



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGAGAAGTTGGGGGAGGGGTCGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGGTAACCTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGAGCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

SphI (560)

AgeI (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGTCCAGCATGCGATGCTCTCAGTCCCGCTCGCCAAACTACTGTAG
13▶ A K C Q K K A W P D H K R E C K C L K S C K P R Y P P D S V R L L
701 GGCAAGTGTCTTCAAATTTGATGGAGCACCTTCAGAATCAGAGAAGCTTTACTCATTTTATGATCTGGAGTCAAATATTAACAACTGACTGAAG
47▶ G R V V F K L M D G A P S E S E K L Y S F Y D L E S N I N K L T E
801 ATAAGAAAGAGGGCCTCAGGCAACTCGTAATGACATTTCAACATTTTCATGAGAGAAGAAATACAGGATGCCTCTCAGTCCACCTGCCTTTGACCTTTT
80▶ D K K E G L R Q L V M T F Q H F M R E E I Q D A S Q L P P A F D L F
901 TGAAGCCTTTGAAAAGTATCTGCAACTCTTCCACATCTGTAATGCGGAGATGCAGGAAGTTGGTGTGGCCTATATCCAGTATCTCTTTGCTCAAT
113▶ E A F A K V I C N S F T I C N A E M Q E V G V G L Y P S I S L L N
1001 CACAGCTGTGACCCCACTGTTGATTGTTCATGGGCCACCTTACTGCGAGCAGTCCGAGACATCGAGGTGGGAGAGGAGCTCACCATCTGCT
147▶ H S C D P N C S I V F N G P H L L L R A V R D I E V G E E L T I C
1101 ACCTGGATATGCTGATGACCACTGAGGAGCGCCGGAAGCAGCTGAGGGACCACTGCTTTGAATGTGACTGTTCCGTTGCCAAACCCAGGACAAGGA
180▶ Y L D M L M T S E E R R K Q L R D Q Y C F E C D C F R C Q T Q D K D
1201 TGCTGATATGCTAACTGGTATGAGCAAGTATGGAAGGAAGTTCAAGAATCCCTGAAAAAATTGAAGAACTGAAGGCACACTGGAAGTGGGAGCAGGTT
213▶ A D M L T G D E Q V W K E V Q E S L K K I E E L K A H W K W E Q V
1301 CTGGCCATGTGCCAGGCAATCATAAGCAGCAATTCTGAACGGCTTCCCGATATCAACATCTACCAGCTGAAGGTGCTCGACTGCGCCATGGATGCCTGCA
247▶ L A M C Q A I I S S N S E R L P D I N I Y Q L K V L D C A M D A C
1401 TCAACCTCGCCCTGTTGGAGGAAGCCTTGTCTATGGTACTCGGACCATGGAGCCATACAGGATTTTTTCCAGGAAGCCATCCCGTCCAGAGGGTTCA
280▶ I N L G L L E E A L F Y G T R T M E P Y R I F F P G S H P V R G V Q
1501 AGTGATGAAAGTTGGCAAACCTGAGCTACATCAAGGCATGTTTCCCAAGCAATGAAGAATCTGAGACTGGCTTTTATATTATGAGAGTGACACATGGC
313▶ V M K V G K L Q L H Q G M F P Q A M K N L R L A F D I M R V T H G

NheI (1694)

1601 AGAGAACACAGCCTGATTGAAGATTTGATTCTACTTTTAGAAGAATGCGAGCCAACATCAGAGCATCCTAAGGGAACGCAGTCAGAGGGAATGCTAGC
347▶ R E H S L I E D L I L L L E E C D A N I R A S •
1701 TGGCCAGACATGATAAGATACATTGATGAGTTTGGACAACCAACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTT
1801 TATTTGTAACCATTATAAGTGAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTACAGGGGAGGTGTGGGAGGTTTTTAAAG
1901 CAAGTAAACCTCTACAAATGTGGTATGGAATCTAAAATACAGCATAGCAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGA
2001 TGAATAAGGCATAGGCATCAGGGGCTGTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTCTTTTCATGGAGTTTAAAGATATAGTATTTTCCCAAG
2101 GTTTGAAGTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCACATTCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCA
2201 ATGAAAATAAATGTTTTTTATTAGGCAGAAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTT
2301 AATAGAAATGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCTGGTGTACTTGGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATT
141▶ • N R T Y K L P I L E E I T T K V L K G N
2401 CATCTCAATGAGCACAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTGACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCA
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
2501 GAGTAGGGGTGCCTGACAGCCCAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCAATGGCAATGGCTTCAGCACAGACAGTACCCTGC
86▶ 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 G
2601 CAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCACTGTTGGTCTGATGGCCGCCCGACATGGTCTTGTGCTCATAGAGCATGGTGAT
53▶ 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
2701 CTTCTCAGTGGCAGCTCCACAGCTCCAGATCCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGAT
20▶ K E T A V E V L E L D Q Q S I N F T K M
2801 ATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCACCGT
2901 ACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAAACAACTCCATTGACGTCAA
3001 TGGGGTGGAGACTTGAAATCCCGTGGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAACCGCATCATCATGGTAATAGCGATGACTAATAC
3101 GTAGATGACTGCCAAGTAGGAAAGTCCATAAGGTGATGACTGGGCATAATGCCAGCGGGCCATTTACCGTCAATTGACGTCAATAGGGGGCTACTT

3201 GGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTACTATGGGAA
3301 CATA CGTCATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAAC
3401 ATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATC
3501 GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTTCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCC
3601 GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCC
3701 AAGCTGGGCTGTGTGCACGAACCCCCGTTTCAGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
3801 CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
3901 AACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGT
4001 TTTTTTGTTCGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAAT
4101 CACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTG
4201 TGTGAATCGTAACATAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACAT
4301 TTCTCTATCGAA