



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGGTCGGCAATTGAACGGGTGCCTA  
101 GAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC  
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC  
301 GCCATCCACGCCGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGCTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC  
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCGCTACCTAGACTCAGCCGGCTCTCCACGCTTTCCTGACCCCTGCTTGTCAACTCTACGCTTTTGTTCGTTT

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

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGGCGCTACCTGAGATCACCGGTGGAGCCTCCAGGACCAGCAGCCACAGCCTCTGC  
1 M E P P G P S T P T A S A  
601 CGTGCCCGGGCAGATCACTACACCCAGGCCCTCCGGCCACTCCCGAAGCGCAGACTTCTATATAGCTTTGCGTTGCTGCTGCTGCTACAGGCTGTC  
13 A A R A D H Y T P G L R P L P K R R L L Y S F A L L L A V L Q A V  
701 TTTGTTCCAGTAACAGCTAACCCAGCCATAATCGTCCAGCTGGCCTACAGCGCCGGAGGAGGCCATCAAGAGGACCCTGTCTAGCAGGCCAGTACC  
47 F V P V T A N P A H N R P A G L Q R P E E S P S R G P C L A G Q Y  
801 TGTCAAAGGAACTGCAAGCCTTGCAAGAGGGTATTGACTACACAGCCATTCCAACCTTCTCTGGATTCATGTATTCTCTGCACAGTCTGTAAGGA  
80 L S E G N C K P C R E G I D Y T S H S N H S L D S C I L C T V C K E  
901 AGATAAAGTCGTAGAAAACCCGATGCAACATAACCACAAATACGGTGTGTCGATGCAAAACAGGCACCTTTGAAGATAAAGACTCCCTGAGATCTGCCAG  
113 D K V V E T R C N I T T N T V C R C K P G T F E D K D S P E I C Q  
1001 TCATGCTCTAACTGCACTGACGGGGAAGAGGAACTGACTTCTGTACCCCGAGAGAAAACCGAAGTGTGTCTCCAAAACGGTTGGGCATCTTGGCATA  
147 S C S N C T D G E E E L T S C T P R E N R K C V S K T A W A S W H  
1101 AGTAGGCTCTGGATAGGACTCTGGTTCAGTAGTCTGCTGATTGGAGCTCTGCTTGTCTGGAAGACTGGAGCATGGAGGCAATGTTGCTCTGTAT  
180 K L G L W I G L L V P V V L L I G A L L V W K T G A W R Q W L L C I  
1201 AAAAAGAGGCTGTGAACGGGATCCCGAAAGTGCGAACCTGTGTCATTGCTCTCTTTGGACCGACAGACATCTAGCACGACAAATGACTCTAACCCACAAC  
213 K R G C E R D P E S A N S V H S S L L D R Q T S S T T N D S N H N  
1301 ACGGAACCTGGCAAGACTCAGAAAACAGGAAAGTGTGCTTCCGGTAAACGAAACAGCTCAGCTGACGACCTGAAGTTTATCTTCAGTATTGTT  
247 T E P G K T Q K T G K K L L V P V N G N D S A D D L K F I F E Y C  
1401 CGGACATAGTGCCTTTGACTCCTGGAACCGTCTCATGCGGCAGTTGGGCCTCACAGACAATCAAATCAAATGGTCAAAGCCGAAACACTGGTCACAG  
280 S D I V P F D S W N R L M R Q L G L T D N Q I Q M V K A E T L V T R  
1501 TGAGGCCCTGTACCAAATGCTGCTCAAGTGGCGCCACAGACTGGCGAAGTGCCTCCATCAACCATCTGCTGGATGCCTTGAAGCCGTGGAAGAGAGA  
313 E A L Y Q M L L K W R H Q T G R S A S I N H L L D A L E A V E E R  
1601 GATGCCATGGAGAAAATTGAAGACTACGCGAGTAAATCCGGGAGGTTACTTATCAGAACGCTGCAGCCCAACAGAGACAGGGCCAGGAGGATCTCAGT  
347 D A M E K I E D Y A V K S G R F T Y Q N A A A Q P E T G P G G S Q

**NheI (1718)**

1701 GCGTTTGAAGTCAGCCTGGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTT  
380 C V •  
1801 GTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTATTTATGTTTCAGGTTACGGG  
1901 GGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATCTAAAATACAGCATAGCAAACTTTAACCTCAAATCAAGCCT  
2001 CTAATTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTT  
2101 TAAGATATAGTGTATTTTCCCAAGTGTGAACTAGCTCTTATTCTTTATGTTTTAAATGCACTGACCTCCACATTCCTTTTTAGTAAAATATTCAG  
2201 AAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTCATAATATCCCCAGTTTAGTAGTTG  
2301 GACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGGATGAGTTCCTCAAT  
141 • N R T Y K L P I L E E I  
2401 GGTGGTTTTGACCAGCTTGCATTCTCAATGAGCACAAGCAGTCAAGAGCATAGTCAAGATGAGCTCTGCACATGCCAGGGGCTGACCACC  
128 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 V  
2501 CTGATGGATCTGTCCACCTCATCAGAGTAGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCCGTTGCTCACAGCAGACCCAATGGCAATGG  
94 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 I A  
2601 CTTGACACAGACAGTACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCGAGTCTTGGTCTGATGGCCGCCCGGACATGGTGTCT  
61 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 K  
2701 GTTGTCTCATAGAGCATGGTGTCTCTCAGTGGCGACCTCCACCAGCTCCAGATCTGCTGAGAGATGTTGAAGGTCTTCATGGTGGCCCTCTATAG  
28 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  
2801 TGAGTCGATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACTAAACGAGCT  
2901 CTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCT  
3001 AAAACAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGACTGCCAAAACCGCATCAT  
3101 CATGGTAATAGCGATGACTAATACGTAGATGTACTGCAAGTAGGAAAGTCCCATAAAGTTCATGACTGGGCATAATGCCAGCGGGCCATTTACCGTCA

3201 TTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGAAAG  
3301 TCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGT  
3401 AACGCCTGCAGGTTAAITAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGCC  
3501 CCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGCTCCCTCGT  
3601 GCGCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGTTTCTCATAGCTCACGCTGTAGGTATCTC  
3701 AGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGACGAACCCCCGTTACGCCGACCGCTGCGCTTATCCGGTAACTATCGTCTTGAGTCCA  
3801 ACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTG  
3901 GCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAA  
4001 CAAACCACCGCTGGTAGCGGTGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT  
4101 CTGACGCTCAGTGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCCAATAAAATATCTTTATTTTCA  
4201 TTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAACTAGCAAAATAGGCTGTCCCA  
4301 GTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA