



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTCTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTCTGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

NcoI (560)
BstEII (555)
AgeI (552)

501 TCTGTTCTGCGCCGTACAGATCCAAGCTGTGACCGGCGCTACCTGAGATCACCGGTACCATGGATGATGGGCATCAAGAGTCAGCTCTGTACGATGG
601 GCACTACGAGGGAGATTTCTGGCTCTTCAACAATTCAGTGATAACAGCCAGGAGAACAACGCTTCTAAAGTTCAAGGAGGCTTTTTGCCTGTGTG
13▶ H Y E G D F W L F N N S S D N S Q E N K R F L K F K E V F L P C V
701 TACCTGGTAGTGTGTTGCTTTGGACTGCTAGGAACTCCCTGGTCTGATTATATACATTTTCTACCAGAAGCTGAGGACTCTGACAGATGTGTTTCTGC
47▶ Y L V V F V F G L L G N S L V L I I Y I F Y Q K L R T L T D V F L
801 TGAAGTTGCCCTGGCTGACCTGGTGTGCTGACTCTGCCCTTTGGCCTATGCAGGCACCTATGAGTGGGTCTTTGGCACAGTCATGTGCAAAAC
80▶ L N L P L A D L V F V C T L P F W A Y A G T Y E W V F G T V M C K T
901 TCTTCGAGGCATGTATAAATGAAGTCTACGTCATGCTACTCTCACCTGCATCACAGTGGATGTTTCATTGTAGTGGTCCAGGCTACCAAGGCC
113▶ L R G M Y T M N F Y V S M L T L T C I T V D R F I V V V Q A T K A
1001 TTCAACGGCAGGCTAAGTGAAGATCTGGGGCAAGTCATTTGCTTGTCTCATTTGGGTGGTCTCCCTGTTGGTTTCTTTGCCACAGATCATCTATGGCC
147▶ F N R Q A K W K I W G Q V I C L L I W V V S L L V S L P Q I I Y G
1101 ATGTTCAAGATATTGACAAGCTTATCTGTGACTACCAGTGGAGATATCCACTATGGTCTTGTATACAGATGACTCTGGGGTCTTCTGCTGCCATT
180▶ H V Q D I D K L I C Q Y H S E E I S T M V L V I Q M T L G F F L P L
1201 GCTCACTATGATTCTGTGCTACTCAGGCATTATCAAGACCTTGCTTGTGCTGAACTTCCAGAAGCACAATCTCTAAAGATCATCTTCTTGTAGTG
213▶ L T M I L C Y S G I I K T L L H A R N F Q K H K S L K I I F L V V
1301 GCTGTGTTCTGTGACCCAGACACCCTTCAACCTTGCATGTTAATCAAAGTACAAGCTGGGAGTACTATACCATAACCAGCTTAAAGTATGCCATCG
247▶ A V F L L T Q T P F N L A M L I Q S T S W E Y Y T I T S F K Y A I
1401 TAGTGACAGAGGCTATAGCATACTTTCCGGCTTGCCTTAAACCTGTACTTTATGCCTTTGTTGGCTTAAAGTCCGGAAGAAGCTCTGAAACTTATGAA
280▶ V V T T E A I A Y F R A C L N P V L Y A F V G L K F R K N V W K L M K
1501 GGATATCGGCTGCTCTCACCTGGGAGTCTCAAGTCAATGGAAGTCTTCTGAGGACAGTCCAAGACTTGTCTGCCTCCACAATGTAGAGACCACC
313▶ D I G C L S H L G V S S Q W K S S E D S S K T C S A S H N V E T T
1601 AGTATGTTCAATTGTAGTAGGCCTTGCACACTTAGAGAAGTTAATAACAGAATTCTAGGAGCATGGCTGTATCATTTGGATGCAACAAGAAAAGCTTT
347▶ S M F Q L •

NheI (1748)

1701 GCTTATAGCATGTGGAGTATCATGGAGAAGTCACTGAACACCATGGCTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAA
1801 CTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAA
1901 TTGCATTCATTTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAAACCTTACAAATGTGGTATGGAATTCTAAAATACAGCA
2001 TAGCAAACTTTAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCT
2101 GTTTGCAGCCTCACCTTCTTTTATGAGTAAAGATATAGTGTATTTTCCCAAGGTTTGAAGTACTCTTCTTCTTTATGTTTAAATGCACTGACCT
2201 CCCACATTCCTTTTTAGTAAAATATTCAGAAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGC
2301 CCTCATAATATCCCCAGTTTAGTAGTTGACTTAGGGAACAAAGGAACCTTAAATAGAAATTTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCT
2401 GGTGACTTGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCATTCTCAATGAGCACAAGCAGTCAGGAGCATAGTCAGAGATGAGC
138▶ 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 Y D S I L
2501 TCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCC
104▶ 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 F D K T Q G
2601 CGTTGCTCACAGCAGACCCAATGGCAATGGCTTCCAGCAGACAGTACCCTGCCAATGTAGGCTCAATGTGGACAGCAGAGATGATCTCCCAGTCTT
71▶ 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 I E G T K
2701 GGTCTGATGGCCGCCCGACATGGTGTGTTGTCCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGAGATG
38▶ 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 D Q Q S I
2801 TTGAAGTCTTCAATGGTGGCCCTCTATAGTGAAGTCTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTC
4▶ N F T K M
2901 CAGCTTATCTGACGGTCACTAAACGAGCTCTGCTTATATAGACTCCACCGTACACGCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGAC
3001 ATTTTGGAAAGTCCCGTTGATTTACTAGTCAAAACAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCAGC

3101 CCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCAT AAGGCATGTACTGG
3201 GCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTA
3301 AATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTCA
3401 GCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
3501 CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATA
3601 CCAGGCGTTTCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCG
3701 CTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTAGCCCGACCGCTGCG
3801 CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGT
3901 AGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA
4001 AAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTC
4101 AAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATC
4201 AGCGGCCCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAACGAAAC
4301 AAAACAACTAGCAAAATAGGCTGTCCCAGTGAAGTGCAGGTGCCAGAACATTTCTCTATCGAA