



NotI (-1)
1 GCGGCCGCAATAAAATATCTTTATTTTTCATTACATCTGTGTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACA
PvuI (172)
SgfI (171)
101 AAACAACTAGCAAAATAGGCTGTCCCGAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAAGGATCTGCGATCGCTCCGGTGCCCGTCACTGGGCA

MfeI (247)AgeI (254)EcoNI (261)
201 GAGCGCACATCGCCACAGTCCCCGAGAAGTTGGGGGAGGGTCCGCAATTGAACCGGTGCTAGAGAAGTGGCGGGGTAAGTGGAAAGTGATG

Psp1406I (368)
301 TCGTGTACTGGCTCCGCTTTTTCCCGAGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCCGCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAG

HindIII (410) Bsu36I (456)
PvuII (404) EcoNI (452)
401 AACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTACCGCGCCCGCCCTACCTGAGCGCCCATCCACGCCGTTGAGTCCGCTTCTGCCGCT

501 CCGCCTGTGGTCCCTCTGAACTCGTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTGAGACCGGCCCTTTGTCCGGCTCCCTTGGAGCTACCTA

NaeI (606)
601 GACTCAGCGGCTCTCCAGCTTTGCTGACCCTGCTTCAACTCTACGCTTTTGTTCGTTTTCTGTTCTGCCGCTTACAGATCCAAGCTGTGACC

KasI (700) AgeI (717)
701 GCGCCTACTGAGATCACCGGTAGGAGGCCATCATGAGACTCATCAGAAACATTTACATTTTTGTAGTATTGTTATGACAGCAGAGGGTGTGCTCC
1 M R L I R N I Y I F C S I V M T A E G D A P

XcmI (880)
801 AGAGTGCAGAAAGAAAGGAACTGATGACCAACTGCTCCAACATGTCTCTAAGAAAGTTCCCGCAGACTTACCCACCAACGACTGGATTTA
22 E L P E E R E L M T N C S N M S L R K V P A D L T P A T T T L D L

PvuII (990)
901 TCCTATAACCTCTTTTTCAACTCCAGAGTTCAGATTTTCTGCTCCAACTGAGAGTTTTGATTCTATGCCATAACAGAATTCACAGCTGGATC
56 S Y N L L F Q L Q S S D F H S V S K L R V L I L C H N R I Q Q L D

1001 TCAAAACCTTTGAATCAACAAGGAGTTAAGATATTTAGATTTGTCTAATAACAGACTGAAGAGTGAACCTTGGTATTACTGGCAGGTCTCAGGTATTT
89 L K T F E F N K E L R Y L D L S N N R L K S V T W Y L L A G L R Y L

BglII (1100) **AvrII (1172)**
1101 AGATCTTTCTTTAATGACTTTGACACCATGCCTATCTGTGAGGAAGCTGGCAACATGTCACACCTGGAAATCTAGGTTTGGAGTGGGCAAAAATACAA
122 D L S F N D F D T M P I C E E A G N M S H L E I L G L S G A K I Q
1201 AAATCAGATTTCCAGAAAATGCTCATCTGCATCTAAATACTGTCTTCTTAGGATTCAGAACTTCTCCTCATTATGAAGAAGGTAGCCTGCCATCTTAA
156 K S D F Q K I A H L H L N T V F L G F R T L P H Y E E G S L P I L
1301 ACACAACAAAAGTGCACATGTTTTACCAATGGACACAAATTTCTGGGTTCTTTTGCCTGATGGAATCAAGACTTCAAAAATATTAGAATGACAAATAT
189 N T T K L H I V L P M D T N F W L L R D G I K T S K I L E M T N I
1401 AGATGGCAAAAGCAATTTGAAGTTATGAATGCAACGAAATCTTAGTTTGAAAAATGCTAAGACATCCGTTCTATTGCTTAATAAGGTTGATTTACTC
222 D G K S Q F V S Y E M Q R N L S L E N A K T S V L L L N K V D L L

XmnI (1548)
1501 TGGGACGACCTTTTCTTATCTTACAATTTGTTGGCATAATCAGTGGAACTTTTTCAGATCCGAAATGTGACTTTTGGTGGTAAAGCTTATCTTGACC
256 W D D L F L I L Q F V W H T S V E H F Q I R N V T F G G K A Y L D

BsrGI (1652)
1601 ACAATTCATTTGACTACTCAAATACTGTAATGAGAACATAAAAATGGAGACTGTACATTTTTCAGAGTGTTTTACATTCAACAGGATAAAAATCTATTTGCT
289 H N S F D Y S N T V M R T I K L E H V H F R V F Y I Q Q D K I Y L L
1701 TTTGACCAAAATGGACATAGAAAACCTGACAAATCAAATGCACAAATGGCCACACATGCTTTTCCGAAATATCTACGAAATCCAATATTTAAATTTT
322 L T K M D I E N L T I S N A Q M P H M L F P N Y P T K F N L N F
1801 GCCAATAATATCTTAACAGACGAGTTTAAAAAGAACTATCCAACCTGCCTCACTTGAACCTCTCATTTTGAATGGCAATAAACTGGAGACACTTTCTT
356 A N N I L T D E F F K R T I Q L P H L K T L I L N G N K L E T L S

Ball (1989)
1901 TAGTAAGTTGCTTTGCTAACAACACACCTTGGAACTTGGATCTGAGTCAAATCTATTACAACATAAAAATGATGAAAATTTGCTCATGGCCAGAAAC
389 L V S C F A N N T P L E H L D L S Q N L L Q H K N D E N C S W P E T
2001 TGTGGTCAATATGAATCTGCATACAATAAATGTCTGATTCTGCTTTCAGGTGCTTCCCAAAAGTATTCAAATACTTGACCTAAATAAACCAGAAATC
422 V V N M N L S Y N K L S D S V F R C L P K S I Q I L D L N N N Q I
2101 CAAACTGTACTAAGAGACTATTCTGATGGCCTTACGAGAACTAAATATTGCATTAATTTTCTAAGTATCTCCCTGGATGCGATCATTTCAGTA
456 Q T V P K E T I H L M A L R E L N I A F N F L T D L P G C S H F S
2201 GACTTTCAGTTCTGAACATTGAAATGAACTTCTCAGCCATCTCTGGATTTTGTTCAGAGCTGCCAGGAAATTAACCTTAAATGGGGAAGAAA
489 R L S V L N I E M N F I L S P S L D F V Q S C Q E V K T L N A G R N
2301 TCCATTCGGGTGACCTGTGAATTAATAAATTTTCACTCAGCTTGAACATATTCAGAGGTGATGATGGTGGATGGTGGATTCATACACCTGTGAATAC
522 P F R C T C E L K N F I Q L E T Y S E V M M V G W S D S Y T C E Y

Bsu36I (2408)
2401 CCTTTAAACCTAAGGGGAAC TAGGTTAAAAGAGCTTCACTCCACGAATATCTTGAACACAGCTCTGTTGATTGTCACCATTTGTTGTTATTTGCTAG
556 P L N L R G T R L K D V H L H E L S C N T A L L I V T I V I M L
2501 TTCTGGGTTGGCTGTGGCTTCTGCTGCTCCACTTTGATCTGCCCTGGTATCTCAGGATGCTAGGCAATGCACACAAACATGGCACAGGGTTAGGAA
589 V L G L A V A F C C L H F D L P W Y L R M L G Q C T Q T W H R V R K

XbaI (2699)
2601 AACAAACCAAGAACAACCTCAAGAGAAATGTCGATTCCACGATTTATTTTATACAGTGAACATGATTCTGTGGGTGAAGAATGAATGATCCCAAT
622 T T Q E Q L K R N V R F H A F I S Y S E H D S L W V K N E L I P N

HindIII (2783)
2701 CTAGAGAAGGAAGATGGTTCTATCTTATTGCTTTTATGAAAGCTACTTTGACCCTGGCAAAAGCATTAGTGAATAATGTAAGCTTCAATGAGAAAA
656 L E K E D G S I L I C L Y E S Y F D P G K S I S E N I V S F I E K
2801 GCTATAAGTCCATCTTTGTTTGTCTCCCACTTTGTCAGAAATGAGTGGTCCATTAATGATTTCTACTTTGCCACCACAATCTCTTCCATGAAAATTC
689 S Y K S I F V L S P N F V Q N E W C H Y E F Y F A H H N L F H E N S

XmnI (2925)
2901 TGATCATATAATCTTATCTTACTGGAACCCATTCCTATTCTATTGCATTCCACCAGGTATCATAAACTGAAAGCTCTCTGGAAAAAAGCATACTTG
722 D H I I L I L L E P I P F Y C I P T R Y H K L K A L L E K K A Y L

AseI (3057) **PstI (3096)**
3001 GAATGGCCCAAGGATAGCGTAAATGTGGGCTTTTCTGGGCAAACTTCGAGCTGCTATTAATGTTAATGTATTAGCCACCAGAGAAATGTATGAACTGC
756 E W P K D R R K C G L F W A N L R A A I N V N V L A T R E M Y E L

XhoI (3127)
3101 AGACATTCACAGAGTTAAATGAAGAGTCTCGAGGTTCTACAATCTCTGATGAGAACAGATTGTCTATAAAATCCACAGTCTTGGGAAGTTGGGGAC
789 Q T F T E L N E E S R G S T I S L M R T D C L •

Ball (3214)
3201 CACATACAGTCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACAGTGAAGTGAAGGAAATGCTTTATTTGTAATTTG

NheI (3208)
3301 TGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTATGTTTCAGGTTCAAGGGGAGGTGTGG

HpaI (3346) **MfeI (3357)**

3401 GAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACCTTAACTCCAAATCAAGCCTCTACTTGAAT
3501 CCTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTAAGATATAG
3601 TGTATTTCCCAAGGTTTGAAGTACTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCTTTTTAGTAAAATATTCAGAAATAATTTA
EcoO109I (3756)
3701 AATACATCATTGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGA
3801 ACAAGGAACCTTTAATAGAATTGGACAGCAAGAAAGCGAGCTTCTAGCGAATTCTCGACTCATTCTTTGGCCCTGGACGAGTGTGGGGCGTCGGT
342 • E K A R P R T S P R R N
ScaI (3913)
3901 TCCACTATCGGCGAGTACTTCTACACAGCCATCGGTCCAGACGGCCGCTTCTCGGGGCGATTGTGTACGCCGACAGTCCCGGCTCCGGATCGGACG
329 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 T G A G S R V
4001 ATTGGTCGCATCGACCTGCGCCAAAGTGCATCATCGAAATGCGTCAACCAAGCTGTATAGAGTTGGTCAAGACCAATCGGAGCATATACGCC
295 I A D C R G Q A W A A D C A D F N G D V L S Q Y L Q D L G I R L M Y A R
SacII (4104) BsrBI (4134)
4101 GGAGCCGCGGCGATCTGCAAGCTCCGGATGCCTCCGCTCGAAGTAGCGCTGTCTGCTCCATACAAGCCAACACGGCCTCCAGAAGAAGATGTTGGC
262 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 P R W F F I N A
4201 GACCTCGTATTGGGAATCCCGAACATCGCTCGCTCCAGTCAATGACCGCTGTATGCGCCATTGTCGTCAGGACATTGTTGGAGCCGAAATCCGGC
229 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 N N S G F D A
DraIII (4302) Tth111I (4377)
4301 TGCACGAGTGGCCGACTTCGGGGCAGTCTCGGCCAAAGCATCAGCTCATCGAGAGCTGCGGACGGACGCACTGACGGTGTCTCCATCACAGTTT
195 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 V T D D M V T Q
NdeI (4432) RsrII (4473) BsrBI (4492)
4401 GCCAGTGATACACATGGGATCAGCAATCGGCATATGAAATCACGCCATGTAGTGTATTGACCGATTCCTTGGGTCCGAATGGCCGAACCCGCTCGT
162 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 G F P G F G S T
PvuI (4520) SgfI (4519) SaeI (4533) DraIII (4595)
4501 CTGGCTAAGATCGGCCGAGCGATCGCATCCATGAGCTCCGCGACGGTTCGAGAACAGCGGGCAGTTCGGTTTCAGGACAGTCTTGAACGTGACACC
129 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 L D Q L T V G
4601 TGTGCACGGCGGAGATGCAATAGGTGAGGCTCTCGTGAATCCCAATGTCAAGCACTTCGGAATCGGGAGCGGGCCGATGCAAGTCCGATAAA
95 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 A S A F H R Y V
4701 CATAACGATCTTTGTAGAACCATCGGCGAGTATTTACCGCAGGACATCCACGCCCTCTACATCGAAGTGAAGCAGGAGATTCTTCGCCCTC
62 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 F A R S E E G E
PshAI (4854) NruI (4850)
4801 CGAGAGTGCATCAGGTGGAGACGCTGTGCAACTTTTCGATCAGAACTTCGCGACAGACTGCGGGTGGTTCAGGCTTTTTCATGATGGCCCTCTCA
29 S L Q M L D S V S D F K E I L F K A V S T A T L E P K K M
AseI (4941) SaeI (4998)
4901 TAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCCTAAACGA
SpeI (5096)
5001 GCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTA
5101 GTCAAAACAACTCCATTGACGTCATGGGGTGGAGACTTGGAAATCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTAAGTCCAAACCCGAT
SnaBI (5224)
5201 CATCATGGTAATAGCGATGACTAATACGTAGATGACTGCCAAGTAGGAAAGTCCCATAAAGTCACTGACTGGGCATAATGCCAGGCGGGCCATTACCG
NdeI (5329)
5301 TCATTGACGTCATAGGGGCGTACTTGGCATATGATACACTTGTACTGCTCAAGTGGGCGAGTTTACCGTAAATACTCCACCCATTGACGTCATGGA
5401 AAGTCCCTATTGGCGTACTACTGGGAACATACGTCATTATTGACGTCATGGGCGGGGCTGTTGGCGGTCAGCCAGGCGGGCCATTACCGTAAGTTA
PaeI (5515) PstI (5508) SdaI (5507)
5501 TGTAACGCTGCAGGTTAATTAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTTCATAGGCTCC
5601 GCGCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAAGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCGTGGAAAGTCCCT
5701 CGTGCCTCTCCTGTTCCGACCTGCCCTTACCGGATACCTGTCCGCTTCTCCCTTCGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTAT
5801 CTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTCAGCCGACCGCTGCGCTTATCCGGTAACTATCGTCTTGTAGT
5901 CCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTCTTGAAGTG
6001 GTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAGTGGTAGCTCTTGTATCCGGC
6101 AAACAAACCCCGCTGGTAGCGGTGGTTTTTTTGGTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGG
PaeI (6255)
6201 GGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTGGTCTAGGCTAGTTAATTAACATTTAAATCA