



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA  
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTGATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC  
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGGCGCCGCCCTACCTGAGGCC  
301 GCCATCCACGCCGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC  
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCAACTCTACGCTTTTGTTCGTTT

**BspHI (560)**

**AgeI (552)**

501 TCTGTTCTGCGCGTTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGTCATCATGACCGCGCCGGGCGCCGCCGGGCGCTGCCCTCCCAC  
1 M T A P G A A G R C P P T  
601 GACATGGCTGGGCTCCCTGCTGTTGTTGGTCTGTCTCCTGGCAGCAGGAGTATCACCGAGGAGGTGTCGGAGTACTGTAGCCACATGATTGGGAGTGGGA  
13 T W L G S L L L L V C L L A S R S I T E E V S E Y C S H M I G S G  
701 CACCTGCAGTCTCTGCAGCGGCTGATTGACAGTCAAGTGGAGACCTCGTGCCAAATTACATTTGAGTTTGTAGACCAGGAACAGTTGAAAGATCCAGTGT  
47 H L Q S L Q R L I D S Q M E T S C Q I T F E F V D Q E Q L K D P V  
801 GCTACCTTAAGAAGGCATTTCTCCTGGTACAAGACATAATGGAGGACACCATGCGCTTCCAGAGATAACACCCCAATGCCATCGCCATTGTGCAGTGC  
80 C Y L K K A F L L V Q D I M E D T M R F R D N T P N A I A I V Q L Q  
901 GGAATCTCTTTGAGGCTGAAGAGCTGCTTACCAAGGATTATGAAGAGCATGACAAGCGCTGCGTCCGAACCTTTCTATGAGACACCTCCAGTTGCTG  
113 E L S L R L K S C F T K D Y E E H D K A C V R T F Y E T P L Q L L  
1001 GAGAAGGTCAAGAATGTCTTTAATGAAACAAAGAATCTCCTTGACAAGGACTGGAATATTTTCAGCAAGAAGTGAACAACAGCTTTGCTGAATGCTCCA  
147 E K V K N V F N E T K N L L D K D W N I F S K N C N N S F A E C S  
1101 GCCAAGATGTGGTACCAAGCCTGATTGCAACTGCCTGTACCCCAAAGCCATCCCTAGCAGTACCCGGCCTCTGTCTCCCCTCATCAGCCCTCGCCCC  
180 S Q D V V T K P D C N C L Y P K A I P S S D P A S V S P H Q P L A P  
1201 CTCATGGCCCTGTGGCTGGCTGACCTGGGAGGACTCTGAGGGAAGTGGGGCAGCTCCCTCTGCTGGTGGAGCAGCCCTGCACACAGTGGATCCA  
213 S M A P V A G L T W E D S E G T E G S S L L P G E Q P L H T V D P  
1301 GGCAGTGCCAAGCAGCGGCCACCCAGGAGCACCTGCCAGAGCTTTGAGCCGCCAGAGACCCCAAGTGTCAAGGACAGCACCATCGGTGGCTCACCACAGC  
247 G S A K Q R P P R S T C Q S F E P P E T P V V K D S T I G G S P Q  
1401 CTCGCCCTCTGTGGGGCCTTCAACCCGGGATGGAGGATATTCTTACTCTGCAATGGGACTAATTGGGTCCCAGAAGAAGCCTCTGGAGAGGCCAG  
280 P R P S V G A F N P G M E D I L D S A M G T N W V P E E A S G E A S  
1501 TGAGATCCCGTACCCCAAGGGACAGAGCTTTCCCCCTCCAGGCCAGGAGGGGCGAGCATGCAGACAGAGCCCGCCAGACCCAGCAACTTCCCTCAGCA  
313 E I P V P Q G T E L S P S R P G G G S M Q T E P A R P S N F L S A  
1601 TCTTCTCCACTCCCTGCATCAGCAAAGGGCCACAGCCGGCAGATGTAAGTGGTACCGCTTGGCCAGGGTGGGCCCCGTGAGGCCACTGGCCAGGACT  
347 S S P L P A S A K G Q Q P A D V T G T T A L P R V G P V I L V L A V G  
1701 GGAATCACACCCCCAGAAGACAGACCATCTGCCTGCTCAGAGACCCCGGAGCCAGGCTCTCCAGGATCTCATCTGCGTCCCGCCAGGGCCT  
380 W N H T P Q K T D H P S A L L R D P P E P G S P R I S S L R P Q G L  
1801 CAGCAACCCCTCCACCCTCTGCTCAGCCACAGCTTTCCAGAAGCCACTCCTCGGGCAGCGTGTGCCCTTGGGGAGTGGAGGGCAGGAGGAGCACC  
413 S N P S T L S A Q P Q L S R S H S S G S V L P L G E L E G R R S T  
1901 AGGGATCGGAGGAGCCCGCAGAGCCAGAAGGAGGACCAGCAAGTGAAGGGGCGAGCCAGGCCCTGCCCCGTTTAACTCCGTTCTTTGACTGACACAG  
447 R D R R S P A E P E G G P A S E G A A R P L P R F N S V P L T D T  
2001 GCCATGAGAGGAGTCCGAGGATCTCCAGCCGAGCTCCAGGAGTGTCTTCCACCTGCTGGTGGCCAGTGTATCTGGTCTGCTGGCCGTCGG  
480 G H E R Q S E G S S P Q L Q E S V F H L L V P S V I L V L A V G  
2101 AGGCCTCTGTTCTACAGGTGGAGGCGGAGCCATCAAGAGCCTCAGAGAGCGATTCTCCCTTGGAGCAACCAGAGGGCAGCCCTGACTCAGGAT  
513 G L L F Y R W R R R S H Q E P Q R A D S P L E Q P E G S P L T Q D

**NheI (2261)**

2201 GACAGACAGGTGGAAGTCCAGTGTAGAGGAAATTCTAAGCTGGACGCACAGAACAGTCTCGTAGCTGGCCAGACATGATAAGATACATTGATGAGTTT  
547 D R Q V E L P V •  
2301 GGACAACCAACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAG  
2401 TTAACAACAACAATTGCATTATTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATT  
2501 CTAANAATACAGCATAGCAAACTTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCA  
2601 ATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTCATGGAGTTAAGATATAGTGTATTTTCCAAGTGTGAACTAGCTCTTCATTTCTTTATGTTTTA  
2701 AATGACTGACCTCCACATTCCCTTTTTAGTAAAATATTCAGAAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCC  
2801 AGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCT  
2901 AGCTTTAGTTCCTGGTGTACTTGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCATTATCTCAATGAGCACAAGCAGTCCAGGAGCATA  
141 • N 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  
3001 GTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGCTGACAGCCACAATGGTGTCA  
109 D S I 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  
3101 AAGTCTTCTGCCGTTGCTCACAGCAGACCAATGGCAATGGCTTCAGCAGACAGTACCCTGCAATGTAGGCTCAATGTGGACAGCAGAGATGA  
75 F D K 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

3201 TCTCCCCAGTCTTGGTCCTGATGGCCGCCCGACATGGTGCTTGTTCCTCATAGAGCATGGTGATCTTCTCAGTGGCGACCTCCACCAGCTCCAGATC  
42 E G T 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  
3301 CTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGG  
9 Q Q S I N F T K M  
3401 TGGATGGCGTCTCCAGCTTATCTGACGGTTCACATAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGGCG  
3501 GAGTTGTTACGACATTTTGAAAGTCCCGTTGATTTACTAGTCAAACAACCTCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAA  
3601 ACCGCTATCCACGCCCATTTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAA  
3701 GGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGG  
3801 GCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGGT  
3901 CGTTGGGCGGTACGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAAC  
4001 CGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG  
4101 GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGTCCCTCGTGCCTCTCTGTTCGACCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTC  
4201 GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTTCAG  
4301 CCGACCGCTGCGCTTATCCGTAACATATCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA  
4401 GAGCGAGGTATGTAGGCGGTGTACAGAGTTCTGAAGTGGTGCCTAACTACGGTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCC  
4501 AGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGA  
4601 AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCAATGGCTAGTTAAT  
4701 TAACATTTAAATCAGCGGCCCAATAAAATATCTTTATTTTTCATTACATCTGTGTGGTTTTTTTGTGTAATCGTAACTAACATACGCTCTCCATCAA  
4801 AACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCCCAGTGAAGTGCAGGTGCCAGAACATTTCTCTATCGAA