



1 GGACCTGCAGGGCTGAATAACCTCTGAAAGAGGAACCTGGTTAGGTACCTTCTGAGGCGGAAAGAACAGCTGTGGAATGTGTGTCAGTTAGGGTGTG  
101 GAAAGTCCCAGGCTCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCAGGCTCCCAGCAGGCAG  
201 AAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCACTAGTGTCTCCGGTGCCCGTCAGTGGCAGAGCGCACATCGCCACAGTCCCCGA  
301 GAAGTTGGGGGAGGGGTGCGCAATTGAACGGGTGCCTAGAGAAGGTGGCGGGGTAACCTGGGAAAGTGTGCTGTACTGGTCCGCTTTTTCC  
401 GAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCGGTGAACGTTCTTTTTCGCAACGGGTTTGGCCGAGAACACAGCTGAAGCTTCGAGGGGCTC  
501 GCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCCGCCATCCACGCGGTTGAGTCGCGTTCTGCCGCTCCCGCTGTGGTGCCTCCTGAACTGC  
601 GTCCGCGCTTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCCTCACGCTTTGC

NcoI (799)

Agel (791)

701 CTGACCCTGCTTCTCAACTCTACGTCTTTGTTTCTGTTTCTGTCGCGGTTACAGATCCAAGCTGTGACCGGGCCTACCTGAGATCACCGGTAC

XhoI (864)

801 CATGGAGATCAAGGTGCTGTTTGCCTCATCTGTATTGCTGTTGCTGAGGCAAAACCCACTGAGCTCGAGGTTAATCTCACCACCCGAACTCAACTCCCA  
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 CCCGCATATACAAATTCCTTACCAGAGGAGTGTACTATCCTGACAAAGTGTTCGGTCAAGTGTCTCCACTTACTCAGGACCTCTTTCTGCCTTTCT  
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 TTTCTAACGTTACATGGTTTTCATGCAATCCATGTGTCTGGGACAAACGGCACAAACGCTTCGACAACCTGTATTGCCATTCAATGATGGGGTGTACTT  
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 TGCCTCCACAGAGAAAATCCAACATCATTGAGGATGGATTTTCGGGACTACTCTGGACTCAAAGACACAGAGCCTGCTGATCGTTAACACGCCACAAC  
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 GTTGTGCATCAAAGTGTGCGAATTCAGTTTTGCAATGATCCCTTCTGGGAGTGTACTATCACAAGAATAACAAGTCTGGATGGAGAGCGAATTTCCGG  
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 TCTACAGCAGCGCAAACACTGCACCTTCGAGTACGTGAGTCAACCTTTCTGATGGACTGGAAGGGAAACAGGGAACTTCAAGAACCTGAGAGAGTT  
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 TGTCTTTAAGAACATCGACGGCTATTTAAGATCTATAGTAAGCATAACGCTTCAACCTGGAAGGGATCTTCCCAGGGCTTTTTCAGCCCTGGAACCT  
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 TTGTTGACTTGCCTATTGGTATCAATATCACCAGATTTTCAGACCCCTTGGCATTGCAICGGTCTTATCTTACCAGGTATTCCTCCCGGGTGGGA  
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 CTGCCGCGCGCTGCCTACTATGTCGGCTATCTGCAACCAAGAAGCTTCTGCTCAAGTACAACGAAACGGCACTATTACGGATGCTGTTGATTGTGC  
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 CCTGGACCTCTGTCTGAGACTAAATGCACCTCAAGAGCTTTACCGTTGAGAAGGGGATTTACCAAACAGTAATTTCCGGGTCCAACCCACCGAAAGC  
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 ATTTGCGGGTCCCAATATCACAATCTGTGTCCCTTTGGCGAAGTGTTCATGCTACAAGTGTGCTTCTGTGTACGCATGGAATAGGAAACGCATCT  
311 I V R F P N I T N L C P F G E V F A N A T R F A S V Y A W N R K R I  
1901 CCAATTGTGTCGCTGATTACTCCGTGCTGTACAATTCGCGCTTTTCTCAACCTTCAAGTGTATGGCGTTTCACTACCAACTAACCRKRTGTGCTT  
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 CACTAATGTGTATGCCGACTCTTTTGTGATACGAGGCGATGAAGTGAAGACAGATTGCACCAGGGCAGACCGGCAAAATGGCGACTACAACACTACAAGCTT  
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 CCAGATGACTTTACCGGATGTGTTATTGCATGGAACCTCAAACAATCTGGATTCCAAGTGGGTGGCAACTATAACTACCTGTATAGACTGTTCCAGGAAAT  
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 I T N L C P F G E V F A N A T R F A S V Y A W N R K R I  
2301 GCAGAGTTACGGATTCCAGCCTACAACGGGGTGGGTTACCAACCCTATCGTGTCTGAGTCTGAGTTTTGAGCTCCTCCATGCCCGACCCAGTCTGT  
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 GGCCCCAAGAAAAGCACCAATCTGGTGAAGAACAATCGGTGAACCTTAACTTAAACGGACTCACAGGAACCGCGTATTGACGGAGAGTAACAAGAAGT  
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 TCCTGCCATTCCAGCAGTTCCGGTTCGCGATATTGCCGACACTACCGACGCTGTCCGAGATCCCCAGACATTGGAGATTCTTGATATCACACCCTGTAGTTT  
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 CGCGGAGTGAAGTGTGATTACGCCCGCAACCAATACCAGCAATCAGGTTGCCGCTGATCAGGACGTGAATTGCACCGAGGTACCTGTCCCATCCAC  
577 G V G Y T P G T N T S Y Q A G V L Y Q C N G V C E G F V N C Y A I H  
2701 GCTGACCAACTTACACCCACATGGCGAGTATATTCCACCGGCTCCAACGCTTTTTCAGACACGTGCTGGATGTCTGATCGGTGCAGAACACGTTAATAATA  
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  
2801 GCTACGAGTGTGATATCCCATCGGTGCTGGAATATGCGCCTTATCAAACCTCAAACCAACTCTCCTAGGCGGGCACGTAGTGTAGCATCCCAAAGTAT  
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 S V A S Q S I  
2901 CATTGCCTACACAATGAGCCTCGGTGCTGAGAATCTGTGCGCTACAGCAACAACCTCATTGCTATCCCTACTAATCTCACAATCAGTGTGACAACCTGAA  
677 I A Y T M S L G A E N S V A Y S N N S I A I P T N F T I S V T T E  
3001 ATTCGCGGTATCTATGACCAAAAACGTTGACTGCACCATGTACATCTGTGGCATTCTACCGAATGTAGCAATCTCCTCCTGCAATACGGATCAT  
711 I L P V S M T K T S V D C T M Y I C G D S T E C S N L L L Q Y G S  
3101 TCTGCACTCAGCTGAATCGTGCCTCACAGGATTGCAAGTGTGAGCAGGACAAGAATACGAGGAAGTGTTCGCCAGGTGAAGCAATCTACAAAACCTC  
744 F C T Q L N R A L T G I A V E Q D K N T Q E V F A Q V K Q I Y K T P

3201 ACCATAAAAGACTTTGGCGGATTCAATTTCTCACAGATCCTGCCGATCCCTCAAACCTCAAGCGTAGCTTTATCGAGGATCTGCTCTTCAACAAG  
777▶ P I K D F G G F N F S Q I L P D P S K P S K R S F I E D L L F N K  
3301 GTAACCTCGCAGATGCCGGTTTCATCAAGCAGTATGGCGATTGTCTGGGAGACATCGCCGCTCGGGACCTGATCTGTGCACAGAAGTTCAATGGACTGA  
811▶ V T L A D A G F I K Q Y G D C L G D I A A R D L I C A Q K F N G L  
3401 CCGTGTGCTCCCTTGTGACCGACGAGATGATGCCCAATACATAGCGCCTGCTGGCCGGCACCATCACTTCTGGGTGGACATTCGGAGCTGGCGC  
844▶ T V L P L P L L T D E M I A Q Y T S A L L A G T I T S G W T F G A G A  
3501 TGCCCTTCAGATTCTTTTGTATGCAGATGGCTACCCTTTAACGGCATCGGTGTGACACAAAACGTTCTGTATGAAAACAGAACTCATCGCCAAC  
877▶ A L Q I P F A M Q M A Y R F N G I G V T Q N V L Y E N Q K L I A N  
3601 CAGTTCAACAGTGTATCGGTAAGATACAGGATAGCCTGTCATCCACTGCCAGCGCATTGGGAAAGTGCAGGATGTAGTGAACCAGAATGCCAGGCAC  
911▶ Q F N S A I G K I Q D S L S S T A S A L G K L Q D V V N Q N A Q A  
3701 TTAACACCTGGTGAACAGCTCTCTTCAAATTTTGGTGCCATTTCTAGCGTGTGAATGACATACTGAGCCGGTTGGACAAGGTGGAGGCTGAAGTGA  
944▶ L N T L V K Q L S S N F G A I S S V L N D I L S R L D K V E A E V Q  
3801 GATTGATAGGCTGATAACTGGCGCCTTCACTCTTCCAGACCTATGTGACCCAGCAGCTCATCCGCGCTGTGAAATTCGCGCATCCGCTAACCTGGCA  
977▶ I D R L I T G R L Q S L Q T Y V T Q Q L I R A A E I R A S A N L A

NcoI (3995)

3901 GCAACCAAATGTCCGAGTGTGTGCTGGGTGAGTCTAAGAGAGTGGACTTTTGGCGGAAGGGGTATCACTGATGTCTTTTCTCAGTCTGCACCCCATG  
1011▶ A T K M S E C V L G Q S K R V D F C G K G Y H L M S F P Q S A P H  
4001 GTGTGGTCTTTCTGCACGTGACTTATGTCCCAGCTCAGGAAAAGAACTTCACTACAGCCCCAGCCATCTGCCACGATGGGAAAGCCCACTTTCCAGGGA  
1044▶ G V V F L H V T Y V P A Q E K N F T T A P A I C H D G K A H F P R E  
4101 AGCGGTATTCGTGTTCAATGGTACTTGGTTGCTCACTCAGAGAAATTTCTACGAGCCCCAGATTATAACCACTGACAATACATTTGTATCCGGCAAT  
1077▶ G V F V S N G T H W F V T Q R N F Y E P Q I I T T D N T F V S G N  
4201 TGTGATGTGGTTATCGGGATTGTGAATAACTGTTTACGATCCTTTGCGACCCAGAGCTGGACTCCTTCAAGGAGGAGCTTGACAAATATTTAAGAATC  
1111▶ C D V V I G I V N N T V Y D P L Q P E L D S F K E E L D K Y F K N  
4301 ACACATCACTGACGTGACCTCGGAGATATTTCAAGAAATCAATGCTTCCGTGGTCAATATTCAGAAGGAGATAGACAGGCTGAATGAGGTTGCCAAGAA  
1144▶ H T S P D V D L G D I S G I N A S V V N I Q K E I D R L N E V A K N

BsiWI (4461)

BamHI (4491)

4401 CCTCAACGAGTCTCTGATCGATCTGCAGGAGTTGGGCAAGTACGAACAGTATATCAAATGGCGTACGGAGAACCTGTACTTCCAGGGCTCTGGATCCGAG  
1177▶ L N E S L I D L Q E L G K Y E Q Y I K W R T E N L Y F Q G S G S E

4501 CCAAATCTAGTGACAAAACCTCACACATGCCACCGTCCCGAGCACCTGAAGCCGAGGGGGACCGTCACTTCTCTTCCCCCAAACCAAGGACC  
2▶ 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 P P K P K D

4601 AACTGATGATCTCCCGACCCCTGAGGTCACATGCGTGGTGGTGGAGCTGAGCCACGAAGACCTGAGGTCAAGTCAACTGGTACGTGGACGGCGTGA  
35▶ 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 W Y V D G V E

4701 GGTGCATAATGCCAAGACAAAGCCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAAGCGTCCCTACCCTCCTGCACCAGGACTGGCTGAATGCC  
68▶ 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 Q D W L N G

4801 AAGGAGTACAAGTGAAGTCTCCAACAAAGCCCTCCAGCCICCATCGAGAAAACCATCTCAAAGCAAAGGGCAGCCCCGAGAACCACAGGTGTACA  
102▶ 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 R E P Q V Y

4901 CCCTGCCCCATCCGGGAGGAGATGACCAAGAACCAGGTGACCTGACCTGCTGGTCAAAGGCTTCTATCCAGCGACATCGCCGTGGAGTGGGAGAG  
135▶ 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 I A V E W E S

5001 CAATGGGACCCGGAGAACAACCTACAAGACCACGCCTCCCGTGTGGACTCCGACGGCTCCTTCTTCTACAGCAAGCTCACCGTGGACAAGAGCAGG  
168▶ 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 T V D K S R

5101 TGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGCTGCATGAGGCTCTGCACAACCACTACACGAGAAGAGCCTCTCCCTGTCTCCGGTAAATGAGTCC  
202▶ 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 S P G K •

MscI (5204)

5201 TAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATT

5301 GCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTCATTTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGGGAGGTTTTTT

5401 AAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATCTAAAATACAGCATAGCAAACTTTAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGA

5501 GGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTAAGATATAGTGTATTTTCC

5601 CAAGTTTGAAGTACTGCTCTTCTTTTATTGTTTAAATGCACTGACCTCCACATTCCTTTTATAGTAAATATTCAGAAATAATTTAAATACATCAT

5701 TGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAGGAAC

5801 CTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGC

5901 CATTATCTCAATGAGCACAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTGCACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTC  
121▶ 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 I S R D V E

6001 ATCAGAGTAGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCCAGCACAGACAGTGAAC  
88▶ 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 E A C V T V

6101 CTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCAGTCTTGGTCTGATGGCCGCCCGACATGGTGCTTGTGTCCTCATAGAGCATGG  
54 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 N D E Y L M T  
6201 TGATCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGAAGGTCTTCATGGTGGCCCTCTATAGTGAGTCGATTATACTATGC  
21 I K E T A V E V L E L D Q Q S I N F T K M  
6301 CGATATACTATGCCGATGATTAATTGTCAAACACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACATAACGAGCTCTGCTTATATAGACCTCCCA  
6401 CCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAAAACAACTCCCATGACG  
6501 TCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTA  
6601 ATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGT  
6701 ACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCGAGTTTACCGTAAATACTCCACCCATTGACGTCAATGAAAGTCCCTATTGGCGTTACTATG  
6801 GGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAAGTTATGTAACGCTGCAGGTTAATTAA  
6901 GAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAA  
7001 AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGTCCCTCGTGCCTCTCCTGTTCCGACCC  
7101 TGCCGTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGTTGCG  
7201 CTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTA  
7301 TCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA  
7401 GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGG  
7501 TGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAA  
7601 AACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAATCAGCGGCCGAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTT  
7701 TTTGTGTAATCGTAACTAACATACGCTCTCCATCAAAAACAAAACGAAAACAAAACAACTAGCAAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGA  
7801 ACATTTCTCTATCGAA