



1 GGATCTGCATCGCTCCGGTCCCGTCACTGGGCGAGAGCGACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCCGTTGAGTCCGCTTCTGCCGCTCCCGCTGTGGTGCCTCCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCAACTCTACGCTTTTGTTCGTTT

BspEI (558)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGCTCCGGACATCATGATCGTGACAAGCTGCTGGACGACAGCCG
1▶ M I V D K L L D D S R
601 CGGCGGAGAGGGGCTGCTGGACGCGCCGCGACTGCGGCCTCATGACCAGCCGCTCAACCTGGCTTACTTCTACGGCGCCTCGCCTCTTCCGCCCCC
11▶ G G E G L L D A A G D C G L M T S P L N L A Y F Y G A S P P S A P
701 GGCGCCGCGACACCGGCTACCTGTCCGCGTGCCTCCGCGCCGGCTCGCCGGCTCCGACTCCTCGATTCTCTCCACTTCTCCGTGTCTCTCT
45▶ G A G D T G Y L S A V P S A P G S P G S D S S D F S S T S S V S S
801 GCGGGCCGTTGAGTCCCGTCCAGAGTGGCGCCCGCGGAGCGCCTCAAGTTGAGCCCCACATGGGGTGGCAGGCAGAGAGGCCCTTTCA
78▶ C G A V E S R P R G G A R A E R P Q V E P H M G V G R Q Q R G P F Q
901 AGTGTTTCGCTGAAGAACTCGGTGAAGGAGCTCCTGCTACACATCCGAAGCAACAAGCAGAAGGCTTCCGGCCAGCCGTTGATGAGTTAAGACACAG
111▶ G V R V K N S V K E L L L H I R S N K Q K A S G Q P V D E F K T Q
1001 AGTGTGAACATTGAGCAGCTGACAGACCTGAAGAGTGCAGTGTCCGAGTGGGAAAAGAAAGGGACCCGATCCTCTGTCCGATGGACCGTTTGAAGA
145▶ S V N I E Q L T D L K S A V S A V G K R K G P D P L S D G P V C K
1101 GGCAGCTCTGTTGCCTTCTCACTTCTGACATCACCGCAAACGCCTACACCTGGGGAGAGCATGGAAGATGTTGCCACAGCGAATCAAGCTGGATAG
178▶ R P A L L P S H F V T S P Q T P T P G E S M E D V R H S E S K L D S
1201 CAGCGCTGCCCTGCTTCAAGATATTATAAATCAAGAATGAGTGCAACCCAGTTTCACTAAACACTGTTAGGTTAGCTGGATGAGCCCCACGGTCCC
211▶ S A A L L Q N I I N I K N E C N P V S L N T V Q V S W M S P T V P
1301 CAGAACTCGCAAGAGACAGTCCAGGATTTCCATGGAGGCCAGGCTTCTCTCCACTCAGAAATACCAACCGTTCCAAGTCAAGTGGCTCTCCACAGA
245▶ Q N S P R D Q C Q D F H G G Q A F S P P Q K Y Q P F Q V S G S P Q
1401 TGATGGACCAGCGTCAATGTACCAGTATTCGCACAAACACAGAATGCAGCAGCCGCGCGCTGCCACCACAGCAGCAACATCAACAGAATTATCC
278▶ M M D Q A S M Y Q Y S P Q T Q N M Q Q P P P L P P Q Q Q H Q Q N Y P
1501 CCACAACCTACCCCTGCAGTTCAGTCTTACTCCAGATGTCCAGTCTCCCAAGTATGATTCACCTCTTTGATACTCACGAGCCACAGTTCGACAG
311▶ H N S P L Q F S P Y S R M S Q S P K Y D S N L F D T H E P Q F C T
1601 GGCCAAAGTTTGTGCTTTACTGACTGGCCCGGGGAACCCGAGTCTCTTGTGTTCCCGTTCGGCTCCACCAGCATTCCACCCAAACTGAGACTC
345▶ G Q S F V S L L T G P G E E S L A V P V P A P T S I P P Q T E T
1701 AGTGCAGACCTTCAAGCTGATGCAAGCAATGCCTGCGAGGCGCTCGTGGGGTGCACGATGTGGGCTCGACTCTTTGGGACGTCCTGTCACTCCA
378▶ Q L Q T F S L T M P S N A C E A V V G V H D V G S H S L G T S L S L Q
1801 GAACATCATGGGAAGCCGATGAATACCACCCAGTGGGGAAGTCAATTTTTTCAAGTGGCAGGTAGAGCAGGAAGAAAGCAAATTTGGCAAATATTTCCAG
411▶ N I M G S P M N T T Q L G K S F F Q W Q V E Q E E S K L A N I P Q
1901 GACCAGTCTCTGCTAGGAGCAGGATGGTGCACGTTTCTCCATATTGCTGTGCCAAGGAGGCGAGCGCTCTGATGTCTTGAAGAAAGATGA
445▶ D Q F L A R D G D G D T F L H I A V A Q G R R A L S Y V L A R K M
2001 ACGCTCTTCAATGCTGGACATTAAGGAGCACAATGGACAGAGTGCCTTTTCAAGTGGCAGGTAGCTGCAATCAGCAGCTCATTGTGCAAGATCTGGTAAA
478▶ N A L H M L D I K E H N G Q S A F Q V A V A A N Q H L I V Q D L V N
2101 CCTGGTGGCCAGGTGAACACCACGATTGCTGGGGAAGGACACTCTGCATGTCTGTGCGGAGAAGGCCACTCCCAGGTGCTCCAGGCCATCCAGAAG
511▶ L G A Q V N T T D C W G R T P L H V C A E K G H S Q V L Q A I Q K
2201 GGAGCTGTGAGGAGCAATCAGTTTGTGGATCTTGGGCAACGAATATGATGGCCTGACTCCCTACATTGCGCGGTCTGGCTCACAATGCGGTGGTGC
545▶ G A V R S N Q F V D L E A T N Y D G L T P L H C A V V A H N A V V
2301 ACGAACTGCAGAGGAATCGGCAGTCTCATTGCCCCAAGTGCAGGACCTTACTAAGGAATAAGAGCCTGGTAGACACCATTAAGTGTCTGATTAGAT
578▶ H E L Q R N R Q S H S P E V Q D L L L R N K S L V D T I K C L I Q M
2401 GGGAGCCGAGTGGAGGCAAAGGATCGTAAGAGTGGCCGACAGCCCTGCATTGGCAGCGGAAGAAGCAAATCTGGAACCTATTGCGCTCTTTTGGAG
611▶ G A A V E A K D R K S G R T A L H L A A E E A N L E L I R L F L E
2501 TTGCCAGTTGCTGTCTTTCTGTAATGCAAGGCTTACAATGGAAACACCGCCCTCCATGTTGCTGCCAGCTGCAGTATCGGGTGCACAGTTGGAGC
645▶ L P S C L S F V N A K A Y N G N T A L H V A A S L Q Y R V T Q L D
2601 CAGTCCGCTGTTGATGAGGAAGGGAGTACCAGTACTCGGAATGGAGAATGAACAGCCGTCATTTGGTCCCCGATGGCCAGTGGGAGAACA
678▶ A V R L L M R K G A D P S T R N L E N E Q P V H L V P D G P V G E Q

NheI (2762)

2701 GATCCGACGATCCTGAAGGGGAAGTCCATTGAGCAGCGAGCACCACCATACTAGGTCCACTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTT
711▶ I R R I L K G K S I Q Q R A P P Y •
2801 TGGACAAACCACTAGAATGCAGTAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGAACCATTATAAGTGCATAAACAA
2901 GTTAAACAACAATTCATTCTTTTATGTTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAACCTCTACAATGTGGTATGGAAT
3001 TCTAAAATACAGCATAGCAAACTTTAACTCAAATCAAGCTTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGCTGTTGCC
3101 AATGTGCATTAGCTGTTGAGCCTCACCTTTTTCATGGAGTTAAGATATAGTGTATTTTCCCAAGGTTTGAACAGTCTTCAATTTCTTTATGTTTT

3201 AAATGCACTGACCTCCACATTCCTTTTTAGTAAAATATTCAGAAAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATC
3301 CAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTC
3401 TAGCTTTAGTTCTCTGGTGTACTTGAGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCTCAATGAGCACAAAGCAGTCAGGAGCAT
-141 • N R 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
3501 AGTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGC
109 D S I L 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
3601 AAAGTCTTCTGCCGTTGCTCACAGCAGACCAATGGCAATGGCTTCAGCACAGACAGTACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATG
76 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 A E I H V A S I
3701 ATCTCCCCAGTCTTGGTCTGATGGCCGCCGACATGGTGTCTGTTGCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGAT
42 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 T A V E V L E L D
3801 CCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCCTATAGTGTGCTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGC
9 Q Q S I N F T K M
3901 GTGGATGGCGTCTCCAGCTTATCTGACGGTTCATAAACGAGCTCTGCTTATATAGACCTCCACCCTACACGCCTACCGCCATTTCGCTCAATGGGGC
4001 GGAGTTGTTACGACATTTTGAAAGTCCCGTTGATTTACTAGTCAAAAACAACTCCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCA
4101 AACCGCTATCCACGCCATTGATGTAAGTCCGATCATCATGGTAATAGCGATGACTAATACGTAGATGTAAGTCCAGTAAAGTCCGATGTAAGTCCGAT
4201 AGGTCATGTAAGTCCGATGTAAGTCCGATGTAAGTCCGATGTAAGTCCGATGTAAGTCCGATGTAAGTCCGATGTAAGTCCGATGTAAGTCCGATGTAAGT
4301 GGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGAAAGTCCCTATTGGCGTACTATGGGAACATACGTCAATATTGACGTCAATGGGCGGGGG
4401 TCGTTGGCGGTGAGCCAGGCGGGCCATTACCGTAAGTTATGTAACGCTGAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
4501 CCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
4601 GGACTATAAAGATACCAGCGTTTTCCCCTGGAAGCTCCCTCGTGGCTCTCCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTT
4701 CGGGAAGCGTGGCGCTTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGTGTGCAGCAACCCCGTTCA
4801 GCCCGACCGCTGCGCCTTATCCGTAACATATCGTCTTGTAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
4901 AGAGCGAGGTATGTAGCGGTGCTACAGAGTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGC
5001 CAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGTATCCGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAG
5101 AAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAA
5201 TTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCA
5301 AAACAAAACGAAACAAAACAACTAGCAAAATAGGCTGTCCCCAGTGAAGTGCAGGTGCCAGAACATTTCTCTATCGAA