



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA  
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC  
201 GTGAACGTTCTTTTTCGAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC  
301 GCCATCCACGCGGGTTGAGTCGCGTTCTGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC  
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

**AgeI (552)**

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTACCGGGAGCCGGCTGGGGTTGCCGGCACCATGGAGTC  
1 M G E P A G V A G T M E S  
601 ACCTTTTAGCCGGGACTCTTTACAGGCTGGATGAAGATTGGGATTCTGCTCTTTTGTGCTGAACTCGGTTATTTACAGACACTGATGAGCTGCAATTG  
13 P F S P G L F H R L D E D W D S A L F A E L G Y F T D T D E L Q L  
701 GAAGCAGCAAATGAGACGTATGAAAACAATTTGATAATCTTGATTTTGGATTGATGCCTGGGAGTCAGACATTTGGGACATCAACAACAAA  
47 E A A N E T Y E N N F D N L D F D L D L M P W E S D I W D I N N Q  
801 TCTGTACAGTTAAAGATATTAAGGCAGAACCTCAGCCACTTTCTCCAGCCTCCTCAAGTTATTCAGTCTGCTCCTCGGTACGTGGACTCTTATTCTC  
80 I C T V K D I K A E P Q P L S P A S S S Y S V S S P R S V D S Y S S  
901 AACTCAGCATGTTCTGAGGAGTTGGATTTGTCTTCTAGTTCTCAGATGTCTCCCTTTCTTATATGGTAAAACTTAATAGTCTCTTCCAGCGGAG  
113 T Q H V P E E L D L S S S S Q M S P L S L Y G E N S N S L S S A E  
1001 CCACTGAAGGAAGATAAGCCTGTCACTGGTCTAGGAACAAGACTGAAAATGGACTGACTCCAAAAGAAAAAATTCAGGTGAATTCAAAACCTTCAATTC  
147 P L K E D K P V T G P R N K T E N G L T P K K K I Q V N S K P S I  
1101 AGCCCAAGCCTTTATTGCTTCCAGCAGACCCCAAGACTCAAACAACTCCAGTGTCCAGCAAAAACCATCATTATTCAGACAGTACCAACGCTTATGCC  
180 Q P K P L L L P A A P K T Q T N S S V P A K T I I I Q T V P T L M P  
1201 ATGGCAAAGCAGCAACCAATATCAGTTTACAACCTGCACCCACTAAAGGCCAGACGGTTTTGCTGTCTCAGCCTACTGTGGTACAACCTCAAGCACCT  
213 L A K Q Q P I I S L Q P A P T K G Q T V L L S Q P T V V Q L Q A P  
1301 GGAGTTCTGCCCTCTGCTCAGCCAGTCTTGCTGTTGCTGGGGAGTCACACAGCTCCCTAATCACGTGGTGAATGTGGTACCAGCCCTTCAGCGAATA  
247 G V L P S A Q P V L A V A G G V T Q L P N H V V N V V P A P S A N  
1401 GCCCAGTGAATGGAAAATTTCCGTGACTAAACCTGTCTACAAAGTACCATGAGAAATGTCGGTTCAGATATTGCTGTGCTAAGGAGACAGCAACGTAT  
280 S P V N G K L S V T K P V L Q S T M R N V G S D I A V L R R Q Q R M  
1501 GATAAAAAATCGAGAATCCGCTTGTGCTCAGTCTCGCAAGAAGAAGAAAGAAATATATGCTAGGGTTAGAGGCGAGATTAAGGCTGCCCTCTCAGAAAACGAG  
313 I K N R E S A C Q S R K K K K E Y M L G L E A R L K A A L S E N E  
1601 CAACTGAAGAAAAGAAAATGGAACACTGAAGCGGCAGCTGGATGAAGTTGTGTGAGAGAACCAGAGGCTTAAAGTCCCTAGTCCAAAGCGAAGAGTTGTCT  
347 Q L K K K E N G T L K R Q L D E V V S E N Q R L K V P S P K T R R R V V  
1701 GTGTGATGATAGTATTGGCATTATAACTGAACATGAGCTATGAGCATGTTGGAACAGGATTCAGGAGAATGAACCTAGTGTGAGCTGACCTGCAAAA  
380 C V M I V L A F I I L N Y G P M S M L E Q D S R R M N P S V S P A N  
1801 TCAAAGGAGGCACCTTCTAGGATTTTCTGCTAAAGAGGCACAGGACACATCAGATGGTATTATCCAGAAAAACAGCTACAGATATGATCATTCTGTTTCA  
413 Q R R H L L G F S A K E A Q D T S D G I I Q K N S Y R Y D H S V S  
1901 AATGACAAAAGCCCTGATGGTGCTAACTGAAGAACCATTGCTTTACATTCCTCCACCTCCTTGTGAGCCCTAATTAACACAACAGAGTCTCTCAGGTTAA  
447 N D K A L M V L T E E P L L Y I P P P P C Q P L I N T T E S L R L  
2001 ATCATGAACTTCGAGGATGGGTTATAGACATGAAGTGAAGAAGCAAGTCAAGAAGAATGACAAAATCAACAGAAAACCCGATTCTCAGGGTGC  
480 N H E L R G W V H R H E V E R T K S R R M T N N Q Q K T R I L Q G A  
2101 TCTGGAACAGGGCTCAAATTCAGCTGATGGCTGTTCAATACACAGAAAACCTAGTAGTATCAGCAGGAACTCAGGGAGTGAGCTACAAGTGTATTAT  
513 L E Q G S N S Q L M A V Q Y T E T T S S I S R N S G S E L Q V Y Y  
2201 GCTTACCAGAAAGTTATCAAGACTTTTTGAAGCCATCCGCAGAAGGGGAGACACATTTTATGTTGTGCATTTGAAAGGGATCACCTGCTGTTACCAG  
547 A S P R S Y Q D F F E A I R R R G D T F Y V V S F R R D H L L L P  
2301 CTACCACCATAACAAGACCACAAGACCAAAAATGTCAATTGTGTACCAGCAATAAACATAAATGAGAATGTGATCAATGGGAGGACTACGAAGTAT  
580 A T T H N K T T R P K M S I V L P A I N I N E N V I N G Q D Y E V M  
2401 GATGCAGATTGACTGTGAGGTGATGGACACCAGGATCCTCCATATCAAAGTTCGTGAGTTCCTCCTTACCTCCGAGATCAGCAGAGGAATCAAACCAAC  
613 M Q I D C Q V M D T R I L H I K S S S V P P Y L R D Q Q R N Q T N

**NheI (2574)**

2501 ACCTTCTTTGGCTCCCTCCCGCAGCCACAGAGGCAACCCACGTTGTGACGACCATCCCTGAGTCATTACAATAGCTAGCTGGCCAGACATGATAAGATA  
647 T F F G S P P A A T E A T H V V S T I P E S L Q •  
2601 CATTGATGAGTTTGGACAAACCAACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGC  
2701 TGCAATAAACAAGTTAAACAACAATTGCATTCATTTATGTTTCAGGTTGAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAAT  
2801 GTGGTATGGAATTCTAAAATACAGCATAGCAAACTTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCA  
2901 GGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTCTTTCATGGAGTTTAAAGATATAGTGTATTTTCCAAGGTTTGAAGTACTCTTCATT  
3001 TCTTTATGTTTTAAATGCACTGACCTCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTA  
3101 TTAGGCAGAAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATGGACAGCAAG

3201 AAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCATCTCAATGAGCACAAGC  
141 • N R 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  
3301 AGTCAGGAGCATAGTCAGAGATGAGCTCTGACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGC  
113 D P A Y D S I L 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  
3401 CACAATGGTGTCAAAGTCCTTCTGCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCAGCACAGACAGTGACCCTGCCAATGTAGGCCTCAATGTGG  
80 V I T D F D K Q G N S V A S G I A I A E A C V T V R G I Y A E I H  
3501 ACAGCAGAGATGATCTCCCCAGTCTTGGTCTGATGGCCGCCCGACATGGTCTTGTGCTCATAGAGCATGGTATCTTCTCAGTGGCGACCTCCA  
46 V A S I I E G T K 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  
3601 CCAGCTCCAGATCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAAT  
13 L E L D Q Q S I N F T K M  
3701 TGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTCACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTG  
CGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCCTTGATTACTAGTCAAACAACCTCCATTGACGTCAATGGGGTGGAGACTTGAAAT  
3801 CCCCCTGAGTCAAACCGTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAG  
3901 GAAAGTCCATAAGGTCACTGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATG  
4001 TACTGCCAAGTGGGCGAGTTTACCCTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTC  
4101 AATGGGCGGGGTCGTTGGGCGGTGACCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATAAGAACATGTGAGCAAAGGCCAGCA  
4201 AAAGGCCAGGAACCGTAAAAAGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGG  
4301 CGAAACCCGACAGACTATAAAGATACCAGGCGTTTTCCCCTGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCG  
4401 CCTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGTGTGCACGA  
4501 ACCCCCCGTTAGCCCGACCGCTGCGCCTTATCCGTAACATATCGTCTTGTAGTCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT  
4601 AACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCG  
4701 CTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCA  
4801 GATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTC  
4901 ATGGCTAGTTAATTAACATTTAAATCAGCGGCCCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATA  
5001 CGCTCTCCATCAAACAAAACGAAACAAAACAAACTAGCAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA  
5101