



1 GGACCTGCAGGGCTGAATAACCTCTGAAAGAGGAAGCTTGGTTAGGTACCTTCTGAGCGGAAAGAACAGCTGTGGAATGTGTGTCAGTTAGGGTGTG
101 GAAAGTCCCAGGCTCCCAGCAGGCAGAAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCAGGCTCCCAGCAGGCAG
201 AAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCACTAGTCTCCGGTGGCCGTCAGTGGGAGAGCGCACATCGCCACAGTCCCCGA
301 GAAGTTGGGGGAGGGGTGCGCAATTGAACGGGTGCCTAGAGAAGGTGGCGGGGTAACCTGGGAAAGTGTGCTGTACTGGCTCCGCTTTTCC
401 GAGGTGGGGGAGAACCCTATATAAGTCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGGCCGAGAACACAGCTGAAGCTTCGAGGGGCTC
501 GCATCTCTCTTACGCGCCCGCCCTACCTGAGGCCGATCCACGCCGTTGAGTCGCGTTCTGCCGCTCCGCTGTGGTGCCTCCTGAACTGC
601 GTCCGCGCTTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCCTCACGCTTTGC

NcoI (799)
AgeI (791)

701 CTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCTGTTTCTGTTCTGCGCAAGTTACAGATCCAAGCTGTGACCGGCGCTACCTGAGATCACCGGTAC

XhoI (864)

801 CATGGAGATCAAGGTGCTGTTTGCCTCATCTGTATTGCTGTTGCTGAGGCAAAACCCACTGAGATGTCGGATAATGGGCCACAGAACCAGAGA
1 M E I K V L F A L I C I A V A E A K P T E L E M S D N G P Q N Q R
901 AATGCACCCCGGATCACCTTTGGAGGACCTTTCAGATTAACCGGCAGTAACAGAATGGTGAGCGATCTGGAGCAGTTCTAAACAACGCAGACCTCAAG
11 N A P R I T F G G P S D S T G S N Q N G E R S G A R S K Q R R P Q
1001 GGCTCCAAACAATACCCTAGTTGGTTTACTGCTCTCACACAGCATGGGAAGGAGGATCTGAAATTTCCAGGAGGCAAGGCTCCCATCAACACCAA
44 G L P N N T A S W F T A L T Q H G K E D L K F P R G Q G V P I N T N
1101 CTCAGTCCCAGCAGCAGATCGGATACTACAGACGAGCAACCCGGCAATCAGAGGCGGCGATGGCAAAATGAAAGACCTGAGTCCCAGGTGGTACTTC
77 S S P D D Q I G Y Y R R A T R R I R G G D G K M K D L S P R W Y F
1201 TATTATCTCGAAACAGGCTGAAGCCGCTGCCTATGGTCCAAACAAGACGGCATCATCTGGTAGCTACTGAGGAGCCTTGAACACTCCCAAGG
111 Y Y L G T G P E A G L P Y G A N K D G I I W V A T E G A L N T P K
1301 ACCACATTGGTACCCGCAATCCCGCAACAATGCTGCCATTGTGCTCCAGTTGCCACAGGGCACAACTCTTCCAAGGTTTCTATGCCGAAGGCTCTCG
144 D H I G T R N P A N N A A I V L Q L P Q G T T L P K G F Y A E G S R
1401 TGGAGGATCTCAAGCCAGCAGCCGTTCTCCTCTCGGAGTAGGAATAGCTCTCGGAACTCCACACCAGGCAGCTCTCGCGAAACAAGCCCTGCTCGCATG
177 G G S Q A S S R S S S R N S S R N S T P G S S R G T S P A R M
1501 GCCGTAACCGTGGTATGCCGCTCTGGCTCTGCTTCTCTGGACAGGCTGAATCAGTGGAACTCCAAGATGTCTGGTAAGGGACAACAGCAACAGGGGC
211 A G N G D A L L L L D R L N Q L E S K M S G K G Q Q Q Q G
1601 AGACGTTACCAAGTCCGCTGCAGAGGCCAGCAAGAAACCTCGGCAAGAGGACCGCCACTAAAGCCTACAACGTGACACAGGCTTCCGCGAGGAG
244 Q T V T K K S A A E A S K K P R Q K R T A T K A Y N V T Q A F G R R
1701 AGGGCCAGAACAGACACAGGGCAACTTTGGTGTATCAGGAGTTGATACGCCAGGGGACAGACTACAAGCATTGGCCACAGATTGCAGTTCCGACCTTCC
277 G P E Q T Q G N F G D Q E L I R Q G T D Y K H W P Q I A Q F A P S
1801 GCTAGTGCCTTTTGGGATGTACGGATAGGCATGGAGGTGACGCCTAGCGGGACATGGCTGACCTATACTGGAGCTATCAAAGTGGATGACAAAGACC
311 A S A T F F G M S R I G M E V T P S G T W L T Y T G A I K L D D K D
1901 CCAACTCAAGGACAGTGTCTGCTGAATAAGCACATTGTACCTACAAAACGTTTCTCCAATGAGCCCAAGGACAAGAAGAAGGAGCCGA
344 P N F K D Q V I L L N K H I D A Y K T F P P T E P K K D K K K A D
2001 TGAAACCCAGGCACTTCTCAGAGGCGAGAAGAAACAGCAGACCGTCACTCTTCTGCCGCGAGCGATCTGGACGACTTCTCAAAAACAGTTGCAACAGTCC
377 E T Q A L P Q R Q K K Q Q T V T L L P A A D L D D F S K Q L Q Q S

BsiWI (2127)

BamHI (2157)

2101 ATGAGCTCAGCCGATTCCACCAAGCTCGTACGGAGAACTGTACTTCCAGGGCTCTGGATCCGAGCCAAATCTAGTGACAAAACCTCACACATGCCAC
411 M S S A D S T Q A R T E N L Y F Q G S G S E P K S S D K T H T C P

2201 CGTGCCAGCACCTGAAGCCGAGGGGGACCGTCAGTCTTCTCTTCCCCCAAAACCAAGGACCAACTGATGATCTCCCGACCCCTGAGGTCACATG
25 P C P A P E A E G G P S V F L F P P K P K D Q L M I S R T P E V T C

2301 CGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAG
58 V V V D V S H E D P E V K F N W Y V D G V E V H N A K T K P R E E

2401 CAGTACAACAGCAGTACCGTGTGGTACGCTCCTCACCGTCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCC
92 Q Y N S T Y R V V S V L T V L H Q D W L N G K E Y K C K V S N K A

2501 TCCAGCCICCATCGAGAAAACCATCTCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCTGCCCCATCCCGGAGGAGATGACCAAGAA
125 L P A S I E K T I S K A K G Q P R E P Q V Y T L P P S R E E M T K N

2601 CCAGGTACGCTGACCTGCCTGGTCAAAGGCTTCTATCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGAGCCGGAGAACAACATAAGACCAG
158 Q V S L T C L V K G F Y P S D I A V E W E S N G Q P E N N Y K T T

2701 CCTCCCGTGTGGACTCCGACGGCTCCTTCTCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGTCCGTC
192 P P V L D S D G S F F L Y S K L T V D K S R W Q Q G N V F S C S V

2801 TGCATGAGGCTCTGCACAACCTACACGCAGAAGAGCCTCTCCCTGTCTCCGGTAAATGAGTCCTAGCTGGCCAGACATGATAAGATACATTGATGAG
225▶ L H E A L H N H Y T Q K S L S L S P G K •

2901 TTTGGACAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAAC

3001 AAGTTAACACAACAATTGCATTCAATTTATGTTTCAGGTTCCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAATGTGGTATGGA
▶

3101 ATTCTAAAATACAGCATAGCAAACTTAACTCCAATCAAGCTCTACTTGAACTCTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTG
◀

3201 CCAATGTGCATTAGCTGTTTGCAGCCTCACCTCTTTCATGGAGTTAAGATATAGTGATTTTCCCAAGGTTTGAAGTAGCTCTTCATTTCTTTATGTT

3301 TTAATGCACTGACCTCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAAAATGTTTTTTATTAGGCAGAA

3401 TCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTAAATAGAAATTGGACAGCAAGAAAGCGAGCT

3501 TCTAGCTTTAGTTCCTGGTGTACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTTCATCTCAATGAGCACAAAGCAGTCAGGAGC
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

3601 ATAGTCAGAGATGAGCTCTGACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTG
110◀ 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 P H R V A V I T

3701 TCAAAGTCCTTCTGCCGTTGCTCAGCAGACCCAATGGCAATGGCTTCCAGCAGACAGTGACCCTGCAATGTAGGCCTCAATGTGGACAGCAGAGA
76◀ 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 A E I H V A S I

3801 TGATCTCCCAGTCTTGGTCTGATGGCCGCCGACATGGTCTTGTCTCATAGAGCATGGTATCTTCTCAGTGGCGACCTCCACCAGCTCCAG
43◀ 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

3901 ATCCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACA
10◀ D Q Q S I N F T K M ◀

4001 GCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTTGCGTCAATGGG

4101 GCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAACAACCTCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCGTGAGT
◀

4201 CAAACCGTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCA

4301 TAAGGTCACTGACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATGTACTGCCAAG

4401 TGGGCAGTTTACCCTAAATACTCCACCATTGACGTCAATGGAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGG

4501 GGTGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGG
◀

4601 AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGA

4701 CAGGACTATAAGATAACCAGGCGTTTTCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCC

4801 TTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGTGTGCACGAACCCCGTT

4901 CAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGTGTTGAGTCCAAACCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA

5001 GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAA

5101 GCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGGCG

5201 AGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTT

5301 AATTAACATTTAAATCAGCGCCGCAATAAAAATATCTTTATTTTATTACATCTGTGTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCAT

5401 CAAAACAAAACGAAACAAAACAAAACCTAGCAAAATAGGCTGTCCCAGTGCAAGTGAGGTGCCAGAACATTCTCTATCGAA