



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTGATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTCTGCCGCTCCCGCTGTGGTGCCTCTGAACTGCGTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

NcoI (560)
BstEII (555)
AgeI (552)

501 TCTGTTCTGCGCCGTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGTACCATTGGCTACCATTAGCTCTTCCCTGGAGCCTGTCCCTCTA
601 CGTCTTCTCCTGCTCCTGGCTACACCTGGGCATCTGCAGCAGTGAAAACTGTTCCCATCTTGAATGCTTCTACAACCTCAAGACCAATGTCTCTTGC
131 V F L L L L A T P W A S A A V K N C S H L E C F Y N S R A N V S C
701 ATGTGGAGCCATGAAGAGGCTCTGAATGTCACAACCTGCCAGTCCATGCCAAGTCCGAACCTGCGACACTGGAACAAAACCTGTGAGCTAATCTTGTGA
471 M W S H E E A L N V T T C H V H A K S N L R H W N K T C E L T L V
801 GGCAGGCATCCTGGGCTGCAACCTGATCCTCGGGTGTTCAGAGTCCCAGTCACTGACCTCCGTGGACCTCCTTGACATAAATGGTGTGCTGGGA
801 R Q A S W A C N L I L G S F P E S Q S L T S V D L L D I N V V C W E
901 AGAGAAGGTTGGCGTAGGGTAAAGACCTGCGACTTCCATCCCTTTGACAACCTTCGCTGGTGGCCCTCATTCCCTCCAAGTCTGCACATTGATACC
1131 E K G W R R V K T C D F H P F D N L R L V A P H S L Q V L H I D T
1001 CAGAGATGTAACATAAGCTGGAAGTCTCCAGGTCTCTCACTACATTGAACCATACTTGAATTTGAGGCCCGTAGACGCTTCTGGGCCACAGCTGGG
1471 Q R C N I S W K V S Q V S H Y I E P Y L E F E A R R R L L G H S W
1101 AGGATGCATCCGTATTAAGCTCAAGCAGAGACAGCAGTGGCTCTTCTGGAGATGCTGATCCCTAGTACCTCATAGAGTCCAGGTGAGGTCAAAAGC
1801 E D A S V L S L K Q R Q Q W L F L E M L I P S T S Y E V Q V R V K A
1201 TCAACGAAACAATACCGGACCTGGAGTCCCTGGAGCCAGCCCTGACCTTTGCGACAAGGCCAGCAGATCCCATGAAGGAGATCCTCCCATGTCATGG
2131 Q R N N T G T W S P W S Q P L T F R T R P A D P M K E I L P M S W

NcoI (1384)

1301 CTCAGATACCTTCTGCTGGTCTTGGTGTGTTTTCTGGCTTCTTCTCCTCGCTACATTTTGGTCAAGTGCCGGTACCTTGGGCCATGGCTGAAGACAG
2471 L R Y L L L V L G C F S G F F S C V Y I L V K C R Y L G P W L K T
1401 TTCTCAAGTGCCACATCCAGATCCTTCTGAGTCTTCTCCAGCTGAGCTCCAGCATGGGGGAGACCTTCAGAAATGGCTCTCCTCGCTGCCCTTCCCTT
2801 V L K C H I P D P S E F F S Q L S S Q H G G D L Q K W L S S P V P L
1501 GTCCTTCTCAGCCCCAGTGGCCCTGCCCTGAGATCTCTCCGCTGGAAGTCTCGACGGAGATTCCAAGGCCGTGCAGCTGCTCCTGTTACAGAAGGAC
3131 S F F S P S G P A P E I S P L E V L D G D S K A V Q L L L L Q K D
1601 TCTGCCCTTTACCTCGCCAGCGCCACTCACAGGCCAGCTGCTTACCAACCAGGGTACTTCTTCCATCTGCCAATGCCTTGGAGATCGAAT
3471 S A P L P S P S G H S Q A S C F T N Q G Y F F F H L P N A L E I E
1701 CCTGCCAGGTGTACTTACCTATGACCCTGTGTGGAAGAGGAGGTGGAGGAGGTGGTCAAGGCTGCCCGAGGGATCTCCCAACCCACCTCTGCTGCC
3801 S C Q V Y F T Y D D P C V E E E V E D G S R L P E G S P H P L L P
1801 TCTGGCTGGAGAACAGGATGACTACTGTGCCCTTCCCGCCAGGGATGACCTGCTTCTTCTCCCGAGCTCAGCACCCCAACTGCCTATGGGGGC
4131 L A G E Q D D Y C A F P P R D D L L L F S P S L S T P N T A Y G G
1901 AGCAGAGCCCCTGAAGAAAGATCTCCACTCTCCCTGCATGAGGGACTTCCCTCCCTAGCATCCCGTACCTGATGGGCTTACAGCGCCCTCTGGAGCGGA
4471 S R A P E E R S P L S L H E G L P S L A S R D L M G L Q R P L E R
2001 TGCCGAAAGGTGATGGAGAGGGCTGTCTGCCAATAGCTCTGGGGAGCAGGCCAGTGTCCAGAAGGCAACCTTCATGGGCAAGATCAGGACAGAGGCCA
4801 M P E G D G E G L S A N S S G E Q A S V P E G N L H G Q D Q D R G Q

NheI (2198)

2101 GGGCCCCATCCTGACCCTGAACACCGATGCCTATCTGTCTTCAAGAACTACAGGCCAAGATTGAGTCCACCTAATATAGCAGGTGGCCAGGACTGGC
5131 G P I L T L N T D A Y L S L Q E L Q A Q D S V H L I •
2201 TAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAGGAAATGCTTTATTTGTGAAATTTGTGATGCTATT
2301 GCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTAAACAACAACAATTGCATTATTTATGTTTCAGGTTTCAGGGGAGGTGTGGAGGTTTTTT
2401 AAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGA
2501 GGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAAGATATAGTATTTC
2601 CAAGTTTGAAGTACTTCTCATTCTTTATGTTTTAAATGCACTGACCTCCACATTCCTTTTTAGTAAATATTCAGAAATAATTTAAATACATCAT
2701 TGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTATGAGTTGACTTAGGGAACAAGGAAC
2801 CTTAATAGAAATTTGACAGCAAGAAGCGAGCTTCTAGCTTTAGTCTCGGTGACTTGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGC
1411 • N R T Y K L P I L E E I T T K V L K G
2901 CATTCTCAATGAGCACAAGCAGTCAAGGAGCATAGTCAAGATGAGCTCTGACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTC
1211 N 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
3001 ATCAGAGTAGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTACAGCAGACAGTACCT
881 D 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

3101 CTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCAGTCTTGGTCTGATGGCCGCCCGACATGGTGTCTTGTGCTCATAGAGCATGG
54 R 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
3201 TGATCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAAGTGAAGTCTATTATACTATGC
21 I K E T A V E V L E L D Q Q S I N F T K M
3301 CGATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTCACTAAACGAGCTCTGCTTATATAGACCTCCCA
3401 CCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAAAACAACTCCCATGACG
3501 TCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTA
3601 ATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGT
3701 ACTTGGCATATGATACTTGTACTGCTGACTGCCAAGTGGGCGAGTTTACCGTAAATACTCCACCCATTGACGTCAATGAAAGTCCCTATTGGCGTTACTATG
3801 GGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAAGTTATGTAACGCTGCAGGTTAATTAA
3901 GAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAA
4001 AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGTCCCTCGTGCCTCTCTCTGTTCCGACCC
4101 TGCCGTTACCGGATACCTGTCCGCCTTCTCCCTTCGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCCG
4201 CTCCAAGCTGGGCTGTGTGCACGAACCCCGTTCAGCCCGACCGTGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTA
4301 TCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
4401 GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGG
4501 TGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAA
4601 AACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTT
4701 TTTGTGTAATCGTAACTAACATACGCTCTCCATCAAAAACAAAACGAAACAAAACAACTAGCAAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGA
4801 ACATTTCTCTATCGAA