



1 GGATCTGCATCGCTCCGGTGCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

NcoI (560)

AgeI (552)

501 TCTGTTCTGCGCGTTACAGATCCAAGCTGTGACCGCGCTACCTGAGATCACCGGTGCCACCCAGCAGAAAACCTCTGACGAGAGGATCTC
1 M A T Q Q K A S D E R I S
601 CCAGTTTGATCACAATTTGCTGCCAGAGCTGTCTGCTCTTCTGGGCTAGATGCAGTTCAGTTGGCAAAGGAAGTGAAGAAGAGGAGCAGAAGGAGCGA
13 Q F D H N L L P E L S A L L G L D A V Q L A K E L E E E E Q K E R
701 GCAAAAATGCAGAAAGGCTACAACCTCAAATGCGCAGTGAAGCAAAAAGGTTAAAGACTTTTGTGACTTATGAGCCGTACAGCTCATGGATACCACAGG
47 A K M Q K G Y N S Q M R S E A K R L K T F V T Y E P Y S S W I P Q
801 AGATGGCGGCGCTGGGTTTACTTCACTGGGGTAAAATCTGGGATTCAAGTCTTCTGCTGTAGCCTAATCCTCTTTGGTGGCGGCTCACGAGACTCCC
80 E M A A A G F Y F T G V K S G I Q C F C C S L I L F G A G L T R L P
901 CATAGAAGACCACAAGAGTTTCACTCCAGATTGTGGTTCCTTTTGAACAAGGATGTTGGTAACATTGCCAAGTACGACATCAAGGGTGAAGAATCTGAAG
113 I E D H K R F H P D C G F L L N K D V G N I A K Y D I R V K N L K
1001 AGCAGGCTGAGAGGAGTAAATGAGGTACCAAGAAGAGGAGGCTAGACTTGCCTCCTCAGGAAGTGGCCATTTTATGTCCAAGGGATATCCCTTGTG
147 S R L R G G K M R Y Q E E E A R L A S F R N W P F Y V Q G I S P C
1101 TGCTCTCAGAGGCTGGCTTTGTCTTTACAGGTAACAGGACACGGTACAGTGTTTTTCTGTGGTGGATGTTTAGGAAATTGGGAAGAGGAGATGATCC
180 V L S E A G F V F T G K Q D T V Q C F S C G G C L G N W E E G D D P
1201 TTGGAAGGAACATGCCAAATGTTCCCAAATGTGAATTTCTCGAGTAAGAAATCCTCAGAGGAAATTACCCAGTATATCAAAGCTACAAGGGATTT
213 W K E H A K W F P K C E F L R S K K S S E E I T Q Y I Q S Y K G F
1301 GTTGACATAACGGGAGAATTTGTGAATTCCTGGTCCAGAGAAATTACCTATGGCATCAGCTTATTGCAATGACAGCATCTTTGCTTACGAAGAAC
247 V D I T G E H F V N S W V Q R E L P M A S A Y C N D S I F A Y E E
1401 TACGGCTGGACTCTTTAAGGACTGGCCCGGGAATCAGCTGTGGGAGTTCAGCACTGGCCAAAGCAGGCTTTTTCTACACAGGTATAAAGGACATCGT
280 L R L D S F K D W P R E S A V G V A A L A K A G L F Y T G I K D I V
1501 CCAGTGTCTTCTGTGGAGGTTGTTAGAGAAATGGCAGGAAGGTGATGACCATTAGACGATCACACCAGATGTTTTCCCAATTGTCCATTCTCCAA
313 Q C F S C G G C L E K W Q E G D D P L D D H T R C F P N C P F L Q
1601 AATATGAAGTCTTGCAGGAGTACTCAGAGCTTCCAGAGGCTGGTGAATTTGTGAATTTACTGGAACCAAGTGAAGCAATCTTGAAGATCAA
347 N M K S S A E V T P D L Q S R G E L C E L L E T T S E S N L E D S
1701 TAGCAGTTGGTCTATAGTGCCAGAAATGGCACAGGTTGAAGCCAGTGGTTTCAAGAGCAAAAGTCTGAATGAGCAGTGAAGGAGCTTATAACAG
380 I A V G P I V P E M A Q G E A Q W F Q E A K N L N E Q L R A A Y T S
1801 CGCAGTTTCCGCCACATGTCTTTGCTTGATATCTTCCGATCTGGCCACGGACCACTTGTGGGCTGTGATCTGTCTATTGCTTCAAACACATCAGC
413 A S F R H M S L L D I S S D L A T D H L L G C D L S I A S K H I S
1901 AAACCTGTGCAAGAACCTCTGGTGTGCTGAGGTCTTTGGCAACTTGAAGTCTGTATGTGTGGAGGGTGAAGCTGGAAGTGAAGGACGGTCCCTC
447 K P V Q E P L V L P E V F G N L N S V M C V E G E A G S G K T V L
2001 TGAAGAAAATAGCTTTTCTGTGGGACTGTGGATGCTCCCTGTTAAACAGGTTCCAGTGGTTTTCTACCTCTCCCTTAGTCCACCAGACAGACGA
480 L K K I A F L W A S G C C P L L N R F Q S L V F Y L S L S S T R P L F V A
2101 GGGGCTGGCCAGTATCATCTGTGACCAGCTCTAGAGAAAAGGATCTGTTACTGAAATGTGCGTGAGGAACATTATCCAGCAGTTAAAGAATCAGGTC
513 G L A S I I C D Q L L E K E G S V T E M C V R N I I Q Q L K N Q V
2201 TTATTCCTTTAGATGACTACAAAGAAATATGTTCAATCCCTCAAGTCATAGGAAAATGATTCAAAAAACCACTTATCCCGGACCTGCCTATTGATTG
547 L F L L D D Y K E I C S I P Q V I G K L I Q K N H L S R T C L L I
2301 CTGTCCGTACAAACAGGGCCAGGGACATCCGCCGATACCTAGAGACATTCTGGAGATCAAAGCATTCCCTTTTATAACTGTCTGTATATTACGGAA
580 A V R T N R A R D I R R Y L E T I L E I K A F P F Y N T V C I L R K
2401 GCTCTTTTACATAATGACTCGTCTGCGAAAATTTATGGTTTACTTTGAAAGAACCAAAGTTTGCAGAAGATACAGAAAACCTCTCTTTGTGGCG
613 L F S H N M T R L R K F M V Y F G K N Q S L Q K I Q K T P L F V A
2501 GCGATCTGTGCTCATTGGTTTCAAGTATCCTTTTACCCATCCTTTGATGATGTGGCTGTTTTCAAGTCTATATGGAACGCCTTTCTTAAGGAACAAAG
647 A I C A H W F Q Y P F D P S F D D V A V F K S Y M E R L S L R N K
2601 CGACAGCTGAAATTTCAAAGCAACTGTGCTCCTGTGGTGAAGTGGCCTTGAAGGGTTTTTTTTCATGTTGCTTTGAGTTAATGATGATGATCTCGC
680 A T A E I L K A T V S S C G E L A L K G F F S C C F E F N D D D L A
2701 AGAAGCAGGGGTTGATGAAGTGAAGATCTAACCATGTGCTTGTGAGCAATTTACAGCCAGAGACTAAGACCATTCTACAGGTTTTAAGTCTGCTGCC
713 E A G V D E D E D L T M C L M S K F T A Q R L R P F Y R F L S P A
2801 TTCCAAGAATTTCTGCGGGATGAGGCTGATTGAACCTCTGGATTCAAGTATAGGCAAGTCAAGTATGAGGAGTATCATTTGAAACAAATCAACT
747 F Q E F L A G G M R L I E L L D S D R Q E H Q D L G L Y H L K Q I N
2901 CACCCATGATGACTGTAAGCGCTACAACAATTTTTGAACTATGTCTCCAGCCTCCCTTCAACAAAAGCAGGGCCAAAATTTGTGTCTCATTGCTCCA
780 S P M M T V S A Y N N F L N Y V S S L P S T K A G P K I V S H L L H
3001 TTTAGTGATAACAAAGAGTCAATGGAGAATATATCTGAAAATGATGACTACTTAAAGCACCAGCCAGAAATTTCACTGCAGATGCAGTTACTTAGGGGA
813 L V D N K E S L E N I S E N D D Y L K H Q P E I S L Q M Q L L R G
3101 TTGTGGCAAAATTTGTCCAAGCTTACTTTTCAATGGTTTCAAGCAATTTACTGGTCTTGCCTGAAAATGCTTATCAAAGCAACTGTTGCTGCGT
847 L W Q I C P Q A Y F S M V S E H L L V L A L K T A Y Q S N T V A A

3201 GTTCTCCATTTGTTTTGCAATTCCTTCAAGGGAGAACACTGACTTTGGGTGCGCTTAACCTACAGTACTTTTTCGACCACCCAGAAAGCTTGTCATTGTT
880▶ C S P F V L Q F L Q G R T L T L G A L N L Q Y F F D H P E S L S L L
3301 GAGGAGCATCCACTTCCCAATACGAGGAAATAAGACATCACCCAGAGCACATTTTTAGTTCTGGAACATGTTTTGACAAATCACAGGTGCCAACTATA
913▶ R S I H F P I R G N K T S P R A H F S V L E T C F D K S Q V P T I
3401 GATCAGGACTATGCTTCTGCCTTTGAACPTAATGAATGGGAGCGAAATTTAGCTGAAAAAGAGGATAATGTAAAGAGCTATATGGATATGCAGCGCA
947▶ D Q D Y A S A F E C P M N E W E R N L A E K E D N V K S Y M D M Q R
3501 GGGCATCACAGACCTTAGTACTGGCTATTGGAACCTTCTCCAAAGCAGTACAAGATTCCCTGTCTAGAAGTCGATGTGAATGATATTGATGTTGTAGG
980▶ R A S P D L S T G Y W K L S P K Q Y K I P C L E V D V N D I D V V G
3601 CCAGGATATGCTTGAATCTAATGACAGTTTTCTCAGCTTACAGCGCATCGAACCATTAAACCACAGCAGAGGCTTTATAGAAAGCATCCGCCA
1013▶ Q D M L E I L M T V F S A S Q R I E L H L N H S R G F I E S I R P
3701 GCTCTTGAAGTGTCTAAGGCCTGTGACCAAGTGTCCATAAGCAAGTTGGAACCTCAGCGCAGCCGAACAGGAAGTCTTCTCACCTGCCTTCCCTGG
1047▶ A L E L S K A S V T K C S I S K L E L S A A E Q E L L L T L P S L
3801 AATCTCTTGAAGTCTCAGGACAATCCAGTCAAGACCAAATCTTCTAATCTGGATAAGTTCTGTGCCTGAAAGAACTGTCTGTGGATCTGGAGGG
1080▶ E S L E V S G T I Q S Q D Q I F P N L D K F L C L K E L S V D L E G
3901 CAATATAAATGTTTTTTCAGTCACTTCTGAAGAATTTCCAAACTTCCACCATATGGAGAAATATTGATCCAAATTTTCAGTGTAGTATGATCCTTCCAAA
1113▶ N I N V F S V I P E E F P N F H H M E K L L I Q I S A E Y D P S K
4001 CTAGTAAAATTAATTCAAAATTCCTCAACCTTCATGTTTTCCATCTGAAGTGAACCTCTTTTCGGATTTTGGGTCTCTCATGACTATGCTTGTTCCT
1147▶ L V K L I Q N S P N L H V F H L K C N F F S D F G S L M T M L V S
4101 GTAAGAACTCACAGAAATTAAGTTTTTCGGATTCATTTTTCAAGCGTCCCATTTGTTGCCAGTTTGCCAAATTTATTTCTCTGAAGATATTAATCT
1180▶ C K K L T E I K F S D S F F Q A V P F V A S L P N F I S L K I L N L
4201 TGAAGCCAGCAATTTCTGTATGAGGAAACATCAGAAAAATTTGCCTACATTTTAGTTTCTTAGTAACTGGAAGAATTGATCCTTCTACTGGGGAT
1213▶ E G Q Q F P D E E T S E K F A Y I L G S L S N L E E L I L P T G D
4301 GGAATTTATCGAGTGGCCAAACTGATCATCCAGCAGTGTGAGCAGCTTATTGTCTCCGAGTCTCTCATTTTTTCAAGACTTTGAATGATGACAGCGTGG
1247▶ G I Y R V A K L I I Q Q C Q Q L H C L R V L S F F K T L N D D S V
4401 TGGAAATTTGCCAAAGTAGCAATCAGTGGAGGTTTCCAGAACTTGAGAACCTAAAGCTTTCAATCAATCACAGATTACAGAGGAAGGATACAGAAATTT
1280▶ V E I A K V A I S G G F Q K L E N L K L S I N H K I T E E G Y R N F
4501 CTTTCAAGCACTGGACAACATGCCAACTTGCAGGAGTTGGACATCTCCAGGCATTTACAGAGTGTATCAAAGCTCAGGCCACAACAGTCAAGTCTTTG
1313▶ F Q A L D N M P N L Q E L D I S R H F T E C I K A Q A T T T V K S L
4601 AGTCAATGTGTACGACTACCAAGGCTATTAGACTGAACATGTTAAGTTGGCTCTTGGATGCAGATGATATTGCATTGCTTAATGTCATGAAAGAAA
1347▶ S Q C V L R L P R L I R L N M L S W L L D A D D I A L L N V M K E

NheI (4775)

4701 GACATCCTCAATCTAAGTACTTAACTATTCTCCAGAAATGGATACTGCCGTTCTCTCCAATCATTAGAAAATAAGCTAGTGGCCAGACATGATAAGAT
1380▶ R H P Q S K Y L T I L Q K W I L P F S P I I Q K •
4801 ACATTGATGAGTTTGACAAACCACAAC TAGAATGCAGTGAATAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATATAAG
4901 CTGCAATAAACAGTTAACAAACA AATTGCATTCATTTTATGTTTCAGGTT CAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAA
5001 TGTGGTATGGAATTCTAAAATACAGCATAGCAAACCTTAACTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATC
5101 AGGGGCTGTTGCCAATGTGCATTAGCTGTTGCAGCCTCACCTCTTTTCATGGAGTTAAGATATAGTGTATTTTCCCAAGGTTGAACTAGCTCTTCAT
5201 TTCTTTATGTTTTAAATGCAGCTGACCTCCACATTCCTTTTTAGTAAAATATTGAGAAATAATTTAAATACATCATTGCAATGAAAATAATGTTTTTT
5301 ATTAGGCAGAATCCAGATGCTCAAGGCCCTTATAATATCCCCAGTTTAGTAGTTGGAAGTCTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAA
5401 GAAAGCGAGCTTCTAGCTTTAGTTCTGGTGTACTTGAGGGGGATGAGTTCTCAATGGTGGTTTTGACCAGCTTGCCATTCTCAATGAGCACAAAG
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
5501 CAGTCAGGAGCATAGTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCTGACAG
113▶ 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 D S Y P H R V A
5601 CCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCCAAATGGCAATGGCTT CAGCACAGACAGT GACCTGCCAATGTAGGCCCTCAATGTG
80▶ 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 R G I Y A E I H
5701 GACAGCAGAGATGATCTCCCGAGTCTGGTCTGATGGCCGCCCCGACATGGTCTTGTCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCC
47▶ 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 I K E T A V E
5801 ACCAGCTCCAGATCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGTGATTATACTATGCCGATATACTATGCCGATGATTA
13▶ V L E L D Q Q S I N F T K M
5901 TTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTCACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATT
6001 GCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCCTTGATTTACTAGTCAAAAACAACTCCCATTGACGTC AATGGGGTGGAGACTTGAAA
6101 TCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGACTGCCAAGTA
6201 GGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGCGGGCCATTTACCGTCATTGACGTC AATAGGGGGCGTACTTGGCATATGATACACTTGAT
6301 GTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCATTGACGTC AATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGT
6401 CAATGGCGGGGGTCTGTTGGCGGTGAGCCAGCGGGCCATTTACCGTAAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGC
6501 AAAAGGCCAGGAACCGTAAAAAGGCCGCTGCTGGCGTTTTTTCATAGGCTCCGCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAGAGGTG

6601 GCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCC
6701 GCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACG
6801 AACCCCCGTTTACGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGG
6901 TAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGC
7001 GCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGCAAGCAGC
7101 AGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
7201 CATGGCTAGTTAATTAACATTTAAATCAGCGGCCGAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTAATCGTAACTAACAT
7301 ACGCTCTCCATCAAAACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA