



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGGTCGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGGTAACCTGGGAAAGTGATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCTCCCGCTGTGGTGCTCTGAAGCTGCGTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCGCTACCTAGACTCAGCCGGCTCTCCACGCTTTCCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

BstEII (555)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTCACCATGGCAGATTCCAACGGTACTATTACCGTTGAAGAGCT
1 M A D S N G T I T V E E L
601 TAAAAAGCTCCTTGAACAATGGAACCTAGTAATAGTTTCTATTCTTACATGGATTTGTCTTCTACAATTTGCCTATGCCAACAGGAATAGGTTTTTG
13 K K L L E Q W N L V I G F L F L T W I C L L Q F A Y A N R N R F L
701 TATATAATTAAGTTAATTTTCTCTGGCTGTTATGGCCAGTAACTTTAGCTTGTGTTTGTGCTGCTGCTGTTACAGAATAAATTGGATCACCGGTGGAA
47 Y I I K L I F L W L L W P V T L A C F V L A A V Y R I N W I T G G
801 TTGCTATCGCAATGGCTTGTCTTGTAGGCTTGTGTTGCTCAGTACTTATTGCTTCTTTCAGACTGTTTGCGCGTACGCGTTCATGTGGTCATTCAA
80 I A I A M A C L V G L M W L S Y F I A S F R L F A R T R S M W S F N
901 TCCAGAACTAACATTCTCAACGTGCCACTCCATGGCACTATTCTGACCAGCCGCTCTAGAAAGTGAAGTGAATCGGAGCTGTGATCCTTCGT
113 P E T N I L L N V P L H G T I L T R P L L E S E L V I G A V I L R
1001 GGACATCTTCGATTGCTGGACACCATCTAGGACGCTGTGACATCAAGGACCTGCCTAAAGAAATCACTGTTGCTACATCACGAACGCTTCTTATTACA
147 G H L R I A G H H L G R C D I K D L P K E I T V A T S R T L S Y Y
1101 AATTGGGAGCTTCGACGCTGTAGCAGGTGACTCAGGTTTTGTGCATACAGTCGCTACAGGATTGGCAACTATAAATTAACACAGACCATTCCAGTAG
180 K L G A S Q R V A G D S G F A A Y S R Y R I G N Y K L N T D H S S S

NheI (1232)

1201 CAGTGACAATATTGCTTTGCTTGTACAGTAAAGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAAGTGAATGCAGTGAAA
213 S D N I A L L V Q •
1301 AAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAAATTGCATTATTTTATG
1401 TTTACAGTTTACAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACCTTAAAC
1501 TCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCT
1601 TCTTTCATGGAGTTAAGATATAGTGTATTTTCCCAAGGTTTGAAGTACTGCTTTCATTTCTTATGTTTTAAATGCACTGACCTCCACATTCCCTTTTT
1701 AGTAAAATATTAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCC
1801 CAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTAAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGG
141 • N R T Y K L P
1901 ATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCATTCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTCTGCACATGCCAC
132 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 E R C M G C
2001 AGGGGCTGACCACCTGATGGATCTGTCCACCTCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCCGTTGCTCAGCAGAGA
99 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 Q G N S V A S
2101 CCAATGGCAATGGCTTCCAGCACAGACAGTACCCTGCCAATGTAGGCTCAATGTGGACAGCAGAGATGATCTCCCAAGTCTTGGTCTGATGGCCGCC
66 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 T R I A A
2201 CCGACATGGTGTGTTGCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGAAGGTCTTATGG
32 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 N F T K M
2301 TGGCCCTCCTATAGTGAGTCGTATTATACTATGCCATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGT
2401 TCACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCTACCGCCATTGCGTCAATGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCG
2501 TTGATTTACTAGTCAAACAACCTCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCGTGAGTCAAACCGCTATCCACGCCATTGATGACTGC
2601 CAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGACTGCCAAGTAGGAAAGTCCATAAGGTCATGTAAGTGGGCATAATGCCAGGGCG
2701 GCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGTGTAAGTGGGCGAGTTTACCGTAAATACTCCACCCATTG
2801 ACGTCAATGAAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGTCGTTGGCGGTGAGCCAGGCGGGCATT
2901 ACCGTAAGTTATGTAACGCTGACGGTTAATAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAAACCGTAAAAGGCCGCTTGTGGCGTTTTT
3001 CCATAGGCTCCGCCCCCTGACGAGCATCACAATAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCT
3101 GGAAGCTCCCTCGTGGCTCTCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGTTTCTCATAGCTCAC

3201 GCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTAGCCCGACCGCTGCGCCTTATCCGGTAACTA
3301 TCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAG
3401 TTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGATTGGTAGCT
3501 CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGAT
3601 CTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAA
3701 TATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAACTAGCAAA
3801 ATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA