



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGAGAAGTTGGGGGAGGGGTCGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGTGCTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCTCTGAACTGCGTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCGGGCTCTCCACGCTTTGCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

SphI (560)

AgeI (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTCAGCATGCCACCACAGAGCATCATCCGGCCAGGACCACCT
13▶ R A G W E Q R L E R S L P A P R V G L L W M G L G L A L V L A L F
701 GACTCCACGGTCTTTGGGTTCTGCCAGAGCTTATCCTCTTCTTCTGAAGGCCATTCTGTCAAATTCAGTGCCATAGCGCCGCGCTCCAGAGTGCCCT
47▶ D S T V L W V P A R A Y P L P S E G H S V K F S A I A P P L Q S A
801 TTGGTGGCAGAACCTCACTTGCCTCCGCTGCAAAGTCTTACTGCTCTCAACCATGGGCTGAAGAAGGAGCCCAATGTGGCAGGGTAGGCTCTGT
80▶ F G W Q N L T C P A C K V L F T A L N H G L K K E P N V A R V G S V
901 GGCCATCAAGATATGCAAGATGTGAACATAGCACCCTAGATGTGTGCCAGTCAGCCGTCCTCTTTGAGGACGATGTGGTGGAGGTGTGGACACGT
113▶ A I K I C K M L N I A P L D V C Q S A V H L F E D D V V E V W T R
1001 TCAGTTCTGAGCCATCAGAGGCTTGTGGCTTGTCTGGGCTCCTCTTGTGGACACTGGGACATCTTTTCGACTTGAACATCTCTTGGCCATCAGTGC
147▶ S V L S P S E A C G L L L G S S C G H W D I F S T W N I S L P S V
1101 CGAAGCCACCCCAAAGCCACCGAGCCACCAGCCCCGGTGCCCTGTGAGCCGTCTCTTCTTACTGATCTACACTGGGACCATGAGTACCTGGA
180▶ P K P P P K P P S P P A P G A P V S R V L F L T D L H W D H E Y L E
1201 GGCACAGACCCCTACTGTGCAGATCCACTTTGCTGCCGCGGGGGTCCGGATGGCCACCCAACTCCAGAAGGGGCTGGTCTGGGGCGAGTACAGC
213▶ G T D P Y C A D P L C C R R G S G W P P N S Q K G A G F W G E Y S
1301 AAGTGTGACCTGCCCTTGCAACGCTAGAAAGCCTGTTGAAAGGACTGGGCCCTGCCGCGCTTTTGAATGGTGTACTGGACGGGAGATATCCCTGCC
247▶ K C D L P L R T L E S L L K G L G P A G P F E M V Y W T G D I P A
1401 ATGATGTCTGGCAACAGTCTGCCAAGATCAGTGTAGGGCCCTGACCACCATCACAGACCTCGTGAGGAAGTCTTGGGCCAGTGCCAGTGTACCCTGC
280▶ H D V W Q Q S R Q D Q L R A L T T I T D L V R K F L G P V P V Y P A
1501 TGTGGGAACCATGAGAGCACTCCTGTCAATGGCTTCCCTCCTCCCTCATAAAGGAAACCAATCTTCAATGGCTCTATGAAGCCATGGCCAAGGCA
313▶ V G N H E S T P V N G F A P P P F I K G N Q S S Q W L Y E A M A K A
1601 TGGAACCCCTGGTTACCAGCTGATGCCCTCACACCCTAAGATTGGGGCTTCTATGCCCTCACCCACGCCCTGGCTCCGCTCATCTCTCAATA
347▶ W E P W L P A D A L H T L R I G G F Y A L T P R P G L R L I S L N
1701 TGAATTTTTGTTCCCGTGAAGCTTTTGGCTTGTGATCAACTCCACAGATCCTGTGGACAACCTCAGTGGCTGGTGGAGGAGCTTCAAGCTGCTGAGAA
380▶ M N F C S R E N F W L L I N S T D P A G Q L Q W L V E E L Q A A E N
1801 TCGAGGAGACAAAGTGCATATAATTGGCCACATCCCTCCAGGACATTGCTTAAAGAGCTGGAGCTGGAATTATTACAAAATCATAGCCAGGTATGAAAAC
413▶ R G D K V H I I G H I P P G H C L K S W S W N Y Y K I I A R Y E N
1901 ACTCTGGCCGGTCACTTCTTTGGCCACTCACGTGGATGAGTTTGTGATCTTCTATGATGAGGAACTCTGAGCCGACCAGCTAGCTGTAGCCTTCTGG
447▶ T L A G T F F A T H C H T H V D E F E I F Y D E E T L S R P L A V A G C C
2001 CGCCAGTGCTACAACCTTTATCAACCTTAACCTGGCTACCGAGTTTACCAAATAGATGAAAACCTCCCGAAGCTCTCATGTGGTCTTGGACACGA
480▶ A P S A T T F I N L N P G Y R V Y Q I D G N Y P G S S H V V L D H E
2101 GACCTACATCTTGAATCTGACCCAGGCCAACGCAGCAGGAGGCACCAAGCTGGAAGCGCTCTACAGGGCTCGAGAAACCTATGGACTTCCAGATGCG
513▶ T Y I L N L T Q A N A A G G T P S W K R L Y R A R E T Y G L P D A
2201 ATGCTGCCTCCTGGCACAACCTGGTCTACCGCATGAGGGATGATGAGCAACTTCCAGACCTTCTGGTTTCTTACCATAAGGGCCACCCACCTTCA
547▶ M P A S W H N L V Y R M R D D E Q L F Q T F W F L Y H K G H P P S
2301 AGCCCTGCGGCACACCCTGCCGCTGGCCACTCTGTGTGCCAGCTCTCAGCCGAGCAGACGCCCTGCTTGTGTGCCACTTGTGCCAATGGGAG
580▶ E P C G T P C R L A T L C A Q L S A R A D S P A L C R H L M P N G S

NheI (2464)

2401 CCTCCAGATGCTAATAGGTTGTGGTACGCCCCCTGTTGTGCTAGTATTCCATGTTCCAGACTGCTAGCTGGCCAGACATGATAAGATACATTGATGAG
613▶ L P D A N R L W S R P L L C •
2501 TTTGGACAAACCAACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAAC
2601 AAGTTAACAACAACAATTGCATTCAATTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGA
2701 ATTCTAAAATACAGCATAGCAAACTTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTG
2801 CCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAAGATAGTGTATTTTCCCAAGTTTGAAGTACTCTTCAATTTCTTATGTT
2901 TTAAATGCACTGACCTCCACATTCCCTTTTTAGTAAAATATTAGAAAATATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAA
3001 TCCAGATGCTCAAGGCCCTTCAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTAAATAGAAATTGGACAGCAAGAAAGCGAGCT
3101 TCTAGCTTTAGTTCTGGTGTACTTGTAGGGGATGAGTTCTCAATGGTGGTTTTGACCAGCTTGCATTATCTCAATGAGCACAAGCAGTCAGGAGC
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 D P A

3201 ATAGTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGTGCCTGACAGCCACAATGGTG
110 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 V I T
3301 TCAAAGTCTTTCTGCCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCAGCACAGACAGTACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGA
76 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 V A S I
3401 TGATCTCCCAGTCTTGGTCTGATGGCCGCCCGACATGGTGTCTTGTCTCATAGAGCATGGTATCTTCTCAGTGGCGACCTCCACCAGTCCAG
43 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 L E L
3501 ATCCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAAGTCTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACA
10 D Q Q S I N F T K M
3601 GCGTGGATGGCGTCTCCAGCTTATCTGACGGTCACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTGCGTCAATGGG
3701 GCGGAGTTGTTACGACATTTTGGAAAGTCCCCTGATTACTAGTCAAAACAAACTCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCTGAGT
3801 CAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCA
3901 TAAGTCACTGACTGGGCATAATGCCAGGCGGGCCATTACCCTGATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGTACTGCCAAG
4001 TGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGG
4101 GGTCTTGGGCGGTCAGCCAGGCGGGCCATTACCCTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGG
4201 AACCGTAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGA
4301 CAGGACTATAAAGATACCAGGCGTTTTCCCCTGGAAGTCCCTCGTGCCTCTCTGTTCCGACCCTGCCGTTACCCTGATACCTGTCCGCTTTCTCCC
4401 TTCGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGTGTGCACGAACCCCGCTT
4501 CAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
4601 GCAGAGCGAGGTATGTAGGCGGTGTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAA
4701 GCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGTTTTTTGTTTGAAGCAGCAGATTACGCGC
4801 AGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTT
4901 AATTAACATTTAATCAGCGGCCCAATAAAAATATCTTTATTTTATTACATCTGTGTGGTTTTTTGTGTGAATCGTAACATAACGCTCTCCAT
5001 CAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA