



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

PspI406I (203) HindIII (245)
201 GTGAACGTTCTTTTTTCGCAACGGGTTTCCGCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCAGCGCGGTTGAGTCGCGTTTCTGCCGCTCCCGCTGTGGTGCCTCCTGAAGTGCCTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC

NgoMIV (441)
401 GGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCAGCGTTTGGCTGACCCCTGTTGCTCAACTCTACGTCTTTGTTTCGTTT

KasI (535) BstEII (555) AgeI (552)
501 TCTGTTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGGCGCTACCTGAGATCACCGGTACCATGGAAATCAGCTGGCTGGTATCGCTCCTTCCAGAT
1 M G N Q L A G I A P S Q I
XhoI (683)
601 CCTGTCTGTAGAGAGTTATTTCTCAGACATCCATGACTTTGAGTATGATAAGAGCCTGGGAAGCACTCGGTTTTTAAAGTCGCTCGAGCCAAGCACCGG
13 L S V E S Y F S D I H D F E Y D K S L G S T R F F K V A R A K H R

BamHI (735) SpeI (754)
701 GAAGCCTGGTGGTTGTGAAGGCTTTGCAATTCAGGATCCCAGTACCTTAACCTAGTTATAAACAGGAGCTGGAGGAAGTAAAAATCCGGCTCCACT
47 E G L V V V K V F A I Q D P T L P L T S Y K Q E L E E L K I R L H

PvuI (894)
801 CCGCCAGAAGTGCCTCCCTTCCAGAAAGCAGCAGAAAAGGCGTCTGAGAAAGCAGCCATGCTGTTCCGGCAGTATGTGAGAGACAACCTCTACGATCG
80 S A Q N C L P F Q K A A E K A S E K A A M L F R Q Y V R D N L Y D R

Eco47III (935)
901 CATCAGTACCCGCGCTTCTAAACAACATCGAGAAGCCTGGATCGCCTTCCAGATCCTGACAGCCGTGGACAAAACGCACAAATCTGGAGTCCGCCAC
113 I S T R P F L N N I E K R W I A F Q I L T A V D Q T H K S G V R H

BstEII (1026)
1001 GGAGACATCAAGACAGAGAATGTGATGGTCAACAGTTGGAAGTGGTCCCTCTAACCGACTTCGCCAGTTTTAAGCCACATATCTTCCCGAGGACAACC
147 G D I K T E N V M V T S W N W V L L T D F A S F K P T Y L P E D N

XhoI (1128) Eco47III (1160)
1101 CAGCGATTTCAACTATTTCTTTGACACCTCGAGAAGGAGACATGCTACATAGTCCCGAGCGCTTCGTTGACGGTGGGATGTTCCGCCACCGAGTTAGA
180 P A D F N Y F F D T S R R R T C Y I A P E R F V D G G M F A T E L E

SalI (1228)
1201 GTACATGAGAGACCCCTCAACCCCGCTTGTGCGACTGAATAGCAATCAGAGAGCAAGGGGAGAGCTGAAGAGAGCCATGGACATCTTTTCCGCGAGGTTGC
213 Y M R D P S T P L V D L N S N Q R A R G E L K R A M D I F S A G C
1301 GTGATAGCTGAGCTCTTTACGGAAGGCGTACCTTTGTTGATCTCTCTCAGCTGCTGGCGTATAGAAATGGACATTTTTTCCGTAACAAGTGCTAAACA
247 V I A E L F T E G V P L F D L S Q L L A Y R N G H F F P E Q V L N
1401 AAATTGAAGATCGCAGTATCAGAGACTGGTAACCTCAGATGATTAACCGTGGAGCCGAGAAGCCTTTGGAAGTGGAGACTACCTGAAGCAGCAGCGAGG
280 K I E D R S I R D L V T Q M I N R E P E K R L E A E D Y L K Q Q R G

BsrBI (1582)
1501 CAACGCTTCCCTGAGATATTTTATACCCTTCCAGCCTTACATGGCCAGTTCGCCAAGGAAACCTTTCTCTGCGAGATGAGCGGATTTTGGTTATA
313 N A F P E I F Y T F L Q P Y M A Q F A K E T F L S A D E R I L V I
1601 CGAAAGATTTGGCAACATTATTCACAACCTCTGTGGACAGGATTTGCCAGAAAAGCAGAAGGGGAGTCCAGGGCCAGTGGGCTGGTTGTCTGGTGT
347 R K D L G N I I H N L C G H D L P E K A E G E S R A S G L V V L V
1701 CCGTGATAACGTCCTCGCTGCAGACACTGAAGTCTGCGACTCCTGAGCTCCTGAGCTGAGCCGCTTGGAGCTCATCTGCTGCTGGCCCGAGACTGAGCGTGCAGT
380 S V I A T S C L Q T L K S C D S K L A A L E L I L H L A P R L S V E I

Tth111I (1843)
1801 CCTTCTGGATCGCATCACTCCCTACCTGTTGCATTTGAGCAATGACTCTGTCCCAAGAGTGGCGGCTGAAGCCCTGAGGACACTACCAAAGTCTTGGCC
413 L L D R I T P Y L L H F S N D S V P R V R A E A L R T L T K V L A
1901 CTTGTCCAAGAGTTCTCGCAACGATGCAATATCTATCCAGAGTATATTCTCCCGCATAGCCCACTGGCCAGGACGACGCTACCATCGTCAGAC
447 L V Q E V P R N D V N I Y P E Y I L P G I A H L A Q D D A T I V R
2001 TGGCTACGCTGAAAACATAGCTCTGTTGGCCGAGACGGCTCTGCGATTCTGGAGCTTGTGCGAGCTGAAAACCTCAACATGGAGAACGAGCCCGATAG
480 L A Y A E N I A L L A E T A L R F L E L V Q L K T L N M E N E P D S

MluI (2119)
2101 CGAGGAGGTAGATGAAGCCACGCGTCTAACGGAGACTATGACACAGAGCTCCAGGCCTTGCATGAGATGGTCCAGCAGAAGTGGTCCAGTTGCTAAGC
513 E E V D E A T R P N G D Y D T E L Q A L H E M V Q Q K V V T L L S
2201 GACCCTGAGAATATCGTGAACAGACGCTGATGGAGAGTGGGATCACACGCTTGTGCGTCTTCTCGGACGCCAGAAGGCCAACGACGTTCTCTTGTCCC
547 D P E N I V K Q T L M E S G I T R L C V F F G R Q K A N D V L L S
2301 ACATGATCACTTTCTCAATGATAAGAAATGATTGGCACCTCGTGGAGCTTTCTTTGATAGCATAGTCCGGTGTGGCTGCCTACGTTGGCTGGCAGAGCTC
580 H M I T F L N D K N D W H L R G A F F D S I V G V A A Y V G W Q S S
2401 CTCATCCTCAAGCCCTGCTCCAGCAAGGCTCAGCGATGCTGAGGAGTTTGTGATCGTGAAGCTCTCAATGCCCTGACCTGCATGTGCCAGTAGG
613 S I L K P L L Q Q G L S D A E E F V I V K A L N A L T C M C Q L G
2501 CTGCTGAGAAAACCCATGTCTACGAGTTCGCCAGTGTATGCTCCCTCCGTGTGACCCCTAACCTGTGGATACGCTATGGCGCTGTGGGATTTATCA
647 L L Q K P H V Y E F A S D I A P F P C H P N L W I R Y G A V G F I

ScaI (2620)
2601 CGGTGGTAGTCTCATCAGATCAGTACTGCTGATGCTACTGCAAGCTCATGCCGTATCTCGACCCGTATATTACACAGCCAGTGATACAGATTGAGAGGAA
680 T V V A H Q I S T A D V Y C K L M P Y L D P Y I T Q P V I Q I E R K

BsrBI (2736)
2701 GCTGGTCTGCTCAGTGTGCTAAAGGAGCCGGTGGAGCCGCTCCATATTTGACTATGCTCTGAGGTCGAAAAGACATCGCCAGCTTGTTCAGACATCTCCAC
713 L V L L S V L K E P V S R S I F D Y A L R S K D I A S L F R H L H

2801 ATGCGCCAGAAGAAGCGAAATGGCTCTCTTCTCGACTGCCGCGCCTGAGGACCCACCATAGCGCAACTTCTGAAGAAGCTGCTCTCACAGGGGATG
747▶ M R Q K K R N G S L L D C P P P E D P T I A Q L L K K L L S Q G M
2900 ACTGAAGAAGAGGAAGATAAAGTCTGGCACTGAAGGACTTCATGAAATCAAATCGAGCCAAAGCTAATCGCGTAGACCAGACCTCCACGACA
780▶ T E E E E D K L L A L K D F M M K S N R A K A N A V D Q S H L H D
3000 GCAGCCAGAAAGGTGTAATCGACTTGGCGCCTTAGGCATCACTGGGAGACAAGTTGATCTTGTAAAAACAAACAAGAACCAGATGAGAAAACGTGCTAG
813▶ S S Q K G V I D L A A L G I T G R Q V D L V K T K Q E P D E K R A R
3100 GAAACATGTGAAACAAGACTCCAATGTAATGAAGAGTGGAAAAGCATGTTTGGGTCAATGGAGCCACGAATATCCACAGGCCCTGCCTAAAACAGT
846▶ K H V K Q D S N V N E E W K S M F G S L E P P N I P Q A L P K T S
3200 GACCATGAGGTTTTCAGCCTGGGAGCCACCTCGCTCTGAGTCTCTGCTGGTATCTGTCCCTTTGTCAACAATCCACAGGCTCAGAAGCAGCA
880▶ D H E V V Q P G K P P R S E S S A G I C V P L S T S P Q V S E A A

DraIII (3372)

BspEI (3365)

BsaBI (3363) BbrPI (3375)

3300 ACATCCCGAGTAAGAAGCCTGTGATCCCGTTGTAAAGTAGCACGGTCTGCCGTCCACTTACCAGATCCGGATCACACGTGCAAGACTGAACTTCCAGC
913▶ H I P S K K P V I P V V S S T V L P S T Y Q I R I T T C K T E L Q
3399 AGTCATACAGCAGAAGCGGGAGAGTGAACCGGGAGCGCATAGCACAAGCAGATGATGAGAATGCCGAGTGGGAGAGCAAGCCACCTCCACCTGGGTG
946▶ Q L I Q Q K R E Q C N A E R I A K Q M M E N A E W E S K P P P P G W
3499 GCGTCTAAAGGCTGCTAGTTGCACATCTTCATGAGCAAAATCTGCTGTGAATCGCATCAGAGTCTCTGATGAACACTTACTTTTTGCAACATGTTCA
979▶ R P K G L L V A H L H E H K S A V N R I R V S D E H L L F A T C S

BglII (3614)

3599 AATGATGGCACAGTGAAGATCTGGAACAGCCAGAAGATGGAGGGGAAGACCACCACAACCGGCTTATTCTCACGTACAGCCGAATTGGAGGACGAGTCA
1013▶ N D G T V K I W N S Q K M E G K T T T T R S I L T Y S R I G G R V

Psp1406I (3705)

MscI (3731)

3699 AGACGCTAACGTTTTGCCAAGGCTCCCACTACTGGCCATAGCATCTGATAATGGTGTCTCCAGCTTCTTGGAAATTGAGGCGTCTAAGTTACCCAAGTC
1046▶ K T L T F C Q G S H Y L A I A S D N G A V Q L L G I E A S K L P K S

Ppu10I (3862)

NsiI (3862)

3799 TCCTAAAATTCACCTCTACAAAGCAGGATTCTGGATCAGAAGGAAGATGGCTGTGTGGTGGACATGCATCACTTCAACTCCGGGGCACAGTCCGTTCTT
1079▶ P K I H P L Q S R I L D Q K E D G C V V D M H H F N S G A Q S V L

MluI (3953)

3899 GCCTATGCCACGGTGAATGGCTCTCTGGTTGGATGGGATCTCAGTCTTCAAGCAACCGTGGACATTAAGCATGACCTAAAGTTCAGGCCTCATCACTT
1113▶ A Y A T V N G S L V G W D L R S S S N A W T L K H D L K S G L I T
3999 CCTTTGCTGTGGACATCCACCAGTCTGGCTGTGCATAGGCACGAGCAGCGGTGCCATGGCGTGTGGGACATGAGGTTCCAGTTGCCCATCTCCAGTCA
1146▶ S F A V D I H Q C W L C I G T S S G A M A C W D M R F Q L P I S S H

KasI (4124)

4099 CTGTCATCCCTCCAGAGCTCGCATCCGGCGCCTCTCCATGCACCCCTGTACCAGTCTGGTAAATCGCAGCTGTTCCAGGCAACAATGAAGTCTCCATG
1179▶ C H P S R A R I R R L S M H P L Y Q S W V I A A V Q G N N E V S M
4199 TGGACATGGAGACTGGCGACAGGAGACTGACTCTGGGCCAGCAGCGCACCCGCACTGTCTGAGTTACAGCCTTACCAGCACAGCGTCCATGGCATCT
1213▶ W D M E T G D R R L T L W A S S A P P L S E L Q P S P H S V H G I
4299 ACTGCAGCCCTGCAGATGAAAACCTATCTGTGACCGTGGCTCAGACATGAAAATAAGGTTCTGGGACTTGGTTTCCCAGAGAGGTCCTATGTTGT
1246▶ Y C S P A D G N P I L L T A G S D M K I R F W D L V S P E R S Y V V

RsrII (4491)

4399 TGGGGAAGTACAGGTTCCCGTCTGTCTCTACTACAAGAAGATAATAGAAGGCACCGAGGTTGTCAGGAAATTCAGAATAAGCAGAAGGTCGGACCC
1279▶ A G S T G S P S V S Y Y K K I I E G T E V V Q E I Q N K Q K V G P

Bsp120I (4518)

4499 AGTGATGACACCCCTCGAGGGGCGCGAGTCTCTGCCTGTGGGACATCATGACATCATACAGACATTGCCACCTTCCAGACCACTCAGGGCTTCATTG
1313▶ S D D T P R R G P E S L P V G H H D I I T D I A T F Q T T Q G F I

MscI (4667)

XbaI (4606)

NheI (4661)

4599 TGAAGTCTTAGAGACGAATTTGAAAGTGTGAAATGAGTCTGCCACTTGATAAAAATGCTAGCTGGCCAGACATGATAAGATACATTGATGAGT
1346▶ V T A S R D G I V K V W K •
4699 TTGGACAACCAACTAGAATGCAAGTGAATAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACA

HpaI (4799) MfeI (4810)

EcoRI (4895)

4799 AGTTAACAACAACAATTGCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAA
4899 TTCTAAAATACAGCATAGCAAACTTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGC
4999 CAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTAAGATATAGTGATTTTCCCAAGGTTTGAAGTCTTCTCATTCTTTATGTTT

SspI (5134)

SwaI (5148)

5099 TAAATGCACTGACCTCCACATTCCTTTTTAGTAAAATATTCAGAAAATATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAAT

5199 CCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTT

5299 CTAGCTTTAGTTCCTGGTACTTGGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCATTCTCAATGAGCACAAGCAGTCCAGGAGCA
141▶ • 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

5399 TAGTCAGAGATGAGTCTCTGCACATGCCAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCTGACAGCCACAATGGTGT
109▶ Y D S 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

5499 CAAAGTCTTCTGCCGTTGCTCACAGCAGACCAATGGCAATGGCTTCCAGCAGACAGTACCCTGCCAATGTAGGCTCAATGTGGACAGCAGAGAT
76▶ F D K 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

5599 GATCTCCCAAGTCTTGGTCTGATGGCCGCCGACATGGTGGTGTGTCCTCATAGAGCATGGTATCTTCTCAGTGGCGACCTCCACAGCTCCAGA
43▶ I E G 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

AseI (5781)

5699 TCCTGCTGAGAGATGTTGAAGGTTCTCATGATGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAG
94▶ D Q Q S I N F T K M ◀

5799 CGTGGATGGCGTCTCCAGC **T**TATCTGACGGTTCACATAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTTCGTCATGCGG

5898 GCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTT**ACTAGT**CAAAACAAACTCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAG

5997 TCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCC

6097 ATAAGTCATGTACTGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATGTACTGCCAA

6197 GTGGGCAGTTTACCGTAAATACTCCACCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGG

6297 GGGTCGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAGTTATGTAAC**GCTGCAGGTTAA**TTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC

6395 AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACC

6495 CGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGTCCCTCGTGGCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCT

6595 CCCTTCGGAAGCGTGGCGCTTTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCC

6695 GTTCAGCCCGACCCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGA

6795 TTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCT

6895 GAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACG

6995 CGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTA

7095 **GTTAATTAACATTTAAATC** AGCGGCCGAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTTGTGTGAATCGTAACTAACATACGCTCTC

7195 CATCAAAACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA