



PvuI (7) SgfI (6) MfeI (82) EcoNI (96)

1 GGATCTGGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCACAGTCCCGAGAAGTTGGGGGAGGGGTGGCAATTGAACGGGTGCCTA

101 GAGAAGTGGCGCGGGTAAACTGGAAAGTGATGTCGTGTACTGGTCCGCCTTTTCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC

Psp1406I (203) HindIII (245) Bsu36I (291)

201 GTGAACGTTCTTTTTCGCAACGGGTTTGGCCGAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTACCGCGCCCGCCCTACCTGAGGCC

301 GCCATCCACGCGGTTGAGTGCCTTCTGCCGCTCCCGCTGTGGTGCTCCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC

NgoMIV (441) NaeI (441)

401 GGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCCTGCTTCTCAACTCTACGTCTTTGTTTCGTTT

BspLU11I (560)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGGCGCTACCTGAGATCACCGGTCAACATGTGCTGTTCTGCAGAGGACAGCTTCAGGAA

KasI (535) AgeI (552)

1 MetSer I l eVal CysSer Al aGl uAspSer PheArgAs

BsrGI (632)

601 TCTCATCTTATTCTTTCAGGCCAGGCTGAAAATGTACATTCAGGTGGACCCAGTGTGGACCCTCATCTTTCTGTCTGCAGAAACCAAGAGCAGATT

13 nLeu l l eLeuPhePheArgP roArgLeuLysMet Tyr l l eGl nVal l Gl uP roVal l LeuAspHi sLeu l l ePheLeuSer Al aGl uThr LysGl uGl n l l e

701 CTTAAAAAGATCAACACCTGTGGTAAACACAGCGCGGAGAACTGCTGCTGAGCACCTTGGAGCAGGGACAATGGCCTCTGGGATGGACGAGATGTTTCG

47 LeuLysLys l l eAsnThr CysGl yAsnThr Ser Al aAl aGl uLeuLeuLeuSer Thr LeuGl uGl nGl yGl nTrpP roLeuGl yTrpThr Gl nMe tPheV

EcoO109I (803) BssHII (835) PshAI (895)

801 TGGAGGCCCTAGAGCACAGTGGCAATCCCTAGCCGCGCTATGTCAAACCCACACTCACTGATCTGCCCTCTCTTCTCTGAGACTGCCCATGACGA

80 a l Gl uAl aLeuGl uHi sSer Gl yAsnP roLeuAl aAl aArgTyrVal LysP roThr LeuThr AspLeuP roSer P roSer Ser Gl uThr Al aHi sAspGl

901 GTGTCTCCACTTGTGACCTCTCCAGCCACTTGGTGGACAACTTCTGATTAACGATGTCTGGACACTTGCTTCGAGAAGGGACTATTGACAGTC

113 uCysLeuHi sLeuLeuThr LeuLeuGl nP roThr LeuVal l AspLysLeuLeu l l eAsnAspVal l LeuAspThr CysPheGl uLysGl yLeuLeuThr Val

1001 GAAGACAGAAATCGGATTTCTGCTGCAGGAAACAGCGGAATGAGTCAGGTGTAAGAGAGCTGCTGAGAAAGATTGTGCAGAAGGAAACTGGTTTTCTA

47 Gl uAspArgAsnArg l l eSer Al aAl aGl yAsnSer Gl yAsnGl uSer Gl yVal l ArgGl uLeuLeuArgArg l l eVal l Gl nLysGl uAsnTrpPheSer T

1101 CCTCCTGGATGTTCTGCGCCAACTGGAAATGATGCACTATTTCAAAGAACTAACAGTGGAGGCTGCCAGAAGACAACACAGACTTGCTAACTCGTC

180 hr PheLeuAspVal l LeuArgGl nThr Gl yAsnAspAl aLeuPheGl nGl uLeuThr Gl yGl yCysP roGl uAspAsnThrAspLeuAl aAsnSer Se

1201 TCACAGAGATGGGCTGCAGCTAATGAGTGTCTTCTGCCTGCTGCGATGAGTCAAGTCTGGAGACAGAGGCCGGAACGTAGACGACATATTACCAGAG

213 r Hi sArgAspGl yP roAl aAl aAsnGl uCysLeuLeuP roAl aVal l AspGl uSer Ser LeuGl uThr Gl uAl aTrpAsnVal l AspAsp l l eLeuP roGl u

BsrGI (1306)

1301 GCTTCTGTACAGATTCGTCTGTGACCACAGAATCAGACACAAGTTTGGCAGAAGGAAGTGTGAGCTGCTTCGATGAAAGTCTTGGACATAACAGCAACA

247 Al aSer CysThr AspSer Ser Val Thr Thr Gl uSer AspThr Ser LeuAl aGl uGl ySer Val Ser CysPheAspGl uSer LeuGl yHi sAsnSerAsnM

NcoI (1418)

1401 TGGGCAGGGATTACGGCACCATGGGAAGTGATTCAGATGAAAGTGTATCCAGACAAAAAGATATCCCCGAGCCAGAACTGCAGCTCAGGCCTTACCA

280 e tGl yArgAspSer Gl yThr MetGl ySer AspSer AspGl uSer Val l l eGl nThr LysArgVal l Ser P roGl uP roGl uLeuGl nLeuArgP roTyrGl

XbaI (1521)

1501 AATGGAAGTGGCCCAACAGCTAGATGGGAAGAATATTATTATCTGCCTCCACGGGGAGTGGGAAAACAGAGTGGCTGTTTACATCACC AAAGAT

313 nMe tGl uVal l Al aGl nP roAl aLeuAspGl yLysAsn l l e l l e l l eCysLeuP roThr Gl ySer Gl yLysThr ArgVal l Al aVal l Tyr l l eThr LysAsp

1601 CACTTAGACAAGAAGAAGCAGGCATCTGAATCCGGGAAGGTTATCGTCTTGTCAATAAGGTAATGTTAGCAGAACAACCTTTCCGAAAAGAGTTCAACC

347 Hi sLeuAspLysLysLysGl nAl aSer Gl uSer Gl yLysVal l l eVal l LeuVal l AsnLysVal l Me tLeuAl aGl uGl nLeuPheArgLysGl uPheAsnP

1701 CATATTTGAAGAAATGGTATCGAATTATTGGATTAAGTGGCGATACCCAGCTGAAAATATCATTTCCAGAAGTGTCAAATCTTACGATGTTATTATCAG

380 r oTyrLeuLysLysTrpTyrArg l l e l l eGl yLeuSer Gl yAspThr Gl nLeuLys l l eSer PheP roGl uVal l Val l LysSer TyrAspVal l l e l l eSe

1801 CACTGCTCAAATCTTGAACCTCTTAAATCTGGAGAGTGGAGAGCTGAGCAGCTGCGGTGTGAGCTGTGAGACTTCTCTCTATTATCATTGATGAGTGC

413 r Thr Al aGl n l l eLeuGl uAsnSer LeuLeuAsnLeuGl uSer Gl yAspAspAspGl yVal l Gl nLeuSer AspPheSer Leu l l e l l e l l eAspGl uCys

1901 CATCACACCAACAAGGAGCAGTCTATAACAACATCATGAGACGATATTTGAAGCAGAAGCTGAGAAAACATGACCTCAAGAAAACAAAACCAAGCCACA

447 Hi sHi sThrAsnLysGl uAl aVal l TyrAsnAsn l l eMe tArgArgTyrLeuLysGl nLysLeuArgAsnAsnAspLeuLysLysGl nAsnLysP roAl a l

DraIII (2033)

2001 TTCCCTGCCGAGATACTAGGACTGACAGCTTCACTGGTGTGGAGCAGCCAAAAAGCAGTCTGAGGCTGAAAAACATATTTTAAATATATGTGCCAA

480 l eP roLeuP roGl n l l eLeuGl yLeuThr Al aSer P roGl yVal l Gl yAl aAl aLysLysGl nSer Gl uAl aGl uLysHi s l l eLeuAsn l l eCysAl aAs

2101 TCTTGATGCTTTTACCATTAAAAACAGTGAAGAGAATCTTGGTCAACTCAAACCAAATAAAGGAACCATGCAAGAAATTTGTATTGCTGATGACACC

513 nLeuAspAl aPheThr l l eLysThr Val l LysGl uAsnLeuGl yGl nLeuLysHi sGl n l l eLysGl uP roCysLysLysPheVal l l eAl aAspAspThr

XbaI (2226)

2201 AGAGAAAAATCCATTTAAAGAGAACTTCTAGAAAATTATGGCAAGCATTAGACTTACTGCCAAAAAAGTCCAATGTGAGATTTTGAACCCCAACATTATG

547 ArgGl uAsnP roPheLysGl uLysLeuLeuGl u l l eMe tAl aSer l l eGl nThr TyrCysGl nLysSer P roMe tSer AspPheGl yThr Gl nHi sTyrG

2301 AGCAGTGGGCCATTCAAATGGAGAAAAAGCTGCTAAAGACGGAATCGCAAAGATCGCGTCTGTGCAGAACATTTGAGGAAAGTACAACGAAGCCCTACA

580 l uGl nTrpAl a l l eGl nMe tGl uLysLysAl aAl aLysAspGl yAsnArgLysAspArgVal l CysAl aGl uHi sLeuArgLysTyrAsnGl uAl aLeuGl

Ppu10I (2426) NsiI (2426)

2401 AATCAACGACACGATCCGAATGATTGATGCATATAGCCACCTGGAGACATTCTACACTGATGAGAAAAGAAAAGATTCGCAGTCTCAATGACAGCGAC

613 n l l eAsnAspThr l l eArgMe t l l eAspAl aTyrSer Hi sLeuGl uThrPheTyrThrAspGl uLysGl uLysLysPheAl aVal l LeuAsnAspSerAsp

2501 AAGAGTGATGACGAGGCCAGCAGCTTGAATGACCAACTTAAGGCGATGTAAGAAAATCTTGAACCTGGACGAAACGGATGAATTTCTCATGAATTTGT

647 LysSer AspAspGl uAl aSer Ser CysAsnAspGl nLeuLysGl yAspVal l LysLysSer LeuLysLeuAspGl uThrAspGl uPheLeuMe tAsnLeuP

2601 TCTTTGATAACAAGAAAAATGTTGAAAAACTAGCTGAAAACCCAAAATACGAGAATGAAAACCTCATTAAATTAAGAAACACGATACTGGAACAATTCAC

680 hePheAspAsnLysLysMe tLeuLysLysLeuAl aGl uAsnP roLysTyrGl uAsnGl uLysLeu l l eLysLeuArgAsnThr l l eLeuGl uGl nPheTh

2701 AAGTCTGAGGAGTCTCCCGAGGAATATTTTCAAAAAACACGACAGACCTACGCCTTTCCAGTGGATCATGGAAAATGCAAAAGTTTGGCGAA

713 r ArgSer Gl uGl uSer ArgGl y l l e l l ePheThr LysThr ArgGl nSer Thr TyrAl aLeuSer Thr nTrp l l eMe tGl uAsnAl aLysPheAl aGl u

2801 GTTGGAGTCAAAGCGCATCACCTGATTGGCGCGGGCAGCAGTGAAGTCAAGCCATGACTCAGACTGAACAAAAAGAGTCAATAGTAAATTTTCGCA

747 Val l Gl yVal l LvsAl aHi sHi sLeu l l eGl vAl aGl vHi sSer Ser Gl uVal l LvsP roMe tThr Gl nThr Gl uGl nLvsGl uVal l l eSer LvsPheAroT

EcoRV (2950)
Eco32I (2950)

2901 CTGGCGAAATAAATCTGCTTATCGCTACGACGGTGGCAGAGGAAGGCCTGGATATCAAAGAGTGCAATATTGTTATTCGTTATGGCCTTGTACGAACGA
780▶hr Gl yGl u l l eAsnLeuLeu l l eAl aThr Thr Val Al aGl uGl uGl yLeuAsp l l eLysGl uCysAsn l l eVal l l eArgTyrGl yLeuVal ThrAsnGl

XmaI (3017)
SmaI (3017)

3001 GATAGCCATGGTCCAGGCCCGGGTCCGAGCCAGAGCTGATGAAAGCACGTATGTCCTGGTCCACCAGCTGGCTCAGGAGTTACCGAACGGGAGATTGTT
813▶ul l eAl aMetVal Gl nAl aArgGl yArgAl aArgGl aAspGl uSer Thr TyrVal l eLeuVal l Thr Ser Ser Gl ySer Gl yVal Thr Gl uArgGl u l l eVal l

3101 AATGATTTCCGAGAGAAGATGATGTATAAAGCTATTAACCGTGTTCAAAAATGAAACAGAGGAGTATGCACATAAGATTTTGAATTGCAGGTGCAAA
847▶AsnAspPheArgGl uLysMetMetTyrLysAl a l l eAsnArgVal l Gl nAsnMetLysProGl uGl uTyrAl aHi sLys l l eLeuGl uLeuGl nVal l Gl nS

SphI (3293)

3201 GTATCCTGGAAAAGAAAATGAAAGTCAAAGAAGCATTGCAAAGCAATACAACGACAATCCATCGTTAATAACACTTCTCTGCAAAAATGTAGCATGCT
880▶er l l eLeuGl uLysLysMetLysVal l LysArgSer l l eAl aLysGl nTyrAsnAspAsnProSerLeu l l eThrLeuLeuCysLysAsnCysSerMetLe

Ppu10I (3337)
NsiI (3337)

3301 GGTCTGCTCGGGAGAAAACATCCATGTCTATTGAGAAGATGCATCATGTCAATATGACACCAGAATCAAGGGACTCTACATTGTAAGAGAAAACAAAGCA
913▶uVal CysSer Gl yGl uAsn l l eHi sVal l l eGl uLysMetHi sHi sVal l AsnMetThrProGl uPheLysGl yLeuTyr l l eVal l ArgGl uAsnLysAl a

MscI (3453)
BalI (3453)

3401 CTGCAAAAGAAATTTGCTGATTATCAGACCAATGGAGAGATTATCTGCAAGTGTGGCCAGGCTGGGGAACAATGATGGTGCACAAAGGTTTAGATTGCT
947▶LeuGl nLysLysPheAl aAspTyrGl nThrAsnGl yGl u l l e l l eCysLysCysGl yGl nAl aTrpGl yThrMetMetVal l Hi sLysGl yLeuAspLeuP

3501 CTTGTCTTAAATAAGGAATTTGTAGTCAATTTCAAATAAATCACCAGAAAGAAACAGTACAAGAAGTGGGTGGAATTGCCTATCAGATTTCTCTGATCT
980▶r oCysLeuLys l l eArgAsnPheVal l ValAsnPheLysAsnAsnSerProLysLysGl nTyrLysLysTrpVal l Gl uLeuPro l l eArgPheProAspLe

MscI (3660)
BalI (3660)

NheI (3654)

3601 TGACTACTCAGAATACTGCTTGTATAGTATGATGAAGATTAGCATTGATTCATGAGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTGGACAAA
1013▶uAspTyrSer Gl uTyrCysLeuTyrSerAspGl uAsp●●●

HpaI (3792)

3701 CCACAAC TAGAATGCAGT GAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAA

MfeI (3803)

3801 CAACAATTGCATTCATTTATGTTTCAGGTTCCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAAATCTAAAT

EcoRI (3888)

3901 ACAGCATAGCAAAACTTTAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCA

4001 TTAGCTGTTGCAGCCTCACCTCTTTTCATGGAGTTTAAAGATATAGTGTATTTTCCCAAGTTTGAAGTAGCTCTTCATTTCTTTATGTTTTAAATGCAC

SwaI (4141)

4101 TGACCTCCACATTCCTTTTTAGTAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCT

EcoO109I (4202)

4201 CAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTAAATAGAAATTGGACAGCAAGAAAGCGACTTCTAGCTTTA
141▶●●●

4301 GTTCTGGTGTACTTGAGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAG
140▶AsnArgThr TyrLysLeuPro l l eLeuGl uGl u l l eThr Thr LysVal l eLeuLysGl yAsnMetGl u l l eLeuVal l PheCysAspProAl aTyrAspSer l

SacI (4402)

4401 ATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCCAATGGTGTCAAAGTCTC
106▶l eLeuGl uArgCysMetGl yCysProSer Val l ValArg l l eSerArgAspVal l Gl uAspSer TyrProHi sArgValAl aVal l l eThrAspPheAspLy

4501 TCTGCCGTTGCTCACAGCAGACCAATGGCAATGGCTCAGCACAGACAGTACCCTGCAATGTAGGCTCAATGTGGACAGCAGAGATGATCTCCCC
73▶sGl nGl yAsnSer ValAl aSer Gl y l l eAl a l l eAl aGl uAl aCysVal l Thr ValArgGl y l l eTyrAl aGl u l l eHi sValAl aSer l l e l l eGl uGl y

4601 AGTCTTGGTCTGATGGCCGCCCGACATGGTGTCTGTGCTCATAGAGCATGGTGTCTCTCAGTGGCGACCTCCACCAGCTCCAGATCCTGCTGA
40▶ThrLysThrArg l l eAl aAl aGl yVal l Hi sHi sLysAsnAspGl uTyrLeuMetThr l l eLysGl uThrAl aVal l Gl uVal l LeuGl uLeuAspGl nGl nS

VspI (4774)
AseI (4774)

4701 GAGATGTTGAAGTCTTCATGATGGCCCTCTATAGTGAAGTCTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGATGG
6▶er l l eAsnPheThrLysMet

SacI (4831)

4801 CGTCTCCAGCTATCTGACGGTTCCTAAACGAGCTCTGCTTATATAGACCTCCACCCTACACGCCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGT

SpeI (4929)

4901 TACGACATTTTGGAAAGTCCCGTTGATTTACTAGTCAAAACAAACTCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCT

SnaBI (5057)
Eco105I (5057)

5000 ATCCACGCCATTGATGTAAGTCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTAAGTCCAAAGTAGGAAAGTCCATAAGGTCAT

NdeI (5162)

5100 GTACTGGGCATAATGCCAGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGTACTGCAAGTGGGCAGTT

5200 TACCCTAAATACTCCACCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGGGGGGTCTGTTGG

SdaI (5340)
PaeI (5348)
BspLU11I (5358)

5300 GCGGTCAGCCAGCGGGCCATTTACCGTAAAGTTATGTAACGCTTGCAGGTTAA TTAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTA

5398 AAAAGGCCGCTTGTGGCGTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAATAAGTACGCTCAAGTCAAGTCAAGGTTGGCGAAACCCGACAGGACT

5498 ATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA

ApaLI (5672)

5598 AGCGTGGCGCTTTTCATAGCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCCG

5698 ACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGC

5798 GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTT

5898 ACCTTGGAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAA

PacI (6088)

5998 AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAAC

EagI (6108)

SwaI (6097) **NotI (6107)**

6098 ATTTAAATC AGCGGCCGAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAACA

6198 AAACGAAACAAAACAACTAGCAAAATAGGCTGTCCCCAGTGAAGTGCAGGTGCCAGAACATTTCTCTATCGAA