



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCGCCCTACCTGAGGCC
301 GCCATCCACGCCGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

BspHI (560)

AgeI (552)

501 TCTGTTCTGCGCGTTACAGATCCAAGCTGTGACCGGGCCTACCTGAGATCACCGGTCATCATGATTCCGGTGACAGAGCTCCGCTACTTTGCGGACAC
601 GCAGCCAGCATACCGGATCCTGAAGCCGTGGTGGGATGTGTTACAGACTACATCTCTATCGTCATGCTGATGATTGCCGCTTTCGGGGGACGCTGCAG
13▶ Q P A Y R I L K P W W D V F T D Y I S I V M L M I A V F G G T L Q
701 GTCACCCAAGACAAGATGATCTGCCTGCCTGTAAGTGGGTACCAAGGACTCCTGCAATGATTGTTCCGGGGCTGGGACGCCCTGGCCCGAGCCCA
47▶ V T Q D K M I C L P C K W V T K D S C N D S F R G W A A P G P E P
801 CCTACCCCAACTCCACATTCTGCCGACCCCTGACACGGGCCACAGGCATCAAGTATGACCTGGACCGGCACCAGTACAACCTACGTGGACGCTGTGTG
80▶ T Y P N S T I L P T P D T G P T G I K Y D L D R H Q Y N Y V D A V C
901 CTATGAAACCGACTGCCTGGTTTGCACGACTTCCCTACCTGGTGTCTTGCACAGCTCATCTTCTGGCTGCAGCAACTTCTGGTCAAATTC
113▶ Y E N R L H W F A K Y F P Y L V L L H T L I F L A C S N F W F K F
1001 CCGCGACCAGCTCGAAGCTGGAGCACTTTGTGTCTATCCTGCTGAAGTGTTCGACTCGCCCTGGACCACGAGGGCCCTGTCGGAGACAGTGGTGGAG
147▶ P R T S S K L E H F V S I L L K C F D S P W T T R A L S E T V V E
1101 AGAGCGACCCCAAGCCGCTTTCAGCAAGATGAATGGGTCCATGGACAAAAAGTCATCGACCTCAGTGAGGACGTTGGAGGCCACCGTGCCATGCTGCA
180▶ E S D P K P A F S K M N G S M D K K S S T V S E D V E A T V P M L Q
1201 GCGGACCAAGTACCGGATCGAGCAGGTATCGTGGACCGCTCAGAGACGGGCGTGTGGACAAGAAGGAGGGGGAGCAAGCCAAGGCGCTGTTTGAGAAG
213▶ R T K S R I E Q G I V D R S E T G V L D K K E G E Q A K A L F E K
1301 GTGAGAAGTTCGGACCATGTGGAGGAGGGGACATTGTGACCGCTCATCATCGGCGAGACCATCAAGGTGATCAAGTTTCAAGTTTCACTCATCATCT
247▶ V K K F R T H V E E G D I V Y R L Y M R Q T I I K V I K F I L I I
1401 GCTACACCGTCTACTACGTGCACAACATCAAGTTGACGCTGGACTGCACCGTGGACATTGAGAGCCTGACGGGCTACCGCACCTACCCTGTGCCACCC
280▶ C Y T V Y Y V H N I K F D V D C T V D I E S L T G Y R T Y R C A H P
1501 CCTGGCCACACTTCAAGATCTGGCGTCTTCTACATCAGCTAGTCATCTTCTACGGCCTCATCTGCATGTATACACTGTGGTGGATGCTACGGCGC
313▶ L A T L F K I L A S F Y I S L V I F Y G L I C M Y T L W W M L R R
1601 TCCCTCAAGAAGTACTCGTTTGAAGTCCGTCGAGGAGGACAGTACAGCGACATCCCCGACGTCGAAGAAGGACTTCGCCTTTCATGCTGCACCTCATTG
347▶ S L K K Y S I R E E S I R E E S Y S D I P D V K N D F A M L H L I
1701 ACCAATACGACCCGCTTACTCCAAGCGCTTCCCGCTTCTCGTGGAGGTGAGTGAGAACAAGCTGCGGACGATGAACCTCAACAACGAGTGGACGCT
380▶ D Q Y D P L Y S K R F A V F L S E V S E N K L R Q L N L N N E W T L
1801 GGACAAGTCCGGCAGCGGCTCACCAAGAACCGCAGGACAAGCTGGAGTGCACCTGTTTCATGCTCAGTGGCATCCCTGACACTGTGTTTGCCTGGTG
413▶ D K L R Q R L T K N A Q D K L E L H L F M L S G I P D T V F D L V
1901 GAGCTGGAGGTCTCAAGCTGGAGTGTATCCCGACGTCACCTCCCGCCAGCATTGCCAGCTCACGGGCTCAAGGAGCTGTGGCTTACCACACAG
447▶ E L E V L K L E L I P D V T I P P S I A Q L T G L K E L W L Y H T
2001 CGGCCAAGATTGAAGCCCGCTTCTGCGGAGAACCTGCGGGCGCTGCACATCAAGTTACAGATCAAGGAGATCCCGCTTGGATGCTGATGATGATG
480▶ A A K I E A P A L F L R E N L R L H I K F T D I K E I P L W I Y
2101 TAGCCTGAAGACTGGAGGAGTGCACCTGACGGGAACCTGAGCGCGGAGAACAACCGCTACATCGTCATCGAGCGGCTGCGGGAGCTCAAACGCCTC
513▶ S L K T L E E L H L T G N L S A E N N R Y I V I D G L R E L K R L
2201 AAGGTGCTGCGGCTCAAGAGCAACCTAAGCAAGCTGCCACAGTGGTACAGATGTGGGCGTGCACCTGCAGAAGCTGCCATCAACAATGAGGGACCA
547▶ K V L R L K S N L S K L P Q V V T D V G V H L Q K L S I N N E G T
2301 AGTCATCGTCTCAACAGCCTCAAGAAGTGGCAACCTGACTGAGCTGGAGCTGATCCGCTGTGACCTGGAGCGCATCCCCACTCCATCTTACGCT
580▶ K L I V L N S L K K M A N L T E L E L I R C D L E R I P H S I F S L
2401 CCACAACCTGCAGGAGATTGACCTCAAGGACAACAACCTCAAGACCATCGAGGAGATCATCAGCTTCCAGCACCTGCACCGCTCACCTGCCTAAGCTG
613▶ H N L Q E I D L K D N N L K T I E E I I S F Q H L H R L T C L K L
2501 TGGTACAACCACATCGCTACATCCCATCCAGATCGGCAACCTCACCAACCTGGAGCGCTTACCTGAACCGCAACAAGATCGAGAAGATCCCCACCC
647▶ W Y N H I A Y I P I Q I G N L T N L E R L Y L N R N K I E K I P T
2601 AGCTCTTCTACTGCGCAAGCTGCGCTACCTGGACCTCAGCCACAACAACCTGACCTTCTCCTGCGGACATCGGCTCCTGCAGAACCTCCAGAACCT
680▶ Q L F Y C R K L R Y L D L S H N N L T F L P A D I G L L Q N L Q N L
2701 AGCCATCACGGCAACCGGATCGAGACGCTCCCTCCGGAGCTTCCAGTGCAGGAGCTGCGGGCCCTGCACCTGGGCAACAACGTGCTGCAGTCACTG
713▶ A I T A N R I E T L P P E L F Q C R K L R A L H L G N N V L Q S L
2801 CCCTCCAGGGTGGGCGAGCTGACCAACCTGACGAGATCGAGTGGGGGCAACCGGCTGGAGTGCCTGCTGTGGAGCTGGGCGAGTGCCCACTGCTCA
747▶ P S R V G E L T N L T Q I E L R G N R L E C L P V E L G E C P L L

NheI (2995)

2901 AGCGCAGCGGCTTGGTGGTGGAGGAGGACCTGTTCAACACACTGCCACCCGAGGTGAAGGAGCGGCTGTGGAGGGTGCACAAGGAGCAGGCTGAGCTAG
780▶ K R S G L V V E E D L F N T L P P E V K E R L W R A D K E Q A •
3001 CTGGCCAGACATGATAAGATACATTGATGAGTTTGACAAACCACAACCTAGAATGCAGTGAATAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCT
3101 TTATTTGTAACCATTATAAGTCGAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTACAGGGGAGGTGTGGAGGTTTTTAA

3201 GCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACCTTAACTCCAATCAAGCCTCTACTTGAATCCTTTCTGAGGG
3301 ATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGACGCCTCACCTTCTTTCATGGAGTTAAGATATAGTGATTTTCCCAA
3401 GGTTTGAAGTACTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCACATTCCCTTTTTAGTAAATATTCAGAATAATTTAAATACATCATTGC
3501 AATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTT
3601 TAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCAT
3701 TCATCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATC
120 M E I 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
3801 AGAGTAGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCAATGGCAATGGCTTCAGCACAGACAGTGACCCTG
87 S Y P 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
3901 CCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCAAGTCTTGGTCTGATGGCCGCCGACATGGTGCTTGTCTCATAGAGCATGGTGA
53 G I Y A 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
4001 TCTTCTCAGTGGCAGCTCCACCAGTCCAGATCCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGA
20 K E T A V E V L E L D Q Q S I N F T K M
4101 TATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACATAACGAGCTCTGCTTATATAGACCTCCACCG
4201 TACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGAAAAGTCCCGTTGATTTACTAGTCAAAAACAACTCCATTGACGTCA
4301 ATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATA
4401 CGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAAGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACT
4501 TGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGA
4601 ACATACGTCATTATTGACGTCAATGGCGGGGGTCTGTTGGCGGTGAGCCAGGCGGGCCATTTACCGTAAAGTTATGTAACGCCTGCAGGTTAATTAAGAA
4701 CATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGCCCTGACGAGCATCACAAAAAT
4801 CGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGC
4901 CGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGAAGCGTGGCGTTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTC
5001 CAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGTGGCGCTTATCCGGTAACTATCGTCTTGTAGTCCAACCCGGTAAGACACGACTTATCG
5101 CCACTGGCAGCAGCCACTGGTAAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCCTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
5201 GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGG
5301 TTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAC
5401 TCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGAATAAAAATATCTTTATTTTATTACATCTGTGTGTTGTTTTT
5501 GTGTGAATCGTAACTAACATACGCTCTCCATCAAACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACA
5601 TTTCTCTATCGAA