



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTCTGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

Agel (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTATCATGAACTTTGGGAGAGAGAGGGCTCTAAGAGATACTG
1 M N F A E R E G S K R Y C
601 CATTCAAACGAAACATGTGGCCATTCTCTGTGCGGTGGTGGTGGTGTAGGATTAATAGTGGACTTGCCGTGGGCTTGACCAGATCGTGTACTCCAGC
13 I Q T K H V A I L C A V V V G V G L I V G L A V G L T R S C D S S
701 GGGGACGGCGGGCCGGGACTGCGCCAGCTCCTCCACCTGCCTTCTCCACGGCCAGCCCTCAGGTCTCTGCCCAGGACCAGGACATCTGCCCGG
47 G D G G P G T A P A P S H L P S S T A S P S G P P A Q D Q D I C P
801 CCAGTGAGGATGAGAGCGGACAGTGGAAAACTTTCGACTGCCGACTTCGTCAACCCAGTCCACTACGACCTGCACGTGAAGCCCTGTTGGAGGAGGA
80 A S E D E S G Q W K N F R L P D F V N P V H Y D L H V K P L L E E D
901 CACCTACACGGGACCGTGCATCCATCAACCTGAGCGCTCCACCCGTAACCTGGCTGCACCTCCGGGAGACCAGGATCACCCGGCTCCCGGAG
113 T Y T G T V S I S I N L S A P T R Y L W L H L R E T R I T R L P E
1001 CTGAAGAGCCCTCTGGGACCAGGTGCAAGTCCGAGGTTGTTTCGAGTACAAAAAGCAGGAGTACGTGGTGGTGCAGGCGGAGGAAGAGCTTACCCCA
147 L K R P S G D Q V Q V R R C F E Y K K Q E Y V V V E A E E E L T P
1101 GCAGTGGAGATGGCCTGTATCTCTGACCATGGAGTTCGCCGGCTGGCTGAACGGCTCCCTGTGGGATTTTATAGAACCCTACACGGAGAACGGAGC
180 S S G D G L Y L L T M E F A G W L N G S L V G F Y R T T Y T E N G R
1201 AGTCAAGAGCATAGTGGCCACCGATCATGAACCAACAGATGCCAGGAAATCTTTCTGTTTGTGAGGCCAACAAAAAGCAACTTATACAATATCT
213 V K S I V A T D H E P T D A R K S F P C F D E P N K K A T Y T I S
1301 ATCACCCATCCCAAAGAATACGGAGCACTTTCAAATATGCCAGTGGCGAAAGAAGAGTCACTGGATGATAAATGGACTCGAACCACTTTGAGAAGCTG
247 I T H P K E Y G A L S N M P V A K E E S V D D K W T R T T F E K S
1401 TCCCATGAGCACGTACCTGGTGTGCTTTGCTGTACATCAATTTGACTCTGTAAGAGAATATCAAATAGTGGAAAACCTCTTACAATTTATGTCCAGCC
280 V P M S T Y L V C F A V H Q F D S V K R I S N S G K P L T I Y V Q P
1501 AGAGCAAAAGCACACAGCCGAATATGCTGCAAAATACTAAAAGTGTGTTTGATTATTTGAAGAATACTTTGCTATGAATTATTCTTCTCAAATTA
313 E Q K H T A E Y A A N I T K S V F D Y F E E Y F A M N Y S L P K L
1601 GATAAAATCGCTATTCAGATTTTGGCACTGGTGCATGGAGAAGTGGGACTCATCAGTACAGAGAAAACCACTGCTTTATGACCCCTAAGGAATCAG
347 D K I A I P D F G T G A M E N W G L I T Y R E T N L L Y D P K E S
1701 CCTCATCAAACCAAGGGGTGGCCACTGTGGTTGCCATGAACCTGTGTCATCAGTGGTTTGGAAATATTGTGACCATGGACTGGTGGGAAGACTTGTG
380 A S S N Q Q R R V A T V V A H E L V H Q W F G N I V T M D W W E D L W
1801 GCTAAATGAAGGATTTGCTTCTTTTGTGATTTTGGGAGTAAACCATGCAGAAACAGACTGGCAAATGCGTGACCAAATGTTACTTGAAGATGATTA
413 L N E G F A S F F E F L G V N H A E T D W Q M R D Q M L L E D V L
1901 CCTGTTCAAGAGGATGATTCTTTGATGTCTTCGCATCAATTATTGTGACTGTGACAACCCTGATGAAATAACATCTGTTTTGATGGAATATCCTATA
447 P V Q E D D S L M S S H P I I V T V T T P D E I T S V F D G I S Y
2001 GCAAGGGATCTTCTATTTGAGAATGCTTGAAGACTGGATAAAACAGAAATTTCAAAGGATGTGAGATGTACTTGGAAAAATACCAATTAAGAA
480 S K G S S I L R M L E D W I K P E N F F Q K G C Q M Y L E K A Y C Q F K N
2101 TGCAAAAACCTTCTGATTTTGGGCGACTGGAAGAGGCTAGGCTACCAGTGAAGAAGTAAATGGACACCTGGACCAGACAGATGGGTTATCCTGTG
513 A K T S D F W A A L E E A S R L P V K E V M D T W T R Q M G Y P V
2201 CTTAACGTGAACGGTGTCAAGAACATCACACAGAAACGCTTTTGTGGACCAAGAGCTAACCTTCTCAGCCCCCTCAGATCTTGGTTATACATGGA
547 L N V N G V K N I T Q K R F L L D P R A N P S Q P P S D L G Y T W
2301 ATATCCCAGTTAAATGGACTGAAGATAATATAACAAGCAGTGTGTTATTAATAGGTGAGAAAAAGGAAGTCACTTTGAACCTCTAATCCTAGTGG
580 N I P V K W T E D N I T S S V L F N R S E K E G I T L N S S N P S G
2401 AAATGCTTTTCTCAAATAAACCCAGATCATATGGGTTTTATCGTGTAAATATGAAGTAGCAACTGGGACTCGATAGCTACACGCTCTCCTTGAAC
613 N A F L K I N P D H I G F Y R V N Y E V A T W D S I A T A L S L N
2501 CACAAGACATTTTCTCAGCAGATCGTGAAGTCTTATTGATGATGCTTTTGCCTTGGCAAGAGCTCAACTTCTAGATTATAAGGTGGCTTTGAACCTGA
647 H K T F S S A D R A S L I D D A F A L A R A Q L L D Y K V A L N L
2601 CCAAGTATCTCAAAGGGAAGAGAATTTTTTACCATGGCAGAGAGTAATTTACAGTGAACCTACATCATTAGCATGTTTGAAGATGATAAAGAGCTATA
680 T K Y L K R E E N F L P W Q R V I S A V T Y I I S M F E D D K E L Y
2701 TCCTATGATTGAGGAATACTTCCAAGTCAAGTGAAGCCTATTGCAGATTCTCTGGATGGAATGATGCTGGAGACCATGTCAAAAGTTACTCCGTTCC
713 P M I E E Y F Q G Q V K P I A D S L G W N D A G D H V T K L L R S
2801 TCCGTGTTAGGGTTTGCCTGCAAGATGGGAGACAGAGAAGCCTTGAACAAATGCTTCTCGTTATTTGAGCAGTGGCTAAATGGGACTGTAAGCCTTCCCG
747 S V L G F A C K M G D R E A L N N A S S L F E Q W L N G T V S L P
2901 TAAATCTCAGGCTTCTGGTGTATCGGTATGGGATGCAGAACTTGGCAATGAGATTTTCAAGTGAACCTACTCTTGGCAATACACTCTTGGCAATACACTCATTAGC
780 V N L R L L V Y R Y G M Q N S G N E I S W N Y T L E Q Y Q K T S L A
3001 TCAAGAAAAAGAAAACTGCTGTATGGATTAGCATCAGTGAAGAAGCTTACTCTTTGTCAAGGATTTTGGATTTGCTCAAGGACACGAACTTATTTAAA
813 Q E K E K L L Y G L A S V K N V T L L S R Y L D L L K D T N L I K
3101 ACTCAGGATGTGTTTACAGTCATTTCGATATATCTCATATAACAGCTATGGGAAGAATGGCCTGGAATTTGGATACAACCTCAACTGGGACTATCTAGTCA
847 T Q D V F T V I R Y I S Y N S Y G K N M A W N W I Q L N W D Y L V
3201 ACAGATATACACTCAATAACAGAAACCTTGGCCGAATTGTCAAAATAGCAGAGCCATTCAACTGAACCTGCAACTGTGGCAGATGGAGAGCTTTTTCG
880 N R Y T L N N R N L G R I V T I A E P F N T E L Q L W Q M E S F F A

3301 AAAATATCCACAAGCTGGAGCAGGAGAAAAACCTAGGGAACAAGTGTGAAACAGTGAAAAACAATATAGAGTGGCTAAAACAACATAGAAACACCATC
913▶ K Y P Q A G A G E K P R E Q V L E T V K N N I E W L K Q H R N T I
NheI (3450)

3401 AGAGAATGGTTTTTAATTTACTTGAGAGTGGTTAATGTATTCAAATGTTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCAC
947▶ R E W F F N L L E S G •

3501 AACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAAACAACA
AATTGCATTCAATTTATGTTTCAGGTTCAAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAG

3601 CATAGCAAACTTTAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAG

3701 CTGTTTGCAGCCTCACCTTCTTTCATGGAGTTAAGATATAGTGTATTTTCCCAAGTTTGAAGTACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGAC

3801 CTCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAG

3901 GCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTT
141◀ • N

4101 CTGGTGTACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGA
138◀ 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 L

4201 GCTCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTG
105◀ 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 Q

4301 CCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCAGCACAGACAGTGACCCTGCCAATGTAGGCCCTCAATGTGGACAGCAGAGATGATCTCCCCAGTC
72◀ 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

4401 TTGGTCTGATGGCCGCCGACATGGTGTGTTGTCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGTCCAGATCCTGCTGAGAGA
38◀ 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 D Q Q S I

4501 TGTGAAGGTCTTCATGTTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTC
5◀ N F T K M

4601 TCCAGCTTATCTGACGGTTCATAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCTACCGCCATTTCGCTCAATGGGGCGGAGTTGTTACG

4701 ACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAAAACAACTCCCATGACGTCAATGGGGTGGAGACTTGGAAATCCCGTGAGTCAAACCGCTATCCA

4801 CGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGTAGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACT

4901 GGGCATAATGCCAGGCGGGCATTACCCTGATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCGAGTTTACCG

5001 TAAATACTCCACCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATATTGACGTCAATGGCGGGGGTCTGTTGGCGGT

5101 CAGCCAGGCGGGCATTACCCTAAGTTATGTAACGCTGAGGTTAATAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGC
CGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGA

5201 TACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGG

5301 CGCTTTTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCCGCTCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTG

5401 CGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT

5501 GTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCTAACTACGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCG

5601 GAAAAAGAGTTGGTAGCTCTTGATCCGGCAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATC

5701 TCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAA

5801 TCAGCGGCGCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAAACAAACGAA
5901 ACAAAAACAACTAGCAAAATAGGCTGTCCCGTGAAGTGCAGGTGCCAGAACATTTCTCTATCGAA
6001