



1 GGATCTGCATCGCTCCGGTCCCGTCACTGGGCGAGAGCGACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGTGCCTA  
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC  
201 GTGAACGTTCTTTTTCGCAACGGGTTTCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCGCCCTACCTGAGGCC  
301 GCCATCCACGCCGTTGAGTCCGCTTCCGCCCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC  
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTCCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

**BstEII (555)**  
**AgeI (552)**

501 TCTGTTCTGCGCGTTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGTCCAGTGGTGCAGCCAGTGTGACTTCCCCTGGCTCCCTAGA  
1 M A A A S V T S P G S L E  
601 ACTGCTACAGCCTGGCTTCTCCAAGACCTCCTGGGACCAGTTAGAAAGCAAGTACCTCTGTTACGCTGCAAAAACATCCTGCGGAGGCCTTTCCAG  
13 L L Q P G F S K T L L G T R L E A K Y L C S A C K N I L R R P F Q  
701 GCCCAGTGTGGCACCCTACTGCTCCTTCTGCCTGACCAGCATCCTCAGCTCTGGGCCCCAGAACTGTGCTGCTGTCTATGAAGCCTGTATGAAG  
47 A Q C G H R Y C S F C L T S I L S S G P Q N C A A C V Y E G L Y E  
801 AAGCATTCTATTTAGAGAGTAGTTCGGCCTTTCCAGATAACGCTGCCCGCAGAGAGGTGGAGAGCCTGCCAGTGTCTGTCCCAATGATGGATGCAC  
80 E G I S I L E S S S A F P D N A A R R E V E S L P A V C P N D G C T  
901 TTGGAAGGGGACCTTGAAGAATACGAGAGCTGCCACGAAGGACTTTGCCATTCTGCTGACGGAGTGCCTGCATGTAAGGCCTGGTCCGCTCAGC  
113 W K G T L K E Y E S C H E G L C P F L L T E C P A C K G L V R L S  
1001 GAGAAGGAGCACCACACTGAGCAGGAATGCCCCAAAAGGAGCCTGAGCTGCCAGCACTGCAGAGCACCTGTAGCCACGTGGACCTGGAGGTACACTATG  
147 E K E H H T E Q E C P K R S L S C Q H C R A P C S H V D L E V H Y  
1101 AGTCTGCCCCAAGTTTCCCTTAACCTGTGATGGCTGTGGCAAGAAGAAGATCCCTCGGGAGACGTTTTCAGGACCATGTTAGAGCATGCAGCAATGCCG  
180 E V C P K F P L T C D G C G K K K I P R E T F Q D H V R A C S K C R  
1201 GGTCTCTGCAGATTCACACCGTTGGCTGTTGAGAGTGGTGGAGACTGAGAACCTGCAGGATCATGAGTGCAGCGGTACGGGAACACCTAGCCCTA  
213 V L C R F H T V G C S E M V E T E N L Q D H E L Q R L R E H L A L  
1301 CTGCTGAGCTCATTCTGGAGGCCCAAGCCTCTCCAGAACCTTGAACCAGTGGGGCCAGAGCTACTCCAGCGGTGCCAGATTTGGAGCAGAAGATAG  
247 L L S S F L E A Q A S P G T L N Q V G P E L L Q R C Q I L E Q K I  
1401 CAACCTTTGAGAACATTGTCTGCGTCTTGAACCGTGAAGTAGAGAGGTAGCAGTGCAGAGGCTTGTAGCCGGCAGCACCGGTAGACCAGGACAA  
280 A T F E N I V C V L N R E V E R V A V T A E A C S R Q H R L D Q D K  
1501 GATTGAGCCCTGAGTAACAAGTGAACAGCTGGAGAGGAGCATCGCCCTCAAGGACCTGGCCATGGCTGACCTGGAGCAGAAGGTCTCCGAGTTGGAA  
313 I E A L S N K V Q Q L E R S I G L K D L A M A D L E Q K V S E L E  
1601 GTATCCACCTATGATGGGCTTCTCATCTGGAAGATCTGACTTACCAGAAAGCGTCAGGAAGCCGTAGCTGGCCGGACACCAGTATCTTCCCCAG  
347 V S T Y D G V F I W K I S D F T R K R Q E A V A G R T P A I F S P  
1701 CCTTACACAAGCAGATATGGCTACAAGATGTGCTACGAGTCACTTGAATGGCGACGGCACTGGGCGGGGAACCTATGTCTCTCTTCTCGTGGT  
380 A F Y T S R Y G Y K M C L R V Y L N G D G T G R G T H L S L F F V V  
1801 GATGAAAGGCCCAATGATGCTCTGTTGCAGTGGCCTTTAATCAGAAGGTAACATTGATGTTGCTGGACCATAACAACCGGGAGCATGTGATCGACGCA  
413 M K G P N D A L L Q W P F N Q K V T L M L L D H N N R E H V I D A  
1901 TTCAGGCCGATGTAACCTGCTCCTCTCCAGAGGCCTGTGAGTGCATGAACATGCCAGTGGCTGCCCTCTTCTGCCCTGTGTCCAAGATGGAGG  
447 F R P D V T S S S F Q R P V S D M N I A S G C P L F C P V S K M E  
2001 CCAAGAATTCCTATGTGCGGGATGATGCATCTTCAAAGCTATTGTGGACCTAACAGGACTTAGCCACCCTGCTAAGAATAGCAGCTCAGTGAGG  
480 A K N S Y V R D A I F I K A I V D L T G L •

**NheI (2122)**

2101 AGCTGTCACATTAGGCCAGCCAGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAATAAATGCTTT  
2201 ATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGTGCATAAACAAGTTAACAACAACAAATTGCATTATTTATGTTTCAGGTTCC  
2301 AGGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTTTAACCTCCAAATCAA  
2401 GCCTTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGG  
2501 AGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAAGTACTGCTTTCATTTCTTTATGTTTTAAATGCAGTGCCTCCACATCCCTTTTTAGTAAATAT  
2601 TCAGAAATAATTTAAATACATCATTGCAATGAAATAAATGTTTTTATTAGGCAGAAATCCAGATGCTCAAGGCCCTCATAATATCCCCAGTTTAGTA  
2701 GTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCTGGTGTACTTGAGGGGGATGAGTTCTCT  
141 • N R T Y K L P I L E E  
2801 CAATGGTGGTTTTGACCAGTTGCCATTCATCTCAATGAGCACAAAGCAGTCAAGGAGCATAGTCAGAGATGAGCTCTGACATGCCACAGGGGCTGAC  
129 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 P S V  
2901 CACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCCGTTGCTCAGAGCAGACCAATGGCA  
96 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 G I A  
3001 ATGGCTTCAAGCAGACAGTGCACCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCAAGTCTTGGTCTGATGGCCGCCCGACATGGT  
62 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 G V H H  
3101 GCTTGTGCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGAAGGCTTTCATGGTGGCCCTCT  
29 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

3201 ATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCCTAAACG  
3301 AGCTCTGCTTATATAGACCTCCACCGTACACGCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAAGTCCCGTTGATTTACT  
3401 AGTCAAAACAACTCCCATTGACGTCAATGGGGTGGAGACTTGAAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCA  
3501 TCATCATGGTAATAGCGATGACTAATACGTAGATGACTGCCAAGTAGGAAAGTCCCATAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACC  
3601 GTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGG  
3701 AAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGGCGGGGTCGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAGTT  
3801 ATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGCCAGCAAAAGCCAGGAACCGTAAAAAGCCGCTTGTGGCGTTTTTCCATAGGCTC  
3901 CGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGCTCCC  
4001 TCGTGGCTCTCCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTA  
4101 TCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTCAGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG  
4201 TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTCTTGAAGT  
4301 GGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG  
4401 CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG  
4501 GGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATT  
4601 TTCATTACATCTGTGTGTTGGTTTTTTGTGTAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAACTAGCAAAATAGGCTGTC  
4701 CCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA