



1 GGACCTGCAGGGCTGAATAACCTCTGAAAGAGGAAGCTGGTTAGGTACCTCTGAGCGGAAAGAACAGCTGTGGAATGTGTGTCAGTTAGGGTGTG
101 GAAAGTCCCAGGCTCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGAAAGTCCCAGGCTCCCAGCAGGCAG
201 AAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCACTAGTCTCGGGTCCCGTCACTGGGAGAGCGCACATCGCCACAGTCCCCGA
301 GAAGTTGGGGGAGGGGTGCGCAATTGAACGGGTGCCTAGAGAAGGTGGCGGGGTAACCTGGGAAAGTGTGCTGTACTGGCTCCGCTTTTTCCC
401 GAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGGCCGAGAACACAGCTGAAGCTTCGAGGGGCTC
501 GCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCCGCCATCCACGCGGTTGAGTCGCGTTCTGCCGCTCCCGCTGTGGTGCCTCTGAACTGC
601 GTCCGCGCTTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGC

NcoI (799)

Agel (791)

701 CTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCTGTTTCTGTCGCGAGTTACAGATCCAAGCTGTGACCGGGCGCTACCTGAGATCACCGGTCA

XhoI (852)

801 CATGGAGATCAAGGTGCTGTTTGCCTCATCTGATTGCTGTTGCTGAGGCACTCGAGGTTAATCTACCACCCGAATCACTCCCACCCGCATATACA
1 M E I K V L F A L I C I A V A E A L E V N L T T R T Q L P P A Y T
901 AATTCCTTACCAGAGGAGTGTACTATCCTGACAAAGTGTTCGGTCAAGTGTCTCCACTCTACTCAGGACCTCTTCTGCCTTTCTTTTCTAACGTTA
15 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 F S N V
1001 CATGGTTTCATGCAATCCATGTGTCTGGGACAAACGGCACAAACGCTTCGACAACCTGTATTGCCATTCAATGATGGGGTGTACTTTGCCTCCACAGA
48 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 A S T E
1101 GAAATCCAACATCATTGAGGATGGATTTTCGGGACTACTCTGGACTCAAAGACACAGAGCCTGCTGATCGTTAACACGCCACAAACGTTGTATCAA
81 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 V V I K
1201 GTGTGCGAATCCAGTTTTGCAATGATCCCTTCTGGGAGTGTACTATCACAAGAATAACAAGTCTGGATGGAGAGCGAATTTGGGGTCTACAGCAGCG
115 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 V Y S S
1301 CAAACAACTGCACCTTCGAGTACGTGAGTCAACCTTTCTGATGGACTGGAAGGAAACAGGGAAACTTCAAGAACCTGAGAGAGTTTGTCTTAAAGAA
148 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 V F K N
1401 CATCGACGGTATTTAAGATCTATAGTAAGCATACGCCTATCAACCTGGTAAGGGATCTTCCCAGGGCTTTTTCAGCCCTGGAACCTTTGGTTGACTTG
181 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 L V D L
1501 CCTATTGGTATCAATACACAGATTCAGACCTTCTGGCATTGCAICGGTCTTATCTTACTCCAGGTGATTCCTCTCCGGTGGACTGCCGGCGCCG
215 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 T A G A
1601 CTGCTACTATGTCGGTATCTGCAACCAAGAAGCTTCTGCTCAAGTACAACGAAACGGCACTATTACGGATGCTGTTGATTGTGCCCTGGACCCTCT
248 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 L D P L
1701 GTCTGAGACTAAATGCACCTCAAGAGCTTTACCGTTGAGAAGGGGATTTACCAAACAGTAATTTCCGGGTCCAACCCAGAAAGCATTGTGCGGTT
281 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 I V R F
1801 CCAAATATCACAATCTGTGTCCCTTTGGCGAAGTGTCAATGCTACAAGGTTTGTCTGTGTACGCATGGAATAGGAAACGCATCTCCAATTGTGTG
315 P N I T N I L C P F T G C A A G E V F T S V Y A W N R A K R I S N A C V
1901 CTGATTACTCCGTGTGACATTCGCTCTTCTCAACCTTAAGTGTATGGCGTTCACCTACCAAACCTTAACACCTGTGCTTCACTAATGTGTA
348 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 T N V Y
2001 TGCCGACTCTTTTGTGATACGAGGCGATGAAGTGAACAGATTGCACCAGGGCAGACCGGCAAAATGCGGACTACAACCTACAAGCTTCCAGATGACTTT
381 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 P D D F
2101 ACCGGATGTGTTATTGCATGGAACCTCAAACAATCTGGATTCCAAGGTGGGTGGCAACTATAACTACCTGTATAGACTGTTCCAGGAAATCCAACCTGAAAC
415 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 S N L K
2201 CATTGAGCGAGATATAAGCACAGAAATCTACCAGGCTGGAAGTACGCCCTGCAACGGCGTGAAGGGTCAACTGCTACTTCCCATTGCAGAGTTACGG
448 P F E R D I S T E I Y Q A G S T Y C N D V G E G F N C V A I H A D Q L
2301 ATTCCAGCTACAACGGGGTGGGTTACCAACCTATCGTGTGCTGAGTCTGAGTTTTGAGCTCCTCCATGCCCCAGCCACAGTCTGTGGCCCAAGAAA
481 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 G P K K
2401 AGCACCAATCTGGTGAAGAACAATGCGTGAACCTTAACTTTAACGGACTCACAGGAACCGCGTATTGACGGAGAGTAACAAGAAGTCTCTGCCATTCC
515 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 F L P F
2501 AGCAGTTCGGTCCGATATTGCCGACACTACCGACGCTGTCCGAGATCCCCAGACATTGGAGATTCTTGATATCACACCCTGTAGTTTCGGCGGAGTGAG
548 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 G G V S
2601 CGTGATTACGCCGGAACCAATACCAGCAATCAGGTTGCCGCTGTATCAGGACGTGAATTCACCCAGGTTACCTGTCCGATCCACGCTGACCAACT
581 V I T P G T N T S N Y Q A V S L Y Q D V N C T E V P V A I H A D Q L
2701 ACACCCACATGGCGAGTATATCCACCGGCTCCAACGCTCTTTCAGACACGCTGATGATGCTGATCGGTGCGAAGACGTTAATAATAGTACGAGTGTG
615 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 S Y E C
2801 ATATCCCATCGGTGCTGGAATATGCGCCTCTTATCAAACCTCAAACCAACTCTCTAGCGGGCAGTGTAGTATCCAAAGTATCATTGCCCTACAC
648 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 I A Y T
2901 AATGAGCCTCGGTGCTGAGAATTCTGTGCTCAGCAACAACCTCCATTGCTATCCCTACTAACTTCACAATCAGTGTGACAACCTGAAATCTGCCGTA
681 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 I L P V
3001 TCTATGACCAAAAAGCGTTGACTGCACCATGTACATCTGTGGCGATTCTACCGAATGTAGCAATCTCCTCCTGCAATACGGATCATTCTGCACTCAGC
715 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 F C T Q
3101 TGAATCGTGCCCTCACAGGATTGCAAGTGTGAGCAGGACAAGAATACGAGGAAGTGTTCGCCAGGTGAAGCAAACTACAAAACCTCACCCATAAAAAG
748 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 P I K D

3201 CTTTGGCGGATTCAATTTCTCACAGATCTGCCCAGTCCCTCAAACCTCCAAGCGTAGCTTTATCGAGGATCTGCTCTTCAACAAGTAACCTCGCA
781▶ 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 V T L A
3301 GATGCCGGTTTTCATCAAGCAGTATGGCGATTGTCTGGGAGACATCGCCGCTCGGGACCTGATCTGTGCACAGAAGTTCAATGGACTGACCGTCTGCCTC
815▶ 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 T V L P
3401 CCTTGGTACCACGAGATGATGCCAATACTAGCCCTGCTGGCCGACCATCACTTCTGGTGGACATTCGGAGCTGGCGCTGCCCTTCAGAT
848▶ 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 A L Q I
3501 TCCTTTTGTATGCGATGGCCTACCGCTTTAACGGCATCGGTGTGACACAAAACGTTCTGTATGAAAACAGAACTCATCGCCAACAGTTCAACAGT
881▶ 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 Q F N S
3601 GCTATCGGTAAGATACAGGATAGCCTGTCCACTGCCAGCGCATTGGGAAAGTTGAGGATGTAGTGAACCAGAATGCCAGGCACTTAACACCCTGG
915▶ 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 L N T L
3701 TGAACAGCTCTCTTCAAATTTTGGTGCATTTCTAGCGTGTGAATGACATACTGAGCCGTTGGACAAGGTGGAGGCTGAAGTGCAGATTGATAGGCT
948▶ 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 I D R L
3801 GATAACTGGGCGCTTTCAGTCTCTTCAGACCTATGTGACCCAGCAGCTCATCCGCGTGTGAAATTCGCGCATCCGCTAACCTGGCAGCAACCAAATG
981▶ 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 A T K M

NcoI (3983)

3901 TCCGAGTGTGTGCTGGGTGAGTCTAAGAGAGTGGACTTTTTCGCGGAAGGGTATCACCTGATGTCTTTTCTCAGTCTGCACCCCATGGTGTGGTCTTTT
1015▶ 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 G V V F
4001 TGCACGTGACTTATGTCCAGCTCAGGAAAAGAACTTCACTACAGCCCAGCCATCTGCCACGATGGGAAAGCCACTTTCCAGGGAAGGCGTATTCGT
1048▶ 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 G V F V
4101 GTCCAATGGTACTCATTGGTTGCTCAGAGAAAATTTCTACGAGCCCAGATTATAACCACTGACAATACATTTGTATCCGGCAATGTGATGTGGTT
1081▶ 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 C D V V
4201 ATCGGGATTGTGAATAACTGTTTACGATCCTTTCAGCCAGAGCTGGACTCCTTCAAGGAGGAGCTTGACAATAATTTAAGAATCACACATCACCTG
1115▶ 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 H T S P
4301 ACGTCGACCTCGGAGATATTTAGGAATCAATGCTTCCGTGGTCAATATTGAGAAGGAGATAGACAGGCTGAATGAGGTTGCCAAGAACCTCAACGAGTC
1148▶ 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 L N E S

BamHI (4449)

NheI (4480)

4401 TCTGATCGATCTGCAGGAGTTGGGCAAGTACGAACAGTATATCAAATGGGATCCGGCCATCATCATCACCATCACTAAAGCTAGCTGGCCAGACATGAT
1181▶ L I D L Q E L G K Y E Q Y I K W G S G H H H H H H •
4501 AAGATACATTGATGAGTTTGACAAACCACAAC TAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCAT

4601 ATAAGCTGCAATAAACAAGTTAAACAACAACAATTGCATTCAATTTATGTTTCAGGTTACAGGGGAGGTGTGGGAGTTTTTTAAAGCAAGTAAACCTCT

4701 ACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACCTTAACTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAG

4801 GCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTGCAGCCTCACCTCTTTCATGGAGTTAAGATATAGTGTATTTTCCAAGGTTTGAAGTACTGCTC

4901 TTCATTTCTTTATGTTTTAAATGCAGTACCTCCACATTCCTTTTTAGTAAAATATTCAGAAATAATTTAATACATCATTGCAATGAAAATAAATGT

5001 TTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATGGAC

5101 AGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCACTCAATGAGCA

5201 CAAAGCAGTCAGGAGCATAGTCAGAGTGAAGCTCTCTGCACATGCCACAGGGCTGACCACCTGATGGATCTGTCCACTCATCAGAGTAGGGGTGCTC

5301 GACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCAATGGCAATGGCTTTCAGCACAGACAGTACCCTGCCAATGTAGGCCTCA

5401 ATGTGGACAGCAGAGATGATCTCCAGTCTTGGTCTGATGGCCGCCGACATGGTGTCTGTCTCATAGAGCATGGTGTCTCTCAGTGGCGA

5501 CCTCCACAGCTCCAGATCCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGATTATACTATGCCGATATACTATGCCGATG

5601 ATTAATTGTCAAACAGCGTGGATGGCTCTCCAGCTTATCTGACGGTTCATAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCC

5701 CATTTGCGTCAATGGGGCGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAAAACAACTCCCATGACGTCAATGGGGTGGAGACTT

5801 GGAAATCCCCTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCC

5901 AAGTAGGAAAGTCCATAAAGTCATGTACTGGCATAATGCCAGGCGGGCCATTTACCCTGATTGACGTCAATAGGGGGCTACTTGGCATATGATACAC

6001 TTGATGTACTGCCAAGTGGGAGTTTACCGTAAATACTCCACCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTT

6101 GACGTCAATGGGGGGGTGCTTGGGCGGTGAGCCAGGCGGGCCATTTACCCTGAAAGTATGTAACCGCTGAGGTTAATTAAGAACATGTGAGCAAAAGG

6201 CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGGTGGCGTTTTTCCATAGGCTCCGCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAG

6301 AGGTGGCGAAACCCGACAGGACTATAAAGATACAGGCGTTTTCCCTGGAAGCTCCCTCGTGGCTCTCTGTTCCGACCTGCCGCTTACCGGATACC

6401 TGTCCGCTTTCTCCCTCGGGAAGCGTGGCGCTTTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGTGTGCTTCCAAAGTGGGCTGTGT

6501 GCACGAACCCCGTT CAGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
6601 ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTA
6701 TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTGCAA
6801 GCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACTCACGTTAAGGGATT
6901 TTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACT
7001 AACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA