



1 GGACCTGCAGGGCTGAATAACCTCTGAAAGAGGAAGCTGGTTAGGTACCTCTGAGCGGAAAGAACAGCTGTGGAATGTGTGTCAGTTAGGGTGTG
101 GAAAGTCCCAGGCTCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCAGGCTCCCAGCAGGCAG
201 AAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCACTAGTCTCCGGTGGCCGTGAGTGGGAGAGCGCACATCGCCACAGTCCCCGA
301 GAAGTTGGGGGAGGGGTGCGCAATTGAACGGGTGCCTAGAGAAGGTGGCGGGGTAACCTGGGAAAGTGTGCTGTACTGGCTCCGCTTTTTCCC
401 GAGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGGCCGAGAACACAGCTGAAGCTTCGAGGGGCTC
501 GCATCTCTCTTACGCGCCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCTCCCGCTGTGGTGCCTCCTGAACTGC
601 GTCCGCGCTTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCCTCACGCTTTGC

NcoI (799)
AgeI (791)

701 CTGACCCTGCTTGCTCAACTCTACGCTTTTGTTCGTTTTCTGTTCTGCGCAGTTACAGATCCAAGCTGTGACCGGCGCTACCTGAGATCACCGGTAC

XhoI (864)

801 CATGGAGATCAAGGTGCTGTTTGCCTCATCTGTATTGCTGTTGCTGAGGCAAAACCCACTGAGCTCGAGGTTAATCTCACCACCCGAAGTCAACTCCCA
1 M E I K V L F A L I C I A V A E A K P T E L E V N L T T R T Q L P
901 CCCGCATATACAAATCCTTACCAGAGGAGTGTACTATCCTGACAAAGTGTTCGGTCAAGTGTCTCCACTCTACTCAGGACCTCTTTCTGCCTTTCT
11 P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L F L P F
1001 TTTCTAACGTTACATGGTTTCATGCAATCCATGTGTCTGGGACAAACGGCACAAACGCTTCGACAACCTGTATTGCCATTCAATGATGGGGTGTACTT
44 F S N V T W F H A I H V S G T N G T K R F D N P V L P F N D G V Y F
1101 TGCTCCACAGAGAAATCCAACATCATTGAGGATGGATTTTCCGGACTACTCTGGACTCAAAGACACAGAGCCTGTGATCGTTAAACAGCCACAAAC
77 A S T E K S N I I R G W I F G T T L D S K T Q S L L I V N N A T N
1201 GTTGTCAACAAAGTGTGCGAATCCAGTTTTCGCAATGATCCCTTCTGGGAGTGTACTATCACAAGAATAACAAGTCTGGATGGAGAGCGAATTTCCGG
111 V V I K V C E F Q F C N D P F L G V Y Y H K N N K S W M E S E F R
1301 TCTACAGCAGCGCAAACAACTGCACCTTCGAGTACGTGAGTCAACCTTTCTGATGGACCTGGAAGGGAAACAGGGAACTTCAAGAACCTGAGAGAGTT
144 V Y S S A N N C T F E Y V S Q P F L M D L E G K Q G N F K N L R E F
1401 TGTCTTTAAGAACATCGACGGCTATTTAAGATCTATAGTAAGCATACGCCTATCAACCTGTAAGGGATCTTCCCAGGGCTTTTACGCCCTGGAACCT
177 V F K N I D G Y F K I Y S K H T P I N L V R D L P Q G F S A L E P
1501 TTGGTTGACTTGCTATTGGTATCAATATCACCAGATTTTCAGACCTTCTGGCATTGCAICGGTCTTACTCTCAGGTGATTCCTCCCGGGTGGGA
211 L V D L P I G I N I T R F Q T L L A L H R S Y L T P G D S S S G W
1601 CTGCCGGCGCGCTACTATGTGCGCTATCTGCAACCAAGAAGCTTCTGCTCAAGTACAACGAAAACCGCACTATTACGGATGCTGTTGATGTTGTC
244 T A G A A A Y Y V G Y L Q P R T F L L K Y N E N G T I T D A V D C A
1701 CCTGGACCCTCTGTGAGACTAAATGCACCCTCAAGAGCTTACCCTGAGAAGGGGATTTACCAAACAGTAATTTCCGGGTCCAACCCACCGAAAGC
277 L D P L S E T K C T L K S F T V E K G I Y Q T S N F R V Q P T E S
1801 ATGTGCGGGTCCCAAATATACCAATCTGTGTCCTTTGGCGAAGTGTCAATGCTACAAGTGTGCTTCTGTGTACGCATGGAATAGGAAACGCATCT
311 I V R T G F P N I T N L C P F G E V F N A T R F A S V Y A A W N R K R I
1901 CCAATGTGTCGATTACTCCGTGCTGACAACTCCGCTCTTTTCAACCTTCAAGTGTATGGCGTTTACCTACCAAACCTAACGACCTGTGCTT
344 S N C V A D Y S V L Y N S A S F S T F K C Y G V S P T K L N D L C F
2001 CACTAATGTGTATGCCGACTCTTTTGTGATACGAGGCGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT
377 T N V Y A D S F V I R G D E V R Q I A P G Q T G K I A D Y N Y K L
2101 CCAGATGACTTTACCGGATGTGTTATTGCATGGAACCTCAAACAATCTGGATTCCAAGTGGGTGGCAACTATAACTACCTGTATAGACTGTTACAGAAAT
411 P D D F T G C V I A W N S N N L D S K V G G N Y N Y L Y R L F R K
2201 CCAACCTGAAACCATTCGAGCGAGATATAAGCACAGAAATCTACCAGGCTGGAAGTACGCCCTGCAACGGCGTGAAGGGTTCAACTGCTACTTCCCATT
444 S N L K P F E R D I S T E I Y Q A G S T P C N G V E G F N C Y F P L
2301 GCAGAGTTACGGATCCAGCCTACAAACGGGGTGGTTACCAACCTATCGTGTGCTAGTCTGAGTTTTGAGCTCCTCATGCCCCAGCCACAGTCTGT
477 Q S Y G F Q P T N G V G Y Q P Y R V V V L S F E L L H A P A T V C
2401 GGCCCCAAGAAAAGCACCAATCTGGTGAAGAACAATGCGTGAACCTTAACTTAAACGGACTCACAGGAACCGGCGTATTGACGGAGAGTAACAAGAAGT
511 G P K K S T N L V K N K C V N F N F N G L T G T G V L T E S N K K
2501 TCCTGCCATTCCAGCAGTTCGGTCGCGATATTGCCGACTACCGACGCTGTCCGAGATCCCAGACATTGGAGATTCTTGATATCACACCCTGTAGTTT
544 F L P F Q Q F G R D I A D T T D A V R D P Q T L E I L D I T P C S F
2601 CGCGGAGTGAGCGTATTACGCCCGGAACCAATACCAGCAATCAGGTTGCCGTCCTGTATCAGGACGTGAATTGCACCGAGGTACCTGTCCGCATCCAC
577 G G V S V I T P G T N T S N Q V A V L Y Q D V N C T E V P V A I H
2701 GCTGACCAACTTACACCACATGGCAGTATATTCACCGGCTCCAACGCTTTTCAGACACGTGCTGGATGTCTGTATGCTGAGTGCAGAACACGTTAATAATA
611 A D Q L T P T W R V Y S T G S N V F Q T R A G C L I G A E H V N N

BsiWI (2880)

2801 GCTACGAGTGTGATATCCCATCGGTGCTGGAATATGCGCCTTATCAAACCTCAAACCAACTCTCCTAGCGGGGCACGTCGTACGGAGAACCTGTACTT
644 S Y E C D I P I G A G I C A S Y Q T Q T N S P R R A R R T E N L Y F

BamHI (2910)

2901 CCAAGGCTCTGGATCCGAGCCAAATCTAGTGACAAAACCTCACACATGCCACCGTGGCCAGCACCTGAAGCCGAGGGGGGACCGTCACTTCTCTCTC
7 Q G S G S E P K S S D K T H T C P P C P A P E A E G G P S V F L F

3001 CCCCCAAAACCAAGGACCAACTGATGATCTCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACT
29▶ P P K P K D Q L M I S R T P E V T C V V V D V S H E D P E V K F N

3101 GGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCTCACCGTCTGCA
62▶ W Y V D G V E V H N A K T K P R E E Q Y N S T Y R V V S V L T V L H

3201 CCAGGACTGGCTGAATGGCAAGGAGTACAAGTGAAGTCTCCAACAAAGCCCTCCAGCCICCATCGAGAAAACCATCTCCAAGCCAAAGGGCAGCCC
95▶ Q D W L N G K E Y K C K V S N K A L P A S I E K T I S K A K G Q P

3301 CGAGAACCACAGGTGTACACCCTGCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTGAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACA
129▶ R E P Q V Y T L P P S R E E M T K N Q V S L T C L V K G F Y P S D

3401 TCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACACTACAAGACCACGCTCCCGTGTGGACTCCGACGGCTCCTTCTCTCTACAGAAAGCT
162▶ I A V E W E S N G Q P E N N Y K T T P P V L D S D G S F F L Y S K L

3501 CACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGCTGCATGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTG
195▶ T V D K S R W Q Q G N V F S C S V L H E A L H N H Y T Q K S L S L

MscI (3623)

3601 TCTCCGGTAAATGAGTCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTG
229▶ S P G K •

3701 TGAAATTTGTGATGCTATTGCTTTATTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAAATTGCATTCATTTTATGTTTCAGGTTCCAGGGG

3801 GAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTTTAACTCCAAATCAAGCCTC

3901 TACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTCATGGAGTTT

4001 AAGATATAGTGTATTTTCCCAAGGTTTGAACACTAGCTCTTCAATTTCTTTATGTTTTAAATGCACTGACCTCCACATCCCTTTTATGATAAATATTAGA

4101 AATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGG

4201 ACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCTGGTGTACTTGAGGGGGATGAGTTCCTCAATG
141▶ • N R T Y K L P I L E E I

4301 GTGGTTTTGACCAGCTTGCCATTCATCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCC
127▶ 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 R

4401 TGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCCGTTGCTCACAGCAGACCAATGGCAATGGC
94▶ 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 A

4501 TTCAGCAGACAGTACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCAAGTCTTGGTCTGATGGCCGCCGACATGGTGCCTTG
61▶ 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 K

4601 TTGTCTCATAGAGCATGGTGTCTTCTCAGTGGCAGCTCCACCAGTCCAGATCTGCTGAGAGATGTTGAAGGTCTTCATGGTGGCCCTCTATAGT
27▶ 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

4701 GAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTACTAAACGAGCTC

4801 TGCTTATATAGACCTCCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCA

4901 AAACAACTCCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTAAGTCCAAAACCGCATCATC

5001 ATGGTAATAGCGATGACTAATACGTAGATGACTGCCAAGTAGAAAGTCCCATAAAGTCAAGTCAAGTACTGGGCATAATGCCAGGCGGGCATTACCCTCAT

5101 TGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATGTAAGTGGGCAAGTGGGCAAGTAAACTCCACCCATTGACGTCAATGAAAAGT

5201 CCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTGAGCCAGGCGGGCATTACCCTAAGTTATGTA

5301 ACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGCC
141▶

5401 CCCTGACGAGCATCACAATAATGACGCTCAAGTCAAGGTTGGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTCCCCTGGAAGCTCCCTCGTG

5501 CGCTCTCCTGTTCCGACCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCA

5601 GTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGTGTGCACGAACCCCGTTGAGCCGACCGCTGCGCTTATCCGGTAACTATCGTCTTGTAGTCCAA

5701 CCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCAGGATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGG

5801 CCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGTATCCGGCAAC

5901 AAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTC

6001 TGACGCTCAGTGGAACGAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGC AATAAAATATCTTTATTTTCAT

6101 TACATCTGTGTGTTGGTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCCCAG

6201 TGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA