

###

TGGAAGGGCTAATTCACCTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAAGTACACACCAGG < 100
 E G L I H S Q R R Q D I L D L W I Y H T Q G Y F P D W Q N Y T P G
 ACCTTCCCGATTAAGTGAGGGTTTCTTCTGTTCTATAGGAACTAGACACCTAGATGGTGTGTCCGATGAAGGGACTAACCGTCTTGATGTGTGGTCC
 10 20 30 40 50 60 70 80 90

GCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGTCTACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAgAGGCCAATAAAGGAGAGAACACCAGC < 200
 P G V R Y P L T F G W C Y K L V P V E P D K V E E A N K G E N T S
 CGGTCCCCAGTCTATAGGTGACTGGAAACCTACCACGATGTTTCGATCATGGTCAACTCGGTCTATTCCATCTTcTCCGGTTATTTCCCTCTCTTGTGGTCCG
 110 120 130 140 150 160 170 180 190

TTGTTACACCCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCCGAG < 300
 L L H P V S L H G M D D P E R E V L E W R F D S R L A F H H V A R E
 AACAAATGTGGGACACTCGGACGTACCTTACCTACTGGGACTCTCTCTTCCACAATCTCACCTCCAAACTGTCGGCGGATCGTAAAGTAGTGCACCGGGCTC
 210 220 230 240 250 260 270 280 290

AGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGG < 400
 L H P E Y F K N C * H R A C Y K G L S A G D F P G R R G L G G T G
 TCGACGTAGGCCTCATGAAGTTCTTGACGACTGTAGCTCGAACGATGTTCCCTGAAAGGGCGACCCCTGAAAGGTCCCTCCGACCCGGACCCGCCCTGACC
 310 320 330 340 350 360 370 380 390

GGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTA < 500
 E W R A L R C C I * A A A F C L Y W V S L V R P D L S L G A L W L
 CCTCACCGCTCGGGAGTCTACGACGTATATTCGTCGACGAAAAACGGACATGACCCAGAGAGACCAATCTGGTCTAGACTCGGACCCTCGAGAGACCGAT
 410 420 430 440 450 460 470 480 490

ACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCCTTGAGTGCTCAAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTC < 600
 T R E P T A * A S I K L A L S A Q S S V C P S V V * L W * L E I P Q
 TGATCCCTTGGGTGACGAATTCGGAGTTATTTGAAACGGAACCTCACGAGTTTCATCACACACGGGCAGACAACACACTGAGACCATTGATCTCTAGGGAG
 510 520 530 540 550 560 570 580 590

AGACCCTTTTGTAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCG < 700
 T L L V S V E N L * Q W R P N R D L K A K V K P E E I S R R R T R
 TCTGGGAAAATCAGTCACACCTTTAGAGATCGTCACCGCGGGCTTGTCCTGAACTTTTCGCTTTTCATTTCCGGTCTCCTCTAGAGAGCTGCGTCTGAGC
 610 620 630 640 650 660 670 680 690

BssHII
McaTI

GCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGCGGGCGACTGGTGTGAGTACGCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAG < 800

L A E A R T A R G E G R R L V S T P K I L T S G G * K E R D G C E
CGAACGACTTCGCGCGTGCCGTTCTCCGCTCCCCGCGCTGACCACTCATGCGGTTTTTAAACTGATCGCCTCCGATCTTCTCTCTACCCACGCTC
710 720 730 740 750 760 770 780 790

AGCGTCGGTATTAAGCGGGGAGAATTAGATAAAATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAAGAAACAATATAAACTAAAACATATAGTATGGGCA < 900
S V G I K R G R I R * M G K N S V K A R G K E T I * T K T Y S M G K
TCGAGCCATAATTCGCCCCCTCTTAATCTATTTACCCTTTTTTAAGCCAATTCCGGTCCCCCTTTCTTTGTTATATTTGATTTTGTATATCATAACCCGT
810 820 830 840 850 860 870 880 890

AGCAGGGAGCTAGAACGATTTCGCAGTTAATCCTGGCCTTTTTAGAGACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAG < 1000
Q G A R T I R S * S W P F R D I R R L * T N T G T A T T I P S D R
TCGTCCCTCGATCTTGCTAAGCGTCAATTAGGACCGGAAAATCTCTGTAGTCTTCCGACATCTGTTTATGACCCTGTGCGATGTTGGTAGGGAAGTCTGTC
910 920 930 940 950 960 970 980 990

GATCAGAAGAACTTAGATCATTATATAATACAATAGCAGTCCTCTATTGTGTGCATCAAAGGATAGATGTAAAAGACACCAAGGAAGCCTTAGATAAGAT < 1100
I R R T * I I I * Y N S S P L L C A S K D R C K R H Q G S L R * D
CTAGTCTTCTTGAATCTAGTAATATATTATGTTATCGTCAGGAGATAACACACGTAGTTTCCCTATCTACATTTTCTGTGGTTCCCTTCGGAATCTATTCTA
1010 1020 1030 1040 1050 1060 1070 1080 1090

AGAGGAAGAGCAAAAACAAAAGTAAGAAAAAGGCACAGCAAGCAGCAGCTGCACAGGAAACAACAGCCAGGTCAGCCAAAATTACCCTATAGTGCAGAAC < 1200
R G R A K Q K * E K G T A S S S * H R K Q Q P G Q P K L P Y S A E P
TCTCCTTCTCGTTTTGTTTTTCAATTCCTTTTTCCGTTGTCGTTTCGTCGACTGTGTCTTTGTTGTGCGGTCCAGTCGGTTTTAATGGGATATCACGTCTTG
1110 1120 1130 1140 1150 1160 1170 1180 1190

CTCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCA < 1300
P G A N G T S G H I T * N F K C M G K S S R R E G F Q P R S N T H
GAGGTCCCCGTTTACCATGTAGTCCGGTATAGTGGATCTTGAAATTTACGTACCCATTTTCATCATCTTCTCTTCCGAAAGTCGGGTCTTCATTATGGGT
1210 1220 1230 1240 1250 1260 1270 1280 1290

TGTTTTTCAGCATTATCAGAAGGAGCCACCCCAAGATTTAAATACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGAC < 1400
V F S I I R R S H P T R F K Y H A K H S G G T S S S H A N V K R D
ACAAAAGTCGTAATAGTCTTCCCTCGGTGGGGTGTCTAAATTTATGGTACGATTTGTGTACCCCCCTGTAGTTTCGTCGGTACGTTTACAATTTTCTCTG
1310 1320 1330 1340 1350 1360 1370 1380 1390

CATCAATGAGGAAGCTGCAGAATGGGATAGATTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCA < 1500
H Q * G S C R M G * I A S S A C R A Y C T R P D E R T K G K * H S R
GTAGTTACTCCTTCGACGTCTTACCCTATCTAACGTAGGTCACGTACGTCCCGGATAACGTGGTCCGGTCTACTCTCTTGGTTCCCTTCACTGTATCGT
1410 1420 1430 1440 1450 1460 1470 1480 1490

SpeI
>**PsrI**
| |
GGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACACATAATCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAA < 1600
N Y * Y P S G T N R M D D T * S T Y P S R R N L * K M D N P G I K

CCTTGATGATCATGGGAAGTCCTTGTATTATCCTACTGTGTATTAGGTGGATAGGGTCATCCTCTTTAGATATTTTCTACCTATTAGGACCCTAATT
1510 1520 1530 1540 1550 1560 1570 1580 1590

>TaqII

ATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCCTTTAGAGACTATGTAGACCGATTCTATAAACTCT < 1700
* N S K N V * P Y Q H S G H K T R T K G T L * R L C R P I L * N S
TATTTTATCATTCTTACATATCGGGATGGTCGTAAGACCTGTATTCTGTTCTGGTTTCCTTGGGAAATCTCTGATACATCTGGCTAAGATATTTTGAGA
1610 1620 1630 1640 1650 1660 1670 1680 1690

AAGAGCCGAGCAAGCTTCCACAAGAGGTAAAAAATTGGATGACAGAAACCTTGTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTG < 1800
K S R A S F T R G K K L D D R N L V G P K C E P R L * D Y F K S I G
TTCTCGGCTCGTTCTGAAGTGTCTCCATTTTTTAACTACTGTCTTTGGAACAACCGGTTTACGCTTGGGTCTAACATTCTGATAAAATTTTCGTAAC
1710 1720 1730 1740 1750 1760 1770 1780 1790

GGACCAGGAGCGACACTAGAAGAAATGATGACAGCATGTGAGGGAGTGGGGGGACCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAA < 1900
T R S D T R R N D D S M S G S G G T R P * S K S F G * S N E P S N
CCTGGTCCCTCGCTGTGATCTTCTTTACTACTGTGCTACAGTCCCTCACCCCTGGGCCGATTTTCGTTCTCAAACCGACTTCGTTACTCGGTTTATT
1810 1820 1830 1840 1850 1860 1870 1880 1890

CAAATCCAGCTACCATAATGATACAGAAAGGCAATTTTTAGGAACCAAAGAAAGACTGTAAAGTGTTCATTTGTGGCAAAGAAGGGCACATAGCCAAAAA < 2000
K S S Y H N D T E R Q F * E P K K D C * V F Q L W Q R R A H S Q K
GTTTAGGTGCGATGGTATTACTATGTCTTTCCGTTAAAATCCTTGGTTTCTTTCTGACAATTCAAAAGTTAACACCGTTTCTTCCCGTGTATCGGTTTTT
1910 1920 1930 1940 1950 1960 1970 1980 1990

TTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTTTTTAGGGAAGATC < 2100
L Q G P * E K G L L E M W K G R T P N E R L Y * E T G * F F R E D L
AACGTCCCGGGGATCCTTTTTCCCGACAACCTTTACACCTTTTCTTCTGTTGTTTACTTTCTAACATGACTCTCTGTCCGATTAAAAAATCCCTTCTAG
2010 2020 2030 2040 2050 2060 2070 2080 2090

TGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAgGTTTGGGGAAGAGACAACAA < 2200
A F P Q G K A R E F S S E Q T R A N S P T R R E L Q V W G R D N N
ACCGGAAGGGTGTTCCTTCCGGTCCCTTAAAAGAAGTCTCGTCTGGTCTCGGTTGTCCGGGTGGTCTTCTCTCGAAGTCAAACCCCTTCTCTGTTGTT
2110 2120 2130 2140 2150 2160 2170 2180 2190

CTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAGGG < 2300
S L S E A G A D R Q G T V S F S F P Q I T L W Q R P L V T I K I G
GAGGGAGAGTCTTCGTCCTCGGCTATCTGTTCTTTGACATAGGAAATCGAAGGGAGTCTAGTGAGAAACCGTCGCTGGGGAGCAGTGTATTCTATCCC
2210 2220 2230 2240 2250 2260 2270 2280 2290

GGGCAATTAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAA < 2400
G Q L K E A L L D T G A D D T V L E E M N L P G R W K P K M I G G I
CCCCTTAATTTCTTTCGAGATAATCTATGTCCTCGTCTACTATGTCATAATCTTCTTTACTTAAACGGTCTTCTACCTTTGGTTTTTACTATCCCCCTT
2310 2320 2330 2340 2350 2360 2370 2380 2390

TTGGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGCGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAA < 2500
G G F I K V R Q Y D Q I L I E I C G H K A I G T V L V G P T P V N
AACCTCCAAAATAGTTTTATTCTGTCTACTAGTCTATGAGTATCTTTAGACGCCTGTATTTTCGATATCCATGTGCATAATCATCCTGGATGTGGACAGTT
2410 2420 2430 2440 2450 2460 2470 2480 2490

CATAATTGGAAGAAATCTGTTGACTCAGATTGGCTGCACTTTAAATTTTCCATTAGTCCTATTGAGACTGTACCAGTAAAATTAAGCCAGGAATGGAT < 2600
I I G R N L L T Q I G C T L N F P I S P I E T V P V K L K P G M D
GTATTAACCTTCTTTAGACAACCTGAGTCTAACCGACGTGAAATTTAAAGGGTAATCAGGATAACTCTGACATGGTCATTTTAAATTTTCGGTCCTTACCTA
2510 2520 2530 2540 2550 2560 2570 2580 2590

GGCCAAAAGTTAAACAATGGCCATTGACAGAAGAAAAATAAAAGCATTAGTAGAAATTTGTACAGAAATGGAAAAGGAAGGAAAAATTTCAAAAATTG < 2700
G P K V K Q W P L T E E K I K A L V E I C T E M E K E G K I S K I G
CCGGGTTTTCAATTTGTTACCGTAACTGTCTTCTTTTTTATTTTCGTAATCATCTTTAAACATGTCTTTACCTTTTCCTTCTTTTTTAAAGTTTTTAAAC
2610 2620 2630 2640 2650 2660 2670 2680 2690

GGCCTGAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAGACAGTACTAAATGGAGAAAATTTAGTAGATTTTCAGAGAACTTAATAAGAGAAC < 2800
P E N P Y N T P V F A I K K K D S T K W R K L V D F R E L N K R T
CCGGACTTTTAGGTATGTTATGAGGTACATAAACGGTATTTCTTTTTCTGTTCATGATTTACCTCTTTTAAATCATCTAAAGTCTCTTGAATTATTCTCTTG
2710 2720 2730 2740 2750 2760 2770 2780 2790

>BtgZI

TCAAGATTTCTGGGAAGTTCAATTAGGAATACCACATCCTGCAGGGTTAAAACAGAAAAATCAGTAACAGTACTGGATGTGGGCGATGCATATTTTTTCA < 2900
Q D F W E V Q L G I P H P A G L K Q K K S V T V L D V G D A Y F S
AGTTCTAAAGACCCTTCAAGTTAATCCTTATGGTGTAGGACGTCCAATTTTGTCTTTTTTAGTCATTGTTCATGACCTACACCCGCTACGTATAAAAAGT
2810 2820 2830 2840 2850 2860 2870 2880 2890

BstZ17I

GTTCCCTTAGATAAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTTCCAC < 3000
V P L D K D F R K Y T A F T I P S I N N E T P G I R Y Q Y N V L P Q
CAAGGGAATCTATTTCTGAAGTCCTTCATATGACGTAAATGGTATGGATCATATTTGTTACTCTGTGGTCCCTAATCTATAGTCATGTTACACGAAGGTG
2910 2920 2930 2940 2950 2960 2970 2980 2990

Tth111I

AGGGATGGAAAGGATCACCAGCAATATTCCAGTGTAGCATGACAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTCATCTATCAATACAT < 3100
G W K G S P A I F Q C S M T K I L E P F R K Q N P D I V I Y Q Y M
TCCCTACCTTTCTAGTGGTCTGTTATAAGGTCACATCGTACTGTTTTTAGAATCTCGGAAAATCTTTTGTTTTAGGTCTGTATCAGTAGATAGTTATGTA
3010 3020 3030 3040 3050 3060 3070 3080 3090

GGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAATAGAGGAACCTGAGACAACATCTGTTGAGGTGGGGATTTACCACACCA < 3200
D D L Y V G S D L E I G Q H R T K I E E L R Q H L L R W G F T T P
CCTACTAAACATACATCCTAGACTGAATCTTTATCCCGTCTGATCTTGTTTTTATCTCCTTACTCTGTTGTAGACAACCTCCACCCCTAAATGGTGTGGT
3110 3120 3130 3140 3150 3160 3170 3180 3190

GACAAAAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTATGAACGCCATCCTGATAAAATGGACAGTACAGCCTATAGTGCTGCCAGAAAAGGACA < 3300
D K K H Q K E P P F L W M G Y E L H P D K W T V Q P I V L P E K D S
CTGTTTTTTGTAGTCTTTCTTTGGAGGTAAGGAAACCTACCCAATACTTGAGGTAGGACTATTTACCTGTCTATGTCGGATATCACGACGGTCTTTTCCCTGT
3210 3220 3230 3240 3250 3260 3270 3280 3290

GCTGGACTGTCAATGACATACAGAAATTAGTGGGAAAAATTGAATTGGGCAAGTCAGATTTATGCAGGGATTAAAGTAAGGCAATTATGTAAACTTCTTAG < 3400
W T V N D I Q K L V G K L N W A S Q I Y A G I K V R Q L C K L L R
CGACCTGACAGTTACTGTATGTCTTTAATCACCCTTTTAACCTTAACCCGTTTCAGTCTAAATACGTCCCTAATTTTCATTCGGTTAATACATTTGAAGAATC
3310 3320 3330 3340 3350 3360 3370 3380 3390

AgeI

GGGAACCAAAGCACTAACAGAAGTAGTACCACCTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAGATTCTAAAAGAACCAGGTACATGGAGTG < 3500
G T K A L T E V V P L T E E A E L E L A E N R E I L K E P V H G V
CCCTTGGTTTTCGTGATTTGTCCTTCATCATGGTGATTTGTCCTTCTTCGTCTCGATCTTGACCGTCTTTTGTCCCTCTAAGATTTTCTTTGGCCATGTACCTCAC
3410 3420 3430 3440 3450 3460 3470 3480 3490

TATTATGACCCATCAAAAGACTTAAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATCTGAAAA < 3600
Y Y D P S K D L I A E I Q K Q G Q G Q W T Y Q I Y Q E P F K N L K T
ATAAATACTGGGTAGTTTTCTGAATTTATCGTCTTTATGTCTTCGTCCCCGTTCCGGTTACCTGTATAGTTTAAATAGTTCTCGGTAAATTTTTTAGACTTTTT
3510 3520 3530 3540 3550 3560 3570 3580 3590

CAGAAAAGTATGCAAGAATGAAGGGTGCCACACTAATGATGTGAAACAATTAACAGAGGCAGTACAAAAATAGCCACAGAAAGCATAGTAATATGGGG < 3700
G K Y A R M K G A H T N D V K Q L T E A V Q K I A T E S I V I W G
GTCCTTTCATACGTTTACTTCCCACGGGTGTGATTTACTACACTTTGTTAATTTGTCCTCCGTCATGTTTTTTTATCGGTGTCTTTTCGTATCATTATACCCC
3610 3620 3630 3640 3650 3660 3670 3680 3690

AAAGACTCCTAAATTTAAATTTACCCATACAAAAGGAAACATGGGAAGCATGGTGGACAGAGTATTTGGCAAGCCACCTGGATTCTTGAGTGGGAGTTTGTCT < 3800
K T P K F K L P I Q K E T W E A W W T E Y W Q A T W I P E W E F V
TTTCTGAGGATTTAAATTTAATGGGTATGTTTCTTTGTACCCTTCGTACCACCTGTCTCATAACCGTTCCGGTGGACCTAAGGACTCACCCCTCAAACAG
3710 3720 3730 3740 3750 3760 3770 3780 3790

>BarI

AATACCCCTCCCTTAGTGAAGTTATGGTACCAGTTAGAGAAAAGAACCATAATAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCCAATAGGGAAACTA < 3900
N T P P L V K L W Y Q L E K E P I I G A E T F Y V D G A A N R E T K
TTATGGGGAGGGAATCACTTCAATACCATGGTCAATCTTCTTTGGGTATTTATCCTCGTCTTTGAAAGATACATCTACCCCGTCGGTTATCCCTTTGAT
3810 3820 3830 3840 3850 3860 3870 3880 3890

AATTAGGAAAAGCAGGATATGTAACCTGACAGAGGAAGACAAAAGTTGTCCCTAACGGACACAACAAATCAGAAGACTGAGTTACAAGCAATTCATCT < 4000
L G K A G Y V T D R G R Q K V P L T D T T N Q K T E L Q A I H L
TTAATCCTTTTTCGTCTATACATTTGACTGTCTCTCTTTCTGTTTTTCAACAGGGGATTCCTGTGTTGTTTAGTCTTCTGACTCAATGTTTCGTTAAGTAGA
3910 3920 3930 3940 3950 3960 3970 3980 3990

AGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTGACAGACTCACAAATATGCATTTGGGAATCATTCAAGCACAAACCAGATAAGAGTGAATCAGAGTTA < 4100

A L Q D S G L E V N I V T D S Q Y A L G I I Q A Q P D K S E S E L
TCGAAACGTCCTAAGCCCTAATCTTTCATTGATCACTGTCTGAGTGTATACGTAACCCCTTAGTAAGTTCGTGTTGGTCTATTCTCACTTAGTCTCAAT
4010 4020 4030 4040 4050 4060 4070 4080 4090

GTCAGTCAAATAATAGAGCAGTTAATAAAAAAGGAAAAAGTCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAAT < 4200
V S Q I I E Q L I K K E K V Y L A W V P A H K G I G G N E Q V D K L
CAGTCAGTTTATTATCTCGTCAATTATTTTTTCTTTTTTTCAGATGGACCGTACCCATGGTCTGTGTTTCTTAACCTCTTTACTTGTTCATCTATTTA
4110 4120 4130 4140 4150 4160 4170 4180 4190

TGGTCAGTGTCTGGAATCAGGAAAGTACTATTTTTAGATGGAATAGATAAGGCCCAAGAAGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAG < 4300
V S A G I R K V L F L D G I D K A Q E E H E K Y H S N W R A M A S
ACCAGTCACGACCTTAGTCCTTTTCATGATAAAAATCTACCTTATCTATTCCGGGTCTTTCTTGTACTCTTTATAGTGTCAATTAACCTCTCGTTACCGATC
4210 4220 4230 4240 4250 4260 4270 4280 4290

TGATTTTTAACCTACCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGGGAAGCCATGCATGGACAAGTAGACTGTAGC < 4400
D F N L P P V V A K E I V A S C D K C Q L K G E A M H G Q V D C S
ACTAAAATTGGATGGTGGACATCATCGTTTTCTTTATCATCGGTGCGACACTATTTACAGTGCATTTTTCCCTTCGGTACGTACCTGTTTCATCTGACATCG
4310 4320 4330 4340 4350 4360 4370 4380 4390

CCAGGAATATGGCAGCTAGATTGTACACATTTAGAAGGAAAAGTTATCTTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTAATTCCAG < 4500
P G I W Q L D C T H L E G K V I L V A V H V A S G Y I E A E V I P A
GGTCCTTATACCGTCGATCTAACATGTGTAAATCTTCTTTTCAATAGAACCATCGTCAAGTACATCGGTACCTATATATCTTCGTCTTCATTAAGGTC
4410 4420 4430 4440 4450 4460 4470 4480 4490

CAGAGACAGGGCAAGAAACAGCATACTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAGTACATACAGACAATGGCAGCAATTTACCAGTAC < 4600
E T G Q E T A Y F L L K L A G R W P V K T V H T D N G S N F T S T
GTCTCTGTCCCCTTTCTTTGTCGTATGAAGGAGAATTTAATCGTCCTTCTACCGGTCAATTTTTGTCTATGTATGTCTGTTACCGTCGTTAAAGTGGTCATG
4510 4520 4530 4540 4550 4560 4570 4580 4590

TACAGTTAAGGCCCGCCTGTTGGTGGGCGGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAAGGAGTAATAGAATCTATGAATAAAGAA < 4700
T V K A A C W W A G I K Q E F G I P Y N P Q S Q G V I E S M N K E
ATGTCAATTCGGGCGGACAACCACCCGCCCTAGTTTCGTCTTAAACCGTAAGGGATGTTAGGGGTTTTCAGTTCCTCATTATCTTAGATACTTATTTCTT
4610 4620 4630 4640 4650 4660 4670 4680 4690

TTAAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGA < 4800
L K K I I G Q V R D Q A E H L K T A V Q M A V F I H N F K R K G G I
AATTTCTTTTAAATATCTCTGTCCATTCTCTAGTCCGACTTGTAGAATTCTGTGTCATGTTTACCGTCATAAGTAGGTGTTAAAATTTTCTTTTCCCCCT
4710 4720 4730 4740 4750 4760 4770 4780 4790

TTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAATTTTCG < 4900
G G Y S A G E R I V D I I A T D I Q T K E L Q K Q I T K I Q N F R
AACCCCATGTACAGTCCCCTTTCTTATCATCTGTATTATCGTTGTCTGTATGTTTGAATTTCTTAATGTTTTTGTAAATGTTTTTAAAGTTTTAAAAGC

4810 4820 4830 4840 4850 4860 4870 4880 4890

GGTTTATTACAGGGACAGCAGAGATCCAGTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATA < 5000
V Y Y R D S R D P V W K G P A K L L W K G E G A V V I Q D N S D I
CCAAATAATGTCCCTGTCTCTAGGTCAAACCTTTCCCTGGTCGTTTCGAGGAGACCTTTCCACTTCCCGTGCATCATTATGTTCTATTATCACTGTAT
4910 4920 4930 4940 4950 4960 4970 4980 4990

AAAGTAGTGCCAAGAAGAAAAGCAAAGATCATCAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAACACA < 5100
K V V P R R K A K I I R D Y G K Q M A G D D C V A S R Q D E D * H M
TTTCATCACGGTTCTTCTTTTCGTTTCTAGTAGTCCCTAATAACCTTTTGTCTACCGTCCACTACTAACACACCGTTCATCTGTCTACTCCTAATTGTGT
5010 5020 5030 5040 5050 5060 5070 5080 5090

TGGAAAAGATTAGTAAAACACCATATGTATATTTCAAGGAAAGCTAAGGACTGGTTTTATAGACATCACTATGAAAGTACTAATCCAAAAATAAGTTCAG < 5200
E K I S K T P Y V Y F K E S * G L V L * T S L * K Y * S K N K F R
ACCTTTTCTAATCATTTTTGTGGTATAACATATAAAGTTCCTTTTCGATTCCGACAAAATATCTGTAGTGATACTTTCATGATTAGGTTTTTATTCAAGTC
5110 5120 5130 5140 5150 5160 5170 5180 5190

AAGTACACATCCCCTAGGGGATGCTAAATTAGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGACTGGCATTGGGGTCAGGGAGTCTCCAT < 5300
S T H P T R G C * I S N N N I L G S A Y R R K R L A F G S G S L H
TTCATGTGTAGGGTATCCCTACGATTTAATCATTATTGTTGTATAACCCAGACGTATGTCCTCTTTCTCTGACCGTAAACCCAGTCCCTCAGAGGTA
5210 5220 5230 5240 5250 5260 5270 5280 5290

AGAATGGAGGAAAAAGAGATATAGCACACAAGTAGACCCTGACCTAGCAGACCAACTAATTCATCTGCACTATTTTGATTGTTTTTTCAGAATCTGCTATA < 5400
R M E E K E I * H T S R P * P S R P T N S S A L F * L F F R I C Y K
TCTTACCTCCTTTTTCTCTATATCGTGTGTTTCATCTGGGACTGGATCGTCTGGTTGATTAAGTAGACGTGATAAACTAACAAAAAGTCTTAGACGATAT
5310 5320 5330 5340 5350 5360 5370 5380 5390

AGAAATACCATATTAGGACGTATAGTTAGTCCTAGGTGTGAATATCAAGCAGGACATAACAAGGTAGGATCTCTACAGTACTTGGCACTAGCAGCATTA < 5500
K Y H I R T Y S * S * V * I S S R T * Q G R I S T V L G T S S I N
TCTTTATGGTATAATCCTGCATATCAATCAGGATCCACACTTATAGTTTCGTCTGTATTGTTCCATCCTAGAGATGTCATGAACCGTGATCGTTCGTAATT
5410 5420 5430 5440 5450 5460 5470 5480 5490

TAAACCAAACAGATAAAGCCACCTTTGCCTAGTGTAGGAAACTGACAGAGGACAGATGGAACAAGCCCCAGAAGACCAAGGGCCACAGAGGGAGCCA < 5600
K T K T D K A T F A * C * E T D R G Q M E Q A P E D Q G P Q R E P
ATTTTGGTTTTGTCTATTTTCGGTGGAAACGGATCACAATCCTTTGACTGTCTCCTGTCTACCTTGTTCGGGGTCTTCTGGTTCCCGGTGTCTCCCTCGGT
5510 5520 5530 5540 5550 5560 5570 5580 5590

TACAATGAATGGACACTAGAGCTTTTAGAGGAACTTAAGAGTGAAGCTGTTAGACATTTTCCTAGGATATGGCTCCATAACTTAGGACAACATATCTATG < 5700
Y N E W T L E L L E E L K S E A V R H F P R I W L H N L G Q H I Y E
ATGTTACTTACCTGTGATCTCGAAAATCTCCTTGAATTCTCACTTCGACAATCTGTAAAAGGATCCTATACCGAGGTATTGAATCCTGTTGTATAGATAC
5610 5620 5630 5640 5650 5660 5670 5680 5690

EcoRI

Sali

AAACTTACGGGGATACTTGGGCAGGAGTGGGAAGCCATAATAAGAATTCTGCAACAACCTGCTGTTTATCCATTTTCAGAATTGGGTGTCGACATAGCAGAAT < 5800
T Y G D T W A G V E A I I R I L Q Q L L F I H F R I G C R H S R I
TTTGAATGCCCCCTATGAACCCGTCCTCACCTTCGGTATTATTCTTAAGACGTTGTTGACGACAAATAGGTAAAGTCTTAACCCACAGCTGTATCGTCTTA
5710 5720 5730 5740 5750 5760 5770 5780 5790

AGGCGTTACTCGACAGAGGAGCAAGAAATGGAGCCAGTAGATCCTAGACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAAACCTGCTTGTACCAA < 5900
G V T R Q R R A R N G A S R S * T R A L E A S R K S A * N C L Y Q
TCCGCAATGAGCTGTCTCCTCTCGTTCTTTACCTCGGTTCATCTAGGATCTGATCTCGGGACCTTCGTAGGTCTTCAGTCGGATTTTTCGCAACATGGTT
5810 5820 5830 5840 5850 5860 5870 5880 5890

TTGCTATTGTAAAAAGTGTGGCTTTTCATTTGCCAAGTTTGTTCATGACAAAAGCCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGA < 6000
L L L * K V L L S L P S L F H D K S L R H L L W Q E E A E T A T K S
AACGATAACATTTTTTCAACGAAAGTAACGGTTCAAACAAAGTACTGTTTTTCGGAATCCGTAGAGGATACCGTCCTTCTTCGCCTCTGTGCTGCTTCT
5910 5920 5930 5940 5950 5960 5970 5980 5990

GCTCATCAGAACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAAGTAGTACATGTAATGCAACCTATAATAGTAGCAATAGTAGCATTAGTAGTAG < 6100
S S E Q S D S S S F S I K A V S S T C N A T Y N S S N S S I S S S
CGAGTAGTCTTGTCAGTCTGAGTAGTTTCAAGAGATAGTTTCGTTCATTCATCATGTACATTACGTTGGATATTATCATCGTTATCATCGTAATCATCATC
6010 6020 6030 6040 6050 6060 6070 6080 6090

CAATAATAATAGCAATAGTTGTGTGGTCCATAGTAATCATAGAATATAGGAAAATATTAAGACAAAGAAAATAGACAGGTTAATTGATAGACTAATAGA < 6200
N N N S N S C V V H S N H R I * E N I K T K K N R Q V N * * T N R
GTTATTATTATCGTTATCAACACACCAGGTATCATTAGTATCTTATATCCTTTTATAATTCTGTTTCTTTTTTATCTGTCCAATTAACATCTGATTATCT
6110 6120 6130 6140 6150 6160 6170 6180 6190

AAGAGCAGAAGACAGTGGCAATGAGAGTGAAGGAGAAGTATCAGCACTTGTGGAGATGGGGGTGGAATGGGGCACCATGCTCCTTGGGATATTGATGAT < 6300
K S R R Q W Q * E * R R S I S T C G D G G G N G A P C S L G Y * * S
TTCTCGTCTTCTGTACCCGTTACTCTCACTTCTCCTTTCATAGTCGTGAACACCTTACCCTTACCCCGTGGTACGAGGAACCCTATAACTACTA
6210 6220 6230 6240 6250 6260 6270 6280 6290

CTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCACCCTCTATTTTGTGCATCAGATGCTAAAGCA < 6400
V V L Q K N C G S Q S I M G Y L C G R K Q P P L Y F V H Q M L K H
GACATCAGATGCTTTTTTAACACCCAGTGTGAGATAATACCCCATGGACACACCTTCTTTCGTTGGTGGTGGAGATAAAACAGTAGTCTACGATTTTCGT
6310 6320 6330 6340 6350 6360 6370 6380 6390

TATGATACAGAGGTACATAATGTTTGGGCCACACATGCCGTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTA < 6500
M I Q R Y I M F G P H M P V Y P Q T P T H K K * Y W * M * Q K I L
ATACTATGCTCCATGTATTACAAACCCGGTGTGTACGGACACATGGGTGTCTGGGGTGGGTGTTCTTTCATCATAACCATTTACACTGTCTTTTAAAT
6410 6420 6430 6440 6450 6460 6470 6480 6490

ACATGTGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAAATTAACCCCACTCTGTGT < 6600
T C G K M T W * N R C M R I * S V Y G I K A * S H V * N * P H S V L
TGTACACCTTTTTACTGTACCATCTTGTCTACGTACTCCTATATTAGTCAAATACCCTAGTTTTCGGATTTTCGGTACACATTTTAATTTGGGGTGAGACACA
6510 6520 6530 6540 6550 6560 6570 6580 6590

TAGTTTAAAGTGCACCTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGGGAGAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAAT < 6700
V * S A L I * R M I L I P I V V A G E * * W R K E R * K T A L S I
ATCAAATTTACAGTGACTAAACCTTACTATGATTATGGTTATCATCATCGCCCTCTTACTATTACCTCTTTCCCTCTCTATTTTTTGGAGAGAAAGTTA
6610 6620 6630 6640 6650 6660 6670 6680 6690

ATCAGCACAAGCATAAGAGATAAGGTGCAGAAAGAATATGCATTCTTTTTATAAACTTGATATAGTACCAATAGATAATACCAGCTATAGGTTGATAAGTT < 6800
S A Q A * E I R C R K N M H S F I N L I * Y Q * I I P A I G * * V
TAGTCGTGTTCCGATTTCTCTATTTCCAGTCTTTCTTATACGTAAGAAAATATTTGAACATATATCATGGTTATCTATTTATGGTTCGATATCCAACATTTCAA
6710 6720 6730 6740 6750 6760 6770 6780 6790

GTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAAATCCCATACATTATTGTGCCCGGCTGGTTTTGCGATTCTAAAAATGTAA < 6900
V T P Q S L H R P V Q R Y P L S Q F P Y I I V P R L V L R F * N V I
CATTTGTGGAGTCAGTAATGTGTCCGGACAGGTTTCCATAGGAACTCGGTTAAGGGTATGTAATAACACGGGGCCGACCAAAACGCTAAGATTTTACATT
6810 6820 6830 6840 6850 6860 6870 6880 6890

TAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGGAATCAGGCCAGTAGTATCAACTCAACTGCTGTTAAAT < 7000
I R R S M E Q D H V Q M S A Q Y N V H M E S G Q * Y Q L N C C * M
ATTATTTCTGCAAGTTACCTTGTCCCTGGTACATGTTTACAGTCGTGTCATGTTACATGTGTACCTTAGTCCGGTCATCATAGTTGAGTTGACGACAATTTA
6910 6920 6930 6940 6950 6960 6970 6980 6990

GGCAGTCTAGCAGAAGAGATGTAGTAATTAGATCTGCCAAATTCACAGACAATGCTAAAACCATAATAGTACAGCTGAACACATCTGTAGAAATTAAT < 7100
A V * Q K K M * * L D L P I S Q T M L K P * * Y S * T H L * K L I
CCGTCAGATCGTCTTCTTCTACATCATTAATCTAGACGGTTAAAGTGTCTGTTACGATTTTGGTATPATCATGTGACTTGTGTAGACATCTTTAATTA
7010 7020 7030 7040 7050 7060 7070 7080 7090

GTACAAGACCCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGGGACCAGGGAGAGCATTGTTTACAATAGGAAAAATAGGAAATATGAGACAAGC < 7200
V Q D P T T I Q E K V S V S R G D Q G E H L L Q * E K * E I * D K H
CATGTTCTGGGTTGTTGTTATGTTCTTTTTCATAGGCATAGGTTCTCCCTGGTCCCTCTCGTAAACAATGTTATCCTTTTTTATCCTTTTATACTCTGTTTCG
7110 7120 7130 7140 7150 7160 7170 7180 7190

NheI
BmtI

ACATTGTAACATTAGTAGAGCAAAATGGAATGCCACTTTAAAACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAAACAATAATCTTTAAG < 7300
I V T L V E Q N G M P L * N R * L A N * E N N L E I I K Q * S L S
TGTAACATTGTAATCATCTCGTTTTACCTTACGGTGAAATTTTGTCTATCGATCGTTAATTTCTCTTGTAAACCTTTATTATTTTGTATTAGAAATTC
7210 7220 7230 7240 7250 7260 7270 7280 7290

CAATCCTCAGGAGGGGACCCAGAAATGTAACGCACAGTTTTAATTTGTGGAGGGGAATTTTTCTACTGTAATTC AACACA ACTGTTTAAATAGTACTTGGT < 7400

N P Q E G T Q K L * R T V L I V E G N F S T V I Q H N C L I V L G
GTTAGGAGTCCCTCCCCTGGGTCTTTAACATTGCGTGTCAAATTAACACCTCCCCTTAAAAAGATGACATTAAGTTGTGTTGACAAATTATCATGAACCA
7310 7320 7330 7340 7350 7360 7370 7380 7390

TTAATAGTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTATAAACATGTGGCAGGA < 7500
L I V L G V L K G Q I T L K E V T Q S H S H A E * N N L * T C G R K
AATATATCATGAACCTCATGACTTCCCAGTTTATTGTGACTTCCCTTCACTGTGTTAGTGTGAGGGTACGTCTTATTTTGTAAATATTTGTACACCGTCCCT
7410 7420 7430 7440 7450 7460 7470 7480 7490

BsaBI

AGTAGGAAAAGCAATGTATGCCCTCCCATCAGTGGACAAATTAGATGTTTCATCAAATATTACTGGGCTGCTATTAACAAGAGATGGTGGTAATAACAAC < 7600
* E K Q C M P L P S V D K L D V H Q I L L G C Y * Q E M V V I T T
TCATCCTTTTTCGTTACATACGGGGAGGGTAGTCACCTGTTTAACTACAAGTAGTTTATAATGACCCGACGATAATTGTTCTCTACCACCATTATTGTTG
7510 7520 7530 7540 7550 7560 7570 7580 7590

AATGGGTCCGAGATCTTCAGACCTGGAGGAGCGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAG < 7700
M G P R S S D L E E A I * G T I G E V N Y I N I K * * K L N H * E
TTACCCAGGCTCTAGAAGTCTGGACCTCCTCCGCTATACTCCCTGTTAACCTCTTCACTTAATATAATTTATATTTTCATCATTTTTAACTTGGTAATCCTC
7610 7620 7630 7640 7650 7660 7670 7680 7690

TAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCCTTGGGTTCTTGGGAGCAGCAGGAAGCAC < 7800
* H P P R Q R E E W C R E K K E Q W E * E L C S L G S W E Q Q E A L
ATCGTGGGTGGTTCCGTTTCTCTTCTCACCACGTCTCTCTTTTTTCTCGTCACCCTTATCCTCGAAACAAGGAACCCAAGAACCCTCGTCTCTCTTCTG
7710 7720 7730 7740 7750 7760 7770 7780 7790

TATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGATATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGGCGCAA < 7900
W A Q R Q * R * R Y R P D N Y C L I * C S S R T I C * G L L R R N
ATACCCGCGTGCAGTTACTGCGACTGCCATGTCCGGTCTGTTAATAACAGACTATATCAGTCTGTCGCTTGTGTTAAACGACTCCCGATAACTCCGCGTT
7810 7820 7830 7840 7850 7860 7870 7880 7890

CAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAACAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTT < 8000
S I C C N S Q S G A S N S S R Q E S W L W K D T * R I N S S W G F
GTCGTAGACAACGTTGAGTGTGACACCCCGTAGTTTGTGTCGAGGTCCGTTCTTAGGACCGACACCTTCTATGGATTTCCCTAGTTGTGCGAGGACCCCTAAA
7910 7920 7930 7940 7950 7960 7970 7980 7990

GGGTTGCTCTGAAAACTCATTTGCACCCTGCTGTGCCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATAACATGACCTGGAT < 8100
G V A L E N S F A P L L C L G M L V G V I N L W N R F G I T * P G W
CCCCAACGAGACCTTTTGGAGTAAACGTGGTGACGACACGGAACCTTACGATCAACCTCATTTATTTAGAGACCTTGTCTAAACCTTATTGTACTGGACCTA
8010 8020 8030 8040 8050 8060 8070 8080 8090

GGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAA < 8200
S G T E K L T I T Q A * Y T P * L K N R K T S K K R M N K N Y W N
CCTCACCTGTCTCTTTAATTGTTAATGTGTTTGAATTTATGTGAGGAATTAACCTTCTTAGCGTTTTGGTCTGTTCTTTCTTACTTGTCTTAATAACCTT

8110 8120 8130 8140 8150 8160 8170 8180 8190

TTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAA < 8300
* I N G Q V C G I G L T * Q I G C G I * N Y S * * * * E A W * V *
AATCTATTTACCCGTTCAAACACCTTAACCAAATTGTATTGTTTAAACCGACCCATATATTTTAATAAGTATTACTATCATCTCCGAACCATCCAAATT
8210 8220 8230 8240 8250 8260 8270 8280 8290

GAATAGTTTTTGTCTGACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTTAGACCCACCTCCCAATCCCGAGGGGACCCGACAG < 8400
E * F L L Y F L * * I E L G R D I H H Y R F R P T S Q S R G D P T G
CTTATCAAAAACGACATGAAAGATATCACTTATCTCAATCCGTCCTATAAGTGGTAATAGCAAAGTCTGGGTGGAGGGTTAGGGCTCCCCTGGGCTGTC
8310 8320 8330 8340 8350 8360 8370 8380 8390

BamHI

GCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTTCGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGC < 8500
P K E * K K K V E R E T E T D P F D * * T D P * H L S G T I C G A
CGGGCTTCCCTTATCTTCTTCCACCTCTCTCTGTCTGTCTAGGTAAGCTAATCACTTGCCTAGGAATCGTGAATAGACCCTGCTAGACGCCTCG
8410 8420 8430 8440 8450 8460 8470 8480 8490

CTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATT < 8600
C A S S A T T A * E T Y S * L * R G L W N F W D A G G G K P S N I
GACACGGAGAAGTCGATGGTGGCGAACTCTCTGAATGAGAACTAATGCTCCTAACACCTTGAAGACCTGCGTCCCCACCCTTCGGGAGTTTATAA
8510 8520 8530 8540 8550 8560 8570 8580 8590

HpaI

GGTGAATCTCCTACAGTATTGGAGTCAGGAACATAAGAATAGTGCTGTTAACTTGTCTCAATGCCACAGCCATAGCAGTAGCTGAGGGGACAGATAGGGT < 8700
G G I S Y S I G V R N * R I V L L T C S M P Q P * Q * L R G Q I G L
CCACCTTAGAGGATGTCATAACCTCAGTCCTTGATTTCTTATCAGACAATTGAACGAGTTACGGTGTCCGATCGTCATCGACTCCCCTGTCTATCCCA
8610 8620 8630 8640 8650 8660 8670 8680 8690

mCherry

TATAGAAGTATTACAAGCAGCTTATAGAGCTATTTCGCCACATACC TAGAAGAATAAGACAGGGCTTGAAAGGATTTTGTCTATAAGATGGTGAGCaAGGG < 8800
* K Y Y K Q L I E L F A T Y L E E * D R A W K G F C Y K M V S K G
ATATCTTCATAATGTTTCGTCGAATATCTCGATAAGCGGTGTATGGATCTTCTTATTCTGTCCC GAACCTTTCCTAAAACGATATTCTACCACTCGtTCCC
8710 8720 8730 8740 8750 8760 8770 8780 8790

>BssSI

CGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGC < 8900
E E D N M A I I K E F M R F K V H M E G S V N G H E F E I E G E G
GCTCCTCCTATTGTACCGGTAGTAGTTCCTCAAGTACGCGAAGTTCACGTGTACCTCCCGAGGCACCTTGCCGGTGTCTCAAGCTCTAGCTCCCCTCCC
8810 8820 8830 8840 8850 8860 8870 8880 8890

<AquII

GAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGT < 9000
E G R P Y E G T Q T A K L K V T K G G P L P F A W D I L S P Q F M Y
CTCCC GGCGGGGATGCTCCC GTGGGTCTGGCGGTTGACTTCCACTGGTTCACC GGGGACGGGAAGCGGACCC TGTAGGACAGGGGAGTCAAGTACA
8910 8920 8930 8940 8950 8960 8970 8980 8990

ACGGCTCCAAGGCTACGTGAAGCACCCCGCGACATCCCCGACTACTTGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGA < 9100
G S K A Y V K H P A D I P D Y L K L S F P E G F K W E R V M N F E
TGCCGAGGTTCCGGATGCACTTCGTGGGGCGGCTGTAGGGGCTGATGAACTTCGACAGGAAGGGGCTCCCGAAGTTCACCCTCGCGCACTACTTGAAGCT
9010 9020 9030 9040 9050 9060 9070 9080 9090

GGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGC < 9200
D G G V V T V T Q D S S L Q D G E F I Y K V K L R G T N F P S D G
CCTGCCGCGGCACCACTGGCACTGGGTCTGAGGAGGGACGTCCTGCCGCTCAAGTAGATGTTCCACTTCGACGCGCCGTGGTTGAAGGGGAGGCTGCCG
9110 9120 9130 9140 9150 9160 9170 9180 9190

<BsrBI

CCCCTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGC < 9300
P V M Q K K T M G W E A S S E R M Y P E D G A L K G E I K Q R L K L
GGGCATTACGTCTTCTTCTGGTACCCGACCCTCCGGAGGAGGCTCGCCTACATGGGGCTCCTGCCGCGGGACTTCCCCTCTAGTTCGTCTCCGACTTCG
9210 9220 9230 9240 9250 9260 9270 9280 9290

TGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCCTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGA < 9400
K D G G H Y D A E V K T T Y K A K K P V Q L P G A Y N V N I K L D
ACTTCTGCCGCGCGGTGATGCTGCGACTCCAGTTCGTGGTGGATGTTCCGGTTCCTCGGGCACGTCGACGGGCGCGGATGTTGCAGTTGTAGTTCAACCT
9310 9320 9330 9340 9350 9360 9370 9380 9390

XmaI

SmaI

T2A

CATCACCTCCCACAACGAGGACTACACCATCGTGGAAACAGTACGAACCGCGCGGAGGGCCGCCACTCCACCGCGGCATGGACGAGCTGTACAAGCCCCGG < 9500
I T S H N E D Y T I V E Q Y E R A E G R H S T G G M D E L Y K P R
GTAGTGGAGGGTGTGCTCCTGATGTGGTAGCACCTTGTCTATGCTTGC GCGGCTCCC GCGGTGAGGTGGCCCGGTACCTGCTCGACATGTTCCGGGGCC
9410 9420 9430 9440 9450 9460 9470 9480 9490 9495

SciI

XhoI

PG3 - Luciferase

GAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTCGAGGAGACGCCAAAAACATAAAGAAAGGCCCGCGCCATTCTATC < 9600
E G R G S L L T C G D V E E N P G P R E D A K N I K K G P A P F Y P
CTCCCCTCTCCTCAGAAGATTGTACGCCACTGCACCTCCTCTTAGGGCCGGGAGCTCTCTGCGGTTTTTGTATTTCTTTCCGGGCCGCGGTAAGATAG
9510 9520 9530 9540 9550 9558 9560 9570 9580 9590

CGCTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTTCTGGAACAATTGCTTTTTACAGATGCACATATCGAGGT < 9700
L E D G T A G E Q L H K A M K R Y A L V P G T I A F T D A H I E V
GCGACCTTCTACCTTGGCGACCTCTCGTTGACGTATTCGATACTTCTCTATGCGGGACCAAGGACCTTGTAAACGAAAATGTCTACGTGTATAGCTCCA
9610 9620 9630 9640 9650 9660 9670 9680 9690

BstBI

GGACATCACTTACGCTGAGTACTTCGAAATGTCCGTTCCGTTGGCAGAAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTTCGTATGCAGT < 9800
D I T Y A E Y F E M S V R L A E A M K R Y G L N T N H R I V V C S
CCTGTAGTGAATGCGACTCATGAAGCTTTACAGGCAAGCCAACCGTCTTCGATACTTTGCTATACCCGACTTATGTTTAGTGTCTTAGCAGCATAACGTCA
9710 9720 9730 9740 9750 9760 9770 9780 9790

GAAAAC TCTCTT CAATTC TTTATGCCGGTGT TGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCC GCGAACGACATTTATAATGAACGTGAATTGCTCA < 9900
E N S L Q F F M P V L G A L F I G V A V A P A N D I Y N E R E L L N
CTTTTGAGAGAAGTTAAGAAATACGGCCACAACCCGCGCAATAAATAGCCTCAACGTCAACGCGGGCGCTTGCTGTAAATATTACTTGCACCTAACGAGT
9810 9820 9830 9840 9850 9860 9870 9880 9890

ACAGTATGGGCATTTTCGCAGCCTACCGTGGTGTTCGTTTCCAAAAAGGGGTGCAAAAAATTTTGAACGTGCAAAAAAGCTCCCAATCATCCAAAAAT < 10000
S M G I S Q P T V V F V S K K G L Q K I L N V Q K K L P I I Q K I
TGTCATACCCGTAAGCGTCCGATGGCACCACAAGCAAAGGTTTTTCCCAACGTTTTTTTAAAACCTTGCACGTTTTTTTTCGAGGGTTAGTAGGTTTTTTTA
9910 9920 9930 9940 9950 9960 9970 9980 9990

<ApyPI
<RceI

TATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTG < 10100
I I M D S K T D Y Q G F Q S M Y T F V T S H L P P G F N E Y D F V
ATAATAGTACCTAAGATTTTGCCTAATGGTCCCTAAAGTCAGCTACATGTGCAAGCAGTGTAGAGTAGATGGAGGGCCAAAATTACTTATGCTAAAACAC
10010 10020 10030 10040 10050 10060 10070 10080 10090

CCAGAGTCCCTTCGATAGGGACAAGACAATTCGACTGATCATGAACTCCTCTGGATCTACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCT < 10200
P E S F D R D K T I A L I M N S S G S T G L P K G V A L P H R T A C
GGTCTCAGGAAGCTATCCCTGTTCTGTTAACGTGACTAGTACTTGAGGAGACC TAGATGACCAGACGGATTTCCACAGCGAGACGGAGTATCTTGACGGA
10110 10120 10130 10140 10150 10160 10170 10180 10190

CGGTGAGATTCTCGCATGCCAGAGATCCATTTTTTGGCAATCAAATCATTCGGGATACTGCGATTTTAAAGTGTGTTCCATTCATCACGGTTTTGGAAT < 10300
V R F S H A R D P I F G N Q I I P D T A I L S V V P F H H G F G M
CGCACTCTAAGAGCGTACGGTCTCTAGGATAAAAACCGTTAGTTTTAGTAAGGCCTATGACGCTAAAATTCACAACAAGGTAAGGTAGTGCCAAAACCTTA
10210 10220 10230 10240 10250 10260 10270 10280 10290

GTTTACTACACTCGGATATTTGATATGTGGATTTTCGAGTCGCTTAAATGTATAGATTTGAAGAAGAGCTGTTTCTGAGGAGCCTTCAGGATTACAAGATT < 10400
F T T L G Y L I C G F R V V L M Y R F E E E L F L R S L Q D Y K I
CAAATGATGTGAGCCTATAAACTATAACCTAAAGCTCAGCAGAATTACATATCTAACTTCTTTCGACAAAGACTCCTCGGAAGTCTAATGTTCTAA
10310 10320 10330 10340 10350 10360 10370 10380 10390

CAAAGTGCCTGCTGGTGCCAAACCCTATTTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAAATACGATTTATCTAATTTACACGAAATTCCTTCTGGTG < 10500
Q S A L L V P T L F S F F A K S T L I D K Y D L S N L H E I A S G G
GTTTACGCGACGACCACGGTTGGGATAAGAGGAAGAAGCGGTTTTTCGTGAGACTAACTGTTTATGCTAAAATAGATTTAAATGTGCTTTTAAACGAAGACCAC
10410 10420 10430 10440 10450 10460 10470 10480 10490

CGCTCCCTCTCTAAGGAAGTCCGGGAAGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAAGGATATGGGCTCACTGAGACTACATCAGCTAT < 10600
A P L S K E V G E A V A K R F H L P G I R Q G Y G L T E T T S A I
CGCGAGGGGAGAGATTCCTTCAGCCCCTTCGCCAACGGTTCTCCAAGGTAGACGGTCCATAGTCCGTTCTTATAACCGAGTGACTCTGATGTAGTCGATA
10510 10520 10530 10540 10550 10560 10570 10580 10590

BsiEI

TCTGATTACGCCCGAGGGGGATGATAAACCGGGCGCGGTTCGGTAAAGTTGTTCCATTTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTG < 10700
L I T P E G D D K P G A V G K V V P F F E A K V V D L D T G K T L
AGACTAATGCGGGCTCCCCCTACTATTTGGCCCCGCCAGCCATTTCAACAAGGTAATAAACTTCGCTTCCAACACCTAGACCTATGGCCCTTTTTCGGAC
10610 10620 10630 10640 10650 10660 10670 10680 10690

GGCGTTAATCAAAGAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCGGTTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAGGATG < 10800
G V N Q R G E L C V R G P M I M S G Y V N N P E A T N A L I D K D G
CCGCAATTAGTTTCTCCGCTTGACACACACTCTCCAGGATACTAATACAGGCCAATACATTTGTTAGCCCTTCGCTGGTTGCGGAACTAACTGTTCCCTAC
10710 10720 10730 10740 10750 10760 10770 10780 10790

>SstE37I

GATGGCTACATTTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGT < 10900
W L H S G D I A Y W D E D E H F F I V D R L K S L I K Y K G Y Q V
CTACCGATGTAAGACCTCTGTATCGAATGACCTGCTTCTGCTTGTGAAGAAGTAGCAACTGGCGGACTTCAGAGACTAATTCATGTTTCCGATAGTCCA
10810 10820 10830 10840 10850 10860 10870 10880 10890

>SdeOSI

GGCTCCCGCTGAATTGGAATCCATCTTGTCTCAACACCCCAACATCTTCGACGCAGGTTGTCGAGGTTTCCCGACGATGACGCCGGTGAACCTCCCGCC < 11000
A P A E L E S I L L Q H P N I F D A G V A G L P D D D A G E L P A
CCGAGGGCGACTTAACCTTAGGTAGAACGAGGTTGTGGGGTTGTAGAAGCTGCGTCCACAGCGTCCAGAAGGGCTGCTACTGCGGCCACTTGAAGGGCGG
10910 10920 10930 10940 10950 10960 10970 10980 10990

>BcgI

GCCGTTGTTGTTTTGGAGCACGGAAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTGCGCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAG < 11100
A V V V L E H G K T M T E K E I V D Y V A S Q V T T A K K L R G G V
CGGCAACAACAAACCTCGTGCCTTTCTGCTACTGCCTTTTTCTCTAGCACCTAATGCAGCGGTTCAGTTTCATTTGTTGGCGCTTTTTCAACGCGCTCTCTC
11010 11020 11030 11040 11050 11060 11070 11080 11090

>EciI

TTGTGTTTTGTGGACGAAGTACCGAAAAGTCTTACCGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAAGATCGC < 11200
V F V D E L P P K G L T G K L D A R K I R E I L I K A K K G G K I A
AACACAAACACTGCTTTCATGGCTTTCCAGAATGGCCTTTGAGCTGCGTTCTTTTTAGTCTCTTAGGAGTATTTCCGGTTCTTCCCCTTTCTAGCG
11110 11120 11130 11140 11150 11160 11170 11180 11190

AclI

CGTGTAAGGAGCCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCAT < 11300
V * G A P L P P P P * R Y W P K P L G I R P V C V C L Y V I F H H
GCACATTCCTCGGGGAGAGGGGAGGGGGGGGATTGCAATGACCGCTTCGGCGAACCTTATTCGGCCACACGCAAACAGATATAACAATAAAAGGTGGTA
11210 11220 11230 11240 11250 11260 11270 11280 11290
11211

ATTGCCGCTTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTTGACGAGCATTCCTAGGGGTCTTTCCCTCTCGCCAAAGGAATGCAAGGT < 11400
I A V F W Q C E G P E T W P C L L D E H S * G S F P S R Q R N A R S
TAACGGCAGAAAACCGTTACTCTCCCGGCTTTGGACCGGGACAGAAGAAGTCTCGTAAGGATCCCCAGAAAGGGGAGAGCGGTTTCTTACGTTCCA
11310 11320 11330 11340 11350 11360 11370 11380 11390

CTGTTGAATGTCGTGAAGGAAGCAGTTCCCTCTGGAAGCTTCTTGAAGACAAACAACGTTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCACCTGGCG < 11500

V E C R E G S S S S G S F L K T N N V C S D P L Q A A E P P T W R
G A C A A C T T A C A G C A C T T C C T T C G T C A A G G A G A C C T T C G A A G A A C T T C T G T T T G T T G C A G A C A T C G C T G G G A A A C G T C C G T C G C C T T G G G G G G T G G A C C G C
11410 11420 11430 11440 11450 11460 11470 11480 11490

BglI

ACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCCAAGTGCACAGTTGTGAGTTGGATAGTTGTGGAAAGAGT < 11600
Q V P L R P K A T C I R Y T C K G G T T P V P R C E L D S C G K S
TGTCCACGGAGACGCCGGTTTTTCGGTGCACATATTCTATGTGGACGTTTTCCGCCGTGTTGGGGTACACCTCAACCTATCAACACCTTTCTCA
11510 11520 11530 11540 11550 11560 11570 11580 11590

CAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAGGTACCCCAATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTT < 11700
Q M A L L K R I Q Q G A E G C P E G T P L Y G I * S G A S V H M L Y
GTTTACCAGAGGAGTTTCGCATAAGTTGTTCCCGACTTCTACGGGTCTTCCATGGGGTAACATACCCTAGACTAGACCCCGAGCCACGTGTACGAAA
11610 11620 11630 11640 11650 11660 11670 11680 11690

ACATGTGTTTGTAGTCGAGGTTAAAAACGTCTAGGCCCCCGAACCCACGGGGACGTGGTTTTCTTTGAAAAACACGATGATAATGGCCACAACCAATGGGT < 11800
M C L V E V K K R L G P P N H G D V V F L * K T R * * W P Q P W V
TGTACACAAATCAGCTCCAATTTTTTGCAGATCCGGGGGGCTTGGTGCCCTGCACCAAAGGAACTTTTTGTGCTACTATTACCGGTGTTGGTACCCA
11710 11720 11730 11740 11750 11760 11770 11780 11790

sequencing continues
* IRES
NEF
* from pcherry picker 2
11795

BlpI

<BsmBI

GGCAAGTGGTCAAAAAGTAGTGTGATTGGATGGCCTGCTGTAAGGGAAAGAATGAGACGAGCTGAGCCAGCAGCAGATGGGGTGGGAGCAGTATGTCGAG < 11900
A S G Q K V V * L D G L L * G K E * D E L S Q Q Q M G W E Q Y V E
CCGTTACCAGTTTTTTCATCACACTAACCTACCGGACGACATTTCCCTTTCTACTCTGCTCGACTCGGTGCTGCTTACCCACCCTCGTCATACAGCTC
11810 11820 11830 11840 11850 11860 11870 11880 11890

ACCTAGAAAACATGGAGCAATCACAAGTAGCAATACAGCAGCTAACAAATGCTGCTTGTGCCTGGCTAGAAGCACAAGAGGAGGAAGAGGTGGGTTTTCC < 12000
T * K N M E Q S Q V A I Q Q L T M L L V P G * K H K R R K R W V F Q
TGGATCTTTTTGTACCTCGTTAGTGTTCATCGTTATGTGCTCGATTGTTACGACGAACACGGACCAGATCTTCGTTCTCCTCCTTCTCCACCCAAAAGG
11910 11920 11930 11940 11950 11960 11970 11980 11990

AGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCAC < 12100
S H L R Y L * D Q * L T R Q L * I L A T F * K K R G D W K G * F T
TCAGTGTGGAGTCCATGGAATTCGGTTACTGAATGTTCCGTCGACATCTAGAATCGGTGAAAAATTTCTTTTCCCCCTGACCTTCCCCGATTAAGTG
12010 12020 12030 12040 12050 12060 12070 12080 12090

TCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTGAGATATC < 12200
P K E D K I S L I C G S T T H K A T S L I G R T T H Q G Q G S D I
AGGGTTTTCTTCTGTTCTATAGGAACCTAGACACCTAGATGGTGTGTTCCGATGAAGGGACTAACCCTTGTATGTGTGGTCCCGGTCCCCAGTCTATAG
12110 12120 12130 12140 12150 12160 12170 12180 12190

CACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACCACAGCTTGTACACCCTGTGAG < 12300
H * P L D G A T S * Y Q L S Q I R * K R P I K E R T P A C Y T L * A

GTGACTGGAAACCTACCACGATGTTTCGATCATGGTCAACTCGGTCATTCCATCTTCTCCGGTTATTTCCCTCTCTTGTGGTCCGAACAATGTGGGACACTC
12210 12220 12230 12240 12250 12260 12270 12280 12290

CCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTAC < 12400
C M E W M T L R E K C * S G G L T A A * H F I T W P E S C I R S T
GGACGTACCTTACCTACTGGGACTCTCTTTCACAATCTCACCTCCAAACTGTCGGCGGATCGTAAAGTAGTGCACCGGGCTCTCGACGTAGGCCTCATG
12310 12320 12330 12340 12350 12360 12370 12380 12390

NEF

TTCAAGAACTGCTGACATCGAGCT < 12424
S R T A D I E X
AAGTTCTTGACGACTGTAGCTCGA
12410 12420