



1 GGATCTGCATCGCTCCGGTGCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTCTGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

AgeI (552)

SphI (568)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTAGGAGGGCCAGCATGCCTCTCAGCTTCAAGAACTAGAATG
1 M P P Q L Q E T R M
601 AACCGAAGCATTCTGTGGAGTTGATGAATCAGAACCATAACCAAGTCAGTTGCTGAAACCAATCCCAAGTATTCCCGGAGAGGAAATCAGAACCAC
11 N R S I P V E V D E S E P Y P S Q L L K P I P E Y S P E E E S E P
701 CTGCTCCAAATATAAGGAACATGGCACCAACAGCTTGTCTGCACCCACAATGCTTCACAATCTCCGGAGACTTTTCTCAAGCTCACTCAACCCTGAA
44 P A P N I R N M A P N S L S A P T M L H N S S G D F S Q A H S T L K
801 ACTTGCAAATCACCAGCGCCTGTATCCCGGACAGGTACCTGCCTGCGCACTCAAGTTCTGGAGGACAGTGAAGACAGTTTCTGCAGGAGACACCCAGGC
77 L A N H Q R P V S R Q V T C L R T Q V L E D S E D S F C R R H P G
901 CTGGCAAAGCTTTCCCTTCTGGGTGCTCTGCAGTCAGCGAGCCTGCTGAGTCTGTGGTGGAGCCCTCCCTGCAGAGCATCAGTTTTCACTTTATGG
111 L G K A F P S G C S A V S E P A S E S V V G A L P A E H Q F S F M
1001 AAAAAAGTAAATCAATGGCTGGTATCTCAGCTTTCAGCGGCTTCTCCTGACACTGGCCATGACTCAGACAAATCAGACCAAAGTTTACCTAATGCCTCAGC
144 E K R N Q W L V S Q L S A A S P D T G H D S D K S D Q S L P N A S A
1101 AGACTCCTTGGGCGTAGCCAGGAGATGGTGAACGGCCCGCCTCAGGAAACCGAGCAGGCTGGATCTGCCAACATAGACACGGGATATGATTCC
177 D S L G G S Q E M V Q R P Q P H R N R A G L D L P T I D T G Y D S
1201 CAGCCCCAGGATGCTCTGGGCATCAGGCAGCTGGAAGGCCCTGCCCTCACCTCCGTGTGTACCCCCAGGACCTCCCCAGACCTCTCAGGTCCAGGG
211 Q P Q D V L G I R Q L E R P L P L T S V C Y P Q D L P R P L R S R
1301 AGTCCCTCAGTTTGAACCTCAGAGGTATCCAGCATGTGCACAGATGCTGCCTCCCAATCTTCCACATGCTCCATGGAACATCATTACCATTTGCTC
244 E F P Q F E P Q R Y P A C A Q M L P P N L S P H A P W N Y H Y H C P
1401 TGGAAATCCCGATCACCAGGTGCATATGGCCATGACTACCTCGAGCAGCTACCAGCAAGTATCCAGCCGGCTCTGCCTGGGAGCCCTGCCTGGA
277 G S P D H Q V P Y G H D Y P R A A Y Q Q V I Q P A L P G Q P L P G
1501 GCCAGTGTGAGAGCCTGCACCTGTGCAGAAGTTATCCTGAATTATCCAGCCCTGGGACCAAGAAGAGAGGCCCGCACAGAGAGACTGCTCCTTTC
311 A S V R G L H P V Q K V I L N Y P S P W D Q E E R P A Q R D C S F
1601 CGGGCTTCCAAGGCACAGGACAGCCACATCACCAGCCACCTAATAGAGTGGTGTCTCTGGGAGTCTTGGAGTGCCTGCAGAGTGCAGACCACA
344 P G L P R H Q D Q P H H Q P P N R A G A P G E S L E C P A E L R P Q
1701 GGTCCCGCAGCCTCCGTCAGCTGTGCTTCCAGCCCTAGCAACCCTCCAGCCAGAGGAACTCTAAAAACAAGCAATTTGCCAGAAGAATTGCCG
377 V P Q P P S P A A V P R P P S N P P A R G T L K T S N L P E E L R
1801 AAAGTCTTATCACTTATTCGATGGACACAGCTATGGAGTGGTGAATTCGTGAACCTTTTGTGGTAAATGGCTTCCAACCTGCAATTGACATATTTG
411 K V F I T Y S M D T A M E V V K F V N F L L V N G F Q T A I D I F
1901 AGGATAGAATCCGAGGCATTGATATCATTAAATGGATGGAGCCTACCTTAGGATAAGACCGTGATGATAATCGTAGCAATCAGCCCAAATACAAACA
444 E D R I R G I D I I K W M E R Y L R D K T V M I I V A I S P K Y K Q
2001 GGACGTGGAAGCGCTGAGTCGACAGCTGGACGAGGATGAGCATGGCTTACATACTAAGTACATTATCGAATGATGCAGATTGAGTTCATAAAACAAGGA
477 D V E G A E S Q L D E D E H G L H T K Y I H R M M Q I E F I K Q G G
2101 AGCATGAATTTAGATTATCCCTGTGCTTCCCAAATGCTAAGAAGGAGCATGTGCCACCTGGCTTCCAGAACACTCATGTCTACAGCTGGCCCAAG
511 S M N F R F I P V L F P N A K K E H V P T W L Q N T H V Y S W P K
2201 ATAAAAAAACATCCTGCTGCGGCTGCTGAGAGAGGAAGATGTGGCTCTCCACGGGGCTCTGCCACCTTCAGGTGGTTCCTTGTGACACCG
544 N K K N I L L R L L R E E E Y V A P P R G P L P T L Q V V P L •

NheI (2320)

2301 TTCATCCCAGATCACTGAGGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTAT
2401 TTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTACG
2501 GGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTTTAACCTCAAATCAAGC
2601 CTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTTTGACGCTCACCTTCTTCATGGAG
2701 TTTAAGATATAGTGTATTTTCCCAAGTTTGAAGTACTGCTTTCATTTCTTTATGTTTTAAATGCAGTACCTCCACATTCCTTTTATAGTAAATATTC
2801 AGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGT
2901 TGGACTTAGGGAACAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCTGGTGTACTTGAGGGGGATGAGTTCCTCA
141 • N R T Y K L P I L E E
3001 ATGGTGGTTTTGACCAGCTTGCATTCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGACCA
128 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 G C P S V V
3101 CCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCAATGGCAAT
95 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 A S G I A I

3201 GGCTTCAGCACAGACAGTGACCCTGCCAATGTAGGCCCTCAATGTGGACAGCAGAGATGATCTCCCAGTCTTGGTCTGATGGCCGCCCGACATGGTGC
62 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 A A G V H H
3301 TTGTTGCCTCATAGAGCATGGTGATCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGAAGGTCTTCATGGTGGCCCTCTAT
28 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 M
3401 AGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACATAACGAG
3501 CTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAG
3601 ICAAAAACAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATC
3701 ATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGT
3801 CATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCATTGACGTCAATGGAA
3901 AGTCCCTATTGGCGTTACTATGGGAACATACGTATTATTGACGTCAATGGGGGGGGTCTGGTGGCGGTGAGCCAGGCGGGCCATTTACCGTAAAGTTAT
4001 GTAACGCTGCAGGTTAAITAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGTTGCTGGCGTTTTTCCATAGGCTCCG
4101 CCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
4201 GTGCGCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTTTCATAGCTCACGCTGTAGGTATC
4301 TCAGTTCGGTGTAGGTGCTTCGCTCCAAGTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTC
4401 CAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGG
4501 TGGCCTAACTACGGTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
4601 AACAAACCACCGCTGGTAGCGGTGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGG
4701 GTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTT
4801 CATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCC
4901 CAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA