

Non-Coding Sequences

Name of Variant	# of Alignment	Alignment	Size (bp)	# Var	Sources*
AmpR_prom-009	967	CGCGGAACCCCTATTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	105	0	In,No,Or,Pr, etc
AmpR_prom-001	58	CGCGGAACCCCTATTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	104	0/1d	MCSG
AmpR_prom-006	32	CGCGGAACCCCTATTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	103	0/2d	Ag
AmpR_prom-003	20	-----TTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	92	0	No, BR, Or, In
AmpR_prom-005	11	-----TTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	93	1/1i	No
AmpR_prom-008	10	CGCGGAACCCCTATTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	105	1	Or
AmpR_prom-010	5	-----TTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	92	1	MCSG
AmpR_prom-004	2	-----CATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	72	0	Lu
AmpR_prom-011	2	CGCGGAACCCCTATTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	105	1	Ta
AmpR_prom-002	1	CGCGGAACCCCTATTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	106	0/1i	Ag
AmpR_prom-007	1	-----CATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	72	2	No
AmpR_prom-012	1	CGCGGAACCCCTATTGGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCT-GATAAATGCTTCAATAAT-ATTCAAAAAGGAAGAGT	105	1	BD

AB – AB Vector
AG – AddGene
Ag – Agilent
BD – BD Biosciences
BR – Bio-Rad
Ca – Cambia
Cl – Clontech
Ev – Evrogen
GB – Geneva Biotech
GE – GE Life Sciences
iGEM – international Genetically Engineered Machine Foundation
IMAGE – I.M.A.G.E. Consortium
In – Invitrogen
Iv – InvivoGen
Lu – Lucigen
MBL – MBL International
MCSG – Midwest Center for Structural Genomics
Mo – MoBiTec
Ne – New England Biolabs
No – Novagen
OB – Open Biosystems
Or – OriGene
Ox – Oxford Genetics
Pr – Promega
Qi – Qiagen
SG – SwitchGear Genomics
Si – Sigma-Aldrich
Ta – TaKaRa
Th – Thermo Scientific

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

[illegible]

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		GGAGTTCCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA	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		TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTAT			
CMV_pro-001		-----			
CMV_pro-003		-----			
CMV_pro-004		-----GTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-009		-----GTGATGCGGTTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-008		-----TGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-010		-----GTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-005		-----TGATGCGGTTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-006		-----GTGATGCGGTTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-002		-----GTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-007		TGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-001		-----TGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA			
CMV_pro-003		-----GTGATGCGGTTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA			

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

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

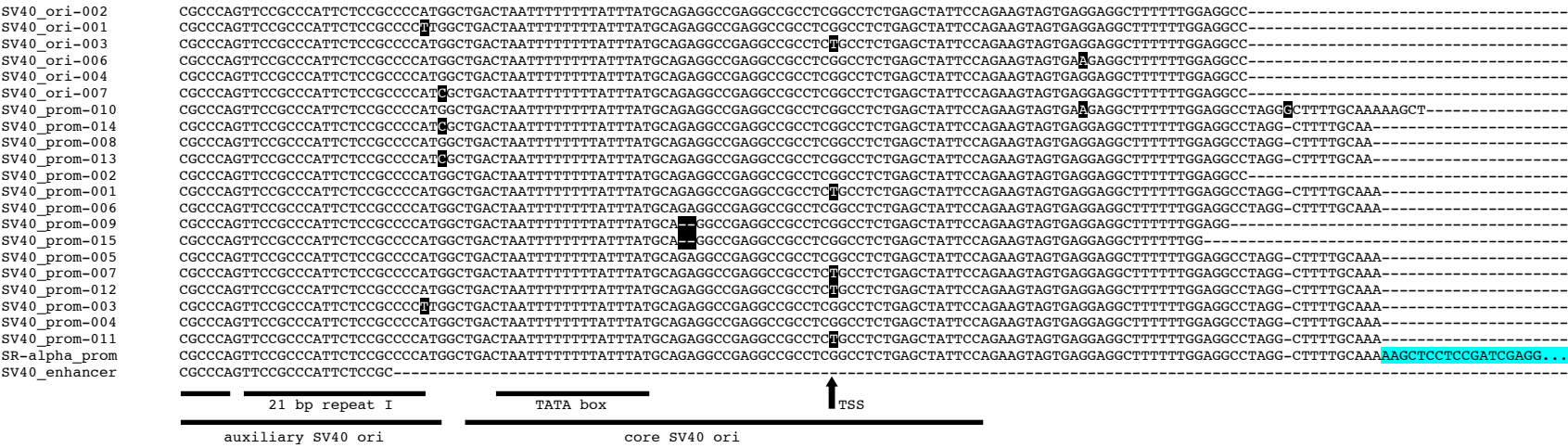
CMV_pro-007 is actually promoter plus enhancer. When split the promoter segment is identical to CMV_pro-004.

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	1/1i	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	-----GGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	303	0	Si
SV40_prom-001	4	-----GGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	317	1	In
SV40_prom-006	1	-----GGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	317	0	?
SV40_prom-009	1	-----GGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	302	1/2d3i	Si
SV40_prom-015	3	-----GGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	299	1	Si
SV40_prom-005	94	-----GTGTGTCAAGTGGGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	330	0	Or,Cl
SV40_prom-007	77	-----GTGTGTCAAGTGGGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	298	1	In,Th,Cl,GE
SV40_prom-012	23	-----GTGTGTCAAGTGGGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	298	1/1i	In,Ne
SV40_prom-003	1	CTGAGGCGGAAAGAACACAGCTGTGGAATGTGTGTCAAGTGGGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	358	1	Lu
SV40_prom-004	182	CTGAGGCGGAAAGAACACAGCTGTGGAATGTGTGTCAAGTGGGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	358	0	Cl,Ag,Ev,Pr,In,Or
SV40_prom-011	48	CTGAGGCGGAAAGAACACAGCTGTGGAATGTGTGTCAAGTGGGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	358	1	Pr,Lu
SR-alpha_prom	1	-----GTGTGTCAAGTGGGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	615	1	?
SV40_enhancer	4	-----GCTGTGGAATGTGTGTCAAGTGGGTGTGGAAGTCCCCAGGCTCCCCAG-CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA	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		-----CAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-014		-----GCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-008		-----TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-013		-----TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-002		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-001		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-006		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-009		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-015		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-005		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-007		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-012		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-003		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-004		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_prom-011		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SR-alpha_prom		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---TAACTC			
SV40_enhancer		GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCC---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

Name of Variant	# of Occur	Alignment	Size (bp)	# Var bp	Sources*
fl_ori-008	342	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	1	Cl,Ag,Or,Ev, etc
fl_ori-007	168	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	0	No, In, Cl
fl_ori-009	48	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	429	0	In, Or, Cl
fl_ori-003	34	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	429	4	In, Ne
fl_ori-006	19	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	2	BD, Cl, Ta
fl_ori-004	8	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	20	In
M13_ori-003	6	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	510	1/1d	7669286
fl_ori-011	4	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	1	In
fl_ori-014	4	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	429	2	In, Mo
fl_ori-017	4	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	473	2/17i	Si
fl_ori-005	2	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	2	19906724
M13_ori-004	2	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	381	1	Nc
fl_ori-001	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	2	Pr
fl_ori-002	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	2	IMAGE
fl_ori-010	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	427	0/2d	In
fl_ori-012	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	2	Pr
fl_ori-013	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	459	2/3i	In
fl_ori-015	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	5	In
fl_ori-016	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	459	1/3i	MSG
M13_ori-001	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	456	2	3323803
M13_ori-002	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	380	5	In
M13_ori-005	1	ACGCGCCCTGTAGCGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCACAGCGCCCTAGCGCCCGCTCCTTCGCTTTCTTCCTTCCTTTCTCGCCA	381	4	Ca

[illegible][illegible]

Primer RNA (- strand synthesis) Stem-Loop D
Gene II nick site (+ strand initiation)

Sequence Alignments

```
f1_ori-008      GGTCCTATTCTTTTGATTATAAGGGATTTGCGCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-007      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-009      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATT-----
f1_ori-003      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATT-----
f1_ori-006      GGCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-004      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
M13_ori-003      GGCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTTA
f1_ori-011      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAATTTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-014      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATT-----
f1_ori-017      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-005      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAATTT---ATTAACGTTTACAATTT-
M13_ori-004      GGCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGG-----
f1_ori-001      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-002      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-010      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATT-----
f1_ori-012      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-013      GGCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-015      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAATTTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
f1_ori-016      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAATCACATTAAACGTTTACAATTT-
M13_ori-001      GGCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGGCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCGAATTTTAACAAAAAT---ATTAACGTTTACAATTT-
M13_ori-002      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGG-----
M13_ori-005      GGTCCTATTCTTTTGATTATAAGGGATTTTGGCCGATTTCGG-----
** ***** **
```

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      AATATTGCTTATACAATCTTCTGTTTTTGGGGCTTTTCTGATTATCAACCGGGGT
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.

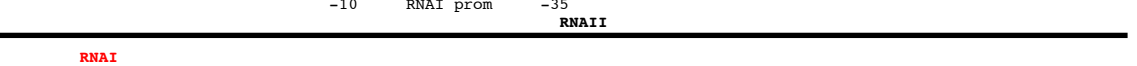
Key **ATGC** = change in nucleotide relative to consensus sequence – non-coding sequence

Sequence Alignments

Name of Variant	# of Occur	Alignment	Size (bp)	# Var	Sources*
ori-014	602	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	1	No,Cl,In,Ca,Ev, etc
ori-013	524	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	0	Cl,In,Or,Pr,Ox, etc
ori-009	220	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	3	In,Qi,MBL,Ne, etc
ori-012	55	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	585	0/4d	Or,Cl,In,GE
ori-008	19	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	2	No,Qi,GE
ori-003	15	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	1	Th
ori-021	12	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	4	24050148
ori-002	8	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	588	1/1d	Lu
ori-016	8	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	4	2659436
ori-010	5	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	1	10890530
ori-011	5	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	588	0/1d	Qi,No
ori-005	4	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	2	Pr
ori-018	4	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	4	15644173
ori-001	1	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	1	Or
ori-004	1	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	2	IMAGE
ori-006	1	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	1	iGEM
ori-007	1	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	588	0/1d	Or
ori-015	1	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	2	In
ori-017	1	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	569	3	In
ori-019	1	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	583	2/6d	No
ori-020	1	TTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCACAAACAAAAAACCCACCGCTACCCAGCGGTGGTTTGTTCGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGTGCTT	589	1	Cl
pBR32ori-t	1	-----	353	0	7655517



ori-014	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-013	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-009	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-012	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-008	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-003	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-021	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-002	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-016	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-010	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-011	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-005	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-018	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-001	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-004	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-006	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-007	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-015	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-017	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-019	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	ori-020	CAGCAGAGGCGAGATACCAAACTACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCACCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGC	pBR32ori-t	-----	CTGC
---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	------------	-------	------




[illegible]

Sequence Alignments

	Hairpin 2	Displaced Strand
ori-014	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-013	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-009	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-012	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-008	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-003	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-021	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-002	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-016	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-010	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-011	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-005	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-018	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-001	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-004	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-006	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-007	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-015	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-017	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-019	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
ori-020	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA
pBR322ori-t	GCTTCAGGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG	TCAGGGGGGCGGAGCCTATGGAAA

RNAII



Structural elements from Selzer et al. (1983) *Cell* 32:119-129. These ori sequences are ColE1/pMB1 derived origins of replication. RNAII (black arrow) is a transcript that hybridizes with single-stranded DNA in the displaced strand (cyan) region to form a “D-loop” and is processed by RNase H to form an RNA primer for DNA pol. RNAII is transcribed from a promoter at the extreme 5’ end of the sequence.

RNAI is an antisense RNA transcribed from RNAI promoter. RNAI and RNAII form small stem loop structures (SL1-3; red) that can interact and facilitate full sense-antisense hybridization. RNAI-RNAII covers the α -stem region (yellow; SL1 overlaps), leaving the free region of RNAII to form a large β - γ stem loop, which inhibits D-loop formation and replication. Rop protein, expressed from a separate locus not part of ori, stabilizes RNAI-RNAII, preventing replication.

Free RNAII forms an α - β stem loop and small hairpin structures (green) that facilitate D-loop formation and replication.

Plasmids with wild-type ColE1/pMB1 origins and Rop exist at ~15-20 copies/cell. An example of this type of medium copy number plasmid is pBR322. High copy number (up to 700/cell) ColE1/pMB1 plasmids, such as the pUC and pET series, lack Rop and also carry a pMB1 derivative in which the A at -1 from the start of the RNAI transcript is changed to G (T-C in the above orientation). This mutation moves the transcription start site +3, which removes the “anti-tail” portion of RNAI, which is crucial to nucleate RNAI-RNAII hybrid formation.

The apparent mutations in ori-007, -012 and -015 in SL3 could destabilize stem-loop formation and reduce the ability of RNAI to anneal to RNAII, resulting in higher copy numbers than pBR322.

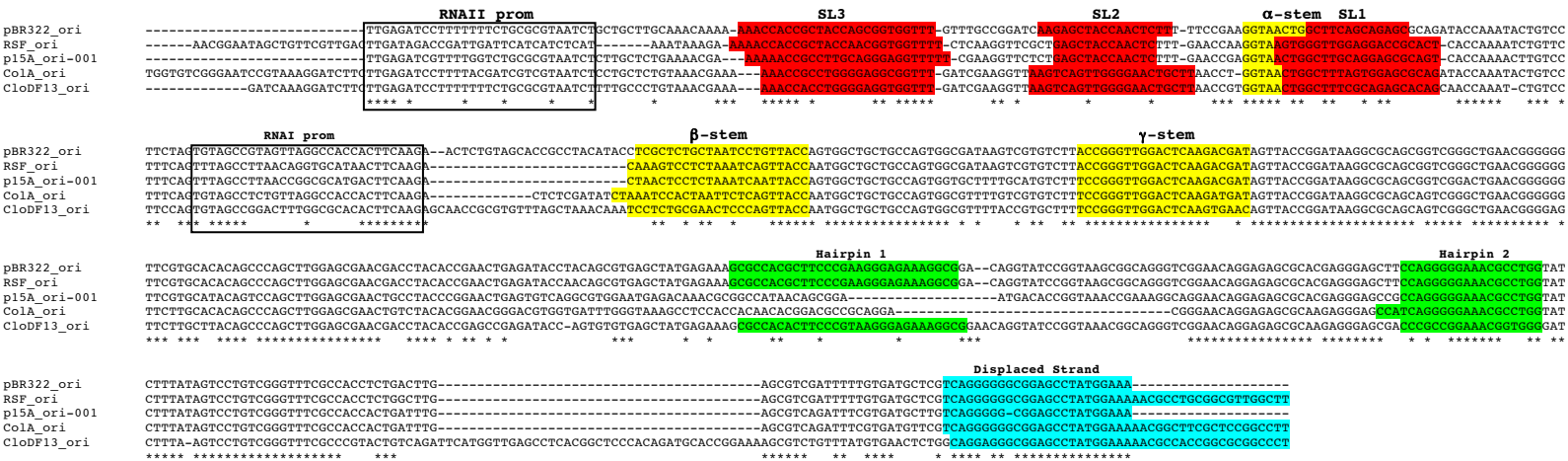
The truncated pBR322 ori is non-functional. The plasmid carrying this (pBINPLUS) uses oriV instead for its replication. pMB1/ColE1 origins can be truncated up to the β -stem, but a strong promoter must be placed in front to make the origin functional.

ori-017 lacks the RNAII promoter. Unless there is a cryptic promoter in SV40 ori immediately 5’ to the sequence shown, this ori should be non-functional. BPROM identifies a potential -35 and -10 promoter about 96 bp 5’ to ori within SV40 ori.

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

Sequence Alignments

ColE1-Related Promoters using a similar RNA based replication initiation mechanism



Colour scheme for structural elements same as for pMB1/ColE1 ori alignment. From Selzer et al. (1983) Cell 32:119-129.

Plasmids using p15A, ColA, RSF1030 or CloDF13 origins can be co-transformed into cells with plasmids carrying pMB1/ColE1 origins because the RNA sequences are sufficiently different to prevent competition between replication origins. Two plasmid systems are useful for bacterial co-expression of proteins, especially for purification of protein complexes that are difficult to express as individual proteins.

Three plasmid co-expression can be accomplished with a third plasmid that uses an iteron-based replicon. Iterons are repeat sequences to which a DNA-binding protein binds, bending the origin and facilitating its melting and entry of the replication machinery. These plasmids must carry the gene for the DNA-binding protein to express some of these proteins. Alternatively, some *E. coli* strains are engineered to express an iteron binding protein, and some can do so conditionally to control plasmid copy number.

Examples of iteron origins are ori2/oriS from F plasmid (uses repE protein), oriV from RK2/IncP-α plasmid (uses trfA protein), oriV from pSC101 plasmid (uses repl101/repA protein), oriY from R6K plasmid (uses pi protein), and oriV from pSA plasmid (uses repA protein).

Some plasmids, such as bacterial artificial chromosomes (BACs) based on the F plasmid ori2/repE replicon, are very low copy number (1-2/cell) to reduce toxic effects of cloned genes or recombination in the insert. These plasmids require partitioning mechanisms (SopA/B/C) to ensure the plasmid is divided between daughter cells. Some plasmids carry two compatible promoters and use conditional expression of iteron proteins to increase copy number when plasmid isolation is required.

The two annotated ori2 variants in the features list are identical but reverse complements
oriV-003 and 004 are identical but reverse complements
The two annotated pSC101 ori variants are identical but reverse complements

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

Name of Variant	# of Occur	Alignment	Size (bp)	# Var bp	Sources*
IRES-014	4	-----	464	1/1i	No
IRES-001	4	-----	463	2	No
IRES-013	2	-----	499	1/1d	15232106
IRES-012	1	-----	485	1/1d	No
IRES-004	4	-----CGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	451	2/1d	Ox
IRES-008	1+1	-----ACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	553	0	Or
IRES-015	3	-----ACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	552	1	Or
IRES-010	2	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	569	0	OB
IRES-002	4	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	573	1/1d	C1
IRES-005	3	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	573	1d	C1
IRES-003	35	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	574	0	C1, Or
IRES-006	1	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	574	1	C1
IRES-011	2	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	575	1/1i	C1
IRES2-001	3	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	588	1/1i	C1
IRES2-002	11	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	587	1	C1
IRES2-003	4	CCCCCTCCCTCCCCCCCCCCC-TAACGTTACTGGCCGAAGCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTATTATTTCCACCATATATGCCGCTCTTTTGGCAATGTGAGGGCCC	587	0	C1, Or

[illegible]

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

Sequence Alignments

IRES-014	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	TCTGATCTGGG	GCCTCGG
IRES-001	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-013	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-012	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-004	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-008	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-015	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-010	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-002	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-005	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-003	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-006	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES-011	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES2-001	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES2-002	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG
IRES2-003	ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTC	ACCTCAAGCGTATTCAACA	GGGGCTGAAGGATGCCCAGAAAGTACCCCATTTGATGGGA	-TCTGATCTGGG	GCCTCGG

	Stem-Loop I		Stem-Loop J	Stem-Loop K
IRES-014	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-001	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-013	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-012	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-004	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-008	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-015	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-010	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-002	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-005	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-003	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-006	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES-011	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	-----
IRES2-001	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	ATGGCCACAACCATC
IRES2-002	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	ATGGCCACAACCATC
IRES2-003	TGCACATGCTTTACATGTGTTAGTCGAGGTTAAAAAAACGCTC	TAGGCCCCC	CCGAACCCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAA	ATGGCCACAACCATC
	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	CGCGGAACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-001	CGCGGAACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-006	CGCGGAACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-008	CGCGGAACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-011	CGCGGAACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-002	CGCGGAACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-012	CGCGGAACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCMV-Cypridina_Luc	ACACTGACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pTK-Cypridina	ACACTGACCCCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-43.1a(+)	TGGTCATGACCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-44a(+)	TGGTCATGACCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-45b(+)	TGGTCATGACCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-51b(+)	TGGTCATGACCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pET-52b(+)	TGGTCATGACCTATTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-005	-----TTTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-003	-----TTTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-010	-----TTTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-004	-----TTTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
AmpR_prom-007	-----TTTGTATTATTTTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCR4-TOPO	TCTACGGGGTCTGACGCTCAGTGGACGAAACGACGCTTAAGGGGATTGGTTCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCAS-Guide	GCCCGTGTCTCAAATCTCTGATGTACATGTGCACAAGATAAAATAATACATCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCMV_SPORT	AATTATGTGCTGTGTAGGGATCGCTGGTATCAAAATATGTGTGCCACCCTGGCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pCMV_SPORT2	AATTATGTGCTGTGTAGGGATCGCTGGTATCAAAATATGTGTGCCACCCTGGCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pZL1	AATTATGTGCTGTGTAGGGATCGCTGGTATCAAAATATGTGTGCCACCCTGGCATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pTriEx-1.1	GCGCGTTTATACACATCTTGGGATTGGATTAAAGATGCGAGAAACGCCGGGACATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pTriEx-2	GCGCGTTTATACACATCTTGGGATTGGATTAAAGATGCGAGAAACGCCGGGACATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pQE-TriSystem_5	GCGCGTTTATACACATCTTGGGATTGGATTAAAGATGCGAGAAACGCCGGGACATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pQE-TriSystem_6	GCGCGTTTATACACATCTTGGGATTGGATTAAAGATGCGAGAAACGCCGGGACATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pdream2.1_MCS	GCGCGTTTATACACATCTTGGGATTGGATTAAAGATGCGAGAAACGCCGGGACATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pQE-TriSystem	----GTTTATACACATCTTGGGATTGGATTAAAGATGCGAGAAACGCCGGGACATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pQE-TriSys_His-Str	----GTTTATACACATCTTGGGATTGGATTAAAGATGCGAGAAACGCCGGGACATGAGACAATACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pGL4.70-hrLuc	---GCGGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCCTTGAGAGACGTACTAGTAACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pLightSwitch_Prom	---GCGGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCCTTGAGAGACGTACTAGTAACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pNL1.1-Nluc	---GCGGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCCTTGAGAGACGTACTAGTAACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pNL2.1-Nluc	---GCGGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCCTTGAGAGACGTACTAGTAACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
pGL4.10-Luc2	---GCGGCCTGTCCAATACCTCCCGTACCTTAATATTACTTACTTATCCCTTGAGAGACGTACTAGTAACCCT-GATAAATGCTTCAATAAT-ATTGAAAAAGGAAGAGT
Lafmid_BA	CTTATACAATCTTCTGTTTTTGGGGCTTTTCTGATTATCAACCGGGGTACATATGATTGACACTCTAGTTTTACGATTACCGTTTCATCGATTGAAAAAGGAAGAGT
pLIC-SGC1	AAGCTCTAAATCGGGGGTCCCTTTAGGGTTCGGATTAGTCTTTTACGGCACCTCGACCCCAGAACTTGATTAGGGTGATGGTTTCACATTGAAAAAGGAAGAGT
pIB_V5-His-DEST	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGTCTACTAAATGAAATAATTAGAGCTTCAATTGAATATATCAGTTATTACCCATTGAAAAAGGAAGAGT
pIB_V5-His	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGTCTACTAAATGAAATAATTAGAGCTTCAATTGAATATATCAGTTATTACCCATTGAAAAAGGAAGAGT
pMIB_V5-His_A	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGTCTACTAAATGAAATAATTAGAGCTTCAATTGAATATATCAGTTATTACCCATTGAAAAAGGAAGAGT
pYC6_CT	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGTCTACTAAATGAAATAATTAGAGCTTCAATTGAATATATCAGTTATTACCCATTGAAAAAGGAAGAGT
pYES6_CT	CGAGGAGCAGGACTGACACGTCCTCCGGGAGATCTCGATGTCTACTAAATGAAATAATTAGAGCTTCAATTGAATATATCAGTTATTACCCATTGAAAAAGGAAGAGT
pYC2_CT	-----AAACCTGTATTATAAGTAAATGCATGTATACTAAATGAAATAATTAGAGCTTCAATTGAATATATCAGTTATTACCCATTGAAAAAGGAAGAGT
pYES2_CT	-----AAACCTGTATTATAAGTAAATGCATGTATACTAAATGAAATAATTAGAGCTTCAATTGAATATATCAGTTATTACCCATTGAAAAAGGAAGAGT
pCR2.1-TOPO	-----AAACCTGTATTATAAGTAAATGCATGTATACTAAATGAAATAATTAGAGCTTCAATTGAATATATCAGTTATTACCCATTGAAAAAGGAAGAGT

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 (lef2), but BPROM suggests ATGCAG 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-SGC1 is identical to a portion of *fl* origin. The highlighted areas correspond to -35 and -10 sites for RNAPol σ^{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- CACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTT CT CGAGATTTTCAGGAGCTAAGGAAGCTAAA
CAT_prom-002	TGATCGG- CACGTAAGAGG G TTCCAACTTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
CAT_prom-003	TGATCGG- CACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
CAT_prom-004	TGATCGG- CACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pSMART_HCKan (KanR)	TGATCGG- CACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pSMART_LCKan (KanR)	TGATCGG- CACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCT-----
pSB1C3 (CmR)	TGATCGG G - CACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRANGER-BTB-1 (AmpR)	----- ACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRANGER-BTB-2 (KanR)	----- ACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRANGER-BTB-3 (CmR)	----- ACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRANGER-BTB-5 (TpR)	----- ACGTAAGAGGTTCCAAC TTTACCATAAT GA AAATAAGATCACTACCGGGCGTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pRFP-C-RS (CmR)	ATATGTAT CGCTCATGAC TA TGTTGACAGCTAT CG CAAT AGCTAGCGGCGCTAGC TTTAA T GAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAA
pVP13 (CmR)	----- CCAAGCTAGCTTGG CGAGATTTTCAGGAGCTAAGGAAGCTAAA
pVP16 (CmR)	----- CCAAGCTAGCTTGG CGAGATTTTCAGGAGCTAAGGAAGCTAAA
pVP33K (CmR)	----- CCAAGCTAGCTTGG CGAGATTTTCAGGAGCTAAGGAAGCTAAA
pVP56K (CmR)	----- CCAAGCTAGCTTGG CGAGATTTTCAGGAGCTAAGGAAGCTAAA
pDNR-LIB (CmR)	----- GGATC TTTCAGGAGCTAAGGAAGCTAAA
pDNR-Dual (CmR)	----- GGATC TTTCAGGAGCTAAGGAAGCTAAA

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	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP56K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP68K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP65K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP81K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pVP33K	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GACGGTCGTTTCGC
pSpark_III	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GAGGATCGTTTCGC
pCAMBIA5105	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GAGGATCGTTTCGC
pZerO-2	TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCT CC CGGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGAT---GAGGATCGTTTCGC
pSF-pA-PromMCS-Fluc	GGCAGCGTACCGATCTGTTTAAACCTAGATA TTGAT AGTCTGATCGGTCAACGTATAAA CGAGTCCTAGCTTTTGCAAACATCT -ATCAAGAGACAGGAT CAGCAGGAGGC TTTCGC
pSF-pA-CMVe-Rluc	GGCAGCGTACCGATCTGTTTAAACCTAGATA TTGAT AGTCTGATCGGTCAACGTATAAA CGAGTCCTAGCTTTTGCAAACATCT -ATCAAGAGACAGGAT CAGCAGGAGGC TTTCGC
pSF-pA-PromMCS-Rluc	GGCAGCGTACCGATCTGTTTAAACCTAGATA TTGAT AGTCTGATCGGTCAACGTATAAA CGAGTCCTAGCTTTTGCAAACATCT -ATCAAGAGACAGGAT CAGCAGGAGGC TTTCGC
pSF-PromMCS-Fluc	GGCAGCGTACCGATCTGTTTAAACCTAGATA TTGAT AGTCTGATCGGTCAACGTATAAA CGAGTCCTAGCTTTTGCAAACATCT -ATCAAGAGACAGGAT CAGCAGGAGGC TTTCGC
pSF-PromMCS-Rluc	GGCAGCGTACCGATCTGTTTAAACCTAGATA TTGAT AGTCTGATCGGTCAACGTATAAA CGAGTCCTAGCTTTTGCAAACATCT -ATCAAGAGACAGGAT CAGCAGGAGGC TTTCGC
pSF-pA-CMVe-Fluc	GGCAGCGTACCGATCTGTTTAAACCTAGATA TTGAT AGTCTGATCGGTCAACGTATAAA CGAGTCCTAGCTTTTGCAAACATCT -ATCAAGAGACAGGAT CAGCAGGAGGC TTTCGC
pSF-CMVe-Fluc	GGCAGCGTACCGATCTGTTTAAACCTAGATA TTGAT AGTCTGATCGGTCAACGTATAAA CGAGTCCTAGCTTTTGCAAACATCT -ATCAAGAGACAGGAT CAGCAGGAGGC TTTCGC
pSF-CMVe-Rluc	GGCAGCGTACCGATCTGTTTAAACCTAGATA TTGAT AGTCTGATCGGTCAACGTATAAA CGAGTCCTAGCTTTTGCAAACATCT -ATCAAGAGACAGGAT CAGCAGGAGGC TTTCGC

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

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

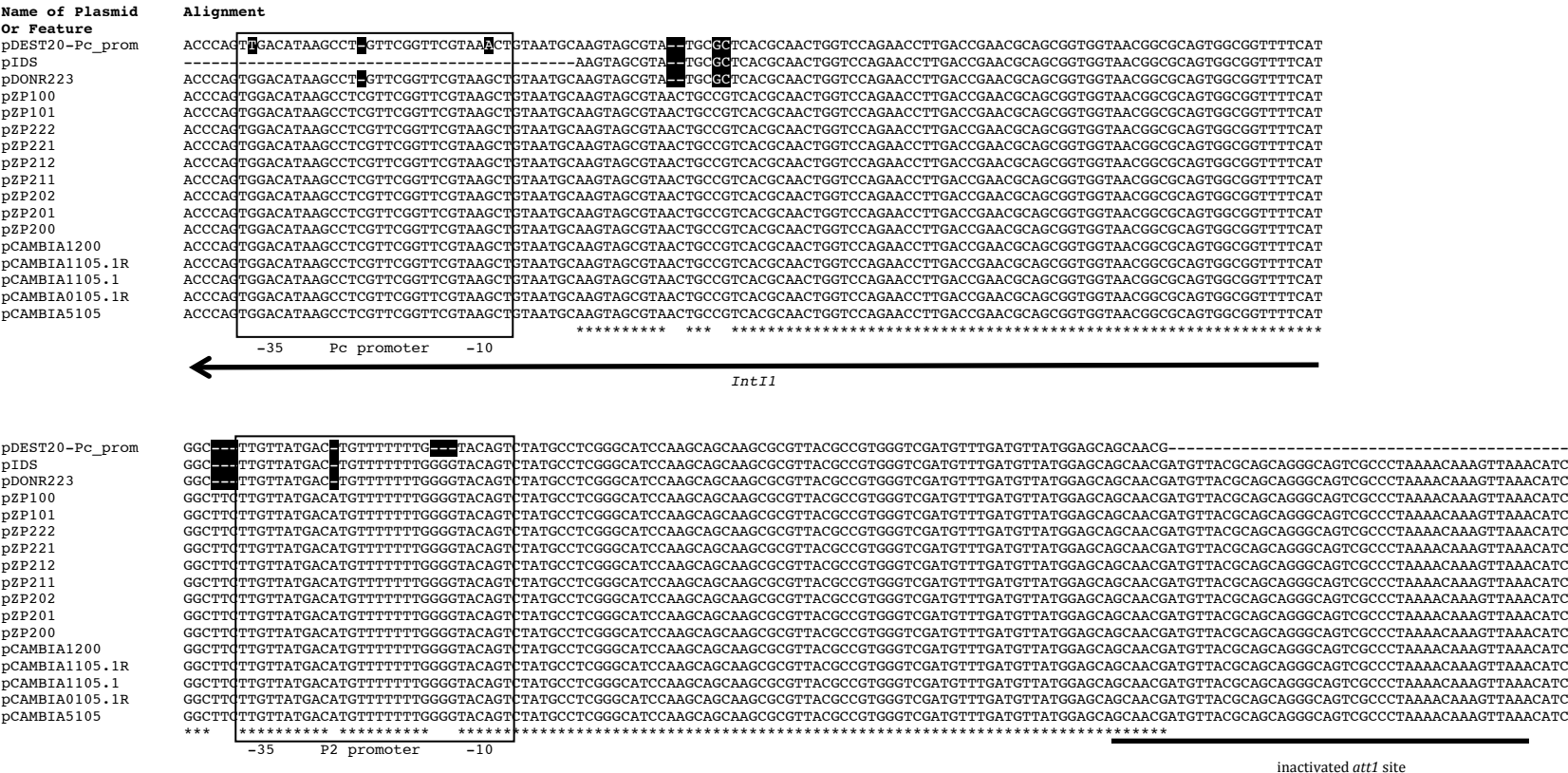
These promoters are the natural *aph(3')*-Ia promoter from *E. coli* Tn5 and all drive expression of KanR (*aph(3')*-IaI). Highlighting shows putative -35 and -10 sequences (BPROM). The pGFP and pCMV series have a hybrid promoter with AmpR promoter (cyan highlighting) from replacing AmpR with KanR during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

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



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.

Un-annotated SmR promoters

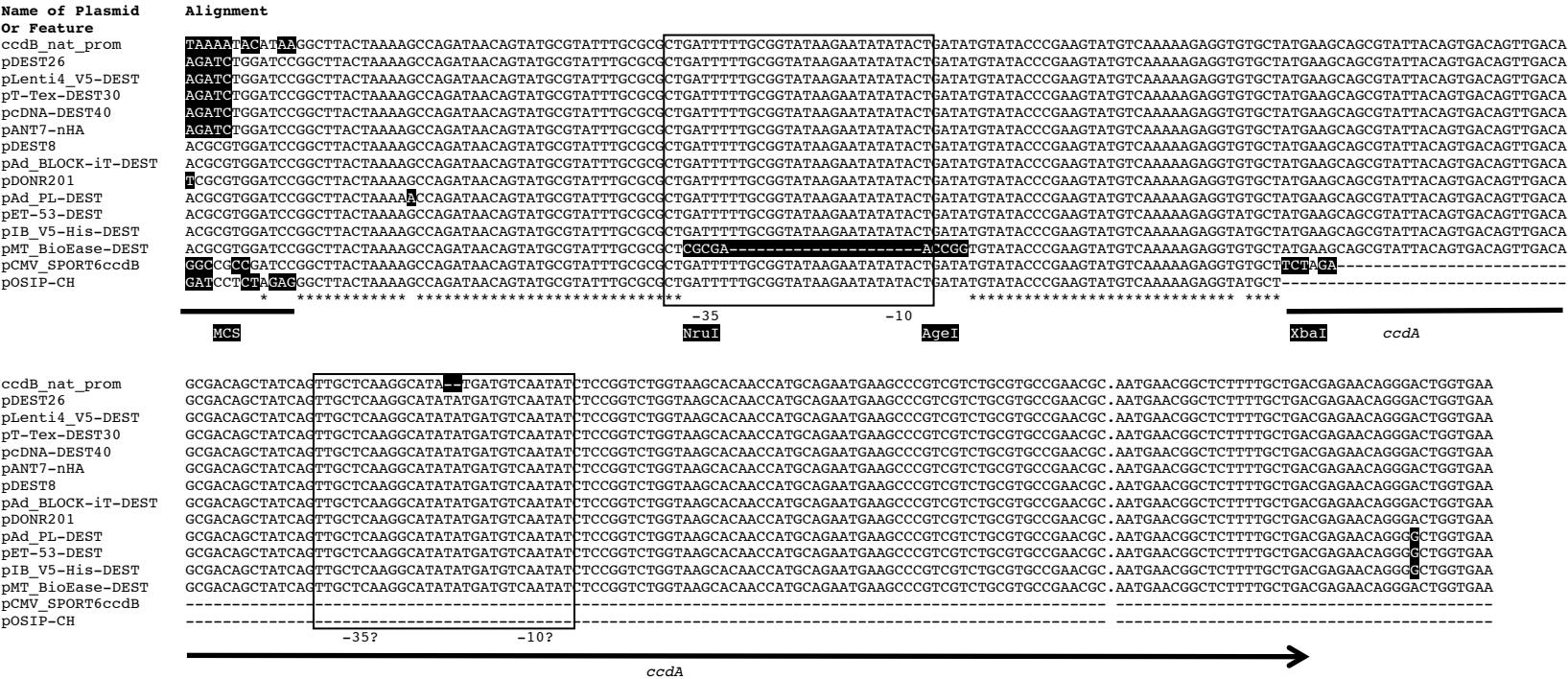
SmR in these plasmids is derived from *E. coli* Tn7 which contains a class 2 integron including *hfrAI* (dihydrofolate reductase conferring resistance to trimethoprim), *sat* (streptothricin acetyltransferase), *aadA1* (SmR - spectinomycin and streptomycin resistance), aminoglycoside adenyltransferase,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

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 *pHELLSGATE* (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



Coding Sequences

Variant	# of Occur	Alignment	Size (bp)	# Var	Sources*
AmpR-013	688	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	0	Most
AmpR-017	78	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	2	In, No, GE, Cl, Ne
AmpR-001	76+2	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	No, GS, MCSG
AmpR-021	69	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	102	Pr, Lu
AmpR-016	45+1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	Cl, Or, In
AmpR-022	28	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	Ev, Qi
AmpR-010	16	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	4	Th, SG
AmpR-004	9+3	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	8	In
AmpR-019	8	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	2	2659436
AmpR-020	8	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	873	0/12i	Cl
AmpR-014	6	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	2	7669286
AmpR-007	5	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	858	0/3d	Ag
AmpR-015	5	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	5	Lu
AmpR-011	4	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	2	Pr
AmpR-002	2	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	9	In
AmpR-018	3	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	7	Ox
AmpR-003	2	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	15782178
AmpR-008	2+1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	No, Or
AmpR-005	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	2	In
AmpR-006	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	3	IMAGE
AmpR-009	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1/3d3i	8242750
AmpR-012	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	In
AmpR-023	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	4	MCSG
bla(M)-001	1	-----CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	789	1/6d	12071693
bla(M)-002	1	-----ACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	795	0	In
bla(M)-003	1	-----CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	789	1/6d	No
bla(M)-004	1	-----ACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	792	0/3d	In
AmpR-007corr	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	870	0	
AmpR-009corr	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	
bla(M)-001corr	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	
bla(M)-003corr	1	ATGAGTATTCAACATTCCCGTGTCCGCCCTATTCCCTTTTTTGGCGCATTTTGCCCTCCTGTTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	861	1	
bla(M)-004corr	1	-----ATG-----ACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCA	837	0	
* * * * *					
AmpR-013	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-017	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-001	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-021	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-016	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-022	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-010	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-004	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-019	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-024	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-007	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-015	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-011	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-002	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-018	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-003	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-008	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-005	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-006	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-009	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-012	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-023	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
bla(M)-001	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
bla(M)-002	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
bla(M)-003	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
bla(M)-004	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-007corr	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
AmpR-009corr	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
bla(M)-001corr	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
bla(M)-003corr	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	
bla(M)-004corr	1	CGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTC-----CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGC	861	1	

[illegible][illegible]

[illegible][illegible]

Sequence Alignments

AmpR-013 GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-017 GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-001 GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
AmpR-021 GAGCGGTGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
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-----ATGC = synonymous
AmpR-017 GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----ATGC = conservative/semi-conservative amino acid substitution
AmpR-001 GAGATAGGTGCCTCACTGATTAAGCATTGGTAA-----ATGC = non-conservative amino acid substitution
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 GAGATAGGTGCCTCACTGATTAAGCATTGGCTGTTATCAACAAGTTGTACAAAAAAGCTGAACGAGAAACGTAA-----
* * * * *

AmpR gene is bla (β -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

Name of Variant	Alignment of Translated Protein	Size (aa)	# Var
AmpR-013	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	0
AmpR-017	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	2
AmpR-001	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	1
AmpR-021	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	0
AmpR-016	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	1
AmpR-022	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	1
AmpR-012	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	0
AmpR-004	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	6
AmpR-019	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	1
AmpR-020	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	290	1/41
AmpR-014	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	2
AmpR-007corr	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	0
AmpR-015	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	3
AmpR-011	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	0
AmpR-002	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	166	5/79d
AmpR-018	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	0
AmpR-003	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	1
AmpR-008	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	0
AmpR-005	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	2
AmpR-006	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	2
AmpR-012	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	1
AmpR-023	MSIQHFRVALIPFFAAFLCPVFAHPETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	286	1
bla(M)-002	-----DPTETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	264	0
bla(M)-004corr	-----DPTETLVKVKDAEDQLGARVGYIELDLSNGKILESFRPEERFP-----MMSTFKVLLCGAVLSRIDAGEQQLGRIIHYSONDLVEYSPVTEKHLTDGMT	278	0
***** : : : : : *****			

[illegible]

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  ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHWPQD-----
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 ERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHLLSTSLYKKAERET
*****

```

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 TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-016 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-015 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-014 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-010 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-004 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-011 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-002 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-012 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-001 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-005 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-003 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-006 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-007 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-008 TC CGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG
CmR-009 **TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG**
CmR-009corr TCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCCCTATTTCCTTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTG

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

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

CmR gene is cat (chloramphenicol actyltransferase) 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	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-016	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-015	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-014	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-010	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-004	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-011	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-002	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-012	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-001	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-005	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	1
CmR-003	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-006	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-007	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	219	0
CmR-008	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY	226	2
CmR-009corr	MEKKITGYTTVDISQWHRKEHFEAFQSVAAQCTYNQTVQLDITAFCLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFFSSLWSEYHDDFRQFLHIY *****;*****	227	1

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

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

Name of Variant	# of Occur	Alignment	Size (bp)	# Var bp	Source*
HygR-010	35	ATG---AAAGAGCCGGAACCTACCCGCTACAGCGTTGAGAAATTTCTGATCGAGAAAGTTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAAGAGCGGCTTCAGCTTC	1038	222	Pr
HygR-012	26	ATG---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1026	8	Ca
HygR-008	11	ATG---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1026	0	In, Cl, OB
HygR-005	7	---AGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1020	0/6d	Cl
HygR-002	4	ATG---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1038	13	Ag
HygR-011	4	---AGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1020	4/6d	No
HygR-004	3	ATGGGTAAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1029	0	16200533
HygR-007	3	---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1023	0/3d	In
HygR-001	2	ATG---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1026	1	Cl
hphMx6	2	ATGGGTAAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1029	0	16200533
HygR-003	1	ATG---AGCAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1026	6	In
HygR-006	1	ATG---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1026	5	10890530
HygR-009	1	ATG---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1026	4	Or
HygR-013	1	ATG---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1023	1/3d	2194165
HygR-013corr		ATG---AAAAAGCCTGGAACCTACCCGCGACGCTGTGTCGAGAAGTTTCTGATCGAAAAGTTTCGACAGCGCTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGTTTC	1023	1/3d	

HygR-010		GATGTGGCGGGCCGGCTATGTCTCGCGGGTAATAGCTGCGCTGATGGCTTCTACAAAGAAGCTAGCTGACGCTACTTGGCGCGCTGCTTCCGATCCGAAGTGTGAC			
HygR-012		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-008		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-005		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-002		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-011		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-004		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-007		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-001		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
hphMx6		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-003		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-006		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-009		GATGTAGGAGGGCGTGATATGTCTCGCGGGTAATAGCTGCGCGGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTGGCATCGGCCGCGCTCCCGATTCCGGAAGTGTCTGAC			
HygR-013		*****			
HygR-013corr		*****			
HygR-010		ATGGGGAATTACGAGGAGGCGTGACCTATTGCATCTCCCGCGGTGCACAGGGGTGCACGTTGCAAGACCTGCCGAAACCGAAGTCCCGCTGTTCTGCAAGCGGTGCGGAGGCCATG			
HygR-012		ATGGGGAATTACGAGGAGGCGTGACCTATTGCATCTCCCGCGGTGCACAGGGGTGCACGTTGCAAGACCTGCCGAAACCGAAGTCCCGCTGTTCTGCAAGCGGTGCGGAGGCCATG			
HygR-008		ATGGGGAATTACGAGGAGGCGTGACCTATTGCATCTCCCGCGGTGCACAGGGGTGCACGTTGCAAGACCTGCCGAAACCGAAGTCCCGCTGTTCTGCAAGCGGTGCGGAGGCCATG			
HygR-005		ATGGGGAATTACGAGGAGGCGTGACCTATTGCATCTCCCGCGGTGCACAGGGGTGCACGTTGCAAGACCTGCCGAAACCGAAGTCCCGCTGTTCTGCAAGCGGTGCGGAGGCCATG			
HygR-002		ATGGGGAATTACGAGGAGGCGTGACCTATTGCATCTCCCGCGGTGCACAGGGGTGCACGTTGCAAGACCTGCCGAAACCGAAGTCCCGCTGTTCTGCAAGCGGTGCGGAGGCCATG			
HygR-011		ATGGGGAATTACGAGGAGGCGTGACCTATTGCATCTCCCGCGGTGCACAGGGGTGCACGTTGCAAGACCTGCCGAAACCGAAGTCCCGCTGTTCTGCAAGCGGTGCGGAGGCCATG			
HygR-004		ATGGGGAATTACGAGGAGGCGTGACCTATTGCATCTCCCGCGGTGCACAGGGGTGCACGTTGCAAGACCTGCCGAAACCGAAGTCCCGCTGTTCTGCAAGCGGTGCGGAGGCCATG			
HygR-007		ATGGGGAATTACGAGGAGGCGTGACCTATTGCATCTCCCGCGGTGCACAGGGGTGCACGTTGCAAGACCTGCCGAAACCGAAGTCCCG			

***** ** ** ***** ** ** ** ***** **** ***** ***** ***** ***** ***** ***** ***** ***** *****

Sequence Alignments

HygR-010 GGCTGTGCTGCACTCTGCCCAGACGGCAACCGCCGCCAGACCGCCGCCAAGGACCTAGGTCTGAGTTTAA
HygR-012 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGAAATAG-----
HygR-008 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-005 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-002 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-011 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-004 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAA-----
HygR-007 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-001 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
hphMX6 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAA-----
HygR-003 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-006 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-009 GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----
HygR-013 **GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATAG-----**
HygR-013corr GGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCAGCTCGTCCGAGGGCAAAGGAATATCGATAA-----

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 aa
HygR-010	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
HygR-012	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	345	1
HygR-008	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
HygR-005	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	342	2
HygR-002	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	339	4
HygR-011	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	2
HygR-004	MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	340	0
HygR-007	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	1
HygR-001	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
hphMX6	MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	345	0
HygR-003	MCKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	339	3
HygR-006	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
HygR-009	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM	341	0
HygR-013corr	-MKKPELTATSVKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLPETELPAVLQPVAEAM *****	342	0
HygR-010	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-012	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-008	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-005	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-002	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-011	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-004	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-007	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-001	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
hphMX6	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-003	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-006	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-009	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL		
HygR-013corr	DAIAAADLSQTSGGFPGFPQGIGQYTTWRDFICAIADPHVYHWQTVMDTIVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMFGDSQYEVANIFFWRPWL *****		
HygR-010	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-012	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-008	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-005	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-002	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-011	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-004	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-007	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-001	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
hphMX6	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-003	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-006	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-009	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEVGRV		
HygR-013corr	ACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCEVLADSGNRRPSTRPRAKEYR-- *****		

Name of Variant	# of Occur	Alignment	Size (bp)	# Var	Sources*
KanR-009	51	ATGAGCCATATTCAACGGGAAACGCTTTGCTCAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	1	No,Qi,19906724
KanR-021	16	ATCGGTAAGAAAGACATCGGTTT-----CGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	19/6d	9717241,7747518
KanR-005	9	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	1	No,C1,In
KanR-010	6	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	2	No
KanR-015	6	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGAG-CTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	30	24050148
KanR-008	5	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	4	10890530
KanR-019	4	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	3	No
KanR-014	3	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	6	Lu
KanR-001	2	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	3	10890530
KanR-002	2	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	7	Lu,16496398
KanR-003	2	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	2	19906724
KanR-004	2	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	2/6d	In,GE
KanR-006	2	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	2	19906724
KanR-007	2	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	810	3/6d	Or
KanR-011	2	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	810	3/6d	MCSG
KanR-012	2	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	2	3327753
KanR-013	1	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	810	3/6d	Or
KanR-017	1	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	3	24057978
KanR-018	1	ATGAGCCATATTCAACGGGAAACGCTTCTGCTCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAAT	816	5	20040575
*** * * * * *					
KanR-009		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-021		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-005		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-010		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-015		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-008		CTACGCGTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGTTGTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-019		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-014		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-001		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-002		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-003		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-004		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-006		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-007		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-011		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-012		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-013		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCC			
KanR-017		CTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTT			

[illegible][illegible][illegible]

[illegible]

KanR-020 (50 occurrences) is *aphA-3* from *Staphylococcus aureus*. There is only one variant of this.

[illegible]

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

[illegible]

[illegible][illegible][illegible]

[illegible][illegible]

NeoR/KanR-010 is identical to NeoR/KanR-002 but missing the START codon because it is in frame with hRLuc

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

[illegible]

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*
PuroR-002	23	ATG---ACCAGGATACAAGCCCACGGTGC	600	1	C1,AG
PuroR-004	15	ATG---ACCAGGATACAAGCCCACGGTGC	600	2	Th
PuroR-009	10	ATG---ACCAGGATACAAGCCCACGGTGC	600	3	C1
PuroR-006	7	ATG---ACCAGGATACAAGCCCACGGTGC	600	1	C1
PuroR-007	6	ATG---ACCAGGATACAAGCCCACGGTGC	600	106	Pr
PuroR-011	6	ATG---ACCAGGATACAAGCCCACGGTGC	600	3	Or,2194165
PuroR-005	2	ATG---ACCAGGATACAAGCCCACGGTGC	600	2	Or
PuroR-008	2	ATG---ACCAGGATACAAGCCCACGGTGC	603	6	OB
PuroR-010	2	ATG---ACCAGGATACAAGCCCACGGTGC	600	2	21706014
PuroR-003	1	ATG---ACCAGGATACAAGCCCACGGTGC	600	4	Or
PuroR-001	1	ATG---ACCAGGATACAAGCCCACGGTGC	597	4	C1
PuroR-001corr		ATG---ACCAGGATACAAGCCCACGGTGC	624	1	

PuroR-002	GACCGCCACATCGAGCGGGTCA
PuroR-004	GACCGCCACATCGAGCGGGTCA
PuroR-009	GACCGCCACATCGAGCGGGTCA
PuroR-006	GACCGCCACATCGAGCGGGTCA
PuroR-007	GACCGCCACATCGAGCGGGTCA
PuroR-011	GACCGCCACATCGAGCGGGTCA
PuroR-005	GACCGCCACATCGAGCGGGTCA
PuroR-008	GACCGCCACATCGAGCGGGTCA
PuroR-010	GACCGCCACATCGAGCGGGTCA
PuroR-003	GACCGCCACATCGAGCGGGTCA
PuroR-001	GACCGCCACATCGAGCGGGTCA
PuroR-001corr	GACCGCCACATCGAGCGGGTCA

PuroR-002	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-004	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-009	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-006	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-007	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-011	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-005	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-008	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-010	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-003	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-001	GAGAGCGTCGAAGCGGGGCGGTGTT
PuroR-001corr	GAGAGCGTCGAAGCGGGGCGGTGTT

PuroR-002	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-004	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-009	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-006	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-007	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-011	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-005	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-008	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-010	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-003	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-001	GAGCCCCGTGGTTCTTGCCACCCTCG
PuroR-001corr	GAGCCCCGTGGTTCTTGCCACCCTCG

PuroR-002	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-004	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-009	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-006	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-007	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-011	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-005	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-008	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-010	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-003	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-001	CTGGAGACCTCCGCGCCCGCAACCTCC
PuroR-001corr	CTGGAGACCTCCGCGCCCGCAACCTCC

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-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-004	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-009	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-006	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-007	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-011	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	5
PuroR-005	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-008	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	200	2
PuroR-010	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	0
PuroR-003	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK	199	5
PuroR-001corr	M-TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHVTDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPEESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPK * *****_*****	207	0
PuroR-002	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-004	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-009	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-006	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-007	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-011	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-005	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-008	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-010	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-003	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGA-----		
PuroR-001corr	EPAWFLATVGVS PDHQKGLGS AVVLPGVEAAERAGVPAFLETSAPRNLPFFYERLGFTVTADVEVPEGPRTWCMTTRKPGATGAASRIK ***** *: *****		

[illegible]

[illegible]

[illegible]

Sequence Alignments

```
lacZ- $\alpha$ -060      TGC TAG-----
lacZ- $\alpha$ -031      TGGCGCGACGCGGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -034      TGA-----
lacZ- $\alpha$ -036      TGGCGCCTGATGCGGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATA-----
lacZ- $\alpha$ -057      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATAATGGTGCACTCTCAGTACAAATCTGCTCTGATGCCGCATAG-----
lacZ- $\alpha$ -073      TGA-----
lacZ- $\alpha$ -076      TGGG-----CGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -032      TGGCGCGAAATTGTAAACGTTAATGTTATCGATACATGA-----
lacZ- $\alpha$ -068      -----
lacZ- $\alpha$ -001      TGCATTTTATTCACAAGCGCGCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGA-----
lacZ- $\alpha$ -004      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATACGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -005      TGG-----GACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -006      -----
lacZ- $\alpha$ -010      TGGG-----ACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -014      TGG-----AAATTGTAA-----
lacZ- $\alpha$ -015      TGG-----GACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -020      TGGCGCTAA-----
lacZ- $\alpha$ -022      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATATGGTGCACTCTCAGTACAAATCTGCTCTGATGCCGCATAG-----
lacZ- $\alpha$ -028      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATATGGTGCACTCTCAGTACAAATCTGCTCTGATGCCGCATAG-----
lacZ- $\alpha$ -039      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATAGATCCGGTACGTCGTTAA-----
lacZ- $\alpha$ -045      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATAGGGTAA-----
lacZ- $\alpha$ -046      TGG-----AAATTGTAA-----
lacZ- $\alpha$ -051      TGG-----AAATTGTAA-----
lacZ- $\alpha$ -052      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATATATCGCTGGGCCATTCTCATGAAGAATATCTTGAATTTATTGTATATTACTAG-----
lacZ- $\alpha$ -036corr  TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATATATCGCTGGGCCATTCTCATGAAGAATATCTTGAATTTATTGTATATTACTAG-----
lacZ- $\alpha$ -002      TGG-----AAATTGTAA-----
lacZ- $\alpha$ -003      TGG-----AAATTGTAA-----
lacZ- $\alpha$ -007      TGGGA-----AATTGTAAACGTTAATATTTTGTAAATATTTTGTAA-----
lacZ- $\alpha$ -008      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATATGGTGCACTCTCAGTACAAATCTGCTCTGATGCCGCATAG-----
lacZ- $\alpha$ -009      TGA-----
lacZ- $\alpha$ -011      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATACGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -012      TGG-----AAATTGTAA-----
lacZ- $\alpha$ -013      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATATGGTGCACTCTCAGTACAAATCTGCTCTGATGCCGCATAG-----
lacZ- $\alpha$ -016      -----
lacZ- $\alpha$ -017      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -018      TGG-----AGATCCAAATTTTAA-----
lacZ- $\alpha$ -019      TGGG-----ACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -021      -----
lacZ- $\alpha$ -023      TGGCGCGAAATTGTAAACGTTAATGTTATCGTACCCCTACTCTCAAAAATGTCAAAGATACAGTCTCAGAAAGACCAAAGGGCTATTGA-----
lacZ- $\alpha$ -041corr  TGGCGCGAAATTGTAAACGTTAATGTTATCGTACCCCTACTCTCAAAAATGTCAAAGATACAGTCTCAGAAAGACCAAAGGGCTATTGA-----
lacZ- $\alpha$ -024      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -025      TGG-----GACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -026      TGG-----AAATTGTAA-----
lacZ- $\alpha$ -027      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -029      TGGG-----ACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -030      -----
lacZ- $\alpha$ -033      TGGG-----AAATTGTAAACGTTAATATTTTGTAAAAATTCGGCTTAAATTTTGTAA-----
lacZ- $\alpha$ -035      TGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCGCATATGGTGCACTCTCAGTACAAATCTGCTCTGATGCCGCATAG-----
lacZ- $\alpha$ -037      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -038      TGGCGCTTCGCTTGGTAA-----
lacZ- $\alpha$ -040      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -041      -----
lacZ- $\alpha$ -042      TGGCGCGACGCGGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -043      TGGCGCTTCGCTTGGTAA-----
lacZ- $\alpha$ -044      -----
lacZ- $\alpha$ -047      TGGCGCGAAATTGTAAACGTTAATGTTTAAACGTTACACCAATAATATATCTGCCAAGATCTCTAAATTCGGGGATCGGAAATCCAGAAGCCCGAGAGGTTGCGCGCTTTTCGGGCTTTTCTTTTCAAAAAAAAATTTATAAAACGATCTGTTTGGCGCGCGCGCGGGTTGTGGCA...
lacZ- $\alpha$ -048      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -049      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -050      TGGCGCTTCGCTTGGTAA-----
lacZ- $\alpha$ -053      TGGCGCTTCGCTTGGTAA-----
lacZ- $\alpha$ -054      TGCATTTTATTCACAAGCGCGCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGA-----
lacZ- $\alpha$ -055      TGA-----AAATTGTAA-----
lacZ- $\alpha$ -056      TGGCGCTTCGCTTGGTAA-----
lacZ- $\alpha$ -058      TGGCGCTTCGCTTGGTAA-----
lacZ- $\alpha$ -059      TGA-----
lacZ- $\alpha$ -061      TGGCGCTTCGCTTGGTAA-----
lacZ- $\alpha$ -062      -----
lacZ- $\alpha$ -063      TGGCGCTTTGCTGGTTTTCGGCACAGAAAGCGGTGCGGAAAGCTGGCTGGAGTGCATCTTCTTGAGGCGGATACTGTCTGCTGTCCTTCAAACTGGCAGATGCACGGTTACGATGCGGCCATCTACACCAACGTGACCTATCCCATTACGGTCAATCGCCGCTTTGTTCCCAAGGAG...
lacZ- $\alpha$ -064      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -065      TGGCGCTTTGCTGGTTTTCGGCACAGAAAGCGGTGCGGAAAGCTGGCTGGAGTGCATCTTCTTGAGGCGGATACTGTCTGCTGTCCTTCAAACTGGCAGATGCACGGTTACGATGCGGCCATCTACACCAACGTGACCTATCCCATTACGGTCAATCGCCGCTTTGTTCCCAAGGAG...
lacZ- $\alpha$ -066      -----
lacZ- $\alpha$ -067      TGGCGCTTAAGAAACCATTTATCATGACATTAACTATAA-----
lacZ- $\alpha$ -069      TGG-----ACGCGCCCTGTAGCGGCGCATTAA-----
lacZ- $\alpha$ -070      TGG-----AAATTGTAA-----
lacZ- $\alpha$ -071      TGGG-----ACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -072      TGGG-----ACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGGTACACTTGCAGCGGCCCTAGCGCCCGGCTCCCTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGGCTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTTC...
lacZ- $\alpha$ -074      TGCATTTTATTCACAAGCGCGCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGA-----
lacZ- $\alpha$ -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- α -040	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCAC TAGT-	pGEM-T
lacZ- α -062	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCAC TAGTA	pSpark Done (linearized)
lacZ- α -064	-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT	pSpark Done (linearized)
lacZ- α -048	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG-----GTCGACCTGCAGGCGGCCGCACTAGT-	pGEM-T Easy
lacZ- α -066	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG-----GTCGACCTGCAGGCGGCCGCACTAGTA	pSpark II (linearized)
lacZ- α -069	-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT	pSpark II (linearized)
lacZ- α -073	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTACTCGAGCTCAACATATG--AAAGTCGACCTGCAGGCGGCCGCACTAGTG	pSpark III, IV and V (linearized)
lacZ- α -076	-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT	pSpark III, IV and V (linearized)

lacZ- α -040	-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -062	-----GAT-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -064	-----ATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -048	-----GATATC-----CCGCGG--CCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -066	-----GAT-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -069	-----ATC-----CCGCGG--CCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -073	GATCCTGAT-----CCGCGG--CCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -076	-----ATCAGGCTGATCCGGAGAATTCGTTTAAAC--CCATGG-----GGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT

lacZ- α -040	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA
lacZ- α -062	-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -064	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA
lacZ- α -048	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA
lacZ- α -066	-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -069	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA
lacZ- α -073	-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCGTTT
lacZ- α -076	TACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTGCCAGCTGGCGTAATAGCGAAGAG...GACGCGCCCTGTAGCGGCGCATTAA

Combined lacZ- α Fragments from Recircularized pSpark Series

Name of Variant	Alignment	Common Restriction Enzyme Sites in MCS
lacZ- α -040	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG--GAGCTCTCCCATATG--GTCGACCTGCAGGCGGCCGCGAATTCAC TA	SacI NdeI SalI PstI NotI EcoRI SpeI
lacZ- α -062+064	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG--GTCGACCTGCAGGCGGCCGCGAATTCAC TA	
lacZ- α -048	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG--GTCGACCTGCAGGCGGCCGCGC-----ACTA	
lacZ- α -066+069	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGG--GAGCTCTCCCATATG--GTCGACCTGCAGGCGGCCGCGC-----ACTA	
lacZ- α -073+076	ATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTACTCGAGCTCAACATATGAAAGTTCGACCTGCAGGCGGCCGCGC-----ACTA	
	*****	*****
	EcoRV SacII EcoRI SacII NotI SphI	
lacZ- α -040	GT-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCG	
lacZ- α -062+064	GT-----GATATC-----GAATTCCTCCGCGGCCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCG	
lacZ- α -048	GT-----GATATCCTCCGCGGCC-----ATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCG	
lacZ- α -066+069	GT-----GATATCCTCCGCGGCC-----ATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCG	
lacZ- α -073+076	GTGGATCCTGATATCAGGCCTGATCCGGA GAATTC-GTTTAAACCCATGG-----GGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC TGGCCGTCG	
	** *****	*****
lacZ- α -040	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
lacZ- α -062+064	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
lacZ- α -048	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
lacZ- α -066+069	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
lacZ- α -073+076	TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC...TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAA	
	*****	*****

lacZ- α -062, lacZ- α -066 and lacZ- α -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- α -064, lacZ- α -069 and lacZ- α -076 missing A at start to complete EcoRV site of linearization (blunt) added in alignment.

Sequence Alignments

Map of Variant	Alignment of Translated Protein	Size (aa)
lacZ-u-041	MTMITPSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRN-----RYQATYRTRPRGGGARYPIRPIVSRTIHWPSFYNVVTGKTALPNLIALQHIP	113
lacZ-u-041_corr	MTMITPSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	141
lacZ-u-023	MTMITPSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	145
lacZ-u-032	MTMITPSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	129
lacZ-u-047	MTMITPSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	216
lacZ-u-012	MTMITPSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	120
lacZ-u-015	MTMITPSAQLTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESYARSLAVVLQRDDWENPGVTQLNRLAAHPP	191
lacZ-u-031	MTMITPSAQLTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESYARSLAVVLQRDDWENPGVTQLNRLAAHPP	192
lacZ-u-005	MTMITPSAQLTLTKGNKSWSSSTAVAAALARADPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESYARSLAVVLQRDDWENPGVTQLNRLAAHPP	191
lacZ-u-025	MTMITPSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	190
lacZ-u-042	MTMITPSSSELTLTKGNKSWSSSTAVAAALELVDPPGCRNISISSLIPSTSRGGP-----VPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	191
lacZ-u-019	MTMITPSSSELTLTKGNKSWVPGPSRSTVSISSLISNCSFGD-----PLVLERPPPRWSSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	190
lacZ-u-037	MTMITPSSSELTLTKGNKSWVPGPSRSTVSISSLISNCSFGD-----PLVLERPPPRWSSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	191
lacZ-u-010	MTMITPSAQLTLTKGNKSWVPGPSRSTVSISSLISNCSFGD-----PLVLERPPPRWSSNS--PYSESYARSLAVVLQRDDWENPGVTQLNRLAAHPP	191
lacZ-u-055	MTMITPSAQLTLTKGNKSWVPGPSRSTVSISSLISNCSFGDQGGQTM-----PLVLERPPPRWSSNS--PYSESYARSLAVVLQRDDWENPGVTQLNRLAAHPP	129
lacZ-u-038	MTMITPSSSELTLTKGTSFAGLNEFALKGEFAAAKF-----NS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	94
lacZ-u-028	MTMITPS--YLGDITIEY---SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	126
lacZ-u-017	MTMITPSSSELTLTKGNKSW--SSRACR-----STLVDPKNSKSFSRVLLERPAHRFSTRVGYQVSVNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	128
lacZ-u-049	MTMITPSSSELTLTKGNK-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	170
lacZ-u-008	MTMITTP-----SLHACR-----FKQSTLDEL-----IKDPAKPRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	116
lacZ-u-030	MTMITTP-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	107
lacZ-u-065	MTMITTP-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	168
lacZ-u-009	MTMITTP-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	76
lacZ-u-057	MTMITTP-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	107
lacZ-u-036	MTMITTP-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	95
lacZ-u-036corr/052	MTMITTP-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	112
lacZ-u-011	MTMITTP-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	384
lacZ-u-001	MTMITTP-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	108
lacZ-u-072	MTMITPSSNTHYRE-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	158
lacZ-u-070	MTMITPSSNTHYRE-----SLHACR-----STLED-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	88
lacZ-u-035	MTMITPSSNTHYRET-----SLHACR-----STLED-----PRVPSNSYSIVSPKS--NSLAVVLQRDDWENPGVTQLNRLAAHPP	127
lacZ-u-033	MTMITPSSNTHYRE-----SLHACR-----STLED-----LLVIWISD-----PRVPSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	110
lacZ-u-075	MTMITPSSNTHYRE-----SLHAGLCRRARDPISRCIRE-----VPSSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	96
lacZ-u-022	MTMITTP-----SLHAGLCRRARDPISRCIRE-----VPSSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	115
lacZ-u-059	MTMITTP-----SLHAPRGVDSRGSPIDG-----VPSSNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	80
lacZ-u-016	MTMITPSYLGDTIEYSSYASSLVPSDDPLVTAASVLEFALKGEFCRYPSSHWRPLEH-----ASRGPN--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	120
lacZ-u-029	MTMITPSYLGDTIEYSSYASSLVPSDDPLVTAASVLEF-----CRYPSSHWRPLEH-----ASRGPN--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	114
lacZ-u-021	MTMITPSYLGDALEYSSYASSLVPSDDPLVTAASVLEFALKGEFCRYPSSHWRPLEH-----ASRGPN--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	120
lacZ-u-027	MTMITTP-----SLVPSDDPLVTAASVLEFALKGEFCRYPSSHWRPLEH-----ASRGPN--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	117
lacZ-u-071	MTMITPSYLGDTIEYSSYASSLVPSDDPLVTAASVLEF-----CRYPSSHWRPLEH-----ASRGPN--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	201
lacZ-u-040	MTMITPSYLGDTIEYSSYASNALGALPYGRPAGGREFTSDIEF--PRPP--WRPGACDV-----GPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	127
lacZ-u-048	MTMITPSYLGDTIEYSSYASNALGALPYGRPAGGR-----TSDI--PR-P--WRPGACDV-----GPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	122
lacZ-u-068	MTMITPSYLGDTIEYSSYASNALGALPYGRPAGGR-----TSDI--PR-P--WRPGACDV-----GPNS--PYSESY--NSLAVVLQRDDWENPGVTQLNRLAAHPP	90
lacZ-u-044	MTMITTP-----SWRAKLACLQVDSRGSFGTELE-----FLINNS-----LAVVLQRDDWENPGVTQLNRLAAHPP	116
lacZ-u-002	MTMITTP-----NLINISGGGGGARDPLGLQERVLAGLHVYTAVRPQASNAYRWYRRRPAGAPVQNSDPRAPGISFWISFLKLNKIILLANQDLSLAVVLQRDDWENPGVTQLNRLAAHPP	86
lacZ-u-003	MTMITTP-----NLINISGGGGGARDPLGLQERVLAGLHVYTAVRPQASNAYRWYRRRPAGAPVQNSDPRAPGISFWISFLKLNKIILLANQDLSLAVVLQRDDWENPGVTQLNRLAAHPP	145
lacZ-u-006	MTMIT-----NSSSVPG-----NSRGSVDLQPSLALAVVLQRDDWENPGVTQLNRLAAHPP	66
lacZ-u-063	MTMIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	168
lacZ-u-004	MTMIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	108
lacZ-u-045	MTMIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	96
lacZ-u-039	MTMIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	100
lacZ-u-013	MTMIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	107
lacZ-u-034	MTMIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	76
lacZ-u-067	M MIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	89
lacZ-u-054	MIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	100
lacZ-u-074	MTMITTP-----NSSSVPG-----DPLESTCRHA-----SL--LAVVLQRDDWENPGVTQLNRLAAHPP	102
lacZ-u-060	MTMIT-----NSSSVPG-----DPLESTCRHA-----SLALAVVLQRDDWENPGVTQLNRLAAHPP	77
lacZ-u-007	MTMIT-----NLIRLTIGNLANLIKARPWIDVNMHDTIRGATSASRSRGRPTGGPYGAAANSSSVPTSRRDVSNNMFKRPGTWDPLESTCRHASFGPDRDNVLSLALAVVLQRDDWENPGVTQLNRLAAHPP	167
lacZ-u-024	MTMITPSSNTHYRESS-----WYACRYRS--GIPGAAAVTSELVD-----SRGSKLTYACMRHRS--SIVSPKF--NSLAVVLQRDDWENPGVTQLNRLAAHPP	122
lacZ-u-018	MTMITPSSNTHYRESSLVFAEAQNKRSKSLATMGVTKETAQAARFERQHIDSGTGGSDDDDKSPGFSKLTQTRYSDPLGDTIEFRYLVDKLLEPRALDHTCGGPPSR-----PHNSLAVVLQRDDWENPGVTQLNRLAAHPP	182
lacZ-u-014	MTMITPSSNTHYRESS-----VPRMLQTRYVSDP-EFVISEF-----VDKILLEPRALDHTCGGPPSRPLYSIVSPKWPHNLSLAVVLQRDDWENPGVTQLNRLAAHPP	126
lacZ-u-026	MTMITPSSNTHYRESS-----VPRMLQTRYVSDP-EFVISEF-----VDKILLEPRALDHTCGGPPSRPLYSIVSPKWPHNLSLAVVLQRDDWENPGVTQLNRLAAHPP	126
lacZ-u-051	MTMITPSSNTHYRESS-----VPRMLQTRYVSDP-EFVISEF-----VDKILLEPRALDHTCGGPPSRPLYSIVSPKWPHNLSLAVVLQRDDWENPGVTQLNRLAAHPP	126
lacZ-u-046	MTMITPSSNTHYRESS-----VPRMLQTRYVSDP-EFVISEF-----PHNSLAVVLQRDDWENPGVTQLNRLAAHPP	116
lacZ-u-043	MTMITPSYVIRLITSGQIFECIARTVRLIEEF-----LQDIWIHEASHGDTVTSRY--LGELWYPLVKALSESYY--GLAVVLQRDDWENPGVTQLNRLAAHPP	123
lacZ-u-050	MTMITPSYVIRLITGQIFECIARTVRLIEEF-----LQDIWIHEASHGDTVTSRY--LGELWYPLVKALSESYY--GLAVVLQRDDWENPGVTQLNRLAAHPP	123
lacZ-u-056	MTMITPSYVIRLITSGARALIEAGWRQAS-----LQDIWIHEFASFGRDASPD--VQACVDPLVKALSESYY--GLAVVLQRDDWENPGVTQLNRLAAHPP	120
lacZ-u-058	MTMITPSYVIRLITGARALIEAGWRQAS-----LQDIWIHEFASFGRDASPD--VQACVDPLVKALSESYY--GLAVVLQRDDWENPGVTQLNRLAAHPP	120
lacZ-u-053	MTMITPSYVIRLITSGYQSSLSGRITGDVSHG-----EAWIHDILQEFLETVRARECIQIFPLVKALSESYY--GLAVVLQRDDWENPGVTQLNRLAAHPP	122
lacZ-u-061	MTMITPSYVIRLITSGSTACLYTPEDAFGRAS-----EFWIHDILQEAFWRLAG--AIEWAPLVKALSESYY--GLAVVLQRDDWENPGVTQLNRLAAHPP	121
lacZ-u-020	MTMITPSLSLTN-----SLAVVLQRDDWENPGVTQLNRLAAHPP	65
lacZ-u-062+064	MTMITPSYLGDTIEYSSYASNALGALPYGRPAGGREFTSRYRIPAAAMAAGSMRRL--AQFAL-----	61
lacZ-u-066+069	MTMITPSYLGDTIEYSSYASNALGALPYGRPAGGR-----TSRY--PAAMAAGSMRRL--AQFAL-----	56
lacZ-u-073+076	MTMITPSYLGDTIEYSSYASNALLELHKMKVDLQA--AALVDPDIRPDENSFKPMGAQFAL-----	60
	:*****	*****

Sequence Alignments

lacZ-<i>α</i>-041	LSPAGVIAKRPAIAPLPHNSCAA -----
lacZ- <i>α</i> -041_corr	FASWRNSEEARTDRPSQQLRSLNGEWREIVNNVIVPLLQKQCRYSLRRPKGY-----
lacZ- <i>α</i> -023	FASWRNSEEARTDRPSQQLRSLNGEWREIVNNVIVPLLQKQCRYSLRRPKGY-----
lacZ- <i>α</i> -032	FASWRNSEEARTDRPSQQLRSLNGEWREIVNNVVIDT-----
lacZ- <i>α</i> -047	FASWRNSEEARTDRPSQQLRSLNGEWREIVNNVNVTPQYILPRSLIPGIGNPEAREVAAFRAFSPFSKKKIYKTIICGRPPGCGQRRWRSTVGNRLRLSTGGAGARSALSTSQGRPIIDIIYIHN
lacZ- <i>α</i> -012	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -015	FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -031	FASWRNSEEARTDRPSQQLRSLNGEWRDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -005	FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -025	FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -042	FASWRNSEEARTDRPSQQLRSLNGEWRDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -019	FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -037	FASWRNSEEARTDRPSQQLRSLNGEWRDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -010	FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -055	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -038	FASWRNSEEARTDRPSQQLRSL-----
lacZ- <i>α</i> -028	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIWCTLSTICSDAA-----
lacZ- <i>α</i> -017	FASWRNSEEARTDRPSQQLRSLNGEWRSNF-----
lacZ- <i>α</i> -049	FASWRNSEEARTDRPSQQLRSLNGEWRDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -008	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIWCTLSTICSDAA-----
lacZ- <i>α</i> -030	FASWRNSEEARTDRPSQQLRSLNGARSFRFLPFLSRHVRLSPSSSKSGAPFRVPI-----
lacZ- <i>α</i> -065	FASWRNSEEARTDRPSQQLRSLNGEWRFAWFPAPAEVPSWLECDLPEADTVVPSNNQMHGVDAPIIYTNVTYPIITVNPFFVPTENPTGYSILTFNVDESWLQEGQTRIIIDGVPVIG-----
lacZ- <i>α</i> -009	FASWRNSEEARTDRPSQQLRSLNGE-----
lacZ- <i>α</i> -057	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIWCTLSTICSDAA-----
lacZ-<i>α</i>-036	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHR -----
lacZ- <i>α</i> -036corr/052	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIRWAILMKNILNLLSY-----
lacZ- <i>α</i> -011	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIRQSNHSTRPVAAH-----
lacZ- <i>α</i> -001	FASWRNSEEARTDRPSQQLRSLNGE LD LFNKAAVPSSQORNALPVLQPINQF-----
lacZ- <i>α</i> -072	FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD-----
lacZ- <i>α</i> -070	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -035	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIWCTLSTICSDAA-----
lacZ- <i>α</i> -033	FASWRNSEEARTDRPSQQLRSLNGEWEIVNNVILLKFALNFC-----
lacZ- <i>α</i> -075	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -022	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIWCTLSTICSDAA-----
lacZ- <i>α</i> -059	FASWRNSEEARTDRPSQQLRSLNGE-----
lacZ- <i>α</i> -016	FASWRNSEEARTDRPSQQLRSL-----
lacZ- <i>α</i> -029	FASWRNSEEARTDRPSQQLRSL-----
lacZ- <i>α</i> -021	FASWRNSEEARTDRPSQQLRSL-----
lacZ- <i>α</i> -027	FASWRNSEEARTDRPSQQLRSLNGEWTTPVAAH-----
lacZ- <i>α</i> -071	FASWRNSEEARTDRPSQQLRSLNGEW-DAPCSGALSAAGVVVTRSVTATLASALAPAPFAFFPSPFLATFAGFPRQALNRGLPLGFRFRALRHLDPKKLDLDGDSRSGPSP-----
lacZ- <i>α</i> -040	FASWRNSEEARTDRPSQQLRSLNGEWTTPVAAH-----
lacZ- <i>α</i> -048	FASWRNSEEARTDRPSQQLRSLNGEWTTPVAAH-----
lacZ- <i>α</i> -068	FASWRNSEEARTDRPSQQLRS-----
lacZ- <i>α</i> -044	FASWRNSEEARTDRPSQQLRSLNGE LS FRFLPFLSRHVRLSPSSSKSGAPFRVPI-----
lacZ- <i>α</i> -002	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -003	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -006	FASWRNSEEARTDRPSQQLRSL-----
lacZ- <i>α</i> -063	FASWRNSEEARTDRPSQQLRSLNGEWRFAWFPAPAEVPSWLECDLPEADTVVPSNNQMHGVDAPIIYTNVTYPIITVNPFFVPTENPTGYSILTFNVDESWLQEGQTRIIIDGVPVIG-----
lacZ- <i>α</i> -004	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIRQSNHSTRPVAAH-----
lacZ- <i>α</i> -045	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIG-----
lacZ- <i>α</i> -039	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIDPVRR-----
lacZ- <i>α</i> -013	FASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIWCTLSTICSDAA-----
lacZ- <i>α</i> -034	FASWRNSEEARTDRPSQQLRSLNGE-----
lacZ- <i>α</i> -067	FASWRNSEEARTDRPSQQLRSLNGEWRLRNHYHIDNL-----
lacZ- <i>α</i> -054	FASWRNSEEARTDRPSQQLRSLNGE LD LFNKAAVPSSQORNALPVLQPINQF-----
lacZ- <i>α</i> -074	FASWRNSEEARTDRPSQQLRSLNGE LD LFNKAAVPSSQORNALPVLQPINQF-----
lacZ- <i>α</i> -060	FASWRNSEEARTDRPSQQLRSLNGE LD -----
lacZ- <i>α</i> -007	FASWRNSEEARTDRPSQQLRSLNGEWEIVNNVILLIFC-----
lacZ- <i>α</i> -024	FASWRNSEEARTDRPSQQLRSLNGEWTTPVAAH-----
lacZ- <i>α</i> -018	FASWRNSEEARTDRPSQQLRSLNGEWEIVNNVILLKFALNFC-----
lacZ- <i>α</i> -014	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -026	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -051	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -046	FASWRNSEEARTDRPSQQLRSLNGEWKL-----
lacZ- <i>α</i> -043	FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- <i>α</i> -050	FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- <i>α</i> -056	FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- <i>α</i> -058	FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- <i>α</i> -053	FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- <i>α</i> -061	FASWRNSEEARTDRPSQQLRSLNGEWRFAW-----
lacZ- <i>α</i> -020	FASWRNSEEARTDRPSQQLRSLNGEWR-----
lacZ- <i>α</i> -062+064	-----
lacZ- <i>α</i> -066+069	-----
lacZ- <i>α</i> -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

Name of Variant	# of Occur	Alignment	Size (bp)	# Var	Sources*
MBP-004	10	-----ATG-----AAAATCGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1101	0	16497515
MBP-010	10	-----AAAATCGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1098	3d	20213425
MBP-001	7	-----AAAATCGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1098	1/3d	16377204
MBP-002	2	-----ATGGGTAAATCGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1104	2	16377204
MBP-009	2	-----ATGGC-----CGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1098	4/3d	AB
MBP-003	1	-----ATG-----AAAATCGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1101	1	19906724
MBP-005	1	-----AAAACGGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1098	3	19906724
MBP-006	1	-----AAAATCGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1074	14	MCSG
MBP-007	1	ATGAAAAATAAAACAGGTGCACGCATCCTCGCATTTATCCGCATTAAACGACGATGATGTTTTCGCCCTCGGCTCTCGCCAAAATCGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1176	3	?
MBP-008	1	-----AAAATCGAAGAAGGTAACCTGGTAATCTGGATTAAACGGCGAT	1074	1/3d	MCSG

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

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

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

MBP-004		CTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGC			
MBP-010		CTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGC			
MBP-001		CTGGATAAAGAACTGAAAGCGAAAGGTAAGAG			

Sequence Alignments

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

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

MBP-004	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-010	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-001	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-002	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-009	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-003	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-005	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-006	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-007	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG
MBP-008	AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCAGAAAAGTGAAATCATGCCGAACATCCCG

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

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	-----MKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	367	0
MBP-010	-----KIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	366	0
MBP-001	-----KIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	366	0
MBP-002	-----MGKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	368	0
MBP-009	-----M ^{ATGC} IEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	366	2
MBP-003	-----MKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	367	0
MBP-005	----- ^{ATGC} KIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	366	1
MBP-006	-----KIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEIT ^{ATGC} PDKAFQDKLYPFTW	358	5
MBP-007	MKIKTGARILALSALTMMFSASALAKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	392	1
MBP-008	-----KIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW	358	0

MBP-004	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-010	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-001	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-002	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-009	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-003	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-005	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-006	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKY ^{ATGC} ENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-007	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		
MBP-008	DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAE		

MBP-004	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-010	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-001	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-002	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-009	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-003	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-005	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-006	AAFNKGETAMTINGPWAWSNIDTS ^{ATGC} VNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-007	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		
MBP-008	AAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFFKGQP...LEAVNKDKPLGAVALKS ^{ATGC} YEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT		

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---TCCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	654	0	GE,No,BD
GST-004	3	ATG GGA TCCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	657	0/3i	No
GST-005	2	---TCCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	651	0/3d	19906724
GST-001	2	ATG---TCCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	654	1	No
GST-003	1	ATG---TCCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA TAA	657	0/3i	#
GST-006	1	ATG--- AG CCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	654	1	AB
GST-007	1	---TCCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	648	6	15232106
GST-008	1	ATG GAA TCCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTT.GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA---	657	3	15782178

Name of Variant	Alignment of Translated Protein	Size (aa)	Tag Pos	# Var aa
GST-002	-MSPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	218	N/C ¹	0
GST-004	M -MSPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	219	N	1i
GST-005	--SPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.KKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK	217	N ²	1d
GST-001	- M -MSPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.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	M -MSPILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYI.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	CAIHost/Non-Host ¹	Sources*
HA-008	10	TACCC T TACGA T GT T CCAGA T TACGC T	27	5	Mammal-Plant	0.783/0.508	Or,C1
HA-007	5	TA T CCGTACGACGT T CCGACTACGC T	27	3	Mammal	0.463/0.510	C1
HA-003	4	TA T CCGT T GA T GTGCC G GACTACGC T	27	5	E. coli	0.424/0.571	MCSG
HA-009	3	TA T CC T TA T ACGCTGCC T GACTA T GCC	27	5	Mammal-S. cer.	0.941-0.071/0.155	C1
HA-002	2	TA T CCGTACGACGT T CCGACTACGC T	27	4	E. coli	0.642/0.395	MCSG
HA-004	2	TACCC T TACGA T GT T CCG G ATACGC T	27	6	Mammal	0.668/0.640	Mo
HA-006	2	TACCC T TACGACGT T CCG G ATACGC T	27	4	S. cer.	0.182/0.478	C1
HA-001	1	TA T CCGT T GA T ACGTGCC G GACTA T GCC	27	4	Mammal	0.784/0.140	Lu
HA-005	1	TACCC T TA T GA T GTGCCG G AT T AGCC	27	5	S. cer.	0.108/0.153	8242750
HA-010	1	TACCCCTACGACGTGCC G GACTACGCC	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
 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