



1 GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGCAGAGCGCACATGCCACAGTCCCCGAGAAGTTGGGGGAGGGTTCGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGTGCTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCCGCTTTCGCCGCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

BspLU11I (561)

SgrAI (551)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTGCAACATGTTATCTGGGAAGAAGCGCGCAGCCGCGCGCGCGG
13▶ M L S G K K A A A A A A
601 CGGCTGCAGCGGCAGCAACCGGACGGAGGCTGGCCCTGGGACAGCAGGGGCTCCGAGAACGGTCTGAGGTGGCCGCGCAGCCCGCGGGCTGTCCGGG
13▶ A A A A A A T G T E A G P G T A G G S E N G S E V A A Q P A G L S G
701 CCCAGCCGAGGTCCGGCCGGGGCGGTGGGGGAGCGCACACCCCGCAAGAAAGAGCCTCCGCGGGCCTCGCCCCCGGGGGCCTGGCGGAACCGCCGGGG
46▶ P A E V G P G A V G E R T P R K K E P P R A S P P G G L A E P P G
801 TCCGAGGGCCTCAGCGCGCCTACTGTGCTGCTGGGTGCGACCCCATGAAACTGGAATAGCAGAGACTCCGGAGGGCGTCCGACCAGCCGGC
80▶ S A G P Q A G P T V V P G S A T P M E T G I A E T P E G R R T S R
901 GCAAGCGGGCGAAGGTAGAGTACAGAGAGATGGATGAAAGCTTGCCAACCTCTCAGAAGATGAGTATTATTTCAGAAGAAGAGAGAAATGCCAAAGCAGA
113▶ R K R A K V E Y R E M D E S L A N L S E D E Y Y S E E E R N A K A E
1001 GAAGGAAAAGAAGCTTCCCCACCACCCCTCAAGCCACCTGAGGAAGAAAATGAAAGTGAAGCCTGAAGAACCATCGGGTGTGGAGGGCGCAGCTTTC
146▶ K E K K L P P P P P Q A P P E E E N E S E P E E P S G V E G A A F
1101 CAGAGCCGACTTCTCATGACCGGATGACTTCTCAAGAAGCAGCCTGTTTTCCAGATATTATCAGTGGACCACAACAGACCCAGAAGGTTTTTCTTTTCA
180▶ Q S R L P H D R M T S Q E A A C F P D I I S G P Q Q T Q K V F L F
1201 TTAGAAACCGCACACTGCAGTTGTGGTGGATAATCCAAGATTGAGTGCATTTGAGGCTACTCTCAACAATTAGAAGCACCTTATAACAGTGATAC
213▶ I R N R T L Q L W L D N P K I Q L T F E A T L Q Q L E A P Y N S D T
1301 TGTGCTTGTCCACCGAGTTCACAGTTATTTAGAGCGTCATGGTCTTATCAACTTCGGCATCTATAAGAGGATAAAAACCCCTACCAACTAAAAAGACAGGA
246▶ V L V H R V H S Y L E R H G L I N F G I Y K R I K P L P T K K T G
1401 AAGTAATTATTATAGGCTCTGGGGTCTCAGGCTTGGCAGCAGCTCGACAGTTACAAAGTTTTGGAATGGATGTCACACTTTTGAAGCCAGGGATCGTG
280▶ K V I I I G S G V S G L A A A R Q L Q S F G M D V T L L E A R D R
1501 TGGGTGGACGAGTTGCCACATTCGCAAAGGAAACTATGTAGCTGATCTTGGAGCCATGGTGGTAACAGGTCTTGGAGGGAATCCTATGGCTGTGGTCAG
313▶ V G G R V A T F R K G N Y V A D L L G A M V T G L G G N P M A V V S
1601 CAAAAGTAAATATGGAATCGCAAGATCAAGCAAAAATGCCACCTTATGAAGCCAACGGACAAGCTGTTCTAAAGAGAAAGATGAAATGGTAGAG
346▶ K Q V N M E L A K I K Q K C P L Y E A N G Q A V P K E K D E M V E
1701 CAAGAGTTTAACCGTGTCTAGAAGCTACATCTTACCTTAGTCATCAACTAGACTTCAATGTCCTCAATAATAAGCCTGTGTCCTTGGCCAGGCATTGG
380▶ Q E F N R L L E A T S Y L S H Q L D F N V L N N K P V S L G Q A L
1801 AAGTTGTCACTCAGTTACAAGAGAAGCATGTCAAAGATGAGCAGATTGAACATTGGAAGAAGATAGTAAAACCTCAGGAAGAATTGAAAGAATCTCTTAA
413▶ E V V I Q L Q E K H V K D E Q I E H W K K I V K T Q E E L K E L L N
1901 TAAGATGGTAAATTTGAAAGAGAAAATTAAGAATCCATCAGCAATACAAAGAAGCATCTGAAGTAAAGCCACCCAGAGATATTACTGCCGAGTTCTTA
446▶ K M V N L K E K I K E L H Q Q Y Y K E A S E V K P P R D I T A E F L
2001 GTGAAAAGCAAACACAGGATCTGACCGCCCTATGCAAGGAATATGATGAATTAGCTGAAACACAAGGAAAGCTAGAAGAAAAAATCAGGAGTTGGAAG
480▶ V K S K H R D L T A L C K E Y D E L A E T Q G K L E E K L Q E L E
2101 CGAATCCCCAAGTGATGTATATCTCTCATCAAGAGACAGACAATACTTATTGGCATTGTTGCAATCTTGAATTTGCTAATGCCACACCTCTCTCAAC
513▶ A N P P S D V Y L S S R D R Q I L D W H F A N L E F A N A T P L S T
2201 TCTCTCCCTTAAGCACTGGGATCAGGATGATGACTTTGAGTTCAGTGGCAGCCACCTGACAGTAAGGAATGGTACTCGTGTGTGCTGTGGCTTTAGCA
546▶ L S L K H W D Q D D D F E F T G S H L T V R N G Y S C V P V A L A
2301 GAAGCCCTAGACATTAACCTGAATACAGCAGTGCAGAGGTTGCTACACGGCTTTCAGGATGTAAGTGTAGCTGTGAATACCCGCTCCAGGAGTCAAA
580▶ E G L D I K L N T A V R Q V R Y T A S G C E V I A V N T R S T S Q
2401 CCTTATTATAAATGCGACGAGTTCTCTGTACCCTTCCCTGGGTGTGCTGAAGCAGCAGCCACCAGCCGTTTTCAGTTTGTGCCACCTCTCCCTGAGTG
613▶ T F I Y K C D A V L C T L P L G V L K Q Q P P A V Q F V P P L P E W
2501 GAAAACATCTGCAGTCCAAGGATGGGATTTGGCAACCTTAACAAGGTGGTGTGTTGTTTGGATCGGGTGTCTGGGATCCAAGTGTCAATTTGTTCCGG
646▶ K T S A V Q R M G F G N L N K V V L C F D R V F W D P S V N L F G
2601 CATGTTGGCAGTACGACTGCCAGCAGGGGTGAGCTTCTCTTCTGGAACCTCTATAAAGCTCCAATACTGTTGGCACTAGTGGCAGGAGAAGCTGTCTG
680▶ H V G S T T A S R G E L F L F W N L Y K A P I L L A L V A G E A A
2701 GTATCATGAAAACATAAGTACGATGTGATTGTTGGCCGATGCTGGCCATTCTCAAAGGATTTTTGGTAGCAGTGCAGTACCTCAGCCAAAGAAAC
713▶ G I M E N I S D D V I V G R C L A I L K G I F G S S A V P Q P K E T
2801 TGTGGTGTCTGTTGGCTGCTGATCCCTGGGCTCGGGGCTTATTCTATGTTGCTGCAGGATCATCTGGAATGACTATGATTTAATGGCTCAGCCA
746▶ V V S R W R A D P W A R G S Y S Y V A A G S S G N D Y D L M A Q P
2901 ATCACTCTGGCCCTCGATTCCAGGTGCCCCACAGCCGATTCCAGACTTCTTTTGGGGGAGAACATACGATCCGTAACCTACCCAGCCACAGTGCATG
780▶ I T P G P S I P G A P Q P I P R L F F A G E H T I R N Y P A T V H
3001 GTGCTCTGCTGAGTGGGCTGCGAGAAGCGGGAAGAATTGCAGACCAGTTTTTGGGGCCATGTATACGCTGCCTCGCCAGGCCACACCAGGTGTTCTGCTG
813▶ G A L L S G L R E A G R I A D Q F L G A M Y T L P R Q A T P G V P A
NheI (3136)
3101 ACAGCAGTCCCCAAGCATGTGAGACAGATGCATTCTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGT
846▶ Q Q S P S M •

3201 GAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATT
3301 TATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTTT
3401 AACCTCCAAATCAAGCCTCTACTTGAATCCTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGAGCCTC
3501 ACCTTCTTTCATGGAGTTAAGATATAGTGATTTTTCCAAGGTTTGAAGTACTCTTCATTTCTTTATGTTTTAAATGACTGACCTCCACATTCCCT
3601 TTTTAGTAAAATATTAGAAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATAT
3701 CCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTGATTCTGGTGTACTTGAG
3801 GGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCATTATCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTCTGCACATG
134 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 E R C M
3901 CCACAGGGGCTGACCACCTGATGGATCTGCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCGTTGCTCACAG
100 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 G N S V A
4001 CAGACCCAATGGCAATGGCTTCAGCACAGACAGTGACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCAGTCTTGGTCTGATGGC
67 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 K T R I A
4101 CGCCCCGACATGGTGTGTTGTCTCATAGAGCATGGTATCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGAGAGATGTTGAAGGTCTTC
34 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 N F T K
4201 ATGGTGGCCCTCTATAGTGAGTCTGATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACACAGCGTGGATGGCGTCTCCAGCTTATCTGA
0 M
4301 CGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCACCCTACACGCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGAAAGT
4401 CCCGTTGATTTACTAGTCAAACAAACTCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGTA
4501 CTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGACTGCCAAGTAGGAAAGTCCATAAGGTATGTACTGGGCATAATGCCAG
4601 GCGGGCCATTTACCGTCATTGACGTCAATAGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCC
4701 ATTGACGTCAATGAAAGTCCCTATTGGCGTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTACGCCAGGCGGGCC
4801 ATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGCCGCTTGTGGCGT
4901 TTTTCATAGGCTCCGCCCCCTGACGAGCATCACAATAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCC
5001 CCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTCGGGAAGCGTGGCGTTTCTCATAGC
5101 TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCCTGCGCTTATCCGGTA
5201 ACTATCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTAC
5301 AGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGT
5401 AGCTCTTGATCCGCAAACAAACCACCGCTGGTAGCGGTGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTT
5501 TGATCTTTTCTACGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGAAT
5601 AAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAACAAACGAAACAAACAAACTAG
5701 CAAAATAGGCTGTCCCAGTGCAAGTGACAGGTGCCAGAACATTTCTCTATCGAA