



PvuI (7)
SgfI (6) 1 GGATCTGGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCACAGTCCCGGAGAAGTTGGGGGAGGGGTGGCAATTGAACGGGTGCCTA
MfeI (82)
101 GAGAAAGTGGCGCGGGTAAACTGGAAAGTGTGTCTGTACTGGCTCCGCCTTTTCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC

Psp1406I (203) 201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
HindIII (245)
301 GCCATCCACGCGGGTTGAGTCGCGTTTCTGCCGCCTCCCGCCTGTGGTGCCTCCTGAAGTGCCTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC

401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCCTGCTTCTCAACTCTACGCTTTGTTTCGTTT

SphI (560)
AgeI (552) 501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCACCGGTGAGTGCACCTCCGGGACGGCCGGGACGCGCTCTGGC
EagI (578) 1 M R P S G T A G A A L L A
Eco47III (588)
601 GCTGCTGGCTGCGCTCTGCCCGCGAGTCCGGGCTCTGGAGGAAAAGAAAGTTTGC AAGGCACGAGTAAACAGCTCACGAGTTGGGCACCTTTTGAAGAT
13 L L A A L C P A S R A L E E K K V C Q G T S N K L T Q L G T F E D
701 CATTTCTCAGCTCCAGAGGATGTTCAATAACTGTGAGGTGGTCTTGGAAATTTGAAATACCTATGTGCAGAGGAATTATGACTTTTCTTCTTAA
47 H F L S L Q R M F N N C E V V L G N L E I T Y V Q R N Y D L S F L
801 AGACCATCCAGGAGGTGGTGGTATGCTCCTATTGCCCTCAACACAGTGGAGCGAATTCCTTTGAAAACTGCAGATCATCAGAGGAAATATGACTA
80 K T I Q E V A G Y V L I A L N T V E R I P L E N L Q I I R G N M Y Y
901 CGAAAATTCCTATGCCTTAGCAGTCTTATCTAACTATGATGCAATAAAAACGGACTGAAGGAGCTGCCCATGAGAAATTTACAGGAAATCCTGCATGGC
113 E N S Y A L A V L S N Y D A N K T G L K E L P M R N L Q E I L H G

1001 GCCGTGCGGTTTCAGCAACAACCTGCCCTGTGCAACGTGGAGAGCATCCAGTGGCGGGACATAGTCAGCAGTGACTTTCTCAGCAACATGTCGATGGACT
147 A V R F S N N P A L C N V E S I Q W R D I V S S D F L S N M S M D
1101 TCCAGAACCACCTGGGACGCTGCCAAAAGTGTGATCCAAGCTGTCCCAATGGGAGCTGCTGGGGTGCAGGAGAGGAGAAGTCCAGAAACTGACCAAAAT
180 F Q N H L G S C Q K C D P S C P N G S C W G A G E E N C Q K L T K I

XmaI (1289)
SmaI (1289)
1201 CATCTGTGCCAGCAGTGTCCGGGCGCTGCCGTGGCAAGTCCCCAGTACTGCTGCCACAACCAGTGTGCTGCAGGCTGCACAGGCCCCCGGGAGAGC
213 I C A Q Q C S G R C R G K S P S D C C H N Q C A A G C T G P R E S

BbrPI (1335)
1301 GACTGCCTGGTCTGCCGCAAATTCGAGACGAAGCCACGTGCAAGGACACCTGCCCCCACTCATGCTCTACAACCCACCACGTACCAGATGGATGTGA
247 D C L V C R K F R D E A T C K D T C P P L M L Y N P T T Y Q M D V
1401 ACCCCGAGGGCAAATACAGCTTTGGTGCCACTCGTGAAGAAGTGTCCCGTAATTATGTGGTGACAGATCACGGCTCGTGCCTCCGAGCCTGTGGGC
280 N P E G K Y S F G A T C C V K K C P R N Y V V T D H G S C V R A C G A
1501 CGACAGCTATGAGATGGAGGAAGACGGCGTCCGCAAGTGAAGAAGTGCGAAGGGCCTTGGCCAAAGTGTGTAACGGAATAGTATTGGTGAATTTAAA
313 D S Y E M E E D G V R K C K K C E G P C R K V C N G I G I G E F K
1601 GACTCACTTCCATAAATGCTACGAATATAAACACTTCAAAAAGTGCACCTCCATCAGTGGCGATCTCCACATCCTGCCGGTGGCATTAGGGGTGACT
347 D S L S I N A T N I K H F K N C T S I S G D L H I L P V A F R G D
1701 CCTTACACATACTCTCCTCTGGATCCACAGGAAGTGGATATTCTGAAAACCGTAAAGGAAATCACAGGGTTTTTGTGATTAGGGTGGCCTGAAA
380 S F T H T P P L D P Q E L D I L K T V K E I T G F L L I Q A W P E N
1801 CAGGACGGACCTCCATGCCTTTGAGAACCTAGAAATCATACGCGGCAGGACCAAGCAACATGGTCAAGTCTTCTTGCAGTCTGCAGCCTGAACATAACA
413 R T D L H A F E N L E I I R G R T K Q H G Q F S L A V V S L N I T
1901 TCCTTGGGATTACGCTCCCTCAAGGAGATAAGTGTGAGAGATGTGATAATTTGAGAAACAAAAATTTGTGCTATGCAAAATACAATAAAGTGAAAAAAC
447 S L G S L K E I S D G D V I I S G N K N L C Y A N T I N W K
2001 TGTTTGGGACCTCCGGTCAGAAAACCAAAATATAAGCAACAGAGGTGAAAACAGCTGCAAGGCCACAGGCCAGGTCTGCCATGCCTTGTCTCCCCGA
480 L F G T S G Q K T K I I S N R G E N S C K A T G Q V C H A L C S P E
2101 GGGCTGCTGGGGCCCGGAGCCAGGGACTGCGTCTTGGCCGAATGTGAGCCGAGGCAAGGAAATGCGTGGACAAGTGAACCTTCTGGAGGGTGGCCA
513 G C W G P E P R D C V S C R N V S R G R E C V D K C N L L E G E P
2201 AGGAGTTTGTGGAAACTCTGAGTGCATACAGTGCACCCAGAGTGCCTGCCTCAGGCCATGAACATCACCTGCACAGGACGGGACAGACAACCTGTA
547 R E F V E N S E C I Q C H P E C L P Q A M N I T C T G R G P D N C

NgoMIV (2398)
2301 TCCAGTGTGCCACTACATTGACGGCCCCACTGCGTCAAGACCTGCCGGCAGGAGTGCATGGGAGAAAAACAACCCCTGGTCTGGAAGTACGCAGAGCC
580 I Q C A H Y I D G P H C V K T C P A G V M G E N N T L V W K Y A D A
2401 CGGCCACTGTGTCACACTGTCACAACTGCAACTACGGATGCAGTGGCCAGGCTTTGAAGGCTGTCCAACGAATGGGCTAAGATCCCATCCATC
613 G H V C H L C H P N C T Y G C T G P G L E G C P T N G P K I P S I
2501 GCCACTGGGATGGTGGGGCCCTCTCTTGTGCTGGTGGTGGCCCTGGGATCGGCCTTTCATGCGAAGGCCACATCGTTCGGAAGCGCACGCTGC
647 A T G M V G A L L L L V V A L G I G L F M R R R H I V R K R T L

DraIII (2639)
2601 GGAGGCTGCTGCAGGAGAGGGAGCTTGTGGAGCCTCTTACACVAGTGAGAACTCCCAACCAAGCTCTTGTGAGGATCTTGAAGGAAACTGAATTCAA
680 R R L L Q E R E L V E P L T P S G E A P N Q A L L R I L K E T E F K
2701 AAAGATCAAAGTGTGGGCTCCGGTGCCTCGGCACGGTGTATAAGGGACTCTGGATCCAGAAGGTGAGAAAGTAAAATCCCGTGCATCAAGGAA
713 K I K V L G S G A F G T V Y K G L W I P E G E K V K I P V A I K E

BbrPI (2878)
2801 TTAAGAGAAGCAACATCTCCGAAAGCAACAAGGAAATCCTCGATGAAGCCTACGTGATGGCCAGCGTGGACAACCCCACTGTGCCGCTGCTGGGCA
747 L R E A T S P K A N K E I L D E A Y V M A S V D N P H V C R L L G
2901 TCTGCCTCACCTCCACCTGTCAGCTCATCAGCAGCTCATGCCCTTCCGCTGCCTCCTGGACTATGTCCGGGAACACAAGACAATATTGGCTCCAGTA
780 I C L T S T V Q L I T Q L M P F G C L L D Y V R E H K D N I G S Q Y

ApaLI (3061)
3001 CCTGCTCAACTGGTGTGTGCAGATCGCAAAGGGCATGAACACTTGGAGGACCGTCGTTGGTGCACCGCAGCTGGCAGCCAGGAACGTACTGGTGAAA
813▶ L L N W C V Q I A K G M N Y L E D R R L V H R D L A A R N V L V K
3101 ACACCGCAGCATGTCAAGATCACAGATTTGGGCTGGCCAAACTGCTGGTGGCGAAGAGAAATACCATGCAGAAGGAGGCAAAGTGCCTATCAAGT
847▶ T P Q H V K I T D F G L A K L L G A E E K E Y H A E G G K V P I K

BstEII (3264)
3201 GGATGGCATTGGAATCAATTTTACACAGAATCTATACCCACCAGAGTGATGTCTGGAGCTACGGGGTACCGTTTGGGAGTTGATGACCTTTGGATCCAA
880▶ W M A L E S I L H R I Y T H Q S D V W S Y G V T V W E L M T F G S K

BglII (3326) ClaI (3382)
3301 GCCATATGACGGAATCCCTGCCAGCGAGATCTCCTCCATCCTGGAGAAGGAGAAGCGCTCCCTCAGCCACCCATATGTACCATCGATGTCTACATGATC
913▶ P Y D G I P A S E I S S I L E K G E R L P Q P P I C T I D V Y M I

Eco47III (3488)
3401 ATGGTCAAGTGCTGGATGATAGACGCAGATAGTCGCCAAAGTTCCTGAGTTGATCATCGAATTCTCCAAAATGGCCCCGAGACCCCCAGCGCTACCTTG
947▶ M V K C W M I D A D S R P K F R E L I I E F S K M A R D P Q R Y L

NsiI (3520)
3501 TCATTGAGGGGGATGAAAGAATGCATTTGCCAAGTCTACAGACTCCAATTCTACCGTGCCTGATGGATGAAGAAGACATGGACGACGTGGTGGATGC
980▶ V I Q G D E R M H L P S P T D S N F Y R A L M D E E D M D D V V D A
3601 CGACGAGTACCTCATCCACAGCAGGGCTTCTCAGCAGCCCCCAGCTCAGGACTCCCCTCTGAGCTCTCTGAGTGCAACCAGCAACAATTCCACC
1013▶ D E Y L I P Q Q G F F S S P S T S R T P L L S S L S A T S N N S T
3701 GTGGCTTGATTGATAGAAATGGGCTGCAAAGCTGTCCATCAAGGAAGCAGCTTCTTGAGCGATACAGCTCAGACCCACAGGCGCCTTGACTGAGG
1047▶ V A C I D R N G L Q S C P I K E D S F L Q R Y S S D P T G A L T E

StuI (3859)
3801 ACAGCATAGACGACACCTTCTCCAGTGCCTGAATACATAAACAGTCCGTTCCCAAAGGCTGCTGGCTCTGTGCAGAATCCTGTCTATCACAATCA
1080▶ D S I D D T F L P V P E Y I N Q S V P K R P A G S V Q N P V Y H N Q
3901 GCCCTGAACCCCGCCAGCAGAGACCCACACTACAGGACCCACAGCAGTGCAGTGGGCAACCCCGAGTATCTCAACACTGTCCAGCCACCTGT
1113▶ P L N P A P S R D P H Y Q D P H S T A V G N P E Y L N T V Q P T C
4001 GTCAACAGCACATTCGACAGCCCTGGCCAGAAAGCAGCCACCAATAGCCTGGACAACCCGTGACTACCAGCAGGACTTCTTTCCCAAG
1147▶ V N S T F D S P A H W A Q K G S H Q I S L D N P D Y Q Q D F F P K
4101 AAGCCAAGCCAAATGGCATCTTTAAGGGCTCCACAGTGAATAATGCAGAATACCTAAGGGTCCGCCACAAAGCAGTGAATTTATTGGAGCATGACCAG
1180▶ E A K P N G I F K G S T A E N A E Y L R V A P Q S S E F I G A •

NheI (4254)
4201 GAGGATAGTATGAGCCCTAAAAATCCAGACTCTTTCGATACCCAGGACCAAGCCGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTGGACAAA

HpaI (4392)
4301 CCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAA

MfeI (4403)
4401 CAACAATTGCATTATTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAATGTGGTATGGAATTCTAAAAT

4501 ACAGCATAGCAAACTTTAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCA

SapI (4670)
4601 TTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAAGATATAGTGTATTTTCCCAAGTTTGAAGCTAGCTCTTCATTCTTTATGTTTTAAATGCAC

SwaI (4741)
4701 TGACCTCCCACATTCCTTTTTAGTAAATATTAGAAATAATTTAAATACATCATTGCAATGAAATAAATGTTTTTTATTAGGCAGAAATCCAGATGCT

4801 CAAGGCCCTTCATAATATCCCCAGTTTGTAGTGTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTA
141▶ •

4901 GTTCTGGTGTACTTGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGTTGCCATTCTCATCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAG
140▶ N R T Y K L P I L E E I T T K V L K G N M E I L V F C D P A Y D S
5001 ATGAGCTCTGACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCT
106▶ I L E R C M G C P S V V R I S R D V E D S Y P H R V A V I T D F D K

StuI (5166)
5101 TCTGCCCGTGTCTCACAGCAGACCAATGGCAATGGCTTCCAGCAGACAGTACCCTGCCAATGTAGCCTCAATGTGGACAGCAGAGATGATCTCCC
73▶ Q G N S V A S G I A I A E A C V T V R G I Y A E I H V A S I I E G
5201 AGTCTTGGTCTGATGGCCGCCGACATGGTGTCTTGTCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGA
40▶ T K T R I A A G V H H K N D E Y L M T I K E T A V E V L E L D Q Q

BspHI (5316) XmnI (5308) AseI (5374)
5301 GAGATGTTGAAGGCTTTCATGATGGCCCTCTATAGTGAGTCTGATTATACTATGCGGATATACTATGCGGATGATTAATTGTCAAACAGCGTGGATGG
6▶ S I N F T K M
5401 CGTCTCCAGCTTATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCTACCGCCATTTCGCTCAATGGGGCGGAGTTGT

SpeI (5529)
5501 TACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAACAAACTCCCATGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCT

SnaBI (5657)
5600 ATCCACGCCATTGATGTAAGTCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTAAGTCCAAAGTAGGAAAGTCCCATAGGTCAT
5700 GTACTGGGCATAATGCCAGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGTACTGCAAGTGGGCAGTT
5800 TACCGTAAATACTCCACCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATATTGACGTCAATGGGGCGGGGTGCTTGG

5900 GCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAA TTAAGACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTA
5998 AAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAGAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACT
6098 ATAAAGATACCAGGCGTTTTCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCTGCCGTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
6198 AGCGTGGCGCTTTTCATAGCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTAGCCCG
6298 ACCGCTGCGCCTTATCCGTAACATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGC
6398 GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTT
6498 ACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAA
6598 AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAAC
EagI (6708)
SwaI (6697) NotI (6707)
6698 ATTTAAATC AGCGGCCCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACATAACGCTCTCCATCAAACA
6798 AAACGAAACAAAACAACTAGCAAAATAGGCTGTCCCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA