



1 GGATCTGCATCGCTCCGGTGCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTCCGCCCTCCCGCTGTGGTGCCTCTGAAGCTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCAACTCTACGCTTTGTTTCGTTT

AgeI (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTCAACATGTCAAGCTCTTCTGGCTCCTTCTCAGCCTTGTTCG
1 M S S S S W L L L S L V A
601 TGTAAGTCTGCTCAGTCCACCATTGAGGAACAGGCCAAGACATTTTTGGACAAGTTTAAACCAGAACCCGAAGACCTGTTCTATCAAAGTCACTTGTCT
13 V T A A Q S T I E E Q A K T F L D K F N H E A E D L F Y Q S S L A
701 TCTTGAATTATAACACCAATATTACTGAAGAGAATGTCCAAAACATGAATAATGCTGGGACAAATGGTCTGCCTTTTTAAAGGAACAGTCCACACTTG
47 S W N Y N T N I T E E N V Q N M N N A G D K W S A F L K E Q S T L
801 CCCAAATGTATCCACTACAAGAAATTCAGAATCTCACAGTCAAGCTTCAAGTCTCAGCTGCAAGGCTTTCAGCAAAAATGGGTCTTCAAGTCTCAGAAGACAAGAG
80 A Q M Y P L Q E I Q N L T V K L Q L Q A L Q Q N G S S V L S E D K S
901 CAAACGGTTGAACACAATCTAAATAACAATGAGCACCATCTACAGTACTGGAAAAGTTGTAAACCAGATAATCCACAAGAATGCTTATTACTTGAACCA
113 K R L N T I L N T M S T I Y S T G K V C N P D N P Q E C L L L E P
1001 GGTGTTGAATGAAATAATGGCAACAGTTTACTACAATGAGAGGCTCTGGGCTTGGGAAAGCTGGAGATCTGAGGTGCGCAAGCAGCTGAGGCCATTAT
147 G L N E I M A N S L D Y N E R L W A W E S W R S E V G K Q L R P L
1101 ATGAAGAGTATGGTCTTGAAGAAATGAGATGGCAAGAGCAATCATTATGAGGACTATGGGATTATTGGAGAGGAGACTATGAAGTAAATGGGGTAGA
180 Y E E Y V V L K N E M A R A N H Y E D Y G D Y W R G D Y E V N G V D
1201 TGGCTATGACTACAGCCGCGCCAGTTGATTGAAGATGTGGAACATACCTTTGAAGAGATTAACCATTTATATGAACATCTTCATGCCTATGTGAGGGCA
213 G Y D Y S R G Q L I E D V E H T F E E I K P L Y E H L H A Y V R A
1301 AAGTTGATGAATGCCTATCCTTCTATATCAGTCCAATTGGATGCCTCCCTGCTCATTGCTTGGTATGATGGGGTAGATTTTGGACAAATCTGTACT
247 K L M N A Y P S Y I S P I G C L P A H L L G D M W G R F W T N L Y
1401 CTTTGACAGTTCCCTTTGGACAGAAACCAAACATAGATGTTACTGATGCAATGGTGGACCAGGCTGGGATGCACAGAGAATATTCAAGGAGGCCGAGAA
280 S L T V P F G Q K P N I D V T D A M V D Q A W D A Q R I F K E A E K
1501 GTTCTTTGTATCTGTTGGTCTTCTAATATGACTCAAGGATCTGGGAAAATCCATGCTAACGGACCAGGAAATGTCAGAAAGCAGTCTGCCATCCC
313 F F V S V G L P N M T Q G F W E N S M L T D P G N V Q K A V C H P
1601 ACAGCTTGGGACTGGGGAAGGGCGACTTCAGGATCCTTATGTGCACAAAGGTGACAAATGGACGACTTCTGACAGTCTCATGAGATGGGGCATATCC
347 T A W D L G K G D F R I L M C T K V T M D D F L T A H H E M G H I
1701 AGTATGATATGGCATATGCTGCACAACCTTTTCTGTAAGAAATGGAGCTAATGAAGGATTCATGAAGCTGTTGGGAAATCATGTCACTTCTGCAGC
380 Q Y D M A Y A A Q P F L L R N G A N E G F H E A V G E I M S L S A A
1801 CACACCTAAGCATTAAAATCCATTGGTCTTCTGTCACCCGATTTTCAAGAAGACAATGAAACAGAAATAAATCTTCTGCTCAAACAAGCACTCAGGATT
413 T P K H L K S I G L L S P D F Q E D N E T E I N F L L K Q A L T I
1901 GTTGGGACTCTGCCATTTACTTACATGTTAGAGAAGTGGAGGTGGATGGTCTTTAAAGGGGAAATCCCAAAGACCAGTGGATGAAAAAGTGGTGGGAGA
447 V G T G L P F T Y M L E K W R W M V F K G E I P K D Q W M K K W W E
2001 TGAAGCGAGAGATAGTTGGGVTGGTGAACCTGTGCCCATGATGAAACATATGTCACCCGCATCTCTGTTCCATGTTTCTAATGATTACTCATTCA
480 M K R E I V G V E P V P H D E T Y C D P A S L F H V S N D Y S F I
2101 TCGATATTACACAAGGCCCTTACCAATCCAGTTTCAAGAAGCATTGTCAAGCAGCTAAACATGAAGGCCCTCTGCACAAATGTGACATCTCAAAC
513 R Y Y T R T L Y Q F Q F Q E A L C Q A A K H E G P L H K C D I S N
2201 TCTACAGAAGCTGGACAGAACTGTTCAATATGCTGAGGCTTGGAAAATCAGAACCCTGGACCCTAGCATTGGAAAATGTTGTAGGAGCAAAGAATGATGA
547 S T E A G Q K L F N M L R L G K S E P W T L A L E N V V G A K N M
2301 ATGTAAGGCCACTGCTCAACTACTTTGAGCCCTATTTACCTGGCTGAAAGACCAGAAACAAGAATCTTTTGTGGGATGGAGTACCGACTGGAGTCCATA
580 N V R P L L N Y F E P L F T W L K D Q N K N S F V G W S T D W S P Y
2401 TGCAGACCAAAGCATCAAAGTGAAGTAAAGCCTAAAATCAGCTTGGAGATAAAGCATATGAATGGAACGACAATGAAATGTACCTGTTCCGATCATCT
613 A D Q S I K V R I S L K S A L G D K A Y E W N D N E M Y L F R S S
2501 GTTGCATATGCTATGAGGAGTACTTTTTAAAAGTAAAAAATCAGATGATTCTTTTTGGGGAGGAGGATGTGCGAGTGGCTAATTTGAAACCAAGAATCT
647 V A Y A M R Q Y F L K V K N Q M I L F G E E D V R V A N L K P R I
2601 CCTTTAATTTCTTGTCACTGCACCTAAAATGTGCTGATATCATTCTAGAACTGAAGTTGAAAAGCCATCAGGATGTCCCGGAGCCGTATCAATGA
680 S F N F F V T A P K N V S D I I P R T E V E K A I R M S R S R I N D
2701 TGCTTTCCGTCTGAATGACAACAGCCTAGAGTTTCTGGGGATACAGCCAACACTTGGACCTCCTAACCAGCCCCCTGTTTCCATATGGCTGATTGTTTT
713 A F R L N D N S L E F L G I Q P T L G P P N Q P P V S I W L I V F
2801 GGAGTTGTGATGGGAGTATAGTGGTTGGCATTGTCTCCTGATCTTCACTGGGATCAGAGATCGGAAGAAGAAAAATAAAGCAAGAAGTGGAGAAAAAT
747 G V V M G V I V V G I V I L I F T G I R D R K K K N K A R S E N

NheI (2979)

2901 CTTATGCCTCCATCGATATTAGCAAAGGAGAAAATAATCCAGGATCCAAAACACTGATGATGTTTCAGACCTCCTTTTAGCTAGTGGCCAGACATGATA
780 P Y A S I D I S K G E N N P G F Q N T D D V Q T S F •
3001 AGATACATTGATGAGTTTGGACAAACCAACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTA
3101 TAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAAACCTCTA

3201 CAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGG
3301 CATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTCTTTCATGGAGTTTAAAGATATAGTGATTTTCCCAAGGTTTGAAGTAGCTCT
3401 TCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTT
3501 TTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACA
3601 GCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTATCTCAATGAGCAC
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
3701 AAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTGTCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTG
115 F 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
3801 ACAGCCACAATGGTGTCAAAGTCCTTCTGCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCAGCACAGACAGTGACCCTGCCAATGTAGGCCTCAA
81 V A 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
3901 TGTGGACAGCAGAGATGATCTCCCCAGTCTTGGTCTGATGGCCGCCGACATGGTCTTGTGCTCATAGAGCATGGTATCTTCTCAGTGGCGAC
48 H 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
4001 CTCCACCAGCTCCAGATCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGA
15 E V L E L D Q Q S I N F T K M
4101 TTAATTGTCAAACACGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCC
4201 ATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCCTTGATTACTAGTCAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTG
4301 GAAATCCCCTGAGTCAAACCGTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCA
4401 AGTAGGAAAGTCCATAAAGTCATGTACTGGGCATAATGCCAGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACT
4501 TGATGTACTGCCAAGTGGGCAGTTTACCCTAAATACTCCACCCATTGACGTCAATGGAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTG
4601 ACGTCAATGGGGGGGGTCTTGGGCGGTGAGCCAGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATAAGAACATGTGAGCAAAGGC
4701 CAGCAAAGGCCAGGAACCGTAAAAAGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGA
4801 GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCT
4901 GTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGTGTG
5001 CACGAACCCCGTTCCAGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGTAGTCCAACCCGGTAAAGACACGACTTATCGCCACTGGCAGCAGCCA
5101 CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTAT
5201 CTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAG
5301 CAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTT
5401 TGGTCATGGCTAGTTAATTAACATTTAAATCAGCGCCGCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTTGTAATCGTAACTA
5501 ACATACGCTCTCCATCAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA