



1 GGATCTGCGATCGCTCCGGTGCCCGTCACTGGGCGAGAGCGACATCGCCACAGTCCCGGAGAAGTTGGGGGAGGGGTCGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTCTGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGAGCTACCTAGACTCAGCGGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

BspLU111 (560)

Agel (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTCAACATGTATTCCCTGGGICAGGAAGTGGGGCTGCTGGTGA
1 M Y S P G S G S G A A G E
601 GAGGAAGCTTTGTCTCTCTCTGCTCCTCATCGGTGCCTGGGCTGTGCTATCTGTACGGGAACCTGTGGACGACATCTGCATAGCGAAGCCCCGA
13 R K L C L L S L L L I G A L G C A I C H G N P V D D I C I A K P R
701 GACATCCCGTGAATCCCTTGTGCATTTACCGTCCCTGGGAAGAAGGCCACCGAGGAGGATGGCTCAGAGCAGAAGGTTCCAGAAGCCACCAACCGGC
47 D I P V N P L C I Y R S P G K K A T E E D G S E Q K V P E A T N R
801 GGTCTGGGAAGTGTCCAAGGCAATTCGCGATTTGCCACTAATCTACCAGCAGCTGGCAGACTCCAAGAATGACAACGACAACATTTTCTGTCCACC
80 R V W E L S K A N S R F A T N F Y Q H L A D S K N D N D N I F L S P
901 CTTGAGCATCTCCACTGCTTTTGTATGACCAAGTGGGTGCCTGTAACGACACTCTCAAGCAGCTGATGGAGGTTTTTAAATTTGATACCATCTCCGAG
113 L S I S T A F A M T K L G A C N D T L K Q L M E V F K F D T I S E
1001 AAGACATCCGACCAGATCCACTTCTTCTTTGCCAAACTGAACTGCCGACTCTATCGAAAAGCCAACAAGTCTCTGACTTGGTATCAGCCAACCGCCTTT
147 K T S D Q I H F F F A K L N C R L Y R K A N K S S D L V S A N R L
1101 TTGGAGACAAATCCCTCACCTTCAACGAGAGCTATCAAGAGCTTAGTGAGGTTGTCTATGGAGCCAAGCTCCAGCCCITGGACTTCAAGGAGAATCCGGA
180 F G D K S L T F N E S Y Q D V S E V V Y G A K L Q P L D F K E N P E
1201 GCAATCCAGAGTGACCATCAACAACCTGGTAGCTAATAAGACTGAAGCCGCATCAAAGATGTCATCCACAGGGCCATTAAACGAGCTCACGGCCCTG
213 Q S R V T I N N W V A N K T E G R I K D V I P Q G A I N E L T A L
1301 GTGCTGGTTAACACCATTTACTTCAAGGCGCTGTGGAAGTCAAAGTTCAGCCCTGAGAACACAAGGAAGAACCGTCTATAAGTTCGATGGGCGATCAT
247 V L V N T I Y F K G L W K S K F S P E N T R K E P F Y K V D G Q S
1401 GCCAGTGCCTATGATGTACCAGGAAGCAAATCAAATACCGCGCGTGGCAGAGGGCACCAGGTGCTAGAGCTGCCCTTCAAGGGGGATGACATCAC
280 C P V P M M Y Q E G K F K Y R R V A E G T Q V L E L P F K G D D I T
1501 CATGGTGTCTATCCTGCCAAGCCTGAGAAGAGCCTGGCAAGGTGGAGCAGGAGCTCACCCAGAGCTGCTCAGGAGTGGCTGGATGAGCTGTCAGAG
313 M V L I L P K P E K S L A K V E Q E L T P E L L Q E W L D E L S E
1601 ACATGCTTGTGGTCCACATGCCCGCTTCCGACCGAGGATGGCTTCACTGTAAGGAGCAGCTGCAAGACATGGGCTCATTGATCTCTCAGCCCTG
347 T M L V V H M P R F R T E D G F S L K E Q L Q D M G L I D L F S P
1701 AAAAGTCCCAACTCCCAGGGATCGTTGCTGGAGGCAGGACGACTCTATGTCTCCGACGATTCCACAAAGCATTCTTGAGGTAATGAGGAAGGCAG
380 E K S Q L P G I V A G G R D D L Y V S D A F H K A F L E V N E E G S
1801 TGAAGCAGCAGGAGTACTTCTGCTGATTACTGGCCGCTCACTGAACCCCAATAGGGTACCTTCAAGGCCAACAGGCCCTTCTGTTTCTATAAGG
413 E A A A S T S V V I T G R S L N P N R V T F K A N R P F L V L I R

NheI (1988)

1901 GAAGTTGCACTGAACACTATTATATTCATGGGAGAGTGGCTAATCCTTGTGTAACATAAATAATCTTTGCACCTTTTCTACTTTGGCTAGCTGGCAA
447 E V A L N T I I F M G R V A N P C V N •
2001 GACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTG
2101 TAACCATTATAAGCTGCAATAAACAAGTAAACAACAACATTCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGTTTTTAAAGCAAGTA
2201 AAACCTCTACAAATGTGGTATGGAATCTAAAATACAGCATAGCAAACTTTAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATA
2301 AGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTTCATGGAGTTTAAAGATATAGTGTATTTTCCAAGGTTTGA
2401 ACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCACATTCCTTTTTAGTAAAATATTCAGAAAATATTTAAATACATCATTGCAATGAAA
2501 ATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGAACCTTAATAGA
2601 AATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCATCTC
141 • N R T Y K L P I L E E I T T K V L K G N M E
2701 AATGAGCACAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAG
118 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 S Y
2801 GGTGCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCCAGCAGACAGTACCCTGCCAATGT
84 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 I Y
2901 AGGCCTCAATGTGGACAGCAGAGATGATCTCCCGAGTCTTGGTCTGATGGCCGCCCCGACATGGTGTCTGTTGTCCTCATAGAGCATGGTATCTTCTC
51 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 K E
3001 AGTGGCGACCTCCACCAGCTCCAGATCTGCTGAGAGATGTTGAAGGCTTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTA
18 T A V E V L E L D Q Q S I N F T K M
3101 TGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTCACTAAACGAGCTCTGCTTATATAGACTCCACCGTACACGC

3201 CTACCGCCATTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAACAACTCCCATTGACGTCAATGGGGT
3301 GGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGAT
3401 GTACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATA
3501 TGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACG
3601 TCATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCTGCAGGTTAAITAAAGAACATGTGA
3701 GCAAAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCT
3801 CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAAGTCCCTCGTGCCTCTCTGTTCCGACCTGCCGTTAC
3901 CGGATACCTGTCCGCTTTCTCCCTTCGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTG
4001 GGCTGTGTGCACGAACCCCGTTCCAGCCGACCGCTGCGCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATGCCACTGG
4101 CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGT
4201 ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTT
4301 GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACCTCACGTT
4401 AAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAA
4501 TCGTAACTAACATACGCTCTCCATCAAACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCT
4601 ATCGAA