



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCGCGTTCTGCCGCTCCCGCTGTGGTGCTCTGAACTGCGTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

NcoI (560)

AgeI (552)

501 TCTGTTCTGCGCGTTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGTGAAAAGATCTATTCCAGAGGGGAGCTTACCACCTT
1 M E K I Y S R G E L H H F
601 CATTGACGGCTTAAATGAAGAGAAAAGCAACTGGATGCGCTATGTGAATCCAGCACACTCTCCCCGGGAGCAAAAACCTGGCTGCGTGTGAGAACGGGATG
13 I D G F N E E K S N W M R Y V N P A H S P R E Q N L A A C Q N G M
701 AACATCTACTTCTACACCATTAAGCCCATCCCTGCCAACCCAGGAACCTTCTGTGTGGTATTGTCGGGACTTTGCAGAAAGGCTTCACTACCTTATCCCG
47 N I Y F Y T I K P I P A N Q E L L V W Y C R D F A E R L H Y P Y P
801 GAGAGCTGACAATGATGAATCTCACACAAAACAGAGCAGTCTAAAGCAACCGAGCACTGAGAAAAATGAACTCTGCCAAAAGAAATGTCCAAAAGAGAGA
80 G E L T M M N L T Q T Q S S L K Q P S T E K N E L C P K N V P K R E
901 GTACAGCGTGAAGAAAATCTAAAATTGGACTCCAACCCCTCAAAGGAAAGGACCTTACCCTTCAACATTTACCCCTCACATCAGAAAAGGACCTC
113 Y S V K E I L K L D S N P S K G K D L Y R S N I S P L T S E K D L
1001 GATGACTTTAGAAGAGCTGGGAGCCCCGAAATGCCCTTCTACCTCGGGTCTTTACCCATCCGGGCCCCTCTGCCAGAAGACTTTTGAAGCTTCCC
147 D D F R R R G S P E M P F Y P R V V Y P I R A P L P E D F L K A S
1101 TGGCTACGGGATCGAGAGACCACGTACATCACTCGTCCCCATTCCATCTCCACCACTCAAGCCCTCTGCAAGAAGCAGCCCCGACCAAAGCT
180 L A Y G I E R P T Y I T R S P I P S S T T P S P S A R S S P D Q S L
1201 CAAGAGCTCCAGCCCTCACAGCAGCCCTGGGAATACGGTGTCCCTGTGGGCCCGGCTCTCAAGAGCACCAGGACTCCTACGCTTACTTGAACGCGTCC
213 K S S S P H S S P G N T V S P V G P G S Q E H R D S Y A Y L N A S
1301 TACGGCACGGAAGTTGGGCTCCTACCTGGCTACGCAACCCCTGCCACCTCCCGCCAGCTTTCATCCCTCGTACAACGCTCACTACCCCAAGTTC
247 Y G T E G L G S Y P G Y A P L P H L P P A F I P S Y N A H Y P K F
1401 TCTTCCCCCTACGGCATGAATTGTAATGGCCTGAGCGTGTGAGCAGCATGAATGGCATCAACAACCTTTGGCCTTCCCGAGGCTGTGCCTGTCTA
280 L L P P Y G M N C N G L S A V S S M N G I N N F G L F P R L C P V Y
1501 CAGCAATCTCCTCGTGGGGGAGCCTGCCACCCCATGCTCAACCCACTTCTCTCCGAGCTCGTGCCTCAGATGGAGCCCGAGGTTGCTCCAG
313 S N L L G G G S L P H P M L N P T S L P S S L P S D G A R R L L Q
1601 CCGGAGCATCCCAGGAGGCTTGTCCCGGCCCCACAGTGCCTTCTCTTTACCGGGGCCCGCCGAGCATGAAGGACAAGGCTGTAGCCCCACAA
347 P E H P R E V L V P A P H S A F S F T G A A A S M K D K A C S P T
1701 GCGGGTCTCCACGCGGGAACAGCCGACGCGAGAACATGTGGTGAGCCCAAAGCTACCTCAGCAGCGATGGCAGCCCCAGCAGCAGCAAGCCAT
380 S G S P T A G T A A T A E H V V Q P K A T S A A M A A P S S D E A M
1801 GAATCTCATAAAAAACAAAAGAAACATGACCGGCTACAAGACCTTCCCTACCCGCTGAAGAAGCAGAACGGCAAGATCAAGTACGAATGCAACGTTTGC
413 N L I K N K R N M T G Y K T L P Y P L K K Q N G K I K Y E C N V C
1901 GCCAAGACTTTCCGGCAGCTCTCAATCTGAAGGTCCACCTGAGAGTGCACAGTGGAGAACGGCTTTCAAATGTCAGACTTCAACAAGGGCTTACTC
447 A K T F G Q L S N L K V H L R V H S G E R P F K C Q T C N K G F T
2001 AGCTCGCCACCTGCAGAAACACTACCTGGTACACACGGGAGAAAAGCCACATGAATGCCAGTCTGCCACAAGAGATTTAGCAGCACCAGCAATCTCAA
480 Q L A H L Q K H Y L V H T G E K P H E C Q V C H K R F S C T S N L K
2101 GACCCACCTGCGACTCCATTCTGGAGAGAAACCATAACCAATGCAAGGTGTGCCTGCCAAGTTCACCCAGTTTGTGCACCTGAAACTGCACAAGCGTCTG
513 T H L R L H S G E K P Y Q C K V C P A K F T Q F V H L K L H K R L
2201 CACACCCGGGAGCGGCCCAAGTGTCCCAGTGCCACAAGAACTACATCCATCTGTAGCCTCAAGGTTACCTGAAAGGAACTGCGCTGCGGCC
547 H T R E R P H K C S Q C H K N Y I H L C S L K V H L K G N C A A A
2301 CGGCGCTGGGCTGCCCTTGAAGATCTGACCCGAATCAATGAAGAAATCGAGAAGTTTGACATCAGTGAATGCTGACCGCTCGAGGACGTGGAGGA
580 P A P G L P L E D L T R I N E E I E K F D I S D N A D R L E D V E D
2401 TGACATCAGTGTGATCTCTGTAGTGGAGAAGGAAATCTGGCGTGGTCAAAAAGAGAAAAGAAACTGGCCTGAAAGTGTCTTTGAAAAGAAACATG
613 D I S V I S V V E K E I L A V V R K E K E E T G L K V S L Q R N M
2501 GGGAAATGGACTCCTCTCAGGCTGACGCTTTATGAGTATCAGATCTACCCCTCATGAAGTTGCCTCCAGCAACCCACTACCTCTGGTACCTGTAA
647 G N G L L S S G C S L Y E S S D L P L M K L P P S N P L P L V P V

NheI (2648)

2601 AGGTCAAACAAGAAACAGTTGAACCAATGGATCCTTAAGATTTTCCAGAGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAA
680 K V K Q E T V E P M D P •
2701 CTAGAATGCAGTGAAGAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGAATAAACAAGTTAACAACAACA
2801 TTGCATTCATTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTCATAAATACAGCA
2901 TAGAAAACCTTTAACCTCAAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTGCCAATGTGCATTAGCT
3001 GTTTCAGCCTCACCTTCTTTCATGAGTAAAGATATAGTGTATTTCCCAAGTGTGAACTAGCTCTTCATTTCTTATGTTTTAAATGCACTGACCT
3101 CCCACATTCCTTTTGTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTATTAGGCAGAAATCCAGATGCTCAAGGC

3201 CCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTAAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCT
141 • N R

3301 GGTGTA CTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCATCTCAATGAGCACAAGCAGTCAGGAGCATAGTCAGAGATGAGC
138 T Y K 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

3401 TCTCTGCACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCCTTCTGCC
104 E R C M 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

3501 CGTTGCTCACAGCAGACCAATGGCAATGGCTT CAGCACAGACAGTGACCCTGCCAATGTAGGCTCAATGTGGACAGCAGAGATGATCTCCCCAGTCTT
71 N S V 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

3601 GGTCTGATGGCCGCCCGACATGGTGTCTGTGCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCTGCTGAGAGATG
38 T R I 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

3701 TTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTC
4 N F T K M

3801 CAGCTTATCTGACGGTTCATAAACGAGCTCTGCTTATATAGACCTCCACCCTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGAC

3901 ATTTTGAAAAGTCCCCTTGATTTACTAGTCAAAAACAACTCCCATTGACGTCAATGGGGTGGAGACTTGAAAATCCCCGTGAGTCAAACCGTATCCAGC

4001 CCCATTGATGTA CTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTA CTGCCAAGTAGGAAAAGTCCATAAGGTCATGTA CTGG

4101 GCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATGTA CTGCCAAGTGGGCAGTTTACCGTA

4201 AATACTCCACCCATTGACGTCAATGAAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGCTGTTGGGCGGTCA

4301 GCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGACGGTTAATTAAGAACATGTGAGCAAAAAGGCCAGCAAAAAGGCCAGGAACCGTAAAAAGGCCG
CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAAACCCGACAGGACTATAAAGATA

4401 CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAAACCCGACAGGACTATAAAGATA

4501 CCAGGCGTTTTCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCG

4601 CTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCTGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTT CAGCCCAGCCGCTGCG

4701 CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGT

4801 AGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA

4901 AAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGGCAGAAAAAAGGATCTC

5001 AAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATC

5101 AGCGGCCCAATAAAATATCTTTATTTTATTACATCTGTGTGGTTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAAACAAAACGAAC

5201 AAAACAACTAGCAAAATAGGCTGTCCCAGTCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA