



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCCGAGAAGTTGGGGGAGGGGTCGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGGTAACCTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGCTGCGTCCGCGCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

SphI (560)

Agel (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTGAGTCCAGCGGAACGGCCCGGGGAGCGGGCTCGGA
13▶ M P A E R P A G S G G S E
601 GGCTCCAGCAATGGTTGAACACTGGACTGCTGTGATTACCCCGGCATGCTAGAAGAGGAAGAACAGCTTGAAGTCTGGACTAGAGAGAGCGG
13▶ A P A M V E Q L D T A V I T P A M L E E E E Q L E A A G L E R E R
701 AAGATGCTGGAAGGCTCGCATGCTTGGGATAGAGAGTGCAGAGAAATTCGGTACCGTAGACTTCAACATTTGCTTAAAAAGCAATATATACTCCA
47▶ K M L E K A R M S W D R E S T E I R Y R R L Q H L L E K S N I Y S
801 AATTTTATTGACGAAATGGAACAGCAACAATTAGAGGAACAGAAGAAGAAAATTTGGAGAGAAAAAGGAGCTTTAAAAGTTAAAAAGGGTAA
80▶ K F L L T K M E Q Q Q L E E Q K K K E K L E R K K E S L K V K K G K
901 AAATTCATTTGATGCAAGTGAAGAGAAGCCAGTTATGAGGAAAAAGAGGAAGAGAAGTGAATCATACAATATTTAGAGGTCATGTCAAAAGAGGAA
113▶ N S I D A S E E K P V M R K K R G R E D E S Y N I S E V M S K E E
1001 ATTTTGTCTGGCTAAAAAATAAAAGGAGAATGAGGATGAAACTCCTCTCTACTAATCTCTGTGGAAGATCTTCAGAAAAATAAGATTCGA
147▶ I L S V A K K N K K E N E D E N S S S T N L C V E D L Q K N K D S
1101 ATAGTATAATTAAGATAGATTGCTGAAACGGTTAGGCAGAACTAAATCTTTTTGACCCAGTCCGGAAGTGAATGGTCAGCCAGTACCTTTTCA
180▶ N S I I K D R L S E T V R Q N T K F F F D P V R K C N G Q P V P F Q
1201 ACAACCAAGCACTTCACTGGAGGATGATGCGATGGTACCAAGTAGAAGGCATGGAATGGCTTAGGATGCTTTGGGAAAATGGAATTAATGGCATTITA
213▶ Q P K H F T G G V M R W Y Q V E G M E W L R M L W E N G I N G I L
1301 GCAGATGAAATGGGATTGGGTAAGACAGTTCAGTGCATTGCTACTATTGATTGATGATTGAGAGAGGAGTACCAGGACCTTTTCTTGTCTGTGGCCTT
247▶ A D E M G L G K T V Q C I A T I A L M I Q R G V P G P F L V C G P
1401 TGCTACACTTCTAAGTGGATGGTGAATCAAAGATTTACACCAGATATCCCTACAATGTTATATCATGGAACCCAGGAGGAACGTCAAAAATGGT
280▶ L S T L P N W M A E F K R F T P D I P T M L Y H G T Q E E R Q K L V
1501 AAGAAATATTTACAAACGGAAAGGGACTTTGCAGATTCATCCTGTGGTAATCACGTCATTTGAAATAGCCATGAGAGACCGAAATGCGTTACAGCATTGC
313▶ R N I Y K R K G T L Q I H P V V I T S F E I A M R D R N A L Q H C
1601 TATTGAAACTTAATAGTAGATGAAGGACACAGGATTAAGAATATGAAGTCCGCTAATCAGGGAGTTAAAACGATTCAATGCTGATAACAACTTC
347▶ Y W K Y L I V D E G H R I K N M K C R L I R E L K R F N A D N K L
1701 TTTTACTGGTACTCCCTTGCAAAACAATTTATCAGAACTTTGGTCTTAACTTTTGTGCCAGATGTATTTGATGACTTGAAGAGCTTTGAGTC
380▶ L L T G T P L Q N N L S E L W S L L N F L L P D V F D D L K S F E S
1801 TTGGTTTGACATCACTAGTCTTTCTGAAACTGCTGAAGATATTATTGCTAAAGAAAGAGAACAAGAAATGATTGCATATGCTGCACCAGATTTTAAACCT
413▶ W F D I T S L S E T A E D I I A K E R E Q N V L H M L H Q I L T P
1901 TTCTTATTGAGAAGACTGAAGTCTGATGTTGCTTGAAGTCTCTAAACGAGAAGTAGTCTGTTTATGCTCCACTTTCAAAGACAGGAGATCTTTT
447▶ F L L R R L K S D V A L E V P C K R E A G V V Y A P L S K K Q E I F
2001 ATACAGCCATTGTGAACCGTACAATTGCAAAATGTTTGGATCCAGTGAAGAAAGAAACAAATTGAGTTAAGTCTACTGGTCCGACCAAAACGACGAAGT
480▶ Y T A I V N R T I A N M F G S S E K E T I E L S P T G R P K R R T R
2101 AAAATCAATAAATACAGCAAAATAGATGATTTCCCTAATGAATTGAAAAACTGATCAGTCAAATACAGCCAGAGGTGGACCGAGAAAGAGCTGTTGTG
513▶ K S I N Y S K I D D F P N E L E K L I S Q I Q P E V D R E R A V V
2201 GAAGTGAATATCCCTGTAGAATCTGAAGTAACTGAAAGTGCAGAATATAATGATGCTACTTCGTAATGTTGTAATCATCATATTTGATTGAATATC
547▶ E V N I P V E S E V N L K L Q N I M M L L R K C C N H P Y L I E Y
2301 CTATAGACCTGTACACAAGAATTAAGATCGATGAAGAATTGGTAACAAATCTGGGAAGTCTTGAATTTGGATCGAATGCTGCCAGAACTAAAAA
580▶ P I D P V T Q E F K I D E E L V T N S G K F L I L D R M L P E L K K
2401 AAGAGGTCACAAGGTGCTGCTTTTTTCACAATGACAAGCATGTTGGACATTTTGTGGATTACTGCCATCTCAGAGATTTCAACTTCAGCAGGCTTGAT
613▶ R G H K V L L F S Q M T S M L D I L M D Y C H L R D F N F S R L D
2501 GGTCCATGTCTTACTCAGAGAGAGAAAAAACATGCACAGCTTCAACACGGATCCAGAGGTGTTTATCTTCTTAGTGAAGTACACGAGCTGGTGGCCTGG
647▶ G S M S Y S E R E K N M H S F N T D P E V F I F L V S T R A G G L
2601 GCATTAATCTGACTGCAGCAGATACAGTTATCATTTATGATAGTATTGGAACCCCGAGTCCGATCTTCCAGGCCAGGATAGATGTCATAGAATTGGTCA
680▶ G I N L T A A D T V I I Y D S D W N P Q S D L Q A Q D R C H R I G Q
2701 GACAAAGCCAGTTGTTTATCGCCTGTTACAGCAAACTATCGATCAGAAAAATGGAAGAGCAGCTGCTAAAAGGAACTGAAAAAGTTGATC
713▶ T K P V V Y R L V T A N T I D Q K I V E R A A A K R K L E K L I
2801 ATCCATAAAAATCATTTCAAAGGTGGTCAAGTCTGGATTAATCTGTCTAAGAATTTCTTAGATCCTAAGGAATTAATGGAATTAATAAATCTAGAGATT
747▶ I H K N H F K G G Q S G L N L S K N F L D P K E L M E L L K S R D
2901 ATGAAAGGGAAATAAAAGGATCAAGAGAGAAGTCAATAGTATAAAGATCTAGAGTTGTTGTTAGATCGAAGTATCTTATTGATCAATGAATGCTTC
780▶ Y E R E I K G S R E K V I S D K D L E L L L D R S D L I D Q M N A S

NheI (3085)

3001 AGGACCAATTAAGAGAAGATGGGGATTTCAAGATATTAGAAAATCTGAAGATTCCAGTCCCTGAATGTTTGTGTTTAAAGTGGAGTAGCTGGCCAGAC
813▶ G P I K E K M G I F K I L E N S E D S S P E C L F •
3101 ATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAGAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTA

3201 CCATTATAAGCTGCAATAAACAAAGTTAACAAACAACAATTGCATTCAATTTATGTTTCAGGTTACAGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAA
3301 CCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAAACTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGG
3401 CATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTGCAGCCTCACCTTCTTTCATGGAGTTAAGATATAGTGATTTTTCCCAAGGTTTGAAC
3501 AGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCACATTCCCTTTTTAGTAAAATATTCAGAAAATAATTTAAATACATCATTGCAATGAAAATA
3601 AATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAT
3701 TGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCACTCAAT
3801 GAGCACAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTGCACATGCCACAGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGG
117 L V F C 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
3901 TGCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCAGCACAGACAGTGACCTGCCAATGTAGG
83 H R V A 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
4001 CCTCAATGTGGACAGCAGAGATGATCTCCCAGTCTTGGTCTGATGGCCGCCCGACATGGTGTCTTGTTCCTCATAGAGCATGGTGTCTTCTCAGT
50 E I H 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
4101 GGCACCTCCACCAGTCCAGATCCTGCTGAGAGATGTTGAAGGTCTTCATGGTGGCCCTCTATAGTGAGTCGATTATACTATGCCGATATACTATGC
17 A V E V L E L D Q Q S I N F T K M
4201 CGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACATAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTA
4301 CCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGAAAGTCCCGTTGATTTACTAGTCAAAACAAACTCCATTGACGTCAATGGGGTGGGA
4401 GACTTGAAATCCCCTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTA
4501 CTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGA
4601 TACACTTGATGTACTGCCAAGTGGGCGAGTTTACCGTAAATACTCCACCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCA
4701 TTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAAGTTATGTAACGCCCTGCAGGTTAATTAAGAACATGTGAGCA
4801 AAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAAAAAATCGACGCTCAA
4901 GTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCCTGCCGTTACCGG
5001 ATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGC
5101 TGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAG
5201 CAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATT
5301 TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGATTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGT
5401 TGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAAACGAAACTCACGTTAAG
5501 GGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTAATCG
5601 TAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATC
5701 GAA