



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCCGGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCGGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

AgeI (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTCCAGTGGCCTCCGGGAGCGTGGCCGAGTGCCTGCAGCAGGA
1 M A S G S V A E C L Q Q E
601 GACCACCTGCCCGTGTGCCTTCACTACTTCGTGGAGCCTATGATGCTCGACTGCGGCCACAACATCTGTTGCGCGTGCCTGGCCCGTGTGGGGTGC
13 T T C P V C L Q Y F V E P M M L D C G H N I C C A C L A R C W G A
701 GCGGAGACTAACGTGTCGTGTCGCGAGTCCCGGAGACCTCCCGCAGCGCCACATGCGACCCAACCGGCACCTGGCCAACGTGACCCAGCTGGTGAAGC
47 A E T N V S C P Q C R E T F P Q R H M R P N R H L A N V T Q L V K
801 AGTGTGCGCACCGAGCGCGCTGCGGGCCCGGCGGCGAGATGGGCGTGTGCGAGAAGCACCGGGAGCCTCTGAAGCTGTACTGCGAGCAGGACAGATGCC
80 Q L R T E R P S G P G G E M G V C E K H R E P L K L Y C E Q D Q M P
901 CATCTGCTGGTGTGCGACCGCTCGCGGAGCAGCAGCGGCGCCACAGCGTGTGCCCTGGAGGAGCGGTGGAGGGCTCAAGGAGCAAATCCAGAACCGA
113 I C V V C D R S R E H R G H S V L P L E E A V E G F K E Q I Q N R
1001 CTGGACCACCTAAGAAGAGTGAAGACTTGAAGAAGAGACGGCGGCACAGGGGAGCAGGCCGAGCTGAGCTTTTGTGAGCTGACTCAGATGGAGAGGG
147 L D H L R R V K D L K K R R R A Q G E Q A R A E L L S L T Q M E R
1101 AAAAAATTGTCTGGGAGTTTGTGAGCAGCTCTACTCGTTGAAGGAGCATGATACCGCTGTTGGCTCGCCTTGGAGGCTGACTTGGCCATCTATAA
180 E K I V W E F E Q L Y H S L K E H E Y R L L A R L E E L D L A I Y N
1201 CAGCATCAACGGGGCCATCACTCAGTTCCTGCAACATCTCCACCTCAGTGGCCTGATTGCCAACTGGAAGAGAAGCAGCAGCAGCCACGCGGGAG
213 S I N G A I T Q F S C N I S H L S G L I A Q L E E K Q Q Q P T R E
1301 CTCTTACAGGACATTGGAGACACTGAGCAGGGCTGAAAGAATCAGGATCCCTGAGCCTGGATCACACCCAGATCTACAAGAGAAAATCCATATTT
247 L L Q D I G D T L S R A E R I R I P E P W I T P P D L Q E K I H I
1401 TTGCCAAAAATGTCTGTTTTTGTGAGAAAAGTCTAAAACAGTTCACAGAAAAGATGCACTGAGATGAGGAGAAGTCAAGAGTTGAGAGAGGCTCAGTT
280 F A Q K C L F L T E S L K Q F T E K M Q S D M E K I Q E L R E A Q L
1501 ATACTCAGTTGATGTAACCTGGACCCAGACACAGCCTATCCAGTCTGATTCTTTCTGATAATCTTCGGCAAGTACGGTACAGTTACCTCCAGCAGGAT
313 Y S V D V T L D P D T A Y P S L I L S D N L R Q V R Y S Y L Q Q D
1601 CTTCCAGACAACCTGAGCGATTCAATCTGTTCCCTGTGCTTGGGCTCTCCATGTTTATGGCTGGGAGACATTACTGGGAGGTAGAGGTGGGAGATA
347 L P D N P E R F N L F P C V L G S P C F M A G R H Y W E V E V G D
1701 AAGCCAAGTGGACCATAGGTGTCTGTAAGACTCGGTGTGCAGAAAAGTGGGGTGCCTCAGCCCCCAGAATGGATTCTGGGAGGTGTCTGTTGGTGA
380 K A K W T I G V C E D S V C R K G G V T S A P Q N G F W A V S L W Y
1801 TGGGAAAGAATATTGGGCTTTACATCCCAATGACTGCACTACCCCTGCGGACCCCTCTCCAGAGAGTAGGGATTTTCTGGACTATGATGCTGGTGGAG
413 G K E Y W A L T S P M T A L P L R T P L Q R V G I F L D Y D A G E
1901 GTTTCCTTCTACAATGTGACAGAGAGATGCCATACCTTACTTTCTCTCACGTACCTTCTGTGGCCAGTCCGGCCCTACTTCCAGCCTAAGTACTCGG
447 V S F Y N V T E R C H T F T F S H A T F C G P V R P Y F S L S Y S
2001 GAGGGAAGAGCGCAGCTCCTTTGATCATCTGCCTATGAGTGGGATTGATGGGTTTTCTGGCCATGTTGGGAATCATGGTCATTCCATGGAGACCTCTCC
480 G G K S A A P L I I C P M S G I D G F S G H V G N H G H S M E T S P

NheI (2137)

2101 ΔTGAGGAAGTGAACGTAGACCACAAGTGTGCTGCTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAG
513 •
2201 TGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGTGAATAAACAAGTTAACAACAACATTGCATTCATT
2301 TTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTT
2401 TAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCT
2501 CACCTTCTTTCATGGAGTTTAAAGATATAGTGTATTTTCCCAAGGTTTGAAGTACTCTTCAATTTCTTTATGTTTTAAATGCACTGACCTCCACATCCC
2601 TTTTTAGTAAAATATTCAGAAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATA
2701 TCCCCAGTTTAGTAGTTGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCTCTGGTGTACTTGA
141 • N R T Y K L
2801 GGGGATGAGTTCCTCAATGGTGGTTTTGACCAGTTCGCCATTCATCTCAATGAGCACAAAGCAGTCAAGGAGCATAGTCAGAGATGAGCTCTCTGCAC L
134 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 E R C M
2901 GCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCCGTTGCTCACA
101 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 N S V
3001 GCAGACCAATGGCAATGGCTTACGACAGACAGTACCGCTGCAATGTAGGCTCAATGTGGACAGCAGAGATGATCTCCCAAGTCTTGGTCTGATGG
67 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 T R I A
3101 CCGCCCCGACATGGTCTTGTGCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACAGCTCCAGATCCTGCTGAGAGATGTTGAAGGCTT
34 A G V H 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

3201 CATGGTGGCCCTCCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACACAGCGTGGATGGCGTCTCCAGCTTATCTG
1 M
3301 ACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCTACCGCCATTTCGCTCAATGGGGCGGAGTTGTTACGACATTTTGAAAG
3401 TCCCGTTGATTTACTAGTCAAAACAAACTCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGT
3501 ACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTAAGTCCCAAGTAGGAAAGTCCCATAGGTCATGTAAGTGGCATAATGCCA
3601 GCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTAAGTGGCAGTTTACCGTAAATACTCCACC
3701 CATTGACGTCAATGGAAAGTCCCTATTGGCGTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTGAGCCAGGCGGGC
3801 CATTACCGTAAGTTATGTAACGCTGCAGGTTAATAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGGCGG
3901 TTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTC
4001 CCCCTGGAAGCTCCCTCGTGGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAG
4101 CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTCAGCCGACCGCTGCGCCTTATCCGGT
4201 AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTA
4301 CAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGATTGG
4401 TAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCT
4501 TTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAA
4601 TAAAAATCTTTATTTTTCATTACATCTGTGTGGTTTTTTTGTGTAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAACTA
4701 GCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA