	Ag
HygR-011	4	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1020	4/6d	No
HygR-004	3	ATGTTGTTAAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1029	0	16200533
HygR-007	3	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1023	0/3d	In
HygR-001	2	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1026	1	Cl
hphMX6*	2	ATGTTGTTAAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1029	0	16200533
HygR-003	1	ATG---CGAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1026	6	In
HygR-006	1	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1026	5	10890530
HygR-009	1	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1026	4	Or
HygR-013	1	ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1023	1/3d	2194165
HygR-013corr		ATG---AAAAGCCTGAACTCACCAGCGAGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGCTCCGACCTGATGCAGCTCTCGGAGGGCCGAAGAATCTCGTCTTCAGCTTC	1023	1/3d	

HygR-010  
HygR-008  
HygR-005  
HygR-002  
HygR-011  
HygR-004  
HygR-007  
HygR-001  
hphMX6  
HygR-003  
HygR-006  
HygR-009  
HygR-013  
HygR-013corr

HygR-010  
HygR-012  
HygR-008  
HygR-005  
HygR-002  
HygR-011  
HygR-004  
HygR-007  
HygR-001  
hphMX6  
HygR-003  
HygR-006  
HygR-009  
HygR-013  
HygR-013corr

HygR-010  
HygR-012  
HygR-008  
HygR-005  
HygR-002  
HygR-011  
HygR-004  
HygR-007  
HygR-001  
hphMX6  
HygR-003  
HygR-006  
HygR-009  
HygR-013  
HygR-013corr



# Sequence Alignments

```

HygR-010  GGCTGTGCTGCACTCTGCCCGAAGCGAACCCGCCGCCACACCGCCGCGAAGGAGCTTAA
HygR-012  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGAAATAG-----
HygR-008  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-005  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-002  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-011  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-004  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-007  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-001  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
hphMX6   GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-003  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-006  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-009  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-013  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-013corr  GGCTGTGCTGAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
          *****

```

HygR gene is *aph(4)-Ia* (aminoglycoside phosphotransferase) from *E. coli*  
hphMX6 is identical to HygR-004 except it includes the *Ashbya gossypii* *TEF1* promoter and terminator, which have been trimmed out of the sequence shown  
**Sequences in bold are incorrectly annotated**  
**HygR-013 missing STOP codon. Corrected sequence still unique variant**

Name of Variant	Alignment of Translated Protein	Size (aa)	# Var
HygR-010	-MKKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
HygR-012	-MKKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	345	1
HygR-008	-MKKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
HygR-005	<b>-M</b> KPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	342	2
HygR-002	<b>-M</b> KPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	339	4
HygR-011	<b>-M</b> KPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	2
HygR-004	<b>M</b> KKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	340	0
HygR-007	<b>-M</b> KPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	1
HygR-001	<b>-M</b> KPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
hphMX6	<b>M</b> KKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	345	0
HygR-003	<b>M</b> KKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	339	3
HygR-006	-MKKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
HygR-009	-MKKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
HygR-013corr	-MKKPELTATSVEKFLIEKFDVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM *** *****	342	0

```

HygR-010  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-012  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-008  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-005  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-002  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-011  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-004  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-007  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-001  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
hphMX6   DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-003  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-006  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-009  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
HygR-013corr  DAIAAADLSQTSGFPGFPQGIQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTNDRITAVIWDSEAMFGDSQYEVANIFFWRPWL
          *****

```

```

HygR-010  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-012  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-008  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-005  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-002  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-011  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-004  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-007  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-001  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
hphMX6   ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-003  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-006  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-009  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV
HygR-013corr  ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDIVRSAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEYR--
          *****

```





# Sequence Alignments

```

KanR-009      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-021      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-005      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-010      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-015      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-008      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-019      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-014      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-001      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-002      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-003      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-004      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-006      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-007      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-011      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-012      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-013      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-017      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
KanR-018      TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAAATTCAGTTTCATTGATGCTCGATGAGTTTTCTAA
*****

```

KanR is *aph(3')-Ia* (aminoglycoside 3'-phosphotransferase) from *E. coli* *Tn5* transposon.  
 KanR-016 (5 occurrences) is *nptII* identical to Neor/KanR but has an additional N-term nos (nopaline synthase) fusion.  
 KanR-020 (50 occurrences) is *aphA-3* from *Staphylococcus aureus*. There is only one variant of this.

Name of Variant	Alignment of Translated Protein	Size (aa)	# Var aa
KanR-009	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	0
KanR-021	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	269	6/2d
KanR-005	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	0
KanR-010	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	0
KanR-015	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	0
KanR-008	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	1
KanR-019	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	0
KanR-014	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	2
KanR-001	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	1
KanR-002	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	2
KanR-003	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	1
KanR-004	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	269	2d
KanR-006	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	0
KanR-007	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	269	2d
KanR-011	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	269	2d
KanR-012	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	1
KanR-013	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	269	1/2d
KanR-017	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	1
KanR-018	MSHIQRETSCSRPRRLNSMMDADLYGKWARDNVQGSATIYRLYGKDPAPFLFKHGKGSVANDVTDENVRLNWLTEFMPLEPTIKHFIRTPDDAWLLTTAIPGKTAFOVLEEYDPSGENI	271	2

```

KanR-009      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-021      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-005      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-010      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-015      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-008      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-019      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-014      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-001      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-002      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-003      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-004      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-006      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-007      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-011      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-012      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-013      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-017      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
KanR-018      VDALAVFLRRLHSIPVCNCPFNDRVFLRLAQAQSRMNNGLVDASDFDDENRNGWPVEQVWKEMHKLFPSPDSVVTTHGDFSLDNLIF...PSLQKRLFKQYGINDPDNKQLQFHLMLDEFF
*****

```

Key ■ ATCC = change in nucleotide relative to consensus sequence – synonymy  
■ ATCC = change in nucleotide relative to consensus sequence – conservative/semi-conservative amino acid substitution  
■ ATCC = change in nucleotide relative to consensus sequence – non-conservative amino acid substitution









# Sequence Alignments

Name of Variant	# of Occur	Alignment	Size (bp)	# Var bp	Sources*
PuroR-002	23	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	1	Cl, AG
PuroR-004	15	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	2	Th
PuroR-009	10	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	3	Cl
PuroR-006	7	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	1	Cl
PuroR-007	6	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	106	Pr
PuroR-011	6	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	3	Or, 2194165
PuroR-005	2	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	2	Or
PuroR-008	2	ATG <b>CGC</b> ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	603	6	OB
PuroR-010	2	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	2	21706014
PuroR-003	1	ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCC <b>GGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	600	4	Or
PuroR-001	1	<b>ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCCGGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG</b>	<b>597</b>	<b>4</b>	<b>Cl</b>
PuroR-001corr		ATG---ACCAGGATACAAGCCACCGTGGCCCTCGCCACCCGCGACGACGTCCCCGGGGCGGTACGACACCTCGCCGCGCGTTCGCGACTACCCGCCACGGCCACACCGTCGATCCG	624	1	

\*\*\*\*\*

PuroR-002 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-004 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-009 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-006 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-007 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-011 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-005 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-008 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-010 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
PuroR-003 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
**PuroR-001 GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC**  
PuroR-001corr GACCGCCACATCGAGCGGGTACCAGGATGCAAGAACTCTTCTCACGCGCGTGGGCTCGACATCGGCAAGGTGTGGTTCGCGGACGACGGCCCGCGGTGGCGGTCTGGACCACGGC  
\*\*\*\*\*

PuroR-002 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-004 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-009 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-006 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-007 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-011 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-005 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-008 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-010 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
PuroR-003 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
**PuroR-001 GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG**  
PuroR-001corr GAGAGCGTGAAGCGGGGGCGGTGTTCCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCCCGCACCGGCCAAG  
\*\*\*\*\*

PuroR-002 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-004 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-009 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-006 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-007 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-011 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-005 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-008 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-010 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
PuroR-003 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
**PuroR-001 GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC**  
PuroR-001corr GAGCCCGGTGGTTCCCTGGCCACCGTCCGGCTCTCGCCGACACACAGGGCAAGGTTCTGGGACGCGCGTCTGTCTCCCGGAGTGGAGGCGCCGAGCGCGCGGGGTGCCCGCTTC  
\*\*\*\*\*

PuroR-002 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-004 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-009 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-006 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-007 CT**GAG**AC**AGC**CGC**CG**CA**AA**CT**CC**CTTCTACGAGCG**CG**CT**CG**GCTTACCGT**AC**CGCCGA**AG**TCGAGGTCGCCGAAGGACCGCGC**CG**ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-011 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC**CG**ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-005 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-008 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-010 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
PuroR-003 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC**CG**ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA-----  
**PuroR-001 CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA**  
PuroR-001corr CTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACCGTACCCGCGACGTCGAGGTCGCCGAAGGACCGCGC-ACCTGGTGCATGACCCGCAAGCCCGTGCCTGA  
\*\*\*\*\*

PuroR gene is *Streptomyces alboniger pac* (puromycin N-acetyltransferase)

Sequences in bold are incorrectly annotated

PuroR-001 missing STOP codon. Corrected sequence still unique variant

# Sequence Alignments

Name of Variant	Alignment of Translated Protein	Size (aa)	# Var aa
PuroR-002	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-004	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-009	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-006	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-007	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-011	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	5
PuroR-005	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-008	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	200	2
PuroR-010	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-003	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	5
PuroR-001corr	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPESEVAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK * *****.*****	207	0
PuroR-002	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-004	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-009	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-006	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-007	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-011	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-005	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-008	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-010	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-003	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-001corr	EPAWFLATVGVSPDHQKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTTRKPGATGAASRIK ***** :.*****		

# Sequence Alignments

Name of Variant	# of Occur	Alignment	Common Restriction Enzyme Sites in MCS							Size (bp)	MCS	Sources*	
			HindIII	KpnI	XhoI	SphI	PstI	EcoRV	SalI	XbaI			
lacZ-u-060	24	ATGACCATGATTAC-----									234	207,328,061,183,195	Ca,7919298
lacZ-u-031	12	ATGACCATGATTACGCCAAGCGCGCAATTAAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									579	115,MCS 2-008	2659436
lacZ-u-034	4	ATGACCATGATTAC-----									231	207,328,061,183,195	Ta,7919298
lacZ-u-036	4	ATGACCATGATTACGCCAAGCTT-----									285		
lacZ-u-057	4	ATGACCATGATTACGCCAAGCTT-----									324	207,328,061,183,195	3073106
lacZ-u-073	4	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATCCAAACGCTTACTCGAGCTCAACCATATGAAA-----									126	327	CB
lacZ-u-076	4	-----									255		CB
lacZ-u-032	3	ATGACCATGATTACGCCAAGCTTCGGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									390	115,MCS 2-008	10890530
lacZ-u-068	3	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATCCAAACGCTTACTCGAGCTCAACCATATGAAA-----									273		20040575
lacZ-u-001	2	ATGACCATGATTACGCCAAGCTT-----									309	207,328,061,183,195	33277753,8566803
lacZ-u-004	2	ATGACCATGATTAC-----									327	207,183	Ta,3323803
lacZ-u-005	2	ATGACCATGATTACGCCAAGCGCGCAATTAAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									576	051,MCS 2-008	Ag
lacZ-u-006	2	ATGACCATGATTAC-----									198	049,117,MCS1-008	Ca
lacZ-u-010	2	ATGACCATGATTACGCCAAGCGCGCAATTAAACCTCCTACTAAAGGGAAACAAAAGCTGGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									576	223,MCS2,MCS 2-008	Ag,2555794
lacZ-u-014	2	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCAACGCGATGCTCGAGCGGTTACGTATC-----									381	077,3'MCS-2	No,Qi
lacZ-u-015	2	ATGACCATGATTACGCCAAGCGCGCAATTAAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									576	115,MCS 2-008	Ag,2555794
lacZ-u-020	2	ATGACCATGATTACGCCAAGCTT-----									198		Lu
lacZ-u-022	2	ATGACCATGATTACGCCAAGCTT-----									348	097	GE
lacZ-u-028	2	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTT-----									381	328,061,195	Pr,Ne
lacZ-u-039	2	ATGACCATGATTAC-----									303	328,061,195	3333305
lacZ-u-045	2	ATGACCATGATTAC-----									291	328,061,195	3333305
lacZ-u-046	2	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCAACGCGATGCTCGAGCGGTTACGTATC-----									198	149,261	Ca
lacZ-u-051	2	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCAACGCGATGCTCGAGCGGTTACGTATC-----									366		In
lacZ-u-052	2+4	ATGACCATGATTACGCCAAGCTT-----									339	207,183	3073106
lacZ-u-036corr	1	ATGACCATGATTACGCCAAGCTT-----									339	207,183	3073106
lacZ-u-002	1	ATGACCATGATTACAAAGCGCGCGCA-----									261		No
lacZ-u-003	1	ATGACCATGATTACAAAGCGCGCAATTAAACCTCCTACTAAAGGGAAACAAAAGCTGGGTGCTCGAGACTCTCTGGCTCTCAGGAGCGAGTTGCGCTGTATACAGCTGTGCGGG-----									438	MCS 2-008	No
lacZ-u-007	1	ATGACCATGATTACG-----AA-----TTTAATACGACTCACTATAGGGAAGTTAAATTTAAATTAAGGCGCGCCATGGATCGATGTTAAACATGCATGATATCACCGCTGGCGGCCAATGATGC-----									504		Mo
lacZ-u-008	1	ATGACCATGATTACGCCAAGCTT-----									351	053	Ne
lacZ-u-009	1	ATGACCATGATTACGCCAAGCTT-----									231	328,061,195	Ta
lacZ-u-011	1	ATGACCATGATTACGCCAAGCTT-----									327	328,061,195	3323803
lacZ-u-012	1	ATGACCATGATTACGCCAAGCTTCGGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									363	115,MCS 2-008	10890530
lacZ-u-013	1	ATGACCATGATTAC-----									324	207,183	2985470
lacZ-u-016	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACCGCTTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									360		In
lacZ-u-017	1	ATGACCATGATTACGCCAAGCTTGGTACCGAGCTC-----									387	088	Ag
lacZ-u-018	1	ATGACCATGATTACGCCAAGCTTCGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCGCGCGCG-----									549	091,3'MCS-2	29
lacZ-u-019	1	ATGACCATGATTACGCCAAGCTTCGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									573	223,MCS2,MCS 2-008	Ny70625
lacZ-u-021	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									360		In
lacZ-u-023	1+1	ATGACCATGATTACGCCAAGCTTCGGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									438	115,MCS 2-008	10890530
lacZ-u-041corr	1	ATGACCATGATTACGCCAAGCTTCGGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									342	115,MCS 2-008	10890530
lacZ-u-024	1	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCGCGGTTACGTATC-----									369		In
lacZ-u-025	1	ATGACCATGATTACGCCAAGCTTCGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									573	115,MCS 2-008	2970625
lacZ-u-026	1	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCAACGCGATGCTCGAGCGGTTACGTATC-----									351	077,3'MCS-2	Ny
lacZ-u-027	1	ATGACCATGATTACGCCAAGCTTGGTACCGAGCTC-----									354		In
lacZ-u-029	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									342	107,301,045	In
lacZ-u-030	1	ATGACCATGATTACGCCAAGCTT-----									324	328,061,195	6095209
lacZ-u-033	1	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCGCGGTTACGTATC-----									333	116	No
lacZ-u-035	1	ATGACCATGATTACGCCAAGCTTCTAATACGACTCACTATAGGGAAGCTCGGTACCGCGGTTACGTATC-----									384		Pr
lacZ-u-037	1	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCGCGGTTACGTATC-----									576	223,MCS2,MCS 2-008	
lacZ-u-038	1	ATGACCATGATTACGCCAAGCTTCTAATACGACTCACTATAGGGAAGCTCGGTACCGCGGTTACGTATC-----AAGCTT-----									282		In
lacZ-u-040	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									384		Pr
lacZ-u-041	1	ATGACCATGATTACGCCAAGCTTCGGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									342	139,MCS 2-008	10890530
lacZ-u-042	1	ATGACCATGATTACGCCAAGCTTCGGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									576	115,MCS 2-008	
lacZ-u-043	1	ATGACCATGATTACGCCAAGCTTCTAATACGACTCACTATAGGGAAGCTCGGTACCGCGGTTACGTATC-----									369	142,157	7669286
lacZ-u-044	1	ATGACCATGATTACGCCAAGCTT-----GCGCGGCCAAGCTT-----									351	141,207,183	7655517
lacZ-u-047	1	ATGACCATGATTACGCCAAGCTTCGGAAATTAACCTCCTACTAAAGGGAAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCG-----									651	115,MCS 2-008	10890530
lacZ-u-048	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									369		Pr
lacZ-u-049	1	ATGACCATGATTACGCCAAGCTTGGTACCGAGCTC-----									513	328,061,195	
lacZ-u-050	1	ATGACCATGATTACGCCAAGCTTCTAATACGACTCACTATAGGGAAGCTCGGTACCGCGGTTACGTATC-----									372	157	7669286
lacZ-u-053	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									369	176,5'MCS-004	7669286
lacZ-u-054	1	ATGACCATGATTAC-----									303	207,183	3327753
lacZ-u-055	1	ATGACCATGATTACGCCAAGCGCGCAATTAAACCTCCTACTAAAGGGAAACAAAAGCTGGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									390	184,MCS2,MCS 2-008	IMAGE
lacZ-u-056	1	ATGACCATGATTACGCCAAGCTTCTAATACGACTCACTATAGGGAAGCTCGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									363	209	7669286
lacZ-u-058	1	ATGACCATGATTACGCCAAGCTTCTAATACGACTCACTATAGGGAAGCTCGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									363	209	7669286
lacZ-u-059	1	ATGACCATGATTACGCCAAGCTT-----									243	208	3327753
lacZ-u-061	1	ATGACCATGATTACGCCAAGCTTCTAATACGACTCACTATAGGGAAGCTCGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									366	224	7669286
lacZ-u-062	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									121	229	CB
lacZ-u-063	1	ATGACCATGATTAC-----									507	328,061,195	Ne
lacZ-u-064	1	-----									263		CB
lacZ-u-065	1	ATGACCATGATTACGCCAAGCTT-----									507	207,183	Ne
lacZ-u-066	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									115	268	CB
lacZ-u-067	1	ATGACCATGATTAC-----									270	328,061,195	Ta
lacZ-u-069	1	-----									254		CB
lacZ-u-070	1	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									267	328,061,195	Th
lacZ-u-071	1	ATGACCATGATTACGCCAAGCTTATTTAGGTGACACTATAGAATACTCAAGCTATGTCATC-----AAGCTTGGTACCGAGCTC-----									606		In
lacZ-u-072	1	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									477	328,061,195	Th
lacZ-u-074	1	ATGACCATGATTACGCC-----									309	328,061,195	8566803
lacZ-u-075	1	ATGACCATGATTACGCCAAGCTT-----CTAATACGACTCACTATAGGGAAGCTCGGTACCGGGCCCCCCCCTCGAGCTCGACGGTATCGATAAGCTTGATATC-----									291	097	Th

\*\*\* \*\*

# Sequence Alignments

	SpeI	BamHI	SpeI	EagI	SmaI	KpnI	SacI	PstI	EcoRI	BamHI	AflIII	EcoRI	PstI	EcoRV	HindIII	EcoRI	SalI	XhoI	SmaI	SpeI	BamHI	NotI	SacI	KpnI	SmaI	XbaI	BamHI		
lacZ- $\alpha$ -060									-GAATTC																				
lacZ- $\alpha$ -031	ACTAGT	GGATCC							CCGGGCTCGAGGAATTC																				
lacZ- $\alpha$ -034									-GAATTC																				
lacZ- $\alpha$ -036		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -057		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -073									CGGCCGC												ACTAGTGGATCCTGAT								
lacZ- $\alpha$ -076									TCAGGCTGTATCCGGAGAATTC																				
lacZ- $\alpha$ -032	ACTAGT	GGATCC							CCGGGCTCGAGGAATTC																				
lacZ- $\alpha$ -068		GGATCC							GACGCACGTGC																				
lacZ- $\alpha$ -001		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -004									-GAATTC																				
lacZ- $\alpha$ -005		CCC	GGG	CGG	ATC	CC			CCGGGCTCGAGGAATTC																				
lacZ- $\alpha$ -006									-GAATTC																				
lacZ- $\alpha$ -010									-GAATTC																				
lacZ- $\alpha$ -014		GGATCC							AGAAATTC																				
lacZ- $\alpha$ -015	ACTAGT	GGATCC							CCGGGCTCGAGGAATTC																				
lacZ- $\alpha$ -020									CGAATTC																				
lacZ- $\alpha$ -022		TGC	ATC						GCGAGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -028		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -039									-GAATTC																				
lacZ- $\alpha$ -045									-GAATTC																				
lacZ- $\alpha$ -046		GGATCC							AGAAATTC																				
lacZ- $\alpha$ -051		GGATCC							AGAAATTC																				
lacZ- $\alpha$ -052		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -036corr		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -002									GGAAATTC																				
lacZ- $\alpha$ -003	CCG	CAAG	CTT	CG	AA	CG	CG	TAT	CG	AT	CG	CT	CG	AG	CG	CG	CT	TG	T	CA	GA	ATTC							
lacZ- $\alpha$ -007	TAG	CAG	ATC						TG	CC	GG	CC	CA	CG	TG	GG	CC	CA	T	AT	CG	CG	CG	CG	CG	CG	CG	CG	CG
lacZ- $\alpha$ -008		CTT	A	A	T	A	A	G	A	T	A	A	G	A	T	A	A	G	A	T	A	A	G	A	T	A	A	G	
lacZ- $\alpha$ -009		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -011		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -012	ACTAGT	GGATCC							CCGGGCTCGAGGAATTC																				
lacZ- $\alpha$ -013									-GAATTC																				
lacZ- $\alpha$ -016		GGATCC							ACTAGTAA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	
lacZ- $\alpha$ -017		GGATCC							ACTAGTAA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	
lacZ- $\alpha$ -018	ACTAGT	GGATCC							AA	GA	A	A	T	T															
lacZ- $\alpha$ -019									AA	GA	A	A	T	T															
lacZ- $\alpha$ -021		GGATCC							ACTAGTAA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	
lacZ- $\alpha$ -023	ACTAGT	GGATCC							CC	GG	G	T	C	G	A	G	A	T	T	C									
lacZ- $\alpha$ -041corr	ACTAGT	GGATCC							CC	GG	G	T	C	G	A	G	A	T	T	C									
lacZ- $\alpha$ -024		GT	A	C	C	G	T	C																					
lacZ- $\alpha$ -025	ACTAGT	GGATCC							CC	GG	G	T	C	G	A	G	A	T	T	C									
lacZ- $\alpha$ -026		GGATCC							AGAAATTC																				
lacZ- $\alpha$ -027		GGATCC							ACTAGTAA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	
lacZ- $\alpha$ -029		GGATCC							ACTAGTAA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	
lacZ- $\alpha$ -030		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -033	CTA	CTA	G	T	A	T	C	A	T	G	G	A	T	C	G	A	T	C	C										
lacZ- $\alpha$ -035		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -037		GT	A	C	C	G	T	C																					
lacZ- $\alpha$ -038		GGATCC							AGAAATTC																				
lacZ- $\alpha$ -040									CGGCCGC																				
lacZ- $\alpha$ -041	ACTAGT	GGATCC							CCGGGCTCGAGGAATTC																				
lacZ- $\alpha$ -042	ACTAGT	GGATCC							CCGGGCTCGAGGAATTC																				
lacZ- $\alpha$ -043		CAG	A	T	C	T	T	C	G	A	T	C	G	C	G	C	A	C	G	T	A	C	G	C	A	C	G	T	
lacZ- $\alpha$ -044		GGATCC							CCGGGTACCGAGCTCGAATTC																				
lacZ- $\alpha$ -047	ACTAGT	GGATCC							CCGGGCTCGAGGAATTC																				
lacZ- $\alpha$ -048									CGGCCGC																				
lacZ- $\alpha$ -049		GGATCC							ACTAGTAA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	
lacZ- $\alpha$ -050		CAG	A	T	C	T	T	C	G	A	T	C	G	C	G	C	A	C	G	T	A	C	G	C	A	C	G	T	
lacZ- $\alpha$ -053		GGATCC							ACTAGTAA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	
lacZ- $\alpha$ -054									-GAATTC																				
lacZ- $\alpha$ -055									-GAATTC																				
lacZ- $\alpha$ -056		TGG	A	T	C				ACGAATTC																				
lacZ- $\alpha$ -058		TGG	A	T	C				ACGAATTC																				
lacZ- $\alpha$ -059		GGATCC							CCC	A	T	C	G	A	T	C	G	A	T	C	G	A	T	C	G	A	T	C	
lacZ- $\alpha$ -061		CT	G	T	A	C	T	C	G	A	G	A	G	C	G														



# Sequence Alignments

```
lacZ-α-060      TGC TAG-----
lacZ-α-031      TGGCGCGACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-034      TGA-----
lacZ-α-036      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATA-----
lacZ-α-057      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATATGTTGCACTCTCAGTACAATCTGCTCTGATGCGGCATAG-----
lacZ-α-073      -----
lacZ-α-076      TGGG-----CGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-032      TGGCGCGAAATTTGTAACGTTAATGTTATCGATCATGA-----
lacZ-α-068      -----
lacZ-α-001      TGGCATTTTATCAACAAGCGCGCTCCCGTCAAGTCAGCGTAAATGCTTGCAGTGTACAAACCAATTAACCAATTTCTGA-----
lacZ-α-004      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATACGTTCAAAGCAACCATAGTACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-005      TGG---GACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-006      -----
lacZ-α-010      TGGG---ACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-014      TGG---AAATTGTAA-----
lacZ-α-015      TGG---GACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-020      TGGCGCTAA-----
lacZ-α-022      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATATGTTGCACTCTCAGTACAATCTGCTCTGATGCGGCATAG-----
lacZ-α-028      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATATGTTGCACTCTCAGTACAATCTGCTCTGATGCGGCATAG-----
lacZ-α-039      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATAGATCCGGTACGCTGTTAA-----
lacZ-α-045      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATAGGGTAA-----
lacZ-α-046      TGG-----AAATTGTAA-----
lacZ-α-051      TGG-----AAATTGTAA-----
lacZ-α-052      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATATATCGCTTGGGCCATTTCTCATGAAGAATATCTTGAATTTATTGTCATATTACTAG-----
lacZ-α-036corr  TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATATATCGCTTGGGCCATTTCTCATGAAGAATATCTTGAATTTATTGTCATATTACTAG-----
lacZ-α-002      TGG-----AAATTGTAA-----
lacZ-α-003      TGG-----AAATTGTAA-----
lacZ-α-007      TGGGA---AATTTGTAACGTTAATATTTTGTAAATATTTTGTAA-----
lacZ-α-008      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATATGTTGCACTCTCAGTACAATCTGCTCTGATGCGGCATAG-----
lacZ-α-009      TGA-----
lacZ-α-011      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATACGTTCAAAGCAACCATAGTACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-012      TGG-----AAATTGTAA-----
lacZ-α-013      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATATGTTGCACTCTCAGTACAATCTGCTCTGATGCGGCATAG-----
lacZ-α-016      -----
lacZ-α-017      TGG---ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-018      TGG---AGATCCAAATTTTAA-----
lacZ-α-019      TGGG---ACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-021      -----
lacZ-α-023      TGGCGCGAAATTTGTAACGTTAATGTTTATCGTACCCTACTCAAATAAGTCTCAAGATACAGCTCTCAGAAGACCAAAGGGCTATTGA-----
lacZ-α-041corr  TGGCGCGAAATTTGTAACGTTAATGTTTATCGTACCCTACTCAAATAAGTCTCAAGATACAGCTCTCAGAAGACCAAAGGGCTATTGA-----
lacZ-α-024      TGG-----ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-025      TGG---GACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-026      TGG-----AAATTGTAA-----
lacZ-α-027      TGG-----ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-029      TGGG---ACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-030      -----
lacZ-α-033      TGGG---AAATTGTAACGTTAATATTTTGTAAATTCGGGTTAAATTTTGTAA-----
lacZ-α-035      TGGCGCCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGGTATTTACACCCGATATGTTGCACTCTCAGTACAATCTGCTCTGATGCGGCATAG-----
lacZ-α-037      TGG-----ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-038      TGGCGCTTCGCTTGGTAA-----
lacZ-α-040      TGG-----ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-041      TGGCGCGACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-042      TGGCGCGACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-043      TGGCGCTTCGCTTGGTAA-----
lacZ-α-044      -----
lacZ-α-047      TGGCGCGAAATTTGTAACGTTAATGTTTAAAGTTACACCACAATATATCTGCCAAGATCTCTAAATTCGGGGATCGGAATCCAGAAGCCGAGAGGTTGCGCGCTTTTCGGGCTTTTCTTTTCAAAAAAAAAAATTTATAAACGATCTGTTGCGGCGGCGCGGGTGTGGGCAA...
lacZ-α-048      TGG-----ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-049      TGG-----ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-050      TGGCGCTTCGCTTGGTAA-----
lacZ-α-053      TGGCGCTTCGCTTGGTAA-----
lacZ-α-054      TGGCATTTTATCAACAAGCGCGCTCCCGTCAAGTCAGCGTAAATGCTTGCAGTGTACAAACCAATTAACCAATTTCTGA-----
lacZ-α-055      TGGG---AAATTGTAA-----
lacZ-α-056      TGGCGCTTCGCTTGGTAA-----
lacZ-α-058      TGGCGCTTCGCTTGGTAA-----
lacZ-α-059      TGA-----
lacZ-α-061      TGGCGCTTCGCTTGGTAA-----
lacZ-α-062      -----
lacZ-α-063      TGGCGCTTCGCTTGGTAA-----
lacZ-α-064      TGG-----ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-065      TGGCGCTTCGCTTGGTAA-----
lacZ-α-066      TGGCGCTTCGCTTGGTAA-----
lacZ-α-067      TGGCGCTTCGCTTGGTAA-----
lacZ-α-069      TGG-----ACGGCGCCCTGTAGCGGCGCATTAA-----
lacZ-α-070      TGG-----AAATTGTAA-----
lacZ-α-071      TGGG---ACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-072      TGGG---ACGGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGGCGAGCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTCCCGCTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTCT...
lacZ-α-074      TGGCATTTTATCAACAAGCGCGCTCCCGTCAAGTCAGCGTAAATGCTTGCAGTGTACAAACCAATTAACCAATTTCTGA-----
lacZ-α-075      TGG-----AAATTGTAA-----
```

Sequences in yellow correspond to the fl origin of replication, which is in frame with lacZ-α in these cases.

lacZ-α-016 and lacZ-α-021 are missing STOP codons because these variants are in frame fusions with ccdB. -016 identical to -053. -021 identical to -029

Sequences in bold are incorrectly annotated

lacZ-α-041 sequence has a single bp deletion that generates a frame-shift and premature STOP codon. Very likely sequence error corrected above. -041 merged with -023

lacZ-α-036 is annotated without a STOP codon. Sequence actually extends another 54 bp and is identical to lacZ-α-052. -036 merged with -052.

# Sequence Alignments

Name of Variant	Alignment	Plasmids
lacZ- $\alpha$ -040	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCACCTAGT-	pGEM-T
lacZ- $\alpha$ -062	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCACCTAGTA	pSpark Done (linearized)
lacZ- $\alpha$ -064	-----	pSpark Done (linearized)
lacZ- $\alpha$ -048	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG-----GTCGACCTGCAGGCGGCCGCACTAGT-	pGEM-T Easy
lacZ- $\alpha$ -066	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG-----GTCGACCTGCAGGCGGCCGCACTAGTA	pSpark II (linearized)
lacZ- $\alpha$ -069	-----	pSpark II (linearized)
lacZ- $\alpha$ -073	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTACTCGAGCTCAACATATG---AAAGTCGACCTGCAGGCGGCCGCACTAGTG	pSpark III, IV and V (linearized)
lacZ- $\alpha$ -076	-----	pSpark III, IV and V (linearized)

lacZ- $\alpha$ -040	-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCGTTT
lacZ- $\alpha$ -062	-----GAT-----
lacZ- $\alpha$ -064	-----ATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCGTTT
lacZ- $\alpha$ -048	-----GATATC-----CCGCGG---CCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCGTTT
lacZ- $\alpha$ -066	-----GAT-----
lacZ- $\alpha$ -069	-----ATC-----CCGCGG---CCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCGTTT
lacZ- $\alpha$ -073	GATCCTGAT-----
lacZ- $\alpha$ -076	-----ATCAGGCTGATCCGGAGAATTCGTTAAAC--CCATGG-----GGGCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCGTTT

lacZ- $\alpha$ -040	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA
lacZ- $\alpha$ -062	-----
lacZ- $\alpha$ -064	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA
lacZ- $\alpha$ -048	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA
lacZ- $\alpha$ -066	-----
lacZ- $\alpha$ -069	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA
lacZ- $\alpha$ -073	-----
lacZ- $\alpha$ -076	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA

## Combined lacZ- $\alpha$ Fragments from Recircularized pSpark Series

Name of Variant	Alignment	Common Restriction Enzyme Sites in MCS
lacZ- $\alpha$ -040	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG--GTCGACCTGCAGGCGGCCGGAATTCACCTA	SacI NdeI SalI PstI NotI EcoRI SpeI
lacZ- $\alpha$ -062+064	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG--GTCGACCTGCAGGCGGCCGGAATTCACCTA	
lacZ- $\alpha$ -048	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG--GTCGACCTGCAGGCGGCCGCGC-----ACTA	
lacZ- $\alpha$ -066+069	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG--GTCGACCTGCAGGCGGCCGCGC-----ACTA	
lacZ- $\alpha$ -073+076	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTACTCGAGCTCAACATATG---AAAGTCGACCTGCAGGCGGCCGCGC-----ACTA	
	*****	*****
	EcoRV SacII EcoRI SacII NotI SphI	
lacZ- $\alpha$ -040	GT-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCG	
lacZ- $\alpha$ -062+064	GT-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCG	
lacZ- $\alpha$ -048	GT-----GATATCCCGCGGCC-----ATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCG	
lacZ- $\alpha$ -066+069	GT-----GATATCCCGCGGCC-----ATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCG	
lacZ- $\alpha$ -073+076	GTGGATCTGATATCAGGCTGATCCGGGAATTC-GTTTAAACCCATGG-----GGGCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCG	
	** *****	*****
lacZ- $\alpha$ -040	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
lacZ- $\alpha$ -062+064	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
lacZ- $\alpha$ -048	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
lacZ- $\alpha$ -066+069	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
lacZ- $\alpha$ -073+076	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGCGAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
	*****	*****

lacZ- $\alpha$ -062, lacZ- $\alpha$ -066 and lacZ- $\alpha$ -073 are linearized plasmids and have added A at end possibly because vector is like Topo with 3' A overhang to make PCR cloning more efficient. A is removed in alignment. Plasmid comes linearized by EcoRV.

lacZ- $\alpha$ -064, lacZ- $\alpha$ -069 and lacZ- $\alpha$ -076 missing A at start to complete EcoRV site of linearization (blunt) added in alignment.

# Sequence Alignments

Name of Variant	Alignment of Translated Protein	Size (aa)
lacZ- <i>u</i> -041	MTMIIPTSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRN----- <b>RYQAYRYRRPGRGARYPIRPIVSRITIHWPFSYVNVVTGKTLALPNLILQHP</b>	113
lacZ- <i>u</i> -041 <sub>corr</sub>	MTMIIPTSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	141
lacZ- <i>u</i> -023	MTMIIPTSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	145
lacZ- <i>u</i> -032	MTMIIPTSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	129
lacZ- <i>u</i> -047	MTMIIPTSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	216
lacZ- <i>u</i> -012	MTMIIPTSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	120
lacZ- <i>u</i> -015	MTMIIPTSAQLTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESYARSLAVVLQRDWDENPGVTQINRLLAAHPP	191
lacZ- <i>u</i> -031	MTMIIPTSAQLTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESYARSLAVVLQRDWDENPGVTQINRLLAAHPP	192
lacZ- <i>u</i> -005	MTMIIPTSAQLTLTKGNKSWSSSTAVAAALARADPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESYARSLAVVLQRDWDENPGVTQINRLLAAHPP	191
lacZ- <i>u</i> -025	MTMIIPTSSKLTTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	190
lacZ- <i>u</i> -042	MTMIIPTSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNLSLISLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	191
lacZ- <i>u</i> -019	MTMIIPTSSKLTTLTKGNKSWVPPGPPSRSTVSIISLISNSCSPGD-----PLVLERPPPRWSSNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	190
lacZ- <i>u</i> -037	MTMIIPTSSSELTLTKGNKSWVPPGPPSRSTVSIISLISNSCSPGD-----PLVLERPPPRWSSNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	191
lacZ- <i>u</i> -010	MTMIIPTSAQLTLTKGNKSWVPPGPPSRSTVSIISLISNSCSPGD-----PLVLERPPPRWSSNS--PYSESYARSLAVVLQRDWDENPGVTQINRLLAAHPP	191
lacZ- <i>u</i> -055	MTMIIPTSAQLTLTKGNKSWVPPGPPSRSTVSIISLISNSCSPGDQQTMD-----PLVLERPPPRWSSNS--PYSESYARSLAVVLQRDWDENPGVTQINRLLAAHPP	129
lacZ- <i>u</i> -038	MTMIIPTSSSELTLTKGTSFAGLNEFALKGEFAAAKF-----NS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	94
lacZ- <i>u</i> -028	MTMIIPTSS--YLGDTIEY--SSLHACR-----STLED--PRVPSNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	126
lacZ- <i>u</i> -017	MTIITPSSKLTTLTKGNKSW--SSRACR-----STLVDPKNSKSFSSVLLERFRAHRFSTRVGVQVSPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	128
lacZ- <i>u</i> -049	MTMIIPTSSSELTLTKGNK--SLHACR-----STLED--PRVPSNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	170
lacZ- <i>u</i> -008	MTMIIPT--SLHACR-----FKQSTLDDL--IKDPARPRVPSNS--PYSESY--LAVVLQRDWDENPGVTQINRLLAAHPP	116
lacZ- <i>u</i> -030	MTMIIPT--SLHACR-----STLED--PRVPSNS--PYSESY--LAVVLQRDWDENPGVTQINRLLAAHPP	107
lacZ- <i>u</i> -065	MTMIIPT--SLHACR-----STLED--PRVPSNS--PYSESY--LAVVLQRDWDENPGVTQINRLLAAHPP	168
lacZ- <i>u</i> -009	MTMIIPT--SLHACR-----STLED--PRVPSNS--PYSESY--LAVVLQRDWDENPGVTQINRLLAAHPP	76
lacZ- <i>u</i> -057	MTMIIPT--SLHACR-----STLED--PRVPSNS--PYSESY--LAVVLQRDWDENPGVTQINRLLAAHPP	107
lacZ- <i>u</i> -036 <sub>corr</sub> /052	MTMIIPT--SLHACR-----STLED--PRVPSNS-- <b>LAVVLQRDWDENPGVTQINRLLAAHPP</b>	95
lacZ- <i>u</i> -011	MTMIIPT--SLHACR-----STLED--PRVPSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	112
lacZ- <i>u</i> -001	MTMIIPT--SLHACR-----STLED--PRVPSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	384
lacZ- <i>u</i> -072	MTMIIPTSSNTTHYRE--SLHACR-----STLED--PRVPSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	108
lacZ- <i>u</i> -070	MTMIIPTSSNTTHYRE--SLHACR-----STLED--PRVPSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	158
lacZ- <i>u</i> -035	MTMIIPTSSNTTHYRET--SLHACR-----STLED--PRVPSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	88
lacZ- <i>u</i> -033	MTMIIPTSSNTTHYRE--SLHACR-----STLED--PRVPSNSYSIVSPKS--NSLAVVLQRDWDENPGVTQINRLLAAHPP	127
lacZ- <i>u</i> -075	MTMIIPTSSNTTHYRE--SLHAGLCSRRARDPISRCIRE-----STLED--LLVIWISD--PRVPSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	110
lacZ- <i>u</i> -022	MTMIIPT--SLHAGLCSRRARDPISRCIRE-----VPSSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	96
lacZ- <i>u</i> -059	MTMIIPT--SLHAPRGVDSRGSPIDG-----VPSSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	115
lacZ- <i>u</i> -016	MTMIIPTSYLGDTIEYSSYASSLVPSSDPLVTAASVLEFALKGEFCRYPSHWRLPEH-----VPSSNS--LAVVLQRDWDENPGVTQINRLLAAHPP	80
lacZ- <i>u</i> -029	MTMIIPTSYLGDTIEYSSYASSLVPSSDPLVTAASVLEF--CRYPSHWRLPEH-----ASRGPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	120
lacZ- <i>u</i> -021	MTMIIPTSYLDALIEYSSYASSLVPSSDPLVTAASVLEFALKGEFCRYPSHWRLPEH-----ASRGPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	114
lacZ- <i>u</i> -027	MTMIIPT--SLVPSDDPLVTAASVLEFALKGEFCRYPSHWRLPEH-----ASRGPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	120
lacZ- <i>u</i> -071	MTMIIPTSYLGDTIEYSSYASSLVPSSDPLVTAASVLEF--CRYPSHWRLPEH-----ASRGPNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	117
lacZ- <i>u</i> -040	MTMIIPTSYLGDTIEYSSYASNALGALPYGRPAGREFTSDIEF--PRPP--WRPGACDV-----GPNNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	201
lacZ- <i>u</i> -048	MTMIIPTSYLGDTIEYSSYASNALGALPYGRPAGGR--TSDI--PR-P--WRPGACDV-----GPNNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	127
lacZ- <i>u</i> -068	MTMIIPTSYLGETIEYSSLHAIIRIVDNGSDARA-----GPNNS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	122
lacZ- <i>u</i> -044	MTMIIPT--SWRAKLAQLQVDSRGSPTGLE-----NS--PYSESY--NSLAVVLQRDWDENPGVTQINRLLAAHPP	90
lacZ- <i>u</i> -002	MTMIIPTP-----FLINNS-----LAVVLQRDWDENPGVTQINRLLAAHPP	116
lacZ- <i>u</i> -003	MTMIIPTP-----NLINISGGGGGARDPLGLQERVLAGLHVYTVAVRQASNAIYRWYRRRPPAGAPVQNSDPRAPGISPWYISFLKLNKIILLANQDLSLAVVLQRDWDENPGVTQINRLLAAHPP	86
lacZ- <i>u</i> -006	MTMIIPT-----NSRGSVDLQPSLALAVVLQRDWDENPGVTQINRLLAAHPP	145
lacZ- <i>u</i> -063	MTMIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	66
lacZ- <i>u</i> -004	MTMIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	168
lacZ- <i>u</i> -045	MTMIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	108
lacZ- <i>u</i> -045	MTMIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	96
lacZ- <i>u</i> -039	MTMIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	100
lacZ- <i>u</i> -013	MTMIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	107
lacZ- <i>u</i> -034	MTMIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	76
lacZ- <i>u</i> -067	MIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	89
lacZ- <i>u</i> -054	IIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	100
lacZ- <i>u</i> -074	MTMIIPT-----DPLESTCRHA-----SL--LAVVLQRDWDENPGVTQINRLLAAHPP	102
lacZ- <i>u</i> -060	MTMIIPT-----DPLESTCRHA-----SLALAVVLQRDWDENPGVTQINRLLAAHPP	77
lacZ- <i>u</i> -007	MTMIIPT-----NLIRLTIGLNLIKARPWIDVNMHDIIRGATSASRSRGTGGPYGPAANSSVPTYSRSDRVSNNWFKRPGTWDPLESTCRHASFGPRDNRVLSLALAVVLQRDWDENPGVTQINRLLAAHPP	167
lacZ- <i>u</i> -024	MTMIIPTSSNTTHYRESS-----WYACRYRS--GIPGAAAVTSELVD-----SRGSKLTYACMRHSS--SIVSPKF--NSLAVVLQRDWDENPGVTQINRLLAAHPP	122
lacZ- <i>u</i> -018	MTMIIPTSSNTTHYRESSLVLFAEAQNKRSITLAVTGVTKTAAAKFERQHIDSGTGGSDDDDKSPGSSKLTQRYVSDPLGDTIEFVRLVDKLLLEPRALDHTCGGSPSSR-----PHNSLAVVLQRDWDENPGVTQINRLLAAHPP	182
lacZ- <i>u</i> -014	MTMIIPTSSNTTHYRESS-----VPRMLQTRYVSDP--EFVISEF--VDKLLLEPRALDHTCGGSPSSRPLYSIVSPKWHNSLAVVLQRDWDENPGVTQINRLLAAHPP	126
lacZ- <i>u</i> -026	MTMIIPTSSNTTHYRESS-----VPRMLQTRYVSDP--EFVISEF--VDKLLLEPRALDHTCGGSPSSRPLYSIVSPKWHNSLAVVLQRDWDENPGVTQINRLLAAHPP	126
lacZ- <i>u</i> -051	MTMIIPTSSNTTHYRESS-----VPRMLQTRYVSDP--EFVISEF--VDKLLLEPRALDHTCGGSPSSRPLYSIVSPKWHNSLAVVLQRDWDENPGVTQINRLLAAHPP	126
lacZ- <i>u</i> -046	MTMIIPTSSNTTHYRESS-----VPRMLQTRYVSDP--EFVISEF--VDKLLLEPRALDHTCGGSPSSR-----PHNSLAVVLQRDWDENPGVTQINRLLAAHPP	116
lacZ- <i>u</i> -043	MTMIIPTSYVIRLITSGQIFECIARTVRLLEEF-----LQDIIWHEASHGVDVTSRY--LGEIWLPLVKALSESYY--GLAVVLQRDWDENPGVTQINRLLAAHPP	123
lacZ- <i>u</i> -050	MTMIIPTSYVIRLITIGQIFECIARTVRLLEEF-----LQDIIWHEASHGVDVTSRY--LGEIWLPLVKALSESYY--GLAVVLQRDWDENPGVTQINRLLAAHPP	123
lacZ- <i>u</i> -056	MTMIIPTSYVIRLITSGARAEAGWRQAS-----LQDIIWHEAFASFGRDASPD--VQACVDPLVKALSESYY--GLAVVLQRDWDENPGVTQINRLLAAHPP	120
lacZ- <i>u</i> -058	MTMIIPTSYVIRLITIGARAEAGWRQAS-----LQDIIWHEAFASFGRDASPD--VQACVDPLVKALSESYY--GLAVVLQRDWDENPGVTQINRLLAAHPP	120
lacZ- <i>u</i> -053	MTMIIPTSYVIRLITSGYQSSLSRGTGDVSHG-----EAWIHDILQEFLETVRARECIQIFPLVKALSESYY--GLAVVLQRDWDENPGVTQINRLLAAHPP	122
lacZ- <i>u</i> -061	MTMIIPTSYVIRLITSGSTACLYTPEDAFGRAS-----EFWIHDILQEFWFRLAG--AIEWAPLVKALSESYY--GLAVVLQRDWDENPGVTQINRLLAAHPP	121
lacZ- <i>u</i> -020	MTMIIPTSLSTN-----SLAVVLQRDWDENPGVTQINRLLAAHPP	65
lacZ- <i>u</i> -062+064	MTMIIPTSYLGDTIEYSSYASNALGALPYGRPAGREFTSRYRIIPAAAMAAGSMRKR--AQFAL-----	61
lacZ- <i>u</i> -066+069	MTMIIPTSYLGDTIEYSSYASNALGALPYGRPAGGR--TSRY--PAAMAAGSMRKR--AQFAL-----	56
lacZ- <i>u</i> -073+076	MTMIIPTSYLGDTIEYSSYASNALLELHKMKVDLQA--AALVDPDIRPDPENSKFPMGAQFAL-----	60

:\*\*\*\*  
\*\*\*\*\*

# Sequence Alignments

```

lacZ- $\alpha$ -041      LSPAGVIAKRPAPIALPNSCAA-----
lacZ- $\alpha$ -041_corr  FASWRNSEEARTDRPSQQLRSLNGEWREI VNNVNI VPLLQKQRYSLRRPKGY-----
lacZ- $\alpha$ -023      FASWRNSEEARTDRPSQQLRSLNGEWREI VNNVNI VPLLQKQRYSLRRPKGY-----
lacZ- $\alpha$ -032      FASWRNSEEARTDRPSQQLRSLNGEWREI VNNVNI VDT-----
lacZ- $\alpha$ -047      FASWRNSEEARTDRPSQQLRSLNGEWREI VNNVNVVTPQYILPRLSLIPGIGNPEAREVAAPRAFSPFKKKIYKTIICGRPPGCGQRRWRSTVGNRLRLSTGGAGARSALSTSQGRPIIDIIYIHN
lacZ- $\alpha$ -012      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -015      FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -031      FASWRNSEEARTDRPSQQLRSLNGEWRDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -005      FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -025      FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -042      FASWRNSEEARTDRPSQQLRSLNGEWRDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -019      FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -037      FASWRNSEEARTDRPSQQLRSLNGEWRDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -010      FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -055      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -038      FASWRNSEEARTDRPSQQLRSL-----
lacZ- $\alpha$ -028      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITWCTLSTICSDAA-----
lacZ- $\alpha$ -017      FASWRNSEEARTDRPSQQLRSLNGEWRSNF-----
lacZ- $\alpha$ -049      FASWRNSEEARTDRPSQQLRSLNGEWRDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -008      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITWCTLSTICSDAA-----
lacZ- $\alpha$ -030      FASWRNSEEARTDRPSQQLRSLNGARSFRFLPLSRHVRRLSPSSSKSGAPFRVPI-----
lacZ- $\alpha$ -065      FASWRNSEEARTDRPSQQLRSLNGEWRFAWFPAPAEVPSWLECDLPEADTVVPSNWMQHGVDAPYITNVITYITVNPFFVPTENPTGCYSLTFNVDESWLQEGQTRIIFDGVPIG
lacZ- $\alpha$ -009      FASWRNSEEARTDRPSQQLRSLNGE-----
lacZ- $\alpha$ -057      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITWCTLSTICSDAA-----
lacZ- $\alpha$ -036      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRIT
lacZ- $\alpha$ -036corr/052 FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITRWAIIIMKNIINLLSY-----
lacZ- $\alpha$ -011      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITRQSNHSTRPVAAH-----
lacZ- $\alpha$ -001      FASWRNSEEARTDRPSQQLRSLNGE DLFNKAAVPSQORNALPVLQPINQF-----
lacZ- $\alpha$ -072      FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFSALRHLDPKKLD
lacZ- $\alpha$ -070      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -035      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITWCTLSTICSDAA-----
lacZ- $\alpha$ -033      FASWRNSEEARTDRPSQQLRSLNGEWEIVNVNILLKPFALNFC-----
lacZ- $\alpha$ -075      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -022      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITWCTLSTICSDAA-----
lacZ- $\alpha$ -059      FASWRNSEEARTDRPSQQLRSLNGE-----
lacZ- $\alpha$ -016      FASWRNSEEARTDRPSQQLRSL-----
lacZ- $\alpha$ -029      FASWRNSEEARTDRPSQQLRSL-----
lacZ- $\alpha$ -021      FASWRNSEEARTDRPSQQLRSL-----
lacZ- $\alpha$ -027      FASWRNSEEARTDRPSQQLRSLNGEWRTPVAAH-----
lacZ- $\alpha$ -071      FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPQALNRGLPLGFRFRALRHLDKPKLDLGDGSRSGPSP
lacZ- $\alpha$ -040      FASWRNSEEARTDRPSQQLRSLNGEWRTPVAAH-----
lacZ- $\alpha$ -048      FASWRNSEEARTDRPSQQLRSLNGEWRTPVAAH-----
lacZ- $\alpha$ -068      FASWRNSEEARTDRPSQQLRS-----
lacZ- $\alpha$ -044      FASWRNSEEARTDRPSQQLRSLNGE SFRFLPFLSRHVRRLSPSSSKSGAPFRVPI-----
lacZ- $\alpha$ -002      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -003      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -006      FASWRNSEEARTDRPSQQLRSL-----
lacZ- $\alpha$ -063      FASWRNSEEARTDRPSQQLRSLNGEWRFAWFPAPAEVPSWLECDLPEADTVVPSNWMQHGVDAPYITNVITYITVNPFFVPTENPTGCYSLTFNVDESWLQEGQTRIIFDGVPIG
lacZ- $\alpha$ -004      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITRQSNHSTRPVAAH-----
lacZ- $\alpha$ -045      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRIG-----
lacZ- $\alpha$ -039      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRIDPVR-----
lacZ- $\alpha$ -013      FASWRNSEEARTDRPSQQLRSLNGEWRMLMRYFLLTHLCCGISHRITWCTLSTICSDAA-----
lacZ- $\alpha$ -034      FASWRNSEEARTDRPSQQLRSLNGE-----
lacZ- $\alpha$ -067      FASWRNSEEARTDRPSQQLRSLNGEWRLRNHYHYHDINL-----
lacZ- $\alpha$ -054      FASWRNSEEARTDRPSQQLRSLNGE DLFNKAAVPSQORNALPVLQPINQF-----
lacZ- $\alpha$ -074      FASWRNSEEARTDRPSQQLRSLNGE DLFNKAAVPSQORNALPVLQPINQF-----
lacZ- $\alpha$ -060      FASWRNSEEARTDRPSQQLRSLNGE D-----
lacZ- $\alpha$ -007      FASWRNSEEARTDRPSQQLRSLNGEWEIVNVNILLKPFALNFC-----
lacZ- $\alpha$ -024      FASWRNSEEARTDRPSQQLRSLNGEWRTPVAAH-----
lacZ- $\alpha$ -018      FASWRNSEEARTDRPSQQLRSLNGEWEIVNVNILLKPFALNFC-----
lacZ- $\alpha$ -014      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -026      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -051      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -046      FASWRNSEEARTDRPSQQLRSLNGEWKLV-----
lacZ- $\alpha$ -043      FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- $\alpha$ -050      FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- $\alpha$ -056      FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- $\alpha$ -058      FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- $\alpha$ -053      FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- $\alpha$ -061      FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- $\alpha$ -020      FASWRNSEEARTDRPSQQLRSLNGEWR-----
lacZ- $\alpha$ -062+064      -----
lacZ- $\alpha$ -066+069      -----
lacZ- $\alpha$ -073+076      -----
*****

```

Key **ATGC** = change in nucleotide relative to consensus sequence - synonymous  
**ATGC** = change in nucleotide relative to consensus sequence - conservative/semi-conservative amino acid substitution  
**ATGC** = change in nucleotide relative to consensus sequence - non-conservative amino acid substitution



# Sequence Alignments

```
MBP-004      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-010      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-001      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-002      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-009      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-003      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-005      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-006      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-007      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
MBP-008      GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCGACCTTCAAGGGTCAACCA
*****

MBP-004      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-010      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-001      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-002      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-009      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-003      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-005      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-006      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-007      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
MBP-008      TCCAAACCGTTTCGTTGGCGTGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAATATCTGTCTGACTGATGAAGGCTTGGAAAGCGGTTAAT
*****

MBP-004      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-010      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-001      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-002      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-009      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-003      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-005      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-006      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-007      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
MBP-008      AAAGACAAACCCGTTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTTATGCGCCACCATTGGAAAACGCCAGAAAAGGTGAAATCATGCCGAACATCCCG
*****

MBP-004      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-010      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-001      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-002      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-009      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-003      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-005      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-006      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-007      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
MBP-008      CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTGCGATGAAGCCCTGAAAAGACGCGCAGACT
*****
```

MBP-010 is identical to MBP-004 but missing START codon for C-terminal in-frame protein tagging

MBP-001 is identical to MBP-003 but missing START codon for C-terminal in-frame protein tagging

Key            **ATGC** = change in nucleotide relative to consensus sequence – synonymous  
              **ATGC** = change in nucleotide relative to consensus sequence – conservative/semi-conservative amino acid substitution  
              **ATGC** = change in nucleotide relative to consensus sequence – non-conservative amino acid substitution

# Sequence Alignments

Name of Variant	Alignment of Translated Protein	Size (aa)	# Var aa
MBP-004	-----MKIEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	367	0
MBP-010	-----KIEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	366	0
MBP-001	-----KIEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	366	0
MBP-002	-----MGIIEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	368	0
MBP-009	-----M <sup>A</sup> IEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	366	2
MBP-003	-----MKIEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	367	0
MBP-005	-----K <sup>A</sup> IEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	366	1
MBP-006	-----KIEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	358	5
MBP-007	MKIKTGARILALASALTMMFSASALAKIEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	392	1
MBP-008	-----KIEEGKLV I W I N G D K G Y N G L A E V G K K F E K D T G I K V T V E H P D K L E E K F P Q V A A T G D G P D I I F W A H D R F G G Y A Q S G L L A E I T P D K A F Q D K L Y P F T W	358	0

MBP-004 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-010 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-001 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-002 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-009 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-003 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-005 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-006 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKY<sup>A</sup>ENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-007 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-008 DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKGYDIKDVGVNDAGAKAGLTFLVDLIIKKNHMNADTDYSIAE

MBP-004 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-010 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-001 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-002 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-009 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-003 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-005 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-006 AAFNKGETAMTINGPWAWSNIDTS<sup>A</sup>VNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-007 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

MBP-008 AAFNKGETAMTINGPWAWSNIDTSKVNYGVTLPFTFKGQP...LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

Key

ATGC = change in nucleotide relative to consensus sequence – synonymous

ATGC = change in nucleotide relative to consensus sequence – conservative/semi-conservative amino acid substitution

ATGC = change in nucleotide relative to consensus sequence – non-conservative amino acid substitution

# Sequence Alignments

Name of Variant	# of Occur	Alignment	Size (bp)	# Var bp	Sources*
GST-002	44	ATG---TCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	654	0	GE, No, BD
GST-004	3	ATG <b>GCA</b> TCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	657	0/3i	No
GST-005	2	---TCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	651	0/3d	19906724
GST-001	2	ATG---TCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	654	1	No
GST-003	1	ATG---TCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA <b>TAA</b>	657	0/3i	#
GST-006	1	ATG--- <b>AG</b> CCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	654	1	AB
GST-007	1	---TCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	648	6	15232106
GST-008	1	ATG <b>GAA</b> TCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	657	3	15782178

Name of Variant	Alignment of Translated Protein	Size (aa)	Tag Pos	# Var aa
GST-002	--MSPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	218	N/C <sup>1</sup>	0
GST-004	<b>M</b> MSPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	219	N	1i
GST-005	--SPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	217	N <sup>2</sup>	1d
GST-001	-- <b>M</b> SPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	218	N	1
GST-003	--MSPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	218	#	0
GST-006	--MSPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	218	N	0
GST-007	---PILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	216	C	2d
GST-008	<b>M</b> SPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	219	N	1i

# from feature file. Not found in any vector in the plasmid library.  
 1 followed by STOP codon after additional in-frame residues or followed by another tag.  
 2 preceded either by another tag or a START codon and additional in-frame residues.

GST-005 is identical to GST-002 but missing START codon for C-terminal in-frame protein tagging

Name of Variant	#of Occur	Alignment	Size (bp)	# Var bp	Express Host	CAI/Host/Non-Host <sup>1</sup>	Sources*
HA-008	10	TACCC <b>A</b> TACGA <b>G</b> T <b>T</b> CCAG <b>A</b> TACGC <b>T</b>	27	5	Mammal-Plant	0.783/0.508	Or, C1
HA-007	5	TA <b>T</b> CCGTACG <b>A</b> CG <b>T</b> CCAG <b>A</b> CTACGC <b>A</b>	27	3	Mammal	0.463/0.510	C1
HA-003	4	TA <b>T</b> CCGT <b>A</b> GA <b>T</b> GTCC <b>G</b> GACTACGC <b>A</b>	27	5	E. coli	0.424/0.571	MCSG
HA-009	3	TA <b>T</b> CC <b>T</b> TA <b>G</b> ACGTGCC <b>T</b> GACT <b>A</b> GCC	27	5	Mammal-S. cer.	0.941-0.071/0.155	C1
HA-002	2	TA <b>T</b> CCGTACG <b>A</b> CG <b>T</b> CC <b>A</b> GACTACGC <b>A</b>	27	4	E. coli	0.642/0.395	MCSG
HA-004	2	TACCC <b>A</b> TACGA <b>G</b> T <b>T</b> CC <b>S</b> GA <b>T</b> ACGC <b>T</b>	27	6	Mammal	0.668/0.640	Mo
HA-006	2	TACCC <b>A</b> TACG <b>A</b> CG <b>T</b> CC <b>A</b> GA <b>T</b> ACGC <b>T</b>	27	4	S. cer.	0.182/0.478	C1
HA-001	1	TA <b>T</b> CCGT <b>A</b> GA <b>T</b> GTCC <b>G</b> GACT <b>A</b> GCC	27	4	Mammal	0.784/0.140	Lu
HA-005	1	TACCC <b>A</b> TAC <b>A</b> GTGCC <b>A</b> TA <b>T</b> AGCC	27	5	S. cer.	0.108/0.153	8242750
HA-010	1	TACCC <b>T</b> ACG <b>A</b> CG <b>T</b> CC <b>G</b> GACTACGC <b>C</b>	27	1	Mammal	1.000/0.096	15232106

1 Non-Host is *E.coli* for Mammal, Plant and *S. cerevisiae*. Non-Host is Mammal (Mouse) for *E. coli*. Carbone et al. (2003) reference set and Eyre-Walker (1996) equation.

Key  
**ATGC** = change in nucleotide relative to consensus sequence – synonymous  
**ATCC** = change in nucleotide relative to consensus sequence – conservative/semi-conservative amino acid substitution  
**ATCG** = change in nucleotide relative to consensus sequence – non-conservative amino acid substitution