



NotI (-1)
1 GCGGCCGCAATAAAAATATCTTTATTTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACA
PvuI (172)
101 AAACAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAAGGATCTGCGATCGCTCCGGTCCCGTCACTGGGCA
SgfI (171)
MfeI (247) AgeI (254)
201 GAGCGCACATCGCCACAGTCCCGGAGAAGTTGGGGGAGGGTCCGCAATTGAACCGGTGCTAGAGAAGGTGGCGGGGTAACTGGGAAAGTGATG
Psp1406I (368)
301 TCGTGACTGGCTCCGCTTTTTCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTCCGCCAG
HindIII (410)
PvuII (404) Bsu36I (456)
401 AACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTACGCGCCCGCCCTACCTGAGGCCGCATCCACGCCGTTGAGTCGCGTTCTGCCGCT
501 CCGCCTGTGGTCCCTCTGAACTCGCTCCGCCGTCTAGGTAAGTTAAAGCTCAGGTCGAGACCGGCCCTTTGTCCGGGCTCCCTGGAGCTACCTA
NaeI (606)
601 GACTCAGCCGGCTCTCCACGCTTTCCTGACCCTGCTTCTCAACTCTACGTCTTTGTTTCTGTTTCTGCGCCGTTACAGATCCAAGCTGTGACC
KasI (700) AgeI (717) SphI (733)
701 GCGCCCTACCTGAGATCACCGGTAGGAGGGCCAGCATGCCACATACTTTGGTGGTGGTGGTCTGGGGTCCATCATCAGCCTTCCAAGGAAGAATC
1 M P H T L W M V W V L G V I I S L S K E E S
Bsu36I (874)
801 CTCCAATCAGGCTTCTGTCTTGTGACCGCAATGGTATCTGCAAGGGCAGCTCAGGATCTTAACTCCCTCAGGGCTCACAGAAGCTGTAAAA
22 S N Q A S L S C D R N G I C K G S S G S L N S I P S G L T E A V K
901 AGCCTTGACCTGTCCAACAACAGGATCACCTACATTAGCAACAGTGACCTACAGAGGTGTGTGAACCTCCAGGCTCTGGTGTGACATCCAATGGAATTA
56 S L D L S N N R I T Y I S N S D L Q R C V N L Q A L V L T S N G I
1001 ACACAATAGAGGAAGATTCTTTTCTCCCTGGGCAGTCTTGAACATTTAGACTTATCTATAATTACTTATCTAATTTATCGTCTTCCCTGGTTCAAGCC
89 N T I E E D S F S S L G S L E H L D L S Y N Y L S N L S S S W F K P
AvrII (1148)
1101 CTTTCTCTTTAACTTCTTAACTTACTGGAAATCCTTACAAAACCTAGGGGAAACATCTCTTTTTCTCATCTCACAAAATTGCAAACTCTGAGA
122 L S S L T F L N L L G N P Y K T L G E T S L F S H L T K L Q I L R
BglIII (1286)
1201 GTGGAAATATGGACACCTTCACTAAGATTCAAAGAAAAGATTTGCTGGACTTACCTTCTTGGAACTTGGAGATTGATGCTTCAGATCTACAGAGCT
156 V G N M D T F T K I Q R K D F A G L T F L E E L E I D A S D L Q S
1301 ATGAGCCAAAAGTTTGAAGTCAATTCAGAATGTAAGTCATCTGATCCTTCATATGAAGCAGCATATTTTACTGCTGGAGATTTTGTAGATGTTACAAG
189 Y E P K S L K S I Q N V S H L I L H M K Q H I L L L E I F V D V T S
1401 TTCCTGGAACTGTTTGAAGTCCGAGATACTGATTTGGACACTTTCATTTTTCAGAATCTCCACTGGTGAACAAAATTCATGATTAATAAGTTTACA
222 S V E C L E L R D T D L D T F H F S E L S T G E T N S L I K K F T
1501 TTTAGAAATGTAAGAAATCACCGATGAAAGTTGTTTTCAGGTTATGAACTTTTGAATCAGATTTCTGGATTGTTAGAATTAGAGTTTGTACTGTACCC
256 F R N V K I T D E S L F Q V M K L L N Q I S G L L E L E F D D C T
HpaI (1667)
Psp1406I (1664)
1601 TTAATGGAGTTGGTAATTTTAGAGCATCTGATAATGACAGAGTTATAGATCCAGGTAAAGTGGAAACGTTAAACAATCCGGAGGCTGCATATTTCAAGGTT
289 L N G V G N F R A S D N D R V I D P G K V E T L T I R R L H I P R F
1701 TTACTTATTTTATGATCTGAGCCTTTATATTTCACTTACAGAAAAGATTAAGAATCACAGTAGAAAAACAGTAAAGTTTTTCTGGTTCCTTGTACTT
322 Y L F Y D L S T L Y S L T E R V K R I T V E N S K V F L V P C L L
1801 TCACAACATTTAAATCATTAGAATACTGGATCTCAGTAAAAATTTGATGGTTGAAGAATACTTGAATAATTCAGCCTGTGAGGATGCCTGGCCCTCTC
356 S Q H L K S L E Y L D L S E N L M V E E Y L K N S A C E D A W P S
EcoRV (1989)
1901 TACAACATTTAATTTTAAAGCAAATCATTGGCATCATTGAAAAAACCGGAGAGACTTTGCTCACTCTGAAAAACTTGACTAACATTGATCAGTAA
389 L Q T L I L R Q N H L A S L E K T G E T L L T L K N L T N I D I S K
SspI (2050)
2001 GAATAGTTTTTCTTCTATGCTGAAACTTGTGAGTGGCCAGAAAAGATGAAATATTTGAACTTATCCAGCACAGAAATACAGAGTGAACAGGCTGCATT
422 N S F H S M P E T C Q W P E K M K Y L N L S S T R I H S V T G C I
BstXI (2101)
2101 CCCAAGACACTGGAAATTTTATGATGTTAGCAACAACAATCTCAATTTATTTTCTTGAATTTGCCGCAACTCAAAGAATTTATATTTCCAGAAAATAAGT
456 P K T L E I L D V S N N N L N L F S L N L P Q L K E L Y I S R N K
SpeI (2237)
2201 TGATGACTTACCAGATGCCTCCCTTACCATGTTACTAGTATTGAAAATCAGTAGGAATGCAATAACTACGTTTCTAAGGAGCAACTTGACTCATT
489 L M T L P D A S L L P M L L V L K I S R N A I T T F S K E Q L D S F
2301 TCACACACTGAAGACTTTGGAAGCTGGTGGCAATAACTTCAATTTGCTCCTGTGAATTCCTCTCCTTCACTCAGGAGCAGAAAGCTGGCCAAAGTCTTG
522 H T L K T L E A G G N N F I C S C E F L S F T Q E Q Q A L A K V L
BstXI (2438)
2401 ATTGATTGGCCAGCAAATACCTGTGTGACTCTCCATCCATGTGCGTGGCCAGCAGGTTGAGGATGTCGCGCTCTCGGTGTGCGAATGTACAGGACAG
556 I D W P A N Y L C D S P S H V R G Q V Q D V R L S V S E C H R T
EcoO109I (2551) NcoI (2570)
2501 CACTGGTGTCTGGCATGTGCTGTGCTCTGTTCTGCTGATCCTGCTCACGGGGTCTGTGCCACCGTTTCCATGGCCTGTGGTATATGAAAATGATGTG
589 A L V S G M C C A L F L L I L L T G V L C H R F H G L W Y M K M M W
NsiI (2659)
2601 GGCTGGCTCCAGGCAAAAAGGAAAGCCAGGAAAGCTCCAGCAGGAACATCTGCATGATGATGATTTGTTTCTTACAGTGAGCAGGATGCCTACTGGGTG
622 A W L Q A K R K P R K A P S R N I C Y D A F V S Y S E Q D A Y W V
2701 GAGAACCTTATGGTCCAGGAGCTGGAGAACTCAATCCCCCTTCAAGTTGTGTCTTATAAGCGGGACTTCACTTCCGCAAGTGGATCATTGACAATA
656 E N L M V Q E L E N F N P P F K L C L H K R D F I P G K W I I D N
2801 TCATTGACTCCATTGAAAAGGCCAAAATGCTTGTGCTTCTGAAAATTTGTGAAGAGTGAAGTGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT
689 I I D S I E K S H K A T V F V L S S E N F V K S E W C K Y E L D F S H F
Eco47III (2974)
2901 CCGTCTTTTATGAGAACAATGATGCTGCCATTCTCATTCTTCTGGAGCCATTGAGAAAAAGCCATTTCCCAAGGCTTCTGCAAGCTGCGGAAGATA
722 R L F D E N N D A A I L L L E P I E K K A I P Q R F C K L R K I
NcoI (3028) AvrII (3085)
3001 ATGAACCAAGACCTACCTGGAGTGGCCATGGACAGGCTCAGCGGGAAGGATTTGGGTAAATCTGAGAGCTGCATAAAGTCTTAGGTTCCATAT
756 M N T K T Y L E W P M D E A Q R E G F W V N L R A A I K S •
NheI (3131)
3101 TTAAGACCAGTCTTGTCTAGTTGGATCTTGTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAACCAACTAGAATGCAGTGAAAA
HpaI (3269) MfeI (3280)
3201 AAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCAATATAAGCTGCAATAAACAAGTTAAACAACAACAATTCATTTCATTATGT
3301 TTCAGTTCCAGGGGAGGTGTGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATCTAAAATACAGCATAGCAAACTTTAACTC

3401 CCAAATCAAGCCTCTACTTGAATCCTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTT
3501 CTTTCATGGAGTTTAAGATATAGTGATTTTCCCAAGGTTTGAAGTACTCTTTCATTTCTTTATGTTTTAAATGCACTGACCTCCACATTCCTTTTTTA
3601 GTAAAAATATTCAGAAATAAATTTAAATACATCATTTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCC
3701 AGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCGAATTCTCGACTCATTCTTTGCCCTC
3801 GGACGAGTGTGGGGCGTCGGTTTCCACTATCGGCGAGTACTCTACACAGCCATCGGTCCAGACGGCCGCGCTTCTGCGGGCGATTTGTGTACGCCGA
3364P R T S P R R N G S D A L V E V C G D T W V A A S R R A I Q T R G V
3901 CAGTCCCGGTCGGGATCGGACGATTGCGTCGCATCGACCTGCGCCAAAGTCGCATCATCGAAATGCGCTCAACCAAGCTCTGATAGAGTTGGTCAAG
3034 T G A G S R V I A D C R G Q A W A A D D F N G D V L S Q Y L Q D L
4001 ACCAATCGGAGCATATACGCCGGAGCCGCGGATCTGCAAGCTCCGGATGCTCCGCTCGAAGTAGCGCGTCTGCTGCTCCATACAAGCCAACAC
2704 G I R L M Y A R L R P S G A L E P H R R E F Y R T Q Q E M C A L W
4101 GGCTCCAGAAGAAGATGTTGGGACCTCGTATTGGGAATCCCGAACATCGCTCGCTCCAGTCAATGACCGCTGTTATGCGGCCATTGTCGCTCAGGA
2364P R W F F I N A V E Y Q S D G F M A E S W D I V A T I R G N D T L V
4201 CATTGTTGGAGCCGAAATCCGCGTGCACGAGTCCGGACTTCCGGGCGAGTCTCGGCCAAAGCATCAGCTCATCGAGAGCTCGCGGACGGACGCAC
2034 N N S G F D A H V L H R V E P C D E A W L M L E D L A Q A V S A S
4301 GACGGTGTCTCCATCACAGTTTCCAGTGATACACATGGGGATCAGCAATCGCCATATGAAATCACGCCATGTAGTGATTGACCGATTCTTTGCGGT
1704 V T D D M V T Q W H Y V H P D A I A C I F D R W T T Y Q G I G Q P
4401 CCGAATGGGCGAACCCGCTCGTCTGGCTAAGATCGGCCGAGCGATCGATCCATGAGCTCCGCGAGGGTTCGAGAACAGCGGGCAGTTCGGTTTCAG
1364 G F P G F G S T Q S L D A A A I A D M L E A V P Q L V A P L E T E P
4501 GCAGGTCTTGAACGTGACACCCTGTGCACGGCGGAGATGCAATAGTTCAGGCTCTCGCTGAATTCGCCAATGTCAAGCACTTCGGGAATCGGGAGCGC
1034 L D Q L T V G Q A R R S I C Y T L S E S F E G I D L V E P I P L A
4601 GGCGATGCAAGTGGCGATAAACATAACGATCTTTGTAGAAACCATCGGCGAGCTATTTACCGCAGGACATCCACGCCCTCCTACATCGAAGCTG
704 A S A F H R Y V Y R D K Y F G D A C S N V R L V Y G R G G V D F S
4701 AAAGCACGAGATCTTCCGCTCCGAGAGTGCATCAGGTCGGAGAGCTGTGCAACTTTTCGATCAGAAATCTTCGACAGACGTCGGCGGTGAGTTCAG
364F A R S E E G E S L Q M L D S V S D F K E I L F K E V S T A T L E P
4801 GCTTTTTTATGATGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCTCTCCAGC
34 K K M
4901 TTATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCACCCTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTT
5001 TGGAAAGTCCCCTTGATTACTAGTCAAAAACAACTCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCA
5101 TTGATGTAAGTCCAAAACCGCATCATCATGTAATAGCGATGACTAATACGTAGATGTAAGTCCAAAGTAGGAAAGTCCATAAGGTCATGTAAGTGGGCAT
5201 AATGCCAGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGGTAAGTGGCATATGATACACTTGTGTAAGTCCAAAGTGGGAGTTTACCGTAAATA
5301 CTCCACCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCTGTTGGGCGGTACGCCA
5401 GCGGGCCATTTACCGTAAGTTATGTAACGCCTGACGTTAATTAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCTT
5501 GCTGGCGTTTTTCCATAGGCTCGCCCCCTGACGAGCATCAAAAATCGAGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATACCAG
5601 GCGTTTTCCCTGGAAGTCCCTCGTGCCTCTCTGTTCCGACCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTT
5701 CTCATAGCTCAGCTGTAGGTATCTCAGTTCCGGTGTAGGTCGTTCCGCTCAAGTGGGCTGTGTGCACGAACCCCGTTACGCCGACCGTGCCTT
5801 ATCCGGTAACTATCGTCTTGTGCAACCCGGTAAGACAGGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC
5901 GGTGCTACAGAGTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAA
6001 GAGTTGTAGCTCTTGTATCGGCAAAACAAACACCGCTGGTAGCGGTGTTTTTTTTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGA
6101 AGATCCTTTGATCTTTTCTACGGGCTGACGCTCAGTGAACGAAAACCTACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCA