



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCCGTTGAGTCCGCTTCTGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTAACTCTACGCTTTGTTTCGTTT

Agel (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTAGGAGGGCCAGCATGCTGCTCACTCTTATCATTCTGTTGCCA
1▶ M L L T L I I L L P
601 GTAGTTTCAAATTTAGTTTTGTAGTCTCTCAGCACCGCAGCACTGGAGCTGCTCTGAAGTACTCTCGCAGGAAATGGGAATTCTACTTGTGTGGGTC
11▶ V V S K F S F V S L S A P Q H W S C P E G T L A G N G N S T C V G
701 CTGACCCCTTCTAATTTTCTCCATGGAAATAGTATCTTTAGGATTGACACAGAAGGAACCAATTATGAGCAATTGGTGGTGGATGCTGGTGTCTCAGT
44▶ P A P F L I F S H G N S I F R I D T E G T N Y E Q L V V D A G V S V
801 GATCATGGATTTTATTATAATGAGAAAAGAACTATTGGGTGGATTTAGAAAAGCAACTTTTGCAAAGAGTTTTCTGAATGGGTCAAGGCAAGAGAGA
77▶ I M D F H Y N E K R I Y W V D L E R Q L L Q R V F L N G S R Q E R
901 GTATGTAATATAGAGAAAATGTTTCTGGAATGGCAATAAATTGGATAAATGAAGAAGTTATTTGGTCAAATCAACAGGAAGGAATCATTACAGTAACAG
111▶ V C N I E K N V S G M A I N W I N E E V I W S N Q Q E G I I T V T
1001 ATATGAAAGGAAATAATCCACATTCTTTAAGTGCTTTAAATATCCTGCAAATGTAGCAGTTGATCCAGTAGAAAAGTTTATATTTTGTCTTTCAGA
144▶ D M K G N N S H I L L S A L K Y P A N V A V D P V E R F I F W S S E
1101 GGTGGCTGGAAGCCTTTATAGACAGATCTCGATGGTGTGGGAGTGAAGGCTCTGTTGGAGACATCAGAGAAAATAACAGCTGTGTCATTGGATGTGCTT
177▶ V A G S L Y R A D L D G V G V K A L L E T S E K I T A V S L D V L
1201 GATAAGCGGCTGTTTTGGATTAGTACAACAGAGAAGGAAGCAATTCTTATTTGCTCCTGTGATTATGATGGAGGTTCTGTCCACATTAGTAAACATC
211▶ D K R L F W I Q Y N R E G S N S L I C S C D Y D G G S V H I S K H
1301 CAACACAGATAATTTGTTGCAATGTCCTTTTGGTGACCGTATCTTCAATCAACATGGAAAATGAAGACAATTTGGATAGCCAAACACACTGG
244▶ P T Q H N L F A M S L F G D R I F Y S T W K M K T I W I A N K H T G
1401 AAAGGACATGGTTAGAATTAACCTCCATTATCATTGTACCCTGGTGAAGTGAAGTAGTGCATCCACTGCACAACCAAGGCAGAAGATGACT
277▶ K D M V R I N L H S S F V P L G E L K V V H P L A Q P K A E D D T
1501 TGGGAGCCTGAGCAGAACTTTGCAAATGAGGAAAGGAACTGCAGCAGCACTGTGTGTGGCAAGACCTCCAGTACACTTGTGCATGTGTGCAGAGG
311▶ W E P E Q K L C K L R K G N C S S T V C G Q D L Q S H L C M C A E
1601 GATACGCCCTAAGTCGAGACCGGAAGTACTGTGAAGATGTTAATGAATGTGCTTTTTGGAATCATGGCTGACTCTTGGGTGTAACCAACCCCTGGATC
344▶ G Y A L S R G R K Y C E D V N E C A F W N H G C T L G C K N T P G S
1701 CTATTACTGACSTGCCCTGTAGGATTTGTTCTGCTCTGTTGGGAAACGATGTCATCAACTGTTTCTGTCACGCAATGTGTGAATGCAGCCAT
377▶ Y Y C T C P V G F V L L P D G K R C H Q L V S C P R N V S E C S H
1801 GACTGTGTTCTGACATCAGAAGTCCCTTATGTTTCTGCTGCTGAAAGGCTCAGTGTGAGAGAGATGGGAAAACATGTAGCGGTTGTTCTCACCCGATA
411▶ D C V L T S E G P L C F C P E G S V L E R D G K T C S G C S S P D
1901 ATGGTGGATGAGCCAGCTCTGCGTTCCTTAGCCAGTATCCTGGGAATGTGATTGCTTTCTGGGTATGACCTACAACCTGGATGAAAAAGCTGTGC
444▶ N G G C G S Q L C V P L S P V S W E C D C F P G Y D L Q L D E K S C A
2001 AGCTTGAGCACCACAACCTTTTGGCTGTTTGCCAAATCTCAAGATATTCGACATATTCGACATATTTGATGGAACAGACTGGAACCTCTGCTCAGCCAGCAG
477▶ A S G Q P P F L L F A N S Q D I R H M H F D G T D Y G T D L S Q Q
2101 ATGGGAATGGTTTATGCCCTAGATCATGACCCCTGTGGAAAATAAGATATACTTTGCCATACAGCCCTGAAGTGGATAGAGAGAGCTAATATGGATGGTT
511▶ M G M V Y A L D H D P V E N K I Y F A H T A L K W I E R A N M D G
2201 CCCAGCGAGAAAGGCTTATTGAGGAAGGAGTAGATGTCCAGAAGGCTTGTGTTGGACTGGATTGGCCGTAGATTCTATTGGACAGACAGAGGAAATC
544▶ S Q R E R L I E E G V D V P E G L A V D W I G R R F Y W T D R G K S
2301 TCTGATTGGAAGGAGTATTTAAATGGGAAACGTTCCAAAATAATCACTAAGGAGAACATCTCTCAACCACAGGGAATTGCTGTTTATCCAATGGCCAAG
577▶ L I G R S D L N G K R S K I I T K E N I S Q P R G I A V H P M A K
2401 AGATTATCTGGACTGATACAGGGATTAATCCACGAATGAAAGTCTTCCCTCAAGGCTTGGCCGTCTGGTTATAGCCAGCTCTGATCTAATCTGGC
611▶ R L F W T D T G I N P R I E S S S L Q G L G R L V I A S S D L I W
2501 CCAGTGAATAACGATTGACTTCTTAACGACAAGTTGACTGGTGCATGCAAGCAGTCTGTGATTGAAATGGCAATCTGGATGGTTCAAACGCCG
644▶ P S G I T I D F L T D K L Y W C D A K Q S V I E M A N L D G S K R R
2601 AAGACTTACCCAGAATGATGTAGGTCACCCATTTGCTGTAGCAGTGTGGAGATTATGTGTGTTCTCAGATTGGGCTATGCCATCAGTAATAAGAGTA
677▶ R L T Q N D V G H P F A V A V F E D Y V W F S D W A M P S V I R V
2701 AACAAAGGACTGGCAAAGATAGAGTACGTCTCAAGGCGAGATGCTGAAGCCCTCATCACTGGTGTGGTTCATCCATTGGCAAACAGGAGCAGATC
711▶ N K R T G K D R V R L Q G S M L K P S S L V V V H P L A K P G A D
2801 CCTGCTTATATCAAACGGAGGCTGTGAACATATTTGCAAAAAGAGGCTTGGAACTGCTTGGTGTCTGCTGCTGAAAGTTTTATGAAAGCCTCAGATGG
744▶ P C L Y Q N G G C E H I C K K R L G T A W C S C R E G F M K A S D G
2901 GAAAACGTGTCTGGCTCTGGATGGTCTCAGCTGTTGGCAGGTGGTGAAGTTGATCTAAAGAACCAAGTAACACCATTGGACATCTTGTCCAAGACTAGA
777▶ K T C L A L D G H Q L L A G G E V D L K N Q V T P L D I L S K T R
3001 GTGTCAGAAGATAACATTACAGAATCTCAACACATGCTAGTGGCTGAAATCATGGTGTGAGTCAAGATGACTGTGCTCCTGTGGGATGCAGCATGTATG
811▶ V S E D N I T E S Q H M L V A E I M V S D Q D D C A P V G C S M Y
3101 CTCGGTGTATTTAGAGGGAGAGGATGCCACATGTCAGTGTGTTGAAAGGATTTGCTGGGGATGGAAAACATGTTCTGATATAGATGAATGTGAGATGGG
844▶ A R C I S E G E D A T C Q C L K G F A G D G K L C S D I D E C E M G
3201 TGTCCCAGTGTGCCCCCTGCCTCCTCAAGTGCATCAACACCGAAGTGGTTATGTCTGCCGTGCTCAGAAGGCTACCAAGGAGATGGGATTCAGTGT
877▶ V P V C P P A S S K C I N T E G G Y V C R C S E G Y Q G D G I H C

3301 CTTGATATTGATGAGTGCCAACTGGGGTGCACAGCTGTGGAGAGAATGCCAGCTGCACAAATACAGAGGGAGGCTATACCTGCATGTGTGCTGGACGCC
911▶ L D I D E C Q L G V H S C G E N A S C T N T E G G Y T C M C A G R
3401 TGCTGAACCAGGACTGATTTGCCCTGACTCTACTCCACCCCTCACCTCAGGGAAGATGACCACCCTATTCGTAAGAAATAGTACTCTGAATGTCC
944▶ L S E P G L I C P D S T P P P H L R E D D H H Y S V R N S D S E C P
3501 CCTGTCCCACGATGGTACTGCCTCCATGATGGTGTGCATGTATATTGAAGCATTGGACAAGTATGCATGCAACTGTGTTGTTGGCTACATCGGGGAG
977▶ L S H D G Y C L H D G V C M Y I E A L D K Y A C N C V V G Y I G E
3601 CGATGTACAGTACCAGACCTGAAGTGGTGGAACTGCGCCACGCTGGCCACGGGACAGCAGAGAAGTGCATCGTGGTGGCTGTCTGCGTGGTGGTGGCTTG
1011▶ R C Q Y R D L K W W E L R H A G H G Q Q Q K V I V V A V C V V V L
3701 TCATGCTGCTCCTCCTGAGCCTGTGGGGGGCCACTACTACAGGACTCAGAAGCTGTATCGAAAAACCAAAGAATCCTTATGAGGAGTCGAGCAGAGA
1044▶ V M L L L L S L W G A H Y Y R T Q K L L S K N P K N P Y E E S S R D
3801 TGTGAGGAGTCGAGGCCTGCTGACACTGAGGATGGGATGTCTCTTGGCCCTCAACCTTGGTTTGTGGTTATAAAAAGAACACCAAGACCTCAAGAATGGG
1077▶ V R S R R P A D T E D G M S S C P Q P W F V V I K E H Q D L K N G
3901 GGTCAACCAGTGGTGGTGGGATGGCAGGCAGCAGATGGGTCAATGCAACCACTTCATGGAGGCAGGAGCCCGAGTTATGTGGAATGGGCACAGAGC
1111▶ G Q P V A G E D G Q A A D G S M Q P T S W R Q E P Q L C G M G T E
4001 AAGGCTGTGGATTCCAGTATCCAGTATAAGGGCTCCTGTCCCAGGTAATGGAGCGAAGCTTTTCATATGCCCTCCTATGGACACAGACCTTGAAGG
1144▶ Q G C W I P V S S D K G S C P Q V M E R S F H M P S Y G T Q T L E G
4101 GGGTGTGAGAAGCCCCATTCTCTCCTATCAGTAACCCATTATGGCAACAAAGGGCCCTGGACCCACCACCAAATGGAGTGTACTCAGTAAAACTG
1177▶ G V E K P H S L L S A N P L W Q Q R A L D P P H Q M E L T Q •

NheI (4244)

4201 GAATTAAGAAAGTCAAGAAGTGAAGTGAATGACTATGTCGATGCACAGCTAGTGGCCAGACATGATAAGATACATTGATGAGTTGGACAAACCACAAGT
4301 AATGCAGTGAATAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGC
4401 ATTCATTTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAATGTGGTATGGAATTCTAAAATACAGCATAGC
4501 AAAACTTTAACCTCCAATCAAGCCTCTACTGAACTCTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTGCCAATGTGCATTAGCTGTTT
4601 GCAGCCTCACCTTCTTCATGGAGTTAAGATATAGTGTATTTTCCAAGGTTTGAAGTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCA
4701 CATTCCCTTTTTAGTAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTT
4801 CATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCTCTGGTG
141▶ • N R T
4901 TACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTC
136▶ 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 E R
5001 TGCACATGCCACAGGGGCTGACCACCCTGATGCTGDCACCTCATCAGATAGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTTCTGCCGTT
103▶ 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 N
5101 GCTCACAGCAGACCCAATGGCAATGGCTTCAGCACAGACAGTACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCCAGTCTGGTC
70▶ 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 T
5201 CTGATGGCCGCCCCGACATGGTCTTGTGTCCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGA
36▶ 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 D Q Q S I N F
5301 AGGTCTTCATGATGGCCCTCTATAGTGAAGTATTACTATGCGGATATACTATGCGGATGATTAATTGTCAAACACGCGTGGATGGCGTCTCCAGC
3▶ T K M
5401 TIATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCATTGCGTCAATGGGGCGGAGTTGTTACGACATTT
5501 TGGAAAGTCCCGTTGATTTACTAGTCAAACAAACTCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCGTGAGTCAAACCGCTATCCACGCCCA
5601 TTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAAGTCACTGACTGGGCAT
5701 AATGCCAGGGGCCATTTACCCTGATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGTACTGCCAAGTGGGCAGTTTACCCTAAATA
5801 CTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTACTATGGGAACATACGTCAATTATTGACGTCAATGGGGGGGGTCTTGGGCGGTGAGCCA
5901 GCGGGCCATTTACCCTAAGTTATGTAACGCTGCAGGTTAATAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGCCGCGTT
6001 GCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAATAATCAGCGCTCAAGTCAAGAGTGGCGAAACCCGACAGGACTATAAGATACCAG
6101 GCGTTTTCCCCTGGAAGCTCCCTGTCGCTCTCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGTTTT
6201 CTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGTGTGCACGAACCCCGCTTACGCCGACCGCTGCGCCTT
6301 ATCCGGTAACTATCGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAA
6401 GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAA
6501 GAGTTGGTAGCTCTTGTCCGCAACAAACCACCGCTGGTAGCGGTGTTTTTTTTGTTTGAAGCAGCAGATTACGCCAGAAAAAAGGATCTCAAGA
6601 AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAACTCACGTTAAGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCG

6701 GCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAACAAAACGAAACAAA
6801 CAAACTAGCAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA