



NotI (-1)
1 GCGGCGCCAATAAAATATCTTTATTTTTCATTACATCTGTGTGTTGGTTTTTGTGTGAATCGTAACATAACGCTCTCCATCAAAACAAAACGAAACA
PvuI (172)
SgfI (171)
101 AAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAAGGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCA

AgeI (254)EcoNI (261)
201 GAGCGCACATCGCCACAGTCCCCGAGAAGTTGGGGGAGGGTTCGCAATTGAACCGGTGCTAGAGAAGTGGCGGGGTAAACTGGGAAAGTGATG

Psp1406I (368)
301 TCGTGTACTGGCTCCGCTTTTTCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAG

HindIII (410) Bsu36I (456)
EcoNI (452)
401 AACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTACGCGCCCGCCCTACCTGAGCGCCATCCACGCCGTTGAGTCCGCTTCTGCCGCT

501 CCGCCTGTGGTCCCTCTGAAGTCCGCTCTAGGTAAGTTAAAGCTCAGGTCGAGACCGGCCCTTTGTCCGGGCTCCCTTGGAGCTACCTA

NaeI (606)
601 GACTCAGCGGCTCTCCAGCTTTGCTGACCCTGCTTCTCAACTCTACGCTTTTGTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACC

KasI (700) AgeI (717)
701 GCGGCTACCTGAGATCACCGGTAGGAGGCCATCATGAGACAGACTTTGCCTGTACTCTACTTTTTGGGGGGCCTTTTGCCTTTGGGATGCTGTGTGC
1 M R Q T L P C I Y F W G G L L P F G M L C A
Acc65I (866)
801 ATCTCCACCACCAAGTGCCTGTTAGCCATGAAGTTGCTGACTGCAGCCACTGAAGTTGACTCAGGTACCCGATGATCTACCCACAACATAACAGTG
22 S S T T K C T V S H E V A D C S H L K L T Q V P D D L P T N I T V
XcmI (991)
901 TTGAACCTTACCATAATCAACTCAGAAGATTACCAGCCGCAACTTCAAGGATATAGCCAGCTAACTAGCTTGGATGAGGATTTAACACCATCTCAA
56 L N L T H N Q L R R L P A A N F T R Y S Q L T S L D V G F N T I S
1001 AACTGGAGCCAGAATTGTCCAGAACTCCCATGTTAAAAGTTTTGAACCTCCAGCACAATGAGCTATCTCAACTTTCTGATAAACTTTGCCTTCTG
89 K L E P E L C Q K L P M L K V L N L Q H N E L S Q L S D K T F A F C
BglII (1190)
1101 CACGAATTTGACTGAAGTCCATCTCATGTCCAACCTCAATCCAGAAAATTAATAATCCCTTTGTCAAGCAGAAGAATTTAATCACATTAGATCTGTCT
122 T N L T E L H L M S N S I Q K I K N N P F V K Q K N L I T L D L S
Eco47III (1286)
AfeI (1286)
1201 CATAATGGCTTGTACATCAAAAATAGGAACTCAGGTTCAAGTGGAAAATCTCCAAGAGCTTCTATTATCAAAACAATAAAATCAAGCGCTAAAAGATG
156 H N G L S S T K L G T Q V Q L E N L Q E L L L S N N K I Q A L K S
EcoRV (1308)
1301 AAGAAGTGGATATCTTTGCCAATTCATCTTTAAAAAATAGAGTTGTCATCGAATCAAAATAAAGAGTTTTCTCCAGGGTGTTCACGCAATTTGGAAG
189 E E L D I F A N S S L K L E L S S N Q I K E F S P G C F H A I G R
SmaI (1433)
1401 ATTATTTGGCCTCTTTCTGAACAATGTCCAGCTGGGTCCAGCCTTACAGAGAAGCTATGTTTGAATTAGCAAAACAAGCATTGGAATCTGTCTCTG
222 L F G L F L N N V Q L G P T L E K L C L E L A N T S I R N L S L
1501 AGTAACAGCCAGCTGTCCACCACAGCAATACTTTCTGGACTAAAGTGGCAAAATCTCACTATGCTCGATCTTTCTACAACAATTTAAATGTGG
256 S N S Q L S T T S N T T F L G L K W T N L T M L D L S Y N N L N V
1601 TTGGTAACGATCTTTGCTTGGCTTCCACAACCTAGAAATTTCTCTAGAGTATAATAATACAGCATTGTTTTCTCACTTTTGACGGGCTTTT
289 V G N D S F L A W L P Q L E Y F F L E Y N N I Q H L F S H L C H G L F
Acc65I (1708)
1701 CAATGTGAGGTACCTGAAATTTGAAACGGCTTTTACTAAACAAGTATTTCCCTTGCCTCACTCCCAAGATTGATGATTTTTCTTTTCAGTGGCTAAAA
322 N V R Y L N L K R S F T K Q S I S L A S L P K I D D F S F Q W L K
1801 TGTTTGGAGCACCTTAACATGGAAGATAATGATATCCAGGCATAAAAAGCAATATGTTTACAGGATTGATAAACCTGAAATACTTAAGTCTATCCAAC
356 C L E H L N M E D N D I P G I K S N M F T G L I N L K Y L S L S N
1901 CCTTACAAGTTTGAAGATTTGACAAAATGATGATGCTTCCACTTGTCTTCCCTTACACATACTCAACCTAACCAAGAATAAAATCTCAAAAAT
389 S F T A S L R T L T N E T F V S L A H S P L H I L N L T K N K I S K I
ScaI (2036) XbaI (2096)
2001 AGAGAGTGATGCTTTCTCTGGTTGGGCCACTAGAAGTACTTGACCTGGCCCTTAATGAAATTTGGCAAGAACTCACAGGCCAGGAATGGAGAGGCTA
422 E S D A F S W L G H L E V L D L G L N E I G Q E L T G Q E W R G L
2101 GAAAATATTTTTGAAATCTATCTTTCTACAACAAGTACCTGCAGCTGACTAGGAACTCCTTTGCCTTGGTCCCAAGCCTTCAACGACTGATGCTCCGAA
456 E N I F E I Y L S Y N K Y L Q L T R N S F A L V P S L Q R L M L R
2201 GGGTGGCCCTTAAAAATGTGGATAGCTCTTCCTTCCACTTCCAGCCTCTTCTGAACTTGACCAATCTGATCTGATCTAAGCAACAACAATGAAACATAAA
489 R V A L K N V D S S P S P F Q P L R N L T I L D L S N N N I A N I N
BspLU11I (2305)
2301 TGATGACATGTTGGAGGCTTGGAGAACTAGAAATTTCTGATTTGACAGCATAACAACCTAGCACGGCTCTGAAACACGCAAAACCTGGTGGTCCCATT
522 D D M L E G L E K L E I L D L Q H N N L A R L W K H A N P G G P I
2401 TATTTCTAAAGGCTGTCTCACCTCCACATCCTTAACCTGGAGTCCAACGGCTTTGACGAGATCCAGTGTGAGGCTTCAAGGATTTATTTGAACTAA
556 Y F L K G L S H L H I L N L E S N G F D E I P V E V F K D L F E L
ClaI (2505)
2501 AGATCATCGATTTAGGATTTGAATAATTAACACACTTCCAGCATCTGTCTTTAATAATCAGGTGTCTCTAAAGTCAATGAACTTCCAGAAGAATCTCAT
589 K I I D L G L N N L N T L P A S V F N N Q V S L K S L N L Q K N L I
XmnI (2615) PmlI (2681)
2601 AACATCCGTTGAGAAGAAGGTTTTCCGGCCAGCTTTGAGAACCTGACTGAGTTAGATATGCGCTTTAATCCCTTTGATTGACAGTGTGAAAGTATTGCC
622 T S V E K K V F G P A F R N L T E L D M R F N P F D C T C E S I A
2701 TGGTTTGTAAATGGATTAACGAGACCCATACCAACATCCCTGAGCTGTCAAGCCACTACCTTTGCAACACTCCACCTCACTATCATGGGTTCCAGTGA
656 W F V N W I N E T H T N I P E L S S H Y L C N T P P H Y H G F P V
BsiBI (2876)
2801 GACTTTTTGATACATCATCTTGCAAGACAGTGCCCCCTTTGAACTCTTTTTCATGATCAATACAGATCTCTGTTGATTTTTATCTTTATTGACTTCT
689 R L F D T S S C K D S A P F E L F F M I N T S I L L I F I F I V L L
EcoRV (2921)
2901 CATCCACTTTGAGGGCTGGAGGATATCTTTTATTGGAATGTTTCACTACATCGAGTTCTTGGTTTCAAAGAAATAGACAGACAGACAGAACAGTTTGA
722 I H F E G W R I S F Y W N V S V H R V L G F K E I D R Q T E Q F E
3001 TATGAGCATATATAATTCATGCCTATAAAGATAAGGATGGGTTGGAACATTTCTCTCAATGAAAAGGAAGCAATCTCTCAAAATTTGTCTGG
756 Y A A Y I I H A Y K D K D W V W E H F S S M E K E D Q S L K F C L
HpaI (3144)
3101 AAGAAAGGACTTTGAGGCGGGTGTTTTTGAACTAGAAGCAATTTGTAACAGCATCAAAAGAAAGCAGAAAATATTTTTGTTATAACACACCATCTATT
789 E E R D F E A G V F E L E A I V N S I K R S R K I I F V I T H H L L
BstXI (3271)
3201 AAAAGACCCATTATGCAAAAGATTCAAGGTACATCATGAGTTCAACAAGCTATTGAACAAAATCTGGATTCATTATATTGGTTTTCTTGGAGGATT
822 K D P L C K R F K V H H A V Q Q A I E Q N L D S I I L V F L E E I

3301 CCAGATTATAAACTGAACCATGCACCTCTGTTTGCAGGAGGAATGTTTAAATCTCACTGCATCTTGAAGTCCAGTTCAGAAAGAACGGATAGGTGCCT
856 P D Y K L N H A L C L R R G M F K S H C I L N W P V Q K E R I G A MscI (3369)

3401 TTCGTCATAAATTGCAAGTAGCACTTGGATCCAAAACTCTGTACATTAATTTTAAATATTCATTAAGCAAAGGAGAACTTCGCTAGCTGGCCA
889 F R H K L Q V A L G S K N S V H • BamHI (3426) BsrGI (3440) MscI (3494) NheI (3488)

3501 GACATGATAAGATACATTGATGAGTTTGGACAAACCACAAC TAGAATGCAGTGAAAAAATGCTTTATTGTGAAATTTGTGATGCTATTGCTTTATTGT
HpaI (3626)

3601 TAACCATTATAAGCTGCAATAAACAGTTAACACAACAATTGCATTCTTTATGTTTCAGGTTCAAGGGGAGGTGTGGGAGTTTTTAAAGCAAGTA
3701 AAACCTCTACAAATGTGGTATGGAATCTAAAATACAGCATAGCAAACCTTAACTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATA
3801 AGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTCATGGAGTTAAGATATAGTGATTTTCCCAAGGTTTGA
3901 ACTAGCTCTTCATTTCTTTATGTTTAAATGCACCTGCCACATCCCTTTTTAGTAAATATTAGAAATAATTTAAATACATCATTGCAATGAAA
4001 ATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGAACCTTTAATAGA
ScaI (4193)

4101 AATTGGACAGCAAGAAAGCGAGCTTAGCGAATTCTCGACTCATTCTTTGCCCTCGGACGAGTGCTGGGGCGTGGTTTTCCACTATCGCGAGTACTT
342 • E K A R P R T S P R R N G S D A L V E
4201 CTACACAGCCATCGGTCCAGACGGCCGCTTCTCGGGGCGATTGTGTACGCCGACAGTCCCGGCTCGGATCGGACGATTGCGTCGCATCGACCCTG
322 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 I A D C R G Q SacII (4384)

4301 CGCCCAAGCTGCATCATGAAATGCGGTCAACCAAGCTCTGATAGAGTTGGTCAAGACCAATCGCGAGCATATACGCCGGAGCCGCGGATCCTGCA
289 A W A A D D F 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 BsrBI (4414)

4401 AGCTCCGGATGCCTCCGCTCGAAGTAGCGCTGCTGCTCCATACAAGCAACCCAGCCCTCCAGAAGAAGATGTTGGCGACCTCGTATTGGGAATCCC
255 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 V E Y Q S D G
4501 CGAATCGCCTCGCTCCAGTCAATGACCGCTGTTATGCGGCCATTGTCGCTCAGGACATTGTTGGAGCCGAAATCCGCTGCACGAGGTGCGCGACTTC
222 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 H V L H R V E Tth111I (4657)

4601 GGGGCGTCTCGGCCAAAGCATCAGCTCATCGAGAGCTGCGCGACGGACGACTGACGGTGTCTGCTCCATCACAGTTTCCAGTGATACACATGGGGA
189 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 W H Y V H P NdeI (4712) RsrII (4753) BsrBI (4772) SgfI (4799)

4701 TCAGCAATCGCGCATATGAAATACCGCATGTAGTGTATTGACCGATTCTTTCGGTCCGAATGGCCGAACCCGCTCGTCTGGCTAAGATCGGCCGCA
155 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 Q S L D A A A PvuI (4800) SaeI (4813)

4801 CGATCGCATCCATGAGCTCCGCGAGGTTGCGAAGCAGCGGCGAGTTCCGTTTCAGGACAGTCTTGAACGTGACACCCCTGTGCACGGCGGGAGATGCA
122 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 Q A R R S I C
4901 ATAGTCAAGCTCTCGTGAATCCCAATGTCAAGCACTTCCGGAATCGGGAGCGGGCCGATGCAAAAGTCCGATAAACATAACGATCTTTGTAGAAA
89 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 Y R D K Y F
5001 CCATCGGCGAGCTATTTACCGCAGGACATCCACGCCCTCTACATCGAAGTGAAGCACGAGATTCTTCCGCTCGAGAGTGCATCAGGTCGG
55 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 S L Q M L D S PshAI (5134)

5101 AGACGCTGTGCAACTTTTCGATCAGAAACTTTCGCGACAGAGCTGCGGGTGAAGTTCAGGCTTTTTCATGATGGCCCTCTATAGTGAGTCTGATTACTA
22 V S D F K E I L F K A V S T A T L E P K K M Tth111I (5101) NruI (5130)

5201 TGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCATAACGAGCTCTGCTTATATAGACCTC
AseI (5221) SacI (5278)

5301 CCACCGTACACGCCTACCGCCATTTGCGTCAATGGGCGGAGTTGTTACGACATTTTGAAAGTCCCGTTGATTACTAGTCAAACAAACTCCATTG
5401 ACGTCAATGGGGTGGAGACTTGGAAATCCCGTGAGTCAAACCGCTATCCACGCCATTGATGACTGCCAAACCGCATCATCATGTAATAGCGATGA
SpeI (5376)

5501 CTAATACGTAGATGACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGCGGGCCATTACCGTCAATGACGTAATAGGGGG
SnaBI (5504)

5601 CGTACTTGGCATATGATACACTTGTACTGCAAGTGGGAGTTTACCCTAAATACCACCCATTGACGTAATGGAAAGTCCCTATTGGCGTTACT
NdeI (5609)

5701 ATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGCTGTTGGGCGGTGAGCCAGCGGGCCATTACCCTAAGTTATGTAACGCTGCAGGTTAAAT
SdaI (5787)

5801 TAAGAATGTTGAGCAAAAGCCAGCAAAAGCCAGGAACCGTAAAAAGCCGTTGCTGGGTTTTTCCATAGGCTCGCCCCCTGACGAGCATCAC
BspLU11I (5805)

5901 AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATACCAGCGTTTCCCTGGAAGCTCCCTCGTGCCTCTCTGTTCCGA
6001 CCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCCGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCCGGTGTAGGTCGT
6101 TCGCTCAAAGCTGGGCTGTGTGCACGAACCCCGTTACGCCGACCGCTGCGCTTATCCGGTAACTATCGTCTTGTAGTCCAACCCGGTAAGCACGAC
6201 TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCTAACTACGGCTACA
6301 CTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAGTTGGTAGCTCTTGTGATCGGCAACAAACCCGCTGGTAG
6401 CGGTGTTTTTTTTTTGTTGCAAGCAGCAGATTACGCGAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAAC
PaeI (6535)

6501 GAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCA