



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCGCGTTTGC CGCCTCCCGCTGTGGTGCTCTGAACTGCGTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

Mlul (558)

501 TCTGTTCTGCGCGTTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGACGCGTCCAGCATGCCAGCGCAACAGCTCCAGCCGAGTGCCT
1 M P A R T A P A R V P
601 TGCCTTGCCTCCCGGAGGCTCGCTCCCGACCATGTCGCGAGCGGCTCAAAGACTTGGAAAGAGATGGCTTAAACAGAAAAGGAGTGTGTGAGGGAG
11 A L A S P A G S L P D H V R R R L K D L E R D G L T E K E C V R E
701 AAATTAACCTTACTGCATGAATCCTGCAAACAGAAATAAAAAGCCAGTTGTGTGACTTGGAAACCAATTACATAAAGAGGAATTATCTGAGGAAGGCT
45 K L N L L H E F L Q T E I K S Q L C D L E T K L H K E E L S E E G
801 ACCTGGCTAAAGTCAAGTCCCTCTAAATAAGGATTTGCTTGGAGAACGGAACACACACTCTCACTCAAAAAGCCAACGGTTGTCCCGCCAACGGGAG
78 Y L A K V K S L L N K D L S L E N G T H T L T Q K A N G C P A N G S
901 CCGGCCAACCTGGAGAGCAGAAATGGCAGACTCAAATAGATCCCAAGATCCAGGCCAACGCTCGGGGACCCAGGAGAAGCAAGTCGGACAGTGACACC
111 R P T W R A E M A D S N R S P R S R P K P R G P R R S K S D S D T
1001 CTTTCAGTTGAAACTTACCTAGTTCGGTGTACGAGGAGAACCACAGGACAGACCACCATCACGGCTCACTTACGAAGGGCCCACTAAACGAAAC
145 L S V E T S P S S V A T R R T T R Q T T I T A H F T K G P T K R K
1101 CCAAGGAAGAGTGGAAAGGGGAACTCGGCTGAGTCGGCTGCAGAGGAGAGAGACCAGGATAAGAAACGAGAGTTGTAGACACAGAGAGTGGTGTGC
178 P K E E S E E G N S A E S A A E E R D Q D K K R R V V D T E S G A A
1201 AGCTGCTGTGGAGAACTGGAAGAGGTAACAGCGGAAACCCAGCTGGGTCCGGAAGAGCCATGTGAACAGGAAGATGACAACAGGAGTCTTCGACGTAC
211 A A V E K L E E V T A G T Q L G P E E P C E Q E D D N R S L R R H
1301 ACCAGAGACTATCATTGAGCGGAAATCAAAGGAGGATCCAGACAGAGAAGCAAGACCGAAACTCACTTGGACGAGGACGAGGACGAAAAAGGATA
245 T R E L S L R R K S K E D P D R E A R P E T H L D E D E D G K K D
1401 AAAGAAGTTCAGACCCAGGAGCCAGCCAGAGATCCAGTGCCAAACGGAGACCCAAAGGAAGCAGAGCCAGAGCAGGTAGTCCAGAGACTCCCGAGGA
278 K R S S R P R S Q P R D P A A K R R P K E A E P E Q V A P E T P E D
1501 CAGAGACGAGGATGAGAGGGAGGAGAAGAGACAAAAACGACCGTAAAAAAGTGGAGTCACACACCGTTCCCGTTCAGAGCAGATCGGAGAGAAAAGCC
311 R D E D E R E E K R R K T T R K K L E S H T V P V Q S R S E R K A
1601 GCTCAAAGCAAAGTGTGATCCCGAAGATCAACTACCAAAGTGCCTCGAGTGTGGCCAGCACCTAGACGACCTAACCTGAAGTACCAGCAGCACCTG
345 A Q S K S V I P K I N S P K C P E C G Q H L D D P N L K Y Q Q H P
1701 AGGATGCTGTGGATGAACCCAGATGTTGACCCAGTGGAAACTGTCCATCTACGACTCCACCTCGACCTGGTTTGATACTTGAAGATTCTCCCATGCA
378 E D A V D E P Q M L T S E K L S I Y D S T S T W F D T Y E D S P M H
1801 TAGGTTCACTTCTTTCAGTGTGACTGCAGTCGCGGCACCTGTGCTGTGACACCGGCTCATTGAGAAGAATGTAGAGCTCTACTTTTCTGGGTGT
411 R F T S F S V Y C S R G H L C P V D T G L I E K N V E L Y F S G C
1901 GCCAAAGCAATTCATGACGAGAATCCATCTATGGAAGTGGTATTAATGGCAAACCTCGGGCAATCAATCAGTGGTGGCTCAGTGGCTTTGATGGTG
445 A K A I H D E N P S M E G G I N G K N L G P I N Q W W L S G F D G
2001 GCGAGAAGGTCTCATTGGCTTCTCCACTGCATTGCTGAATACATTTGATGGAGCCAGCAAGAGTATGAGCCAATATTTGGGCTGATGAGGAGAA
478 G E K V L I G F T S T A F A E Y I L M E P S K E Y E P I F G L M Q E K
2101 AATTTACATCAGCAAGATTGTTGTTGAGTTCCTGCAAAAACCTGATGCTGTATATGAAGACTGATCAATAAGATTGAGACCACTGTTCTCCTTCT
511 I Y I S K I V V E F L Q N N P D A V Y E D L I N K I E T T V P P S
2201 ACCATTAATGTGAACCGTTTACAGAGGACTCCCTCTACGCCAGCCAGTTTGTAGTGAGCCAGGTAGAGATTACGACGAAGCCAAGGACGATGATG
545 T I N V N R F T E D S L L R H A Q F V V S Q V E S Y D E A K D D D
2301 AGACCCCATCTTCTTGTCTCCCTGTATGAGAGCCCTGATCCATTTGGCTGGTGTCTCCCTGGGACAGAGGCGAGCAACAAGCGCGTATGGGTGCTAC
578 E T P I F L S P C M R A L I H L A G V S L G Q R R A T R R V M G A T
2401 CAAGGAGAAGGACAAAGCACCCACGAAAGCCACCACCAAGCTGGTCTATCAGATCTTTGACACTTCTTCTCAGAGCAGATTGAGAAGTATGATAAG
611 K E K D K A P T K A T T T K L V Y Q I F D T F F S E Q I E K Y D K
2501 GAGGACAAGGAGAATGCCATGAAGCGCCCGCTGTGGTGTCTGTGAGGTCTGTGACGAGCTGAGTGTGGGAAGTGAAGGCGTGAAGATATGGTGA
645 E D K E N A M K R R R C G V C E V C Q Q P E C G K C K A C K D M V
2601 AGTTTGGTGGCACTGGACGGAGTAAGCAGGCTTGCCTCAAGAGGAGGTGTCTAACTTGGCGGTGAAGGAGGCAGACGACGATGAAGAGGCTGATGATGA
678 K F G G T G R S K Q A C L K R R C P N L A V K E A D D D E E A D D D
2701 TGTGTCAGAGATGCCATCACCAAAAAGCTGCATCAGGGGAAGAAGAAGAAGCAGAACAAGGACCGCATCTCTGGCTTGGGACGCTATGAAGATTGAA
711 V S E M P S P K K L H Q G K K K K Q N K D R I S W L G Q P M K I E
2801 GAGAATAGAATTACTATCAGAAGGTGAGCATGATGAGGAGATGCTAGAGTGGGCGACTGCGTCTCGGTCATTCCAGATGATTCCTCAAAACCACTCT
745 E N R T Y Y Q K V S I D E E M L E V G D C V S V I P D D S S K P L
2901 ATCTAGCCAGGTCACAGCTCTGTGGGAAGACAAAATGGTCAAGTATGTTCCATGCGCACTGGTCTGCGCTGGGACAGACACAGTCTGGGAGCCAC
778 Y L A R V T A L W E D K N G Q M M F H A H W F C A G T D T V L G A T
3001 CTCCGACCCCTGGAAGTGTCTGGTGGGCGAGTGCAGAAAACATGCAGCTTCTTACATCCACAGCAAGGTCAAGGTCTACAAAAGCCCTTCTGAA
811 S D P L E L F L V G E C E N M Q L S Y I H S K V K V I Y K A P S E
3101 AACTGGCCATGGAGGGAGGCACAGACCTGAGACCACACTGCCTGGGGCTGAGGATGGCAAGACTTACTTCTCCAGCTCTGGTACAACAGGAGTACG
845 N W A M E G G T D P E T T L P G A E D G K T Y F F Q L W Y N Q E Y
3201 CAAGGTTTGAATCCCAACCAAGACCCGACCCGAGGACAACAAGCACAAGTTCTGCCTATCTTGTATCCGGCTGGCTGAGCTGAGACAAAAAGAAAT
878 A R F E S P P K T Q P T E D N K H K F C L S C I R L A E L R Q K E M

3301 GCCCAAGTCTCTGGAACAAATTGAGGAGGTGGATGGCCGGTCTACTGCAGTTCATACCAAGAATGGTGTGTCTACCGACTGGGTGACAGTGTGTAC
911▶ P K V L E Q I E E V D G R V Y C S S I T K N G V V Y R L G D S V Y
3401 CTTCTCCCGAGGCTTTACTTTCAACATCAAAGTGGCTAGCCCCGTGAAACGCCAAAGAAGGATCTGTGAACGAGACCCGTACCTGAGCACTACC
945▶ L P P E A F T F N I K V A S P V K R P K K D P V N E T L Y P E H Y
3501 GCAAGTATTCTGACTACATCAAGGGGAGCACTGGATGCTCCAGACCCCTATCGCATCCGTCGGATAAAAAGATCCACTGTGGCAAGAAAGGCAA
978▶ R K Y S D Y I K G S N L D A P E P Y R I G R I K E I H C G K K K G K
3601 GGTCACGAGGCAGACATCAAGCTGAGGCTCTACAAGTCTACAGGCTGAGAATACCCACAGGTCCTACAACGGATCCTATCACACTGACATCAACATG
1011▶ V N E A D I K L R L Y K F Y R P E N T H R S Y N G S Y H T D I N M
3701 CTTTACTGGAGCAGGAGGAAGTGTGGTGAATTCAGCGACTGCAGGGCCGTGTACCGTGGAGTACGGGAAGACCTACTTGAGAGCATCCAGGATT
1045▶ L Y W S D E E A V V N F S D V Q G R C T V E Y G E D L L E S I Q D
3801 ATTCACAAGGGGGCCCTGACCGTTCTACTTCTCAGGCTACAATTCAAAGACCAAGAAGTTCGAGACCCACCAACCATGCCCCGAGCCCTGGGAA
1078▶ Y S Q G G G P D R F Y F L E A Y N S K T K N F E D P P N H A R S P G N
3901 CAAAGGAAAGGGGAAAGGGGAAAGGGGAAAGGGAAGCATCAGGTGTGAGAGCCAAAGAGCCTGAGGCAGCCATCAAAGTCCGCAAGCTCCGGACC
1111▶ K G K G K G K G K G K G K G K H Q V S E P K E P E A A I K L P K L R T
4001 CTGGATGTGTTTTCCGGCTGTGGAGGTTATCGGAAGGATTCCACCAAGCAGGCATCTCGAAACGCTGTGGCCATCGAGATGTGGGACCCGGCAGCCC
1145▶ L D V F S G C G G L S E G F H Q A G I S E T L W A I E M W D P A A
4101 AGGCATTTCCGGTGAACAACCCCGCACCACAGTGTTCACAGAGGACTGCAACGTGCTTCTTAAGCTGGTCATGGCTGGGAGGTGACCAACTCTCTGGG
1178▶ Q A F R L N N P G T T V F T E D C N V L L K L V M A G E V T N S L G
4201 CCAAAGGCTGCCAAGAAGGGGATGTGGAGATGCTGTGTGGTGGGCCACCCTGCCAGGGCTCAGTGGCATGAACCGCTTCAACTCCCGCACTTACTCC
1211▶ Q R L P Q K G D V E L C G G P C Q G F T S G M N R F A N S R T Y S
4301 AAGTTCAAAAACCTCAGTGGTCTCCTTCCAGTACTGTGACTACTACCGGCTCGGTTCTTCTTGGAGAACGTGAGGAACCTCGTGTCTTACA
1245▶ K F K N S L V V S F L S Y C D Y Y R P R F F L L E N V R N F V S Y
4401 GACGCTCCATGGTGTGAAGCTCACACTGCGCTGCCTGGTCCGCATGGGCTACAGTGCACCTTTGGTGTGCTCCAGGCTGGACAGTATGGCGTGGCCCA
1278▶ R R S M V L K L T L R C L V R M G Y Q C T F G V L Q A G Q Y G V A Q
4501 GACACGAAGGAGGCCATCATCTTGGCTGCAGCCCCAGGAGAAAGCTGCCTCTGTTCCAGAGCCTCTGCATGTGTTGCGCCCCGTGCCTGCCAGCTG
1311▶ T R R R A I I L A A A P G E K L P L F P E P L H V F A P R A C Q L
4601 AGCGTTGTGGTGGATGACAAGAAGTTTGTAGCAACATAACGAGGCTGAGCTCGGGGCCCTCCGAACCATCACCGTGCAGACCATGTCTGACCTCC
1345▶ S V V V D D K K F V S N I T R L S S G P F R T I T V R D T M S D L
4701 CCGAGATCCAGAATGGAGCCTCGAATTCGAGATCCCCATAATGGAGGCCACTGTCTGTTCCAGAGGCAGCTGCGAGGATCACACTACCGCCAT
1378▶ P E I Q N G A S N S E I P Y N G E P L S W F Q R Q L R G S H Y Q P I
4801 CCTCAGGACCATATCTGCAAGGACATGAGCCACTGGTGGCTGCCCGCATGCGGCACATCCACTGTTCCAGGATCAGATTGGCGTACCTGCCAAC
1411▶ L R D H I C K D M S P L V A A R M R H I P L F P G S D W R D L P N
4901 ATACAGGTGCGGCTGGGAGATGGCGTATAGCCATAAGCTACAGTACACCTTTCATGATGTGAAAAATGGCTACAGCAGCACCAGGTCGCCCTGCGTGGAG
1445▶ I Q V R L G D G V I A H K L Q Y T F H D V K N G Y S S T G A L R G
5001 TCTGTTCTGTGCAGAAGGCAAGGCTGCGACCTGAGTCCAGGCAATTGAGCACCCTCATCCCCTGGTGCCTGCCGACACTGGGAACCCGGCACAACCA
1478▶ V C S C A E G K A C D P E S R Q F S T L I P W C L P H T G N R H N H
5101 CTGGGCTGGCCTCTACGGGCTGTGAGTGGGATGGCTTCTTCCAGCACCCTGTACCAACCTGAGCCCATGGGCAAGCAGGGTCCGGTGTCCACCCG
1511▶ W A G L Y G R L E W D G F F S T T V T N P E P M G K Q G R V L H P
5201 GAGCAGCACCAGGCTCGTGTGAGTGTTCGGGAATGTGCCGCTCCAGGGCTTCCAGATAGTACCGGTTCTTCCGCAACATCCTGGACAGACACCGGCAGG
1545▶ E Q H R V V S V R E C A R S Q G F P D S Y R F F G N I L D R H R Q
5301 TGGTAATGCTGTGCCACCACCCTGGCCAAAGCCATTGGCCTGGAGATTAAGCTCTGCTGTGTCAGTGTCCAGTGTCCGGGAGAGCGCATCAGCTGCAGTTAA
1578▶ V G N A V P P P L A K A I G L E I K L C L L S S A R E S A S A A V K

XbaI (5447)

5401 AGCAAAAGAGGAGGCTGCTACCAAGGACTAGTGTCTCACCCAGAGCTAGACTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACC
1611▶ A K E E A A T K D •
5501 ACAAC TAGAATGCAGTGA AAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAACAAGTTAACAACA
5601 ACAATTGCATTATTTTATGTTTCAGGTTCCAGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAATGTGGTATGGAATTCTAAAATAC
5701 AGCATAGCAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATT
5801 AGCTGTTTGCAGCCTCACCTTCTTCATGGAGTTAAGATATAGTGTATTTTCCCAAGGTTTGAAGTACTCTTCATTCTTTATGTTTAAATGCACTG
5901 ACCTCCCACATTCCTTTTTAGTAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAATAAATGTTTTTTATTAGGAGAATCCAGATGCTCA
6001 AGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGT
6101 TCCTGGTGTACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGTTCGCATTCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGAT
139▶ 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 D S I
6201 GAGCTCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCCCTC
106▶ 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 F D K
6301 TGCCGTTGCTCAGCAGACCAATGGCAATGGCTTCCAGCACAGACAGTACCTGCCAATGTAGGCTCAATGTGGACAGCAGAGATGATCTCCCAG
72▶ 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 I E G T
6401 TCTTGGTCTGATGGCCGCCGACATGGTGTGTTGTCCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGA
39▶ K T R I A A G G V H H K N D E Y L M T I K E T A V E V L E L D Q Q S
6501 GATGTTGAAGTCTTCTGTTGGCCCTCTATAGTGTGCTATTATACTATGCCGATATACTATGCCGATTAATTTGTCAAACAGCGTGATGGCG
6▶ I N F T K M
6601 TCTCCAGCTTATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCACCCTACACGCTACCGCCATTTCGCTCAATGGGCGGAGTTGTTA

6701 CGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATC
6801 CACGCCCATTGATGTA CTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTA CTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTA
6901 CTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGCCATATGATACACTTGATGTA CTACTGCCAAGTGGGCAGTTTAC
7001 CGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCTGTTGGGCG
7101 GTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAG
7201 GCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
7301 GATACCAGGCGTTTTCCCCTGGAAGTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGT
7401 GGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTCAGCCCGACCGC
7501 TCGCCTTATCCGTA ACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGT
7601 ATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
7701 CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGA
7801 TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTA
7901 AATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAACAAAACG
8001 AAACAAAACAACTAGCAAATAGGCTGTCCCCAGTGAAGTGCAGGTGCCAGAACATTTCTCTATCGAA