



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGCGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA  
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC  
201 GTGAACGTTCTTTTTCGAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC  
301 GCCATCCACGCCGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC  
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCGGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

**AgeI (552)**

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCTACCTGAGATCACCGGTACCATGTCTGATAATGGACCCCAAATCAGCGAAATGCACC  
1 M S D N G P Q N Q R N A P  
601 CCGCATTACGTTTGGTGGACCCTCAGATTCAACTGGCAGTAACCAAGATGGAGAACGCGAGTGGGGCGCGATCAAAACACGTCGGCCCAAGTTTACCC  
13 R I T F G G P S D S T G S N Q N G E R S G A R S K Q R R P Q G L P  
701 AATAACTGCGTCTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTTAAATTCCTCGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTC  
47 N N T A S W F T A L T Q H G K E D L K F P R G Q G V P I N T N S S  
801 CAGATGACCAAATGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTACCGTAAATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCT  
80 P D D Q I G Y Y R R A T R R I R G G D G K M K D L S P R W Y F Y Y L  
901 AGGAAGTGGCCAGAAGCTGGACTTCCCTATGGTGTAAACAAAGCGCATCATATGGTGTCAACTGAGGGAGCCTTGAATACACCAAAAGATCACATT  
113 G T G P E A G L P Y G A N K D G I I W V A T E G A L N T P K D H I  
1001 GGCACCCGCAATCCTGTAACAATGCTGCAATCGTGTACAACCTTCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCGGCA  
147 G T R N P A N N A A I V L Q L P Q G T T L P K G F Y A E G S R G G  
1101 GTCAAGCCTCTTCTCGTCTCATCAGTAGTCGCAACAGTTCAAGAAATCAACTCCAGGCAGCAGTAGGGAACTTCTCTGCTAGAATGGCTGGCAA  
180 S Q A S S R S S S R S R N S S R N S T P G S S R G T S P A R M A G N  
1201 TGGCGGTGATGCTGCTTGTCTTGTGCTGCTGCTGACAGATTGAACAGCTTGAGAGCAAAATGTCTGGTAAAGGCCAAACAACAAGGCCAAACTGTC  
213 G G D A A L A L L L L D R L N Q L E S K M S G K G Q Q Q Q G Q T V  
1301 ACTAAGAAATCTGCTGAGGCTTCTAAGAAGCCTCGGCAAAAACGTAAGTCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCAG  
247 T K K S A A E A S K K P R Q K R T A T K A Y N V T Q A F G R R G P  
1401 AACAAACCAAGGAAATTTGGGGACCAGGAACATAACAGACAAGGAAGTATTACAAACATTGGCCGCAATTCGACAATTTGCCCCAGCGCTTACGC  
280 E Q T Q G N F G D Q E L I R Q G T D Y K H W P Q I A Q F A P S A S A  
1501 GTTCTTCCGAATGTCGCGCATTGGCATGGAAGTACACCTTCGGGAACGTGGTGGACTACACAGGTGCCATCAAATGGATGACAAAAGATCCAAATTC  
313 F F G M S R I G M E V T P S G T W L T Y T G A I K L D D K D P N F  
1601 AAAGATCAAGTCAATTTGCTGAATAAGCATATTGACGCATACAAAACATTCCACCAACAGAGCCTAAAAAGGACAAAAAGAAAGGCTGATGAAACTC  
347 K D Q V I L L N K H I D A Y K T F P P T E P K K D K K K K A D E T  
1701 AAGCCTTACCGCAGAGACAGAAGAAACAGCAAACTGTACTTCTTCTGCTGCAGATTTGGATGATTTCTCAAACAATTGCAACAATCCATGAGCAG  
380 Q A L P Q R Q K K Q Q T V T L L P A A D L D D F S K Q L Q Q S M S S

**NheI (1823)**

1801 TGCTGACTCAACTCAGGCCTAAAGCTAGCTGGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAAC TAGAATGCAGTGAAAAAATGCTT  
413 A D S T Q A •  
1901 TATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTT  
2001 CAGGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTTTAACCTCAAATCA  
2101 AGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGACGCTCACCTTCTTTATG  
2201 GAGTTAAGATATAGTGTATTTCCCAAGTGTGAAGTACTGCTTTCATTTCTTATGTTTTAATGCAGTACCTCCACATTCCCTTTTATGTAATA  
2301 TTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTATG  
2401 AGTTGACTTAGGGAACAAAGAACCTTAAATAGAAATTTGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTACTTGGGGGATGAGTTCC  
141 • N R T Y K L P I L E  
2501 TCAATGGTGGTTTTGACCAGCTTGCATTATCTCAATGAGCACAAGCAGTCAAGGAGCATAGTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGA  
129 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 P S V  
2601 CCACCCTGATGGATGTCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCCGTGTCTACAGCAGACCAATGGC  
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  
2701 AATGGCTTCAAGCAGACAGTACCTGCAATGTAGGCTCAATGTGGACAGCAGAGATGATCTCCCAAGTCTTGGTCTGATGGCCGCCCGACATGG  
63 I A E A C V T V R G I Y A E I H V A S I E G T K T R I A A G V H  
2801 TGCTTGTTCCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACAGCTCCAGATCCTGCTGAGAGATGTTGAAGTCTTCAATGGTGGCCCTCC  
29 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  
2901 TATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACTAAAC  
3001 GAGCTCTGCTTATATAGACTCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTGTACGACATTTTGGAAAGTCCCGTTGATTTAC  
3101 TAGTCAAACAACTCCCATGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTAAGTCCCAAACCGC

3201 ATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTAC  
3301 CGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATG  
3401 GAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAGT  
3501 TATGTAACGCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCTGGCGTTTTTCCATAGGCT  
3601 CCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTCCCTGGAAGCTCC  
3701 CTCGTGCGCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGT  
3801 ATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA  
3901 GTCCAACCCGGTAAGACACGACTTATGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAG  
4001 TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCG  
4101 GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTAC  
4201 GGGGTCTGACGCTCAGTGAACGAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCCAATAAAATATCTTTAT  
4301 TTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAACGAAACAAACAACTAGCAAAATAGGCTGT  
4401 CCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA