



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTTCGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCAGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTCTGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCAACTCTACGCTTTTGTTCGTTT

BspHI (560)

AgeI (552)

501 TCTGTTCTGCGCGTTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGTCATGAAGATGGCGGCTCCACCGCCAACAAAGGACGCCTC
1 M K M A A P T A N K A A S
601 CCTGGGCTGTAACAACAAGCCTGCGTTCGGGAGCTGGATTTCAGGTCGGGAGCTCGGGTGGAGGAATTGAACAAACTCATCAAGAATTTACGAAGCAC
13 L G C N N K P A F P E L D F R S G A R V E E L N K L I Q E F T K H
701 GACCAGCGGGAATACGACGACCAGAGAGCGCTGGAGATTACACAGCCAAGGATTTTCATCTTTCCATGCTGGGAATGGTCAAAAAGCTGGACCAAAAGC
47 D Q R E Y D D Q R A L E I H T A K D F I F S M L G M V Q K L D Q K
801 TTCAGTGGCTAATGAATACCTGTTGCTCTCTGGAGGTGTCCGGGAAGGCGTGGTGGACCTGGACTTAGATGAGCTTAATGTCTATGCCCGGGTACTGA
80 L P V A N E Y L L L S G G V R E G V V D L D L D E L N V Y A R G T D
901 CTATGATATGGACTTTACCCTCTGGTGCCAGCCTCAAGTCGATGACCGTAAATCAGCCTGTGACACTCGACATCGGCCACTCAGCCTTGCCACTCT
113 Y D M D F T L L V P A L K L H D R N Q P V T L D M R H S A L C H S
1001 TGGCTGAGCCTTCGGCTCTTTGATGAGGGGACAATCAGTAAATGAAAAGACTGCTGCACCATTGTAGATCACATCAATGGTGCACCAACTACTTCTTCT
147 W L S L R L F D E G T I S K W K D C C T I V D H I N G A T N Y F F
1101 CACCTACCAAAGTGGCTGACTGGTCTATGACTCTATCAGCATTGTCTATCAGAAATACAGAAGAAACCCAGCGAGGGATGCCAAAGGTAGAAAAGT
180 S P T K V A D W F Y D S I S I V L S E I Q K K P Q R G M P K V E K V
1201 GGAAAAAATGGACCATCATCTCCATCATTCTGGGTGTAGGGAGTAGTCGCATGTTGTATGATATTGTCCTGTGGTATCTTTCAAAGTTGGCCTGCA
213 E K N G T I I S I I L G V G S S R M L Y D I V P V V S F K G W P A
1301 GTGGCCAGAGTGGCTCATGGAGAACCCTTTGGGATGGGAAGATTACTGAGGAAGAGTTCATCAGTGGGTTTACTGGTGCCTGCTGCTCCTCTACA
247 V A Q S W L M E N H F W D G K I T E E E V I S G F Y L V P A C S Y
1401 AGGTAAGAAGGACAATGAATGGCGGCTGTCTTTGCCAGGAGCGAGGTGCAGTTGAAGAAGTGCATCTCCAGCAGCCTCATGAGGCCTATCAGGCCTG
280 K G K K D N E W R L S F A R S E V Q L K K C I S S S L M Q A Y Q A C
1501 CAAAGCCATCATATAAAGTGTGTCGCGCCCAAGGCTATTAGCCCTATCACCTGCGGAGCATGATGCTCTGGGCTGCGACAGACTTCTGCCAAC
313 K A I I I K L L S R P K A I S P Y H L R S M M L W A C D R L P A N
1601 TACTTGGCTCAAGAAGACTATGCAGCCACTTTTTGCTGGGCTCATGATGACCTGCAACTGTCTGGTCAACAAGATGTGCCCAATTATTTTCATCC
347 Y L A Q E D Y A A H F L G L G L I D L Q H C L V N K M C P N Y F I
1701 CTCAGTGAACATGCTGGAACATCTGTCTGAGGAGACAGTCTGCTTACGDCCGGAAGCTGCTCTGTGCGCTCAGACCCGGCAGAGCACTTGCAGC
380 P Q C N M L E H L S E E T V M L H A R K L S S V R S D P A E H L R T
1801 CGCATTGAGCATGTCAAGGAGCAACCGGCTGACTGAGGCTCCAGAGGCGAGGTAGCACCACCAGCATCCCCTCTCCACAGTCTGACGGAGGGGAC
413 A I E H V K A A N R L T L E L Q R R G S T T S I P S P Q S D G G D
1901 CCCAACAGCCTGATGACCGTTTGGCAAAAAAAGTGCAGCAGTAGTACTGAGAACCCGGGAAAGTCAATCTGTCTTTATCAATCCTGACGATGTCA
447 P N Q P D D R L A K K L Q Q L V T E N P G K S I S V F I N P D D V

NheI (2038)

2001 CAAGGCCATTTTCAAGATTGATGACAAATTTTTCTGAGCTAGTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAAGTGAATGCA
480 T R P H F R I D D K F F •
2101 GTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCA
2201 TTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACT
2301 TTAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCC
2401 TCACCTTCTTTCATGGAGTTAAGATATAGTGTATTTCCCAAGGTTTGAAGTACTCTTCTTTTATGTTTTAAATGCACTGACCTCCACATTCC
2501 CTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAAATCCAGATGCTCAAGGCCCTTCATAAT
2601 ATCCCCAGTTTAGTAGTTGACTTAGGGAACAAGGAACCTTAATAGAATTTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTG
141 • N R T Y K
2701 AGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCATTCTCAATGAGCACAAGCAGTCCAGGAGCATAGTCAGAGATGAGCTCTCTGCACA
134 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 C M
2801 TGGCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCCGTGTCTCAC
101 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 S V
2901 AGCAGACCCAATGGCAATGGCTTCCAGCAGACAGTACCCTGCAATGTAGGCTCAATGTGGACAGCAGAGATGATCTCCCGAGTCTTGGTCTGATG
68 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 R I
3001 GCCGCCCCGACATGGTGTGTTGTCTCATAGAGCATGGTGTCTTCTCAGTGGCGACTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGAAGGTCT
34 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 T K
3101 TCATGGTGGCCCTCTATAGTGTGATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCT
1 M

3201 GACGGTTCATAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAA
3301 GTCCCGTTGATTTACTAGTCAAAAACAACTCCCATTTGACGTCAATGGGGTGGAGACTTGAAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATG
3401 TACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCC
3501 AGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCAC
3601 CCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGGGGGGTCTTGGGCGGTGAGCCAGGCGGG
3701 CCATTTACCGTAAGTTATGTAACGCTGCAGGTTAATAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGCCGCGTTGCTGGC
3801 GTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
3901 CCCCCTGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATA
4001 GCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTCAGCCCGACCGTGCCTTATCCGG
4101 TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT
4201 ACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTG
4301 GTAGTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCC
4401 TTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCA
4501 ATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAACT
4601 AGCAAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA