



NotI (-1)
1 GCGGCGCGCAATAAATATCTTTATTTTTCATTACATCTGTGTGGTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACA
PvuII (172)
SgfI (171)
101 AAACAACTAGCAAAATAGGCTGTCCCCAGTGCAGGTGCAGAGTCCAGAACATTTCTCTATCGAAGGATCTGCGATCGCTCCGGTGCCCGTCACTGGGCA
MfeI (247) EcoNI (261)
201 GAGCGCACATCGCCACAGTCCCGGAGAAAGTTGGGGGAGGGGTGGCAATTGAACGGGTGCTAGAGAAAGTGGCGGGGTAAACTGGGAAAGTGATG
Psp1406I (368)
301 TCGTGTACTGGTCCGCCTTTTTCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAG
HindIII (410)
PvuII (404) EcoNI (452)
401 AACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTACGCGCCCGCCCTACCTGAGGCCGCATCCACGCCGTTGAGTCCGGTCTGCCGCCCT
501 CCGCGCTGTGGTGCCTCCTGAACTCGCTCCGCGCTTAGGTAAGTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCTACCTA
NaeI (606)
601 GACTCAGCCGGCTCTCCACGCTTGGCTGACCCTGCTTGTCTCAACTCTACGCTTTTTGTTTCTGTTTCTGCGCCGTTACAGATCCAAGCTGTGACC
NcoI (725)
BstEII (720)
AgeI (717)
701 GGGCGCTACCTGAGATCACCGGTCCACATGAAAAATGTTCTTCAGTCGTCAATGCTGACCTGCATTTTCTGCTAATATCTGGTTCCTGTGAGTTAT
1 M E N M F L Q S S M L T C I F L L I S G S C E L
SalI (881)
801 GCGCGAAGAAAAATTTTCTAGAAGCTATCCTTGTGATGAGAAAAAGCAAATGACTCAGTATTGCGAGAGTGCAGCAATCGTCGACTACAGGAAGTTCC
25 C A E E N F S R S Y P C D E K K Q N D S V I A E C S N R R L Q E V P
901 CCAAACGGTGGGCAAATATGTGACAGAACTAGACCTGTCTGATAATTTTATCACACACATAACGAATGAATCAATTTCAAGGGCTGCAAAATCTCACTAAA
58 Q T V G K Y V T E L D L S D N F I T H I T N E S F Q G L Q N L T K
BsrGI (1023)
1001 ATAAATCTAAACCACAACCCCAATGTACAGCACCAGAACGAAATCCCGGTATACAATCAAATGGCTTGAATATCACAGACGGGCATTCTCAACCTAA
92 I N L N H N P N V Q H Q N G N P G I Q S N G L N I T D G A F L N L
1101 AAAACCTAAGGGAGTTACTGCTTGAAGCAACACAGTTACCCCAAATACCTCTGGTTTCCAGAGTCTTTGACAGAACTTAGTCTAATTCAAAACAAAT
125 K N L R E L L L E D N Q L P Q I P S G L P E S L T E L S L I Q N N I
1201 ATACAACATAACTAAAGAGGGCATTTCAGACTTATAAACTTAAAAATCTCTATTTGGCCTGGAAGTCTATTTTAAACAAAGTTTGGGAGAAAACCTAAC
158 Y N I T K E G I S R L I N L K N L Y L A W N C Y F N K V C E K T N
1301 ATAGAAGTGGAGTATTTGAAACGCTGACAAATTTGGAGTTGCTATCCTATCTTTCAATTTCTTTTACATGTGCCACCAAACTGCCAAGCTCCCTAC
192 I E D G V F E T L T N L E L L S L S F N S L S H V P P K L P S S L
1401 GAAACCTTTTCTGAGCAACACCCAGATCAAATACATTAGTGAAGAAGATTTCAAGGGATTGATAAATTTAACCATTACTAGATTTAAGCGGGAAGTGTCC
225 R K L F L S N T Q I K Y I S E E D F K G L I N L T L L D L S G N C P
1501 GAGGTGCTTCAATGCCCATTTCCATGCGTGCCTTGTGATGGTGGTCTCAATTAATATAGATCGTTTTGCTTTTCAAAACTTGACCAACTTCCGATAC
258 R C F N A P F P C V P C D G G A S I N I D R F A F Q N L T Q L R Y
1601 CTAACCTCTAGCACTTCCCTCAGGAAGATTAATGCTGCTGGTTAAAAATGCTCATCTGAAGGTGCTGGATCTTGAATCAACTATTTAGTGG
292 L N L S S T S L R K I N A A W F K N M P H L K V L D L E F N Y L V
1701 GAGAAATAGCCTCTGGGGCATTTTAACGATGCTGCCCGCTTAGAAATACTTGACTTGTCTTTAACATATAAAGGGGAGTTATCCACAGCATATTA
325 G E I A S G A F L T M L P R L E I L D L S F N Y I K G S Y P Q H I N
1801 TATTTCCAGAACTCTCAAACTTTTGTCTACGGCATTGCATTTAAGAGTTATGTGTTCCAGGAACCTAGAGAAGATGATTTCCAGCCCTGAT
358 I S R N F S K L L S T L R A L H L R G Y V F Q E L R E D D P C T L M
ClaI (1951)
1901 CAGCTTCAAACCTATCGACTATCAACTTGGGTATTAATTTTATTAAGCAAATCGATTTCAAACCTTTTCAAATTTCTCCAATCTGGAATATTTACT
392 Q L P N L S T I N L G I N F I K Q I D F K L F Q N F S N L E I I Y
2001 TGTGAGAAAACAGAATACCCGTTGGTAAAAGATACCCGGCAGAGTTATGCAAATAGTTCCTCTTTTCAACGTCATATCCGGAACGACGCTCAACAGA
425 L S E N R I S P L V K D T R Q S Y A N S S S F Q R H I R K R R S T D
BstBI (2118)
2101 TTTTGTGTTGACCCACATTCGAACCTTTATCATTTCACCCGCTCTTAATAAAGCCACAATGTGCTGCTTATGAAAAGCCTTAGATTTAAGCCTCAAC
458 F E F D P H S N F Y H F T R P L I K P Q C A A Y G K A L D L S L N
2201 AGTATTTTCTTATTGGGCAACCAATTTGAAAATCTTCTGACATTCCTGCTGTTTAAATCTGCTGCAAAATAGCAATGCTCAAGTGTAAAGTGGAACTG
492 S I F F I G P N Q F E N L P D I A C L N L S A N S N A Q V L S G T
2301 AATTTTCAAGCATTCTCATGCAAAATTTTGGATTTGACAAAACATAGACTAGACTTTGATAATGCTAGTCTTACTGAATTTCCGACTTGGAAAGT
525 E F S A I P H V K Y L D L T N N R L D F D N A S A L T E L S D L E V
BglIII (2403)
2401 TCTAGATCTCAGTATAATTCACACTATTTTCAAGATAGCAGCGTAAACACATCATCTAGAATTTATTCAAATTTTCAAAATCTAAAAGTTTAAACTTG
558 L D L S Y N S H Y F R I A G V T H H L E F I Q N F T N L K V A T L L T G
2501 AGCCACAACAACATTTTAACTTTAAACAGATAAGTATAACCTGGAAGCAAGTCCCTGGTGAATTTAGTTTTCAGTGGCAATCGCTTGCATTTTGTGGAA
592 S H N N I Y T L T D K Y N L E S K S L V E L V F S G N R L D I L W
2601 ATGATGATGACAACAGGTATATCTCCATTTTCAAAGTCTCAAGAACTGACACGCTGCGATTTATCCCTTAAATAGGCTCAAGCACATCCAAATGAAGC
625 N D D D N R Y I S I F K G L K N L T R L D L S L N R L K H I P N E A
XhoI (2794)
2701 ATTCCTTAATTTGCCAGCGAGTCTCACTGAACATACATAAATGATAATATGTTAAAGTTTTTAACTGGACATTACTCCAGCAGTTTCTCTCGTCTCGAG
658 F L N L P A S L T E L H I N D N M L K F F N W T L L Q Q F P R L E
2801 TTGCTTACTGACTGAGTGGAAACAACTACTCTTTTAACTGATAGCCTATCTGACTTTACATCTTCCCTTCCGACACTGCTGCTGAGTCATAACAGGATTT
692 L L D L R G N K L L F L T D S L S D F T S S L R T L L L S H N R I
2901 CCCACCTACCCTCTGGCTTTCTTCTGAAAGTCAAGTCTGAAAGCCTCGATTTAAAGTTCCAATCTGCTAAAAACAATAAACAAATCCGCACTTGAAC
725 S H L P S G T L S E V S S L K H L D L S S N L L K T I N K S A L E T
BsiBI (3088)
3001 TAAGACCACCACAAATATCTATGTTGAACTACACGAAACCCCTTGAATGACCTGTGACATTGGAGATTTCCGAAAGTGGATGGATGAACATCTG
758 K T T T K L S M L E L H G N P F E C T C D I G D F R R W M D E H L
3101 AATGCAAAATTTCCAGACTGGTAGATGTCATTTGTGCCAGTCCCTGGGGATCAAAGAGGGAAGATTTGTGAGTCTGGAGTAAACAACTTGTGTTTCAG
792 N V K I P R L V D V I C A S P G D Q R G K S I V S L E L T T C V S
PstI (3206) NcoI (3245)
3201 ATGCTACTGACGTGATATTATTTTCTTACGCTCTTTATCACACCATGTTTATGTTGGCTGCCCTGGCTCACATTTGTTTACTGGGATGTTTGGTT
825 D V T A V I L F A F F T F F I T T M V M L A A L A H H L F Y W D V W F
3301 TATATAAATGTGTGTTTAGCTAAGGTAAGGTAAGGCTACAGGCTCTTTTCCACATCCCAAACTTTCTATGATGCTTACATTTCTTATGACACCAAGATGCC
858 I Y N V C L A K V K G Y R S L S T S Q T F Y D A Y I S Y D T K D A
SmaI (349)
SandI (3491)
EcoI019I (3491)
Psp1406I (3460)
3401 TCTGTACTGACTGGGTGATAAATGAGCTGCCTACCACCTTGAAGAGAGCCGAGACAAAACGTTTCTCTTTGCTAGAGGAGGGATTGGGACCCGG
892 S V T D W V I N E L R Y H L E E S R D K N V L L C L E E R D W D P

BalI (3503)
3501 GATTGGCCATCATCGACAACCTCATGCAGAGCATCAACCAAAGCAAGAAAACAGTATTGTTTAAACAAAAATATGCAAAAAGCTGGAACCTTAAAC
925▶ G L A I I D N L M Q S I N Q S K K T V F V L T K K Y A K S W N F K T
3601 AGCTTTTACTTGGCTTTCGACAGGCTAATGGATGAGAACATGGATGTGATTATATTTATCCTGCTGGAGCCAGTGTACAGCATTCTCAGTATTTGAGG
958▶ A F Y L A L Q R L M D E N M D V I I F I L L E P V L Q H S Q Y L R
3701 CTACGGCAGCGGATCTGTAAGAGCTCCATCCTCCAGTGGCCGCAACCCGAAGGCAGAAGGCTTGTTTTGGCAAACCTGAGAAAATGTGGTCTTGACTG
992▶ L R Q R I C K S S I L Q W P D N P K A E G L F W Q T L R N V V L T

BalI (3874)
NheI (3868)
3801 AAAATGATTACGGTATAACAATATGTATGTCGATTCCATTAAGCAATACTAACTGACGTTAAGTCATGCTAGCTGGCCAGACATGATAAGATACATTGA
1025▶ E N D S R Y N N M Y V D S I K Q Y •
3901 TGAGTTTGGCAAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAAT

HpaI (4006) MfeI (4017)
4001 AAACAAGTTAAACAACAACAAATTGCATTCATTTATGTTTCAGGTTCCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTA
4101 TGGAAATCTAAAATACAGCATAGCAAACTTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCT
4201 GTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTCTTTCATGGAGTTTAAAGATATAGTGATTTTCCCAAGGTTTGAACAGCTCTTCTTTCTTTA
4301 TGTTTTAAATGCACTGACCTCCCACATCCCTTTTAGTAAATATTCAGAAATAATTAATACATCATTTGCAATGAAAATAAATGTTTTTTATTAGGC

EcoO109I (4416)
4401 AGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTGTGACTTAGGGAACAAAGAACCTTTAATAGAAATTGGACAGCAAGAAAGCG

ScaI (4573)
4501 AGCTTCTAGCGAATTCGACTCATTCCCTTGGCCCTGGACGAGTGTGGGGCTGGGTTTCCACTATCGCCGAGTACTTCTACACAGCCATCGGTCACG
3424 • E K A R P R T S P R R N G S D A L V E V C G D T W
4601 ACGGCCGCGCTTTCGGGGGATTTGTGTACGCCGACAGTCCCGGCTCCGGATCGGACGATTGCGTCGCATCGACCTGCGCCAAAGCTGCATCATCGA
315▶ V A A S R R A I Q T R G V T G A G S R V I A D C R G Q A W A A D D F

SacII (4764) BsrBI (4794)
4701 AATTGCCGTCACCAAGCTCTGATAGAGTTGGTCAAGACCAATGCGGAGCATATACGCCGGAGCCGCGGCGATCCTGCAAGCTCCGGATGCCTCCGCTC
282▶ N G D V L S Q Y L Q D L G I R L M Y A R L R P S G A L E P H R R E
4801 GAAGTAGCGCTGTCTGCTCCATAACAAGCAACACGGCTCCAGAAGAAGATGTTGGCGACCTCGTATTGGGAATCCCGAACATCGCCTCGTCCAG
249▶ F Y R T Q Q E M C A L W P R W F F I N A V E Y Q S D G F M A E S W
4901 TCAATGACCGCTGTTATGCGGCCATTGTCGTCAGGACATTGTTGGAGCCGAAATCCGCTGCACGAGGTCCCGACTTCGGGGCAGTCTCGGCCAAA
215▶ D I V A T I R G N D T L V N N S G F D A H V L H R V E P C D E A W L

Tth111I (5037) NdeI (5092)
5001 GCATCAGCTCATCGAGAGCCTGCGCGACGACGACTGACGGTGTCTGCATCACAGTTTCCAGTGATACACATGGGGATCAGCAATCGCCATATGAA
182▶ M L E D L A Q A V S A S V T D D M V T Q W H Y V H P D A I A C I F

RsrII (5133) BsrBI (5152) PvuI (5180) SgfI (5179)
5101 ATCACGCCATGTAGTGATTGACCGATTCTGCGGTCGGAATGGGCCGAAACCGCTCGCTGGCTAAGATCGGCCGAGCGCATCCATCCATGAGCTCC
149▶ D R W T T Y Q G I G Q P G F G S T Q S L D A A A I A D M L E
5201 GCGACGGTTGCAGAACAGCGGGCAGTTCGGTTTCAGGACGCTTGAACGTCGACACCTGTGCACGGCGGAGATGCAATAGTTCAGGCTCTCGTGA
115▶ A V P Q L V A P L E T E P L D Q L T V G Q A R R S I C Y T L S E S F
5301 ATTCCCAATGTCAAGCACTTCGGGAATCGGAGCGGGCCGATGCAAAAGTCCGATAAACATAACGATCTTTGTAGAACCATCGGCCAGCTATTTC
82▶ E G I D L V E P I P L A A S A A F H R Y V Y R D K Y F G D A C S N V

Tth111I (5481)
5401 CCGCAGGACATATCCACGCCCTCTACATCGAAGCTGAAAGCACGAGATTCTTCCGCTCCGAGAGCTGCATCAGGTCGGAGACGCTGCGAACTTTTCG
49▶ R L V Y G R G G V D F S F A R S E E G E S L Q M L D S V S D F K E

PshAI (5514) NruI (5510) BspHI (5543)
5501 ATCAGAACTTCGCGACAGAGCTCGGGTGGAGTTCAGGCTTTTTCATGATGGCCCTCTATAGTGAGTCGATTATACATATGCCGATATACTATGCCGAT
15▶ I L F K A V S T A T L E P K K M
5601 GATTAATTGTCAAACAGCGTGGATGGCTCCAGCTTATCTGACGGTTCATAACGAGCTCTGTTATATAGACCTCCACCGTACAGCCTACCGC

SpeI (5756)
5701 CCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAAACAACTCCATTGACGTCAATGGGGTGGAGACT

SnaBI (5884)
5801 TGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGC

NdeI (5989)
5901 CAAGTAGGAAAGTCCCATAAGGTGATGACTGGGCATAATGCCAGGCGGGCATTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACA
6001 CTTGATGACTGCCAAGTGGGAGTTTACCCTAAATACTCCACCATTGACGTCAATGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTATTAT

PstI (6168) SdaI (6167)
6101 TGACGTCAATGGGGCGGGTCTGTTGGGCGGTCAGCCAGGCGGGCATTACCGTAAGTTATGTAACGCTGCAGGTTAATTAAAGACATGTGAGCAAAAG

6201 GCCAGCAAAAGGCCAGGAACCGTAAAAGGCCGCTGTTGCTGGGTTTTTCCATAGGCTCCGCCCTGACGAGCATCAAAAATCGACGCTCAAGTCA

6301 GAGTGGCGAAACCCGACAGACTATAAGATACCAGCGTTTTCCCTGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCTGCCGTTACCGGATAC

6401 CTGTCGCTTTCTCCCTTGGGAAGCGTGGCGTTTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCGGGTAGGTGTTGCTCCAAGCTGGGCTGTG

6501 TGCACGAACCCCGTTGAGCCGACCGCTGCGCTTATCCGGTAACTATCGTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGAACAGATTTGGT

6601 CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGAACAGATTTGGT

6701 ATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGTATCGGCAAAACAAACCGCTGGTAGCGGTGTTTTTTGTTTGA
6801 AGCAGCAGATTACCGCGAGAAAAAGGATCTCAAGAAGATCTTTGATCTTTTACGGGTCTGACGCTCAGTGAACGAAAACCTACGTTAAGGGAT
6901 TTTGGTCATGGCTAGTTAATTAACATTAATCA