



1 GGATCTGCATCGCTCCGGTGCCGTCAGTGGGAGAGCGACATGCCACAGTCCCGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTGTGCTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTCCGCCCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCGGCGCTCCCTTGAGCCTACCTAGACTCAGCGGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

AgeI (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTGCGGCTGCTGCTGGCCCTGTTGGGGTCTGCTGAG
1 M R L L L A L L G V L L S
601 TGTGCTGGGCTCCAGTCTTGTCCCTGGAGGCTCTGAGGAAGTGGAGCTTGGCCCTGCCAGCTGGAGCAGCAAGAGCAGGAGCTGACA
13 V P G P P V L S L E A S E E V E L E P C L A P S L E Q Q E Q E L T
701 GTAGCCCTGGGAGCCTGTGCGGCTGTGCTGTGGCGGGCTGAGCGTGGTGGCCACTGGTACAAGGAGGGCAGTCGCTGGCACCTGCTGGCCGTGTAC
47 V A L G Q P V R L C C G R A E R G G H W Y K E G S R L A P A G R V
801 GGGGCTGGAGGGGCGCCTAGAGATTGCCAGCTTCTACCTGAGGATGCTGGCCGCTACCTCTGCCTGGCAGGAGGCTCCATGATGCTCTGCAGAATCT
80 R G W R G R L E I A S F L P E D A G R Y L C L A R G S M I V L Q N L
901 CACCTTGATTACAGTGTACTGACTCCAGCAACGATGATGAGGACCCCAAGTCCCATAGGGACCTCGAATAGGCACAGTTACCCCGCAAGCA
113 T L I T Q P D S L T S S N D D E D P K S H R D L S N R H S Y P Q Q A
1001 CCCTACTGGACACACCCCGAGCATGGAGAAGAACTGCATGCACTACCTGCGGGGAACACCGTCAAGTTCGCTGTCCAGTGCAGGCAACCCACGC
147 P Y W T H P Q R M E K K L H A V P A G N T V K F R C P A A G N P T
1101 CCACCATCGCTGGCTTAAGGATGGACAGGCCCTTTCATGGGAGAACCATTGGAGGCTTCGCTGCGCCATCAGCACTGGAGTCTCGTATGGAGAG
180 P T I R W L K D G Q A F H G E N R I G G I R L R H Q H W S L V M E S
1201 CGTGGTGCCTCGGACCGGACACATACCTGCTGGTAGAGAACGCTGTGGGAGCAGTCCGTTATAACTACCTGCTAGATGTGCTGGAGCGGTCCCG
213 V V P S D R G T Y T C L V E N A V G S I R Y N Y L L D V L E R S P
1301 CACCGCCCATCTGCAGCGGGCTCCCGGCAACACACAGCCGTGGTGGGAGCAGCTGGAGTGTGCAAGGTGTACAGCGATGCCAGCCCG
247 H R P I L Q A G L P A N T T A V V G S D V E L L C K V Y S D A Q P
1401 ACATCCAGTGGTGAAGCACATCGTCATCAACGGCAGCAGTTCGAGCGGACGGTTTCCCTATGTGCAAGTCTAAAGACTGCAGACATCAATAGCTC
280 H I Q W L K H I V I N G S S F G A D G F P Y V Q V L K T A D I N S S
1501 AGAGGTGGAGTCTGTACTGCGGAACGTGTAGCCGAGGACGAGGCGAGTACCTGCTCGCAGGCAATTCATCGGCCTCTCTACCAGTCTGCC
313 E V E V L Y L R N V S A E D A G E Y T C L A G N S I G L S Y Q S A
1601 TGGCTCACGGTGTGCCAGAGGAGGCCACATGGACCGCAGCAGCGCCGAGGCGAGTATACGGACATCATCCTGTACGCGTGGGCTCCCTGGCCT
347 W L T V L P E E D P T W T A A A P E A R Y T D I I L Y A S G S L A
1701 TGGCTGTCTCCTGCTGGCAGGCTGTATCGAGGCGAGGCGCTCCAGCCGCGCACCCCGCCCGCCGACTGTGCAGAAGCTCTCCCGCTTCCC
380 L A V L L L L A R L Y R G Q A L H G R H P R P P A T V Q K L S R F P
1801 TCTGGCCGACAGTCTCCCTGGAGTCAAGGCTCTCCGGCAAGTCAAGCTCATCCCTGGTACGAGGCGTGGTCTCTCTCCAGCGGCCCCGCTTGTCT
413 L A R Q F S L E S G S S G K S S S S L V R G V R L S S S G P A L L
1901 GCCGGCCTCGTGTAGTCTACCTCTGACCCACTATGGGAGTTCGCCGGGACAGGCTGGTGTGGGAAGCCCTAGCGGAGGCTGCTTTGCC
447 A G L V S L D L P L D P L W E F P R D R L V L G K P L G E G C F G
2001 AGGTAGTACGTGCAGAGGCTTTGGCATGGACCTGCCCGGCTGACCAAGCCAGCAGTGTGGCCGTCAAGATGCTCAAAGACAAGCCTCTGACAAGGA
480 Q V V R A E A F G M D P A R P D Q A S T V A V K M L K D N A S D K D
2101 CCTGGCCGACCTGGTCTCGGAGATGGAGGTGATGAAGCTGATCGGCCGACACAAGAATCATCAACCTGCTTGGTGTCTGCACCCAGGAAGGCCCTG
513 L A D L V S E M E V M K L I G R H K N I I N L L G V C T Q E G P L
2201 TACGTGATCGTGGAGTGCAGCGCAAGGAACTGCGGGAGTTCCTGCGGGCCGCGCCCCAGGCCCGACCTCAGCCCCGACGGTCTCGGAGCA
547 Y V I V E C A A K G N L R E F L R A R R P P G P D L S P D G P R S
2301 GTGAGGGCCGCTCTCTCCAGTCCCTGCTCTGCGCCTACCAGTGGCCCGAGGATGCAATCTGGAGTCCCGGAAGTGTATCCACCGGACCT
580 S E G P L S F P V L V S C A Y Q V A R G M Q Y L E S R K C I H R D L
2401 GGCTGCCGCAATGTGCTGGTGTACTGAGGACAATGTGATGAAGATTGCTGACTTTGGGCTGGCCCGGGCTCCACCACATTGACTACTATAAGAAAAC
613 A A R N V L V T E D N V M K I A D F G L A R G V H H I D Y Y K K T
2501 AGCAACGGCCGCTGCTGTGAAGTGGATGGCGCCGAGGCTTGTGGTACCGGGTGTACACACACAGAGTACGCTGGTCTTTGGGATCCTGCTAT
647 S N G R L P V K W M A P E A L F D R V Y T H Q S D V W S F G I L L
2601 GGGAGATCTTACCTCGGGGCTCCCGTATCCTGGCATCCCGGTGGAGGAGCTGTTCTCGTGTGCGGGAGGACATCGGATGGACCGACCCCA
680 W E I F T L G G S P Y P G I P V E E L F S L L R E G H R M D R P P H
2701 CTGCCCCCAGAGCTGTACGGGCTGATGCTGAGTGTGGCACGAGCGCCCTCCAGAGGCTACCTTCAAGCAGCTGGTGGAGGCGCTGGACAAGGTC
713 C P P E L Y G L M R E C W H A A P S Q R P T F K Q L V E A L D K V
2801 CTGCTGGCGTCTGTGAGGAGTACCTGACCTCCGCTGACCTTCGGACCTATTCCCGCTGCTGGGGAGCCAGCAGCAGCTGCTCTCCAGCGATT
747 L L A V S E E Y L D L R L T F G P Y S P S G G D A S S T C S S S D

NheI (2982)

2901 CTGTCTTACGCCACGACCCCTGCCATTGGGATCCAGTCTCTTCCCTTCGGGTGTGGGTGCAGACATGAGCAAGGCTCAAGCTAGTGGCCAGACATG
780 S V F S H D P L P L G S S S F P F G S G V Q T •
3001 ATAAGATACATTGATGAGTTTGGACAAACACAACCTAGAATGCAAGTAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCA
3101 TTATAAGCTGCAATAACAAGTTAACAACAACAAATTGCATTCAATTTATGTTTCAGGTTTCAGGGGAGGTTGGGAGGTTTTTAAAGCAAGTAAACCT

3201 CTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACCTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCAT
3301 AGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTAAGATATAGTGATTTTTCCCAAGGTTTGAACACTAGC
3401 TCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCACATTCCCTTTTTAGTAAAAATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAAT
3501 GTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGG
3601 ACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCACTCAATGAG
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
3701 CACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGC
116 V 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
3801 CTGACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCAGCACAGACAGTGACCCTGCCAATGTAGGCCT
82 R 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
3901 CAATGTGGACAGCAGAGATGATCTCCCACTTGGTCTGATGGCCGCCGACATGGTGTCTGTTGCTCATAGAGCATGGTGTCTTCTCAGTGGC
49 I 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
4001 GACCTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGA
16 V E V L E L D Q Q S I N F T K M
4101 TGATTAATTGTCAAACACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTTACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCTACCG
4201 CCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGAAAGTCCCGTTGATTTACTAGTCAAAAACAACTCCATTGACGTCAATGGGGTGGAGAC
4301 TTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGACTG
4401 CCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATAC
4501 ACTTGATGTACTGCCAAGTGGGCGAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTA
4601 TTGACGTCAATGGGCGGGGTGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCTGCAGGTTAATTAAGAACATGTGAGCAAAA
4701 GGCCAGCAAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGGTGGCTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTC
4801 AGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGCTCCCTCGTGGCTCTCCTGTTCCGACCTGCCGCTTACCGGATA
4901 CCTGTCCGCTTTCTCCCTTCGGAAGCGTGGCGCTTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGGGCTGT
5001 GTGCACGAACCCCCGTTAGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGTAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAG
5101 CCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGG
5201 TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGC
5301 AAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGA
5401 TTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCCAATAAAATATCTTTATTTTTCATTACATCTGTGTGGTTTTTTGTGTGAATCGTAA
5501 CTAACATACGCTCTCCATCAAAAACAAACGAAACAAACAACTAGCAAATAGGCTGTCCCAGTGAAGTGCAGGTGCCAGAACATTTCTCTATCGAA