

gagcctccaaggccttgttccattctgacaaatgtcttcgggtccctcggatctctgggtcatgtggcaaattagtctatgaccggtgacccctgaggc
jgtccatggctacgtgcccgtgtgcttacagtgtccaaccaataaaccactatgcaattctctgtggagctatcagtgccactcccctggataag
>cacaggaagaagaaggcaggaacctctgatgttgactttcccttgattttctgacaccccctccccatataggctctgtttgaacattaat
acacagctacttggctcaaggcagagaacaaagtgacgcttccctcctccctctttataagggaccagggttcccagtcacataagaggatcttt
>gattctaagtttctggaacgagatttcttcaaaagcctggaagcagtagggcagccATGGGGCTGCTGGTTGGAGATGACTTATGGGCTGTGGTCA
M G L L V G D D L W A V V
FTCACAGCCATCTTCCTGCTTCTGGTGGACCTGGTGCACCGGGCCAGCGCTGGACTGCCTGCTACCCACCAGGCCCTGTGCCATTCCCTGGGCTGGG
F T A I F L L L V D L V H R R Q R W T A C Y P P G P V P F P G L
BspMI BspMI NarI
ACCTGCTGCAGGTGGACTTCGAGAACATACCATACAGCTTTTACAAGgtgagagaggtggtggtggaggacagaagaagt cagagtgaccaggcagc
N L L Q V D F E N I P Y S F Y K

1

gtggaggtgtgaagccactggaagacaggaagatggcccccaagagaaaaaagagtgtgtgggaaaacaaagtggagtgtgtg XoX(~1kb?)
>acacacagagagagagagagatagagatagagagagagagagagagagaagcagagacttgtatgaagagtattgatcttccactctgctggaaa
>tacttctatgtgctctgtcctatttcagCTTCAAAACCGCTATGGCAATGTGTTTACAGCTACAGATGGCCTGGAAGCCCGTGGTTGTGGTCAACGG
L Q N R Y G N V F S L Q M A W K P V V V V N

2

FGAAGGCAGTGCAGGAACTGCTGGTGCACCTATGGAGAGGACACTTCTGACCGCCCCCTAATGCCATCTATAATCACATAGGCTATGGGCACAAATCA
L K A V R E L L V T Y G E D T S D R P L M P I Y N H I G Y G H K S
AGgcaaagctctatagtggtggggcaaggcagggagagtggggttgagaggggtgagcagaagatagacagacctaatggagatgatgtggtccg XoX
<
BspMI
>tcagttacacacagatatgtctgcagGTGTGATCCTTGACCTTACGGGCCCCGAGTGGCGAGAGCAGCGGCGATTTTCTGTGTCTACCCTGCGTGAC
G V I L A P Y G P E W R E Q R R F S V S T L R D

3

GGCCTGGGCAAGAAATCACTGGAGCAGTGGGTGACAGAGGAGGCTGGCCACCTCTGCGATGCTTTCACCAAGGAGGCCGgtgggt XoX
F G L G K K S L E Q W V T E E A G H L C D A F T K E A
atgttagagatggagagtgtgagcagggactgaagagcactatcatgacttgcttgcattccaatcctagAACATCCCTTCAATCCCAGCCCCCTCC
SapI E H P F N P S P L

4

AGTAAAGCTGTGAGCAACGTGATCGCCTCCCTCATTTATGCTCGTCGATTTGAGTATGAAGACCCTTTCTTCAACAGGATGCTCAAAACATTGAAGGA
S K A V S N V I A S L I Y A R R F E Y E D P F F N R M L K T L K
rdIII
CTTAGGAGAAGACTGGCTTCGTCGGAGAGgtatgtggctcggggcagtgggcctggccaacttctatgctataaattctatcagagtaagatgaa
S L G E D T G F V G E *MscI*
>aggggtgggctggtgtgatggccaagtgcagggagaaaaggcttgggtgttgacaaacctcagataagtggagggaggagcagaccaaggttaatggc
MscI
>agaaagggggtgagcctgcagggaaaccacacaagcctgttatatttgaaggtagagtatttctttcttccacctgagctacaggagagcatgca
baI PstI SphI
>tagaactggggaagaagcaggtgatgggtagggacagattctgagtcatttgcgaagaggctgagaactgtgactggtccaggtaaactgagca
BspMI
aacagccccgggggcaaaagagctagatggtcagtggttaaatgagcttaccagcaagcctggagccctgagctggattcctgagctccatataag
SmaI-XmaI BpmI
ataaaacttactctataagttgttctctaatagacaaaacatgagtcactatatacaggcaagcactgtgtacatgcatgtgcaacacactggca
BsrGI NsiI

5

aaaattatgaattcttaattagaaaagcaactggacctgtttgtttctttcacagGTGCTGAATGCCATCCCAATGCTACTACACATCCCTGGTTT
EcoRI V L N A I P M L L H I P G
EcoRI BamHI
TGATAAAGCCTTCCCAAGCTGAATTCATTCATAGCCTTAGTGAATAAGATGTTAATTGAGCACGACTCGACCTGGGATCCTGCACAGCCACCCCGA
P D K A F P K L N S F I A L V N K M L I E H D S T W D P A Q P P R
FTTACTGATGCCTTCTGGCTGAGGTGGAGAAGgtgaatggcagctccaagaagcttggccaggggctgtgcaactgaggtcctgtggctgtgatgca
) L T D A F L A E V E K *MscI*

atgggttagagtctctgcatcacciaagtgactcagccctgggtttcggggaaacatgggaggccactgtgggatgactcctcctgctcaaccac

-ttctgtctggccagGCCAAGGGGAATCCTGAGAGCAGCTTCAATGATAAGAACCTGCGCATAGTGGTGATTGACCTGTTTCATGGCAGGGATGGTGA
A K G N P E S S F N D K N L R I V V I D L F M A G M V
ApaI BspHI FspI
ACCTCAACCACACTGTCTGGGCCCTGCTGCTCATGATCCTGCATCCAGATGTGCAGCgtgagcctgggtggggctgggagggagggaagaggcattt
T S T T L S W A L L L M I L H P D V Q

6

jaaactagt XoX aagctggagctccaccggtggcgccgctctagtggctccgcattagcttgacctaggctcccacagtctgatggg
SpeI SacII 7-F 7B-F
-cttatccatcgtattaaagccagaggttcaagttcaagggtagcagctttttctcagaagagatcctatctatccagGCCGTGTCCACCAGGAAAT
R R V H Q E

7

ACGAGGTCATAGGGCACGTGCGGCATCCAGAGATGGCAGACCAGGCCGCATGCCCTACACAAACGCGGTTCATTCATGAGGTGCAACGCTTTGCAGAC
D E V I G H V R H P E M A D Q A R M P Y T N A V I H E V Q R F A D

IGTCCCAACAAATTTACCACATATGACATCCCAGACATTAATTTCAAGACTTCTTCATCCCCAAGgtaggcatggggaccttgggtccgagaatcct
I V P T N L P H M T S R D I K F Q D F F I P K
:ggggcaaacatgggtgataaacccaagaccaccacacactggatttccacatggggacaggatggagctaggtgtgagtacactgcagccgttgct
7B-R 7-R PstI

agagtccttggggaaaaggacggccagtgaaatgagcttaaggggttgtaggagggcactgcagatgggaactttgtcacagggccttggatgggtacaa
8SEQ8Z-F 8SEQ8X-F 8SEQ8Y-F
jaaggtgctcagaggtccacgggtgagagtgaatgcaaggatcccgtgtgtgaggttagaggtgtgctgtgaaacttgggtgaccacgacctaggcag
BamHI

:cacctggcccactcttnacaccatgtgacctgcccagGGGACGACACTCATTCCCAACCTGTCTCCGTGCTGAAGGATGAGACTGTCTGGGAGAA
BstNI BstNI G T T L I P N L S S V L K D E T V W E
BspMI

8

ICCTTCGCTTCTATCCTGAACACTTCTGGATGCCAGGGCCACTTTGTGAAGCACGAGGCCTTCATGCCATTCTCAGCAGgtgactgcaggctactt
P L R F Y P E H F L D A Q G H F V K H E A F M P F S A

:tgctgtcttttctcaggggtgccttggaggggtgaagcccaatgaggtcctagactaactgaaccctctcccaccacagGCCGCAGATCATGCCTG
BspMI G R R S C L
3GAGCCCCTGGCCCGCATGGAGCTTCTCCTTCTTCCCTGCCTCCTGCAGCGCTTTAGCTTCTCAGTGCCCGATGGACAGCCAGCCAGTGGATT
3 E P L A R M E L F L F F T C L L Q R F S F S V P D G Q P R P S D

9

3GCATCTATAAATGCCAGTTACTCCAGAGCCCTATCAGCTCTGTGCAGTGGCTCGATAGCAGGGGCAATGGTCTACCTTCCCCAAGTTTGGCCTG
G I Y T M P V T P E P Y Q L C A V A R *

9SEQ9B-R

9SEQ9A-R

FGGTGTTCAATAAA CAAGTTCTGTGTCTGCAACTCANATTTGGTCCATCACCCACTCTTATCTCAATTGTGTCCCAAGACGACAGTCTATAAACCTT

9SEQ9C-R

FGCCCTCTGTCTCAGCCCAACCCATGAACAACCTTTGATGCTGNGAGAATTAANNTTATTACTCCAGGAGCAGTGGCCTCCAGCAAAGCACATGGGAGG
3AAACAAGACGATGGCAAGTCCAAGNCCTGNATGGGTTATATAAAGTGGACTACTTC