



1 GGATCTGCATCGCTCCGGTGCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTGATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCCGGTTGAGTCGCGTTTGCGCCCTCCCGCTGTGGTGCTCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

BspHI (560)

AgeI (552)

501 TCTGTTCTGCGCGTTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGTCATGAGTTCCTCGCTGTTAATGTA AAAAAGCTGAAGGT
1 M S S S P V N V K K L K V
601 GTCGGAGCTGAAAGAGGAGCTCAAGAAGCGACGCCTTCTGACAAGGGTCTCAAGGCCGAGCTCATGGAGCGACTCCAGGCTGCGCTGGACGACGAGGAG
13 S E L K E E L K K R R L S D K G L K A E L M E R L Q A A L D D E E
701 GCCGGGGCCGCCCGCCATGGAGCCCGGAACGGCAGCCTAGACCTGGCGGGGATTCCGCTGGGCGCTCGGAGCAGGCCCTCGAGCAGGAGGCCCGG
47 A G G R P A M E P G N G S L D L G G D S A G R S G A G L E Q E A A
801 CCGCGCGCATGAAGAGGAGGAGGAAGAGGAAGAGGAGGAGGAAGAACTCCGCTCTGGACGGCGACCAGATGGAGCTAGGAGAGGAGAACGGGGCCG
80 A G G D E E E E E E E E E E E E G I S A L D G D Q M E L G E E N G A A
901 GGGGGCGGCGACTCGGGCCCGATGGAGGAGGAGGAGCCGCTCGGAAGACGAGAACGGCAGCAGGATCAGGGTTCCAGGAAGGGGAAGATGAGCTCGG
113 G A A D S G P M E E E E A A S E D E N G D D Q G F Q E G E D E L G
1001 GACGAAGAGGAAGGCGGGGCGACGAGAACGGGACGGGAGCAGCAGCCTCAACCGCGGCGACGACGAGCAACAGCCCAACAGCAGCGCGGGCCG
147 D E E E G A G D E N G H G E Q Q P Q P P A T Q Q Q Q P Q Q Q R G A
1101 CCAAGGAGGCCGCGGGGAAGAGCAGCGCCCACTCGCTGTTGCGGTGACGGTGGCGCCCGGGGCGAGGAGGCGCAGCAGGCGGGAGGTAA
180 A K E A A G K S S G P T S L F A V T V A P P G A R Q G Q Q Q A G G K
1201 GAAGAAGCGGAAGGCGGCGGAGGCGCGGTGCGCCCGGGGCTCCGGCGGGGGGACGGCAAAACAGAAAGCGGAGATAAAAAGAGGGGTGT
213 K K A E G G G G G R P G A P A A G D G K T E Q K G G D K K R G V
1301 AAAAGACACGAGAAGATCATGGCCGTGGATTTTTGAGTACATTGAAGAGAACAAGTATAGCAGAGCAAATCTCCTCAGCCACTGTTGAAGAAGAA
247 K R P R E D H G R G Y F E Y I E E N K Y S R A K S P Q P P V E E E
1401 ATGAACACTTCGATGACACAGTGGTTTGTCTTGATACTTATAATTGTGATCTACATTTTAAATATCAAGAGATCGTCTCAGTGCTTCTCCCTTACAAT
280 D E H F D D T V V C L D T Y N C D L H F K I S R D R L S A S S L T M
1501 GGAGAGTTTTGCTTTTCTTTGGGCTGGAGGAAGAGCATCCTATGGTGTGCAAAAGGCAAAGTGTGTTTTGAGATGAAGGTTACAGAGAAGATCCAGTA
313 E S F A F L W A G G R A S Y G V S K G K V C F E M K V T E K I P V
1601 AGGCATTTATACAAAAGATATTGACATACATGAAGTTGCTATTGGCTGGTCACTAAGTGAAGTGTACTTGGTGAAGAAGAATTTTCTTATG
347 R H L Y T K A D I D I H E V R I G W S L T T S G M L L G E E E F S Y
1701 GGTATTCTAAAAGGAATAAAAACATGCAACTGTGAGACTGAAGATTATGGAGAAAAGTTTATGATAAATGATGATTACATGTTTTGCTAATTTGA
380 G Y S L K G I K T C N C E T E D Y G E K F D E N D V I T C F A N F E
1801 AAGTGATGAAGTAGAAGTCTCGTATGCTAAGAAATGGACAAGATCTTGGCGTTGCCCTCAAATCAGTAAGGAAGTCTTGGTGGACGGCCACTGTTCCG
413 S D E V E L S Y A K N G Q D L G V A F K I S K E V L A G R P L F P
1901 CATGTTCTCGCCACAAGTGTGAGTGAATTTAATTTTGGTGAAGGAAAAGCCATATTTTCCAATACCTGAAGAGTATACTTTTATCCAGAAGCTCC
447 H V L C H N C A V E F N F G Q K E K P Y F P I P E E Y T F I Q N V
2001 CCTTAGAGGATCGAGTTAGAGGACCAAGGGGCTGAAGAGAAGAAAGATTGTAAGTTGTGATGATGATTGGCTTCCAGGAGTGGAAAACTACCTG
480 P L Q E D R V R G P K G P E K K D C E V V M M I G L P G A G K T T W
2101 GGTACTAAACATGCAGCAGAAAATCCAGGAAAATATAACATTTCTTGGCACAATACTATTATGGATAAGATGATGGTGGCAGGTTTTAAGAAGCAAATG
513 V T K H A A E N P G K Y N I L G T N T I M D K M M V A G F K K Q M
2201 GCAGATACTGAAAAGTGAACACACTGTTGCAGAGAGCCCCAGTGTCTTGGGAAAATTTATTGAGATTGCTGCCGAAAAGAGCAAATTTTATTCTGG
547 A D T G K L N T L L Q R A P Q C L G K F I E I A A R K K R N F I L
2301 ATCAGACAAAATGTGCTGCTGCTGCCAGAGGAGAAAATGTGCTGTTTGCAGGCTTCCAGCGAAAAGCTGTTGTAGTTTGCCTAAAAGATGAAGACTA
580 D Q T N V S A A A Q R R K M C L F A G F Q R K A V V V C P K D E D Y
2401 TAAGCAAAGAACAGAAAGAAAGCAGAAGTAGAGGGGAAAGACCTACCAGAACATGCGGTCTCAAATGAAAGGAACTTTACCCTCCAGAGGTAGCT
613 K Q R T Q K K A E V E G K D L P E H A V L K M K G N F T L P E V A
2501 GAGTGCTTTGATGAAATAACCTATGTTGAACCTCAGAAGGAAGAAGCCAAAAACTCTGGAGCAATATAAGGAAGAAAGCAAAGGCTCTCCACCAG
647 E C F D E I T Y V E L Q K E E A Q K L L E Q Y K E E S K K A L P P
2601 AAAAGAAAACAGAACTGGCTCAAAGAAAAGCAATAAAAATAAGAGTGGAAGAACCAGTTTAAACAGAGGTGGTGGCCATAGAGGACGTGGAGGATCAA
680 E K K Q N T G S K K S N K N K S G K N Q F N R G G G H R G R G G F N
2701 TATGCGTGGTGGAAATTTAGAGGAGGAGCCCTGGGAATCGTGGCGGATATAATAGGAGGGGCAACATGCCACAGAGAGGTGGTGGCGGTGGAGGAAGT
713 M R G G N F R G G A P G P G N R G G Y N R R G N M P Q R G G G G G S
2801 GGTGAATCGCTATCCATACCTCGTCCCTGTTTTCTGGCGGTGGTACTCAAACAGAGGGAACATAACAGAGGTGGAATGCCAACAGAG
747 G G I G Y P Y P R A P V F P G R G S Y S N R G N Y N R G N Y N R
2901 GGAACACAAACAGAACTTACAGAGGACGAGGAAACAATCGTGGCTACAAAATCAATCTCAGGGCTACAACAGTGGCAGCAGGGTCAATTCTGGGTCA
780 G N Y N Q N F R G R G N N R G Y K N Q S Q G Y N Q W Q Q G Q F W G Q

NheI (3062)

3001 GAAGCCATGGAGTCAGCATTATACCAAGGATATTATTGAATACCCAAATAAAAACGAAGTCTAGCTGGCCAGACATGATAAGATACATTGATGAGTT
813 K P W S Q H Y H Q G Y Y •
3101 TGGACAAACCACAACCTAGAATGCAGTGA AAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAA

3201 GTTAAACAACAATTGCATTCATTTTATGTTTCAGGTTCAAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAAT
TCTAAAATACAGCATAGCAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCC
3301
AATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTAAGATATAGTGTATTTTCCCAAGGTTTGAAGTACTCTTCATTTCTTTATGTTTT
3401
AAATGCACTGACCTCCACATTCCTTTTATGATAAATATTAGAAATATTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAAATC
3501
CAGATGCTCAAGGCCCTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAGGAACCTTAAATAGAAATTGGACAGCAAGAAAGCGAGCTTC
3601
TAGCTTTAGTTCCTGGTACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTCCATTCAATGAGCACAAAGCAGTCAGGAGCAT
3701
AGTCAGAGATGAGCTCTGCACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCTGACAGCCACAATGGTGC
3801
AAAGTCTTCTGCCCGTGTCTCACAGCAGACCAATGGCAATGGCTTTCAGCACAGACAGTACCCTGCAATGTAGGCCTCAATGTGGACAGCAGAGATG
3901
ATCTCCCAGTCTTGGTCTGATGGCCGCCGACATGGTGTGTTGTCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGAT
4001
CCTGCTGAGAGATGTTGAAGTCTTCAATGGTGGCCCTCTATAGTGTGCTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGC
4101
GTGGATGGCGTCTCCAGCTTATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCACCCTACACGCTACCGCCATTGCGTCAATGGGGC
4201
GGAGTTGTTACGACATTTTGAAAGTCCCGTTGATTTACTAGTCAAAACAACCTCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCA
4301
AACCGTATCCACGCCATTGATGTAAGTCCGTAAGTATGTAACGCTGAGGTTAATAAGCATGACTAATACGTAGATGACTGCCAAGTAGAAAGTCCATA
4401
AGGTCATGTAAGTGGCATAATGCCAGGCGGGCATTACCCTGATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGTACTGCAAGTG
4501
GGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGAAAGTCCCTATTGGCGTACTATGGAACATACGTCAATATTGACGTCAATGGGCGGGG
4601
TCGTTGGGCGGTGAGCCAGGCGGGCATTACCCTGTAAGTATGTAACGCTGAGGTTAATAAGCATGAGCAAAAGGCCAGCAAAAGGCCAGGAA
4701
CCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGCCCTGACGAGCATCAGAAAAATCGACGCTCAAGTCAAGGTTGGCGAAACCCGACA
4801
GGACTATAAAGATACCAGGCGTTTCCCCTGGAAGTCCCTCGTGCCTCTCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTT
4901
CGGGAAGCGTGGCGTTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGTGTGCACGAACCCCCGTCA
5001
GCCCAGCGCTGCGCTTATCCGTAATATCGTCTTGTAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
5101
AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGC
5201
CAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGTATCCGCAAAACAACCCAGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAG
5301
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGATTTTGGTATGGCTAGTTAA
5401
TTAACATTTAAATCAGCGGCCGAATAAAATATCTTTATTTTTCATTACATCTGTGTGTTGGTTTTTTGTGTAATCGTAACTAACATACGCTCTCCATCA
5501
AAACAAAACGAAACAAAACAACCTAGCAAAATAGGCTGTCCCAGTCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA
5601