



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCCGAGAAGTTGGGGGAGGGTTCGGCAATTGAACGGGTGCCTA  
101 GAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGTGCTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC  
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC  
301 GCCATCCACGCGGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC  
401 GGGCCTTTGTCCGGCGCTCCCTTGAGAGCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

SphI (560)

AgeI (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTCAGCATGCTGTCGGGCTCTCCGGCCAGACTCCGCGCGGCC  
13> M L S G S P G Q T P P A P  
601 GTTCCCCAGCCCGCCACCGCCAGCCCCGGCCAGCCGCGCCGCGCTTCCCCAGTCCACGTCAAGTCGGGCTGCAGATCCGAAAGAACGCCATCACCC  
13> F P S P P P A P A Q P P P P F P Q F H V K S G L Q I R K N A I T  
701 GACGACTACAAGTCAACAGCAAGTGTGGCTGGGCATCAACGGGAAGGTGCTGCGGATCTTCGACAAGAGAACCAGCAAAAATTCGCCCTAAAGA  
47> D D Y K V T S Q V L G L G I N G K V L R I F D K R T Q Q K F A L K  
801 TGCTCCAGACTGTCCGAAGCGCCAGAGAGGTGGAGCTGCCTGGAGGGCTCCAGTGCACACATTGTGCACATCGTGGATGTCTATGAGAACCT  
80> M L Q D C P K A R R E V E L H W R A S Q C P H I V H I V D V Y E N L  
901 GTATGCCGGGAGGAAGTGCCTGCTGATTGTGTCATGGAGTGTCTCGATGGTGGAGAGCTCTTAGTCAATCCAGGACCGAGGAGACCAGGCATTACAGAA  
113> Y A G R K C L L I V M E C L D G G E L F S R I Q D R G D Q A F T E  
1001 AGAGAGGCGTCAGAGATCATGAAGAGCATCGGCAGGCCATCCAGTACCTGCCTCGATCAACATTGCTCACGGGATGTCAAGCCTGAGAACCTCTTAT  
147> R E A S E I M K S I G E A I Q Y L H S I N I A H R D V K P E N L L  
1101 ATACTTCCAAAAGGCCAATGCCATTTTGAAGTCACTGATTTTGGCTTTGCCAAGGAAACCACAGTCAAACTCTTTGACCACTCCGTTTATACACC  
180> Y T S K R P N A I L K L T D F G F A K E T T S H N S L T T P C Y T P  
1201 ATACTATGTGGCTCCGGAAGTCTGGCCCGGAGAAGTATGACAAGTCTGTGACATGTGGTCTTGGTGTGCATGTATATTTTGTGTGGGTAT  
213> Y Y V A P E V L G P E K Y D K S C D M W S L G V I M Y I L L C G Y  
1301 CCCCCCTTCTATTCCAATCACGGCCTTGCCATCTCTCCGGCATGAAGACTCGTATTGAATGGGCCAGTGAATTTCTAACCCGGAGTGGTCAGAAG  
247> P P F Y S N H G L A I S P G M K T R I R M G Q Y E F P N P E W S E  
1401 TATCAGAAGAAGTGAAGATGCTTATCCGGAATCTGCTAAAAACAGAGCCACCCAGAGAATGACCATCACAGAATTCATGAACCACCCCTGGATCATGCA  
280> V S E E V K M L I R N L L K T E P T Q R M T I T E F M N H P W I M Q  
1501 ATCTACGAAGGTCCTCAGACTCCACTGCACACCAGCCGTGCTGCTGAAGGAGGACAAGGAACGATGGGAGGATGTCAAGGAGGAGATGACCAGTGCCTTG  
313> S T K V P Q T P L H T S R V L K E D K E R W E D V K E E M T S A L  
1601 GCCACGATGCGTGTGACTATGAGCAGATCAAGATAAAGAAGATAGAAGACGATCCAACCCCTGCTTCTCAAGAGCGGAAGAAAGCTCGTGTGTGG  
347> A T M R V D Y E Q I K I K K I E D A S N P L L L K R R K K A R A V

NheI (1748)

1701 AGGATGCGGCTCTCGCCACTGAGCCGCTCTGTCTCCACCCACTGGAGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAA  
380> E D A A L A H •  
1801 CTAGAATGCAGTGAATAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACA  
1901 TTGCATTCATTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCA  
2001 TAGCAAACTTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCT  
2101 GTTTCAGCCTCACCTTCTTTCATGAGTAAAGATATAGTGTATTTCCCAAGTGTGAAGTACTGCTTTCATTTCTTATGTTTAAATGCACTGACCT  
2201 CCCACATTCCTTTTTAGTAAAATATTCAGAAAATATTTAAATACATCATTGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGC  
2301 CCTCATAATATCCCCAGTTTAGTAGTTGACTTAGGGAACAAAGAACCTTAAATAGAAATTTGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCT  
141< • N R  
2401 GGTGACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCATTATCTCAATGAGCACAAGCAGTCAGGAGCATAGTCAGAGATGAGC  
138< 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 D P A Y D S I L  
2501 TCTCTGCATGCCACAGGGGCTGACCACCCTGATGGATGTGCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCC  
104< 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 V I T D F D K Q G  
2601 CGTTGCTCACAGCAGACCCAATGGCAATGGCTTCCAGCAGACAGTACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCAAGTCTT  
71< N S V A S 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  
2701 GGTCTGATGGCCGACATGGTCTTGTTCCTCATAGAGCATGGTGTGATCTTCTCAGTGGCGACCTCCACAGCTCCAGATCCTGCTGAGAGATG  
38< T R I A A G V 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  
2801 TTGAAGTCTTCTGTTGGCCCTCTATAGTGTGATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTC  
4< N F T K M  
2901 CAGCTTATCTGACGGTCACTAACGAGCTCTGCTTATATAGACTCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTGTACGAC  
3001 ATTTTGGAAAGTCCCGTTGATTTACTAGTCAAACAAACTCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCAGC  
3101 CCCATTGATGACTGCCAAAACCGCATCATGTGTAATAGCGATGACTAATACGTAGATGACTGCCAAGTAGGAAAGTCCCATAAGGTCATGACTGG

3201 GCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATGTA  
3301 AATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTCA  
3401 GCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCTGCAGGTTAAITAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG  
3501 CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATA  
3601 CCAGGCGTTTCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCTGCCGTTACCGGATACCTGTCCGCTTCTCCCTTCGGGAAGCGTGGCG  
3701 CTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGTGGC  
3801 CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGT  
3901 AGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA  
4001 AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAACACCGCTGGTAGCGGTGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTC  
4101 AAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAACCTACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATC  
4201 AGCGGCCCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAAC  
4301 AAAACAACTAGCAAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA