```
5'  tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgacc
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  70
3'  accgcttaccctgcgcgggacatcgccgcgtaattcgcgccgcccacaccaccaatgcgcgtcgcactgg

5'  gctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  140
3'  cgatgtgaacggtcgcgggatcgcgggcgaggaaagcgaaagaagggaaggaaagagcggtgcaagcggc

5'  gctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcga
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  210
3'  cgaaaggggcagttcgagatttagcccccgagggaaatcccaaggctaaatcacgaaatgccgtggagct

                                          DraIII
                                          |
5'  ccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccct
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  280
3'  ggggttttttgaactaatcccactaccaagtgcatcacccggtagcgggactatctgccaaaaagcggga

     AlOI'                                AlOI
     |                                    |
5'  ttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatct
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  350
3'  aactgcaacctcaggtgcaagaaattatcacctgagaacaaggtttgaccttgttgtgagttgggataga

                          PsiI
                          |
5'  cggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattta
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  420
3'  gccagataagaaaactaaatattccctaaaacggctaaagccggataaccaatttttactcgactaaat

5'  acaaaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  490
3'  tgtttttaaattgcgcttaaaattgttttataattgcgaatgttaaatccaccgtgaaaagcccctttac

5'  tgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataacc
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  560
3'  acgcgccttggggataaacaaataaaaagatttatgtaagtttatacataggcgagtactctgttattgg

5'  ctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttatt
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  630
3'  gactatttacgaagttattataactttttccttctcatactcataagttgtaaaggcacagcgggaataa
                                                 [Ap]

5'  ccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  700
3'  gggaaaaaacgccgtaaaacggaaggacaaaaacgagtgggtctttgcgaccactttcattttctacgac
    [Ap]

5'  aagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttt
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  770
3'  ttctagtcaacccacgtgctcacccaatgtagcttgacctagagttgtcgccattctaggaactctcaaa
    [Ap]
```

```
5'  tcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgt
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  840
3'  agcggggcttcttgcaaaaggttactactcgtgaaaatttcaagacgatacaccgcgccataatagggca
                                   Ap

5'  attgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcac
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  910
3'  taactgcggcccgttctcgttgagccagcggcgtatgtgataagagtcttactgaaccaactcatgagtg
                                   Ap

5'  cagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgag
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  980
3'  gtcagtgtcttttcgtagaatgcctaccgtactgtcattctcttaatacgtcacgacggtattggtactc
                                   Ap

5'  tgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcac
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1050
3'  actattgtgacgccggttgaatgaagactgttgctagcctcctggcttcctcgattggcgaaaaaacgtg
                                   Ap

5'  aacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1120
3'  ttgtacccccctagtacattgagcggaactagcaacccttggcctcgacttacttcggtatggtttgctgc
                                   Ap

                          PstI
5'  agcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactacttac
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1190
3'  tcgcactgtggtgctacggacgtcgttaccgttgttgcaacgcgtttgataattgaccgcttgatgaatg
                                   Ap

5'  tctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1260
3'  agatcgaagggccgttgttaattatctgacctacctccgcctatttcaacgtcctggtgaagacgcgagc
                                   Ap

                                                                  BsaI
5'  gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1330
3'  cgggaaggccgaccgaccaaataacgactatttagacctcggccactcgcacccagagcgccatagtaac
                                   Ap

                                                                AhdI
5'  cagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactat
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1400
3'  gtcgtgaccccggtctaccattcgggagggcatagcatcaatagatgtgctgcccctcagtccgttgata
                                   Ap
```

```
5'  ggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaa
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1470
3'  cctacttgctttatctgtctagcgactctatccacggagtgactaattcgtaaccattgacagtctggtt
o   ==========================================Ap==========================

o
5'  gtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatcc
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1540
3'  caaatgagtatatatgaaatctaactaaattttgaagtaaaaattaaattttcctagatccacttctagg
o

5'  tttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtaga
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1610
3'  aaaaactattagagtactggttttagggaattgcactcaaaagcaaggtgactcgcagtctggggcatct
o

5'  aaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaacca
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1680
3'  tttctagtttcctagaagaactctaggaaaaaaagacgcgcattagacgacgaacgtttgtttttttggt
o   ==========================================Ori=========================

o
5'  ccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttca
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1750
3'  ggcgatggtcgccaccaaacaaacggcctagttctcgatggttgagaaaaaggcttccattgaccgaagt
o   ==========================================Ori=========================

o
5'  gcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgt
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1820
3'  cgtctcgcgtctatggtttatgacaggaagatcacatcggcatcaatccggtggtgaagttcttgagaca
o   ==========================================Ori=========================

o
5'  agcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgt
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1890
3'  tcgtggcggatgtatggagcgagacgattaggacaatggtcaccgacgacggtcaccgctattcagcaca
o   ==========================================Ori=========================

o
5'  cttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgt
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1960
3'  gaatggcccaacctgagttctgctatcaatggcctattccgcgtcgccagcccgacttgccccccaagca
o   ==========================================Ori=========================

o
5'  gcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaag
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  2030
3'  cgtgtgtcgggtcgaacctcgcttgctggatgtggcttgactctatggatgtcgcactcgatactctttc
o   ==========================================Ori=========================

o
5'  cgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgc
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  2100
3'  gcggtgcgaagggcttccctctttccgcctgtccataggccattcgccgtcccagccttgtcctctcgcg
o   ==========================================Ori=========================

o
```

```
5'  acgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2170
3'  tgctccctcgaaggtcccccttttgcggaccatagaaatatcaggacagcccaaagcggtggagactgaac
                                   Ori

5'  agcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttt
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2240
3'  tcgcagctaaaaacactacgagcagtcccccgcctcggatacctttttgcggtcgttgcgccggaaaaa
                                   Ori

5'  acggttcctggccttttgctggcctttttgctcacatgttctttcctgcgttatcccctgattctgtggat
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2310
3'  tgccaaggaccggaaaacgaccggaaaacgagtgtacaagaaaggacgcaatagggactaagacaccta

5'  aaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcag
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2380
3'  ttggcataatggcggaaactcactcgactatggcgagcggcgtcggcttgctggctcgcgtcgctcagtc

                SapI
5'  tgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2450
3'  actcgctccttcgccttctcgcggactacgccataaaagaggaatgcgtagacacgccataaagtgtggc

                                                          BstZ17I
5'  catatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatc
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2520
3'  gtatataccacgtgagagtcatgttagacgagactacggcgtatcaattcggtcatatgtgaggcgatag

5'  gctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgt
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2590
3'  cgatgcactgacccagtaccgacgcggggctgtgggcggttgtgggcgactgcgcgggactgcccgaaca

5'  ctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcac
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2660
3'  gacgagggccgtaggcgaatgtctgttcgacactggcagaggccctcgacgtacacagtctccaaaagtg

5'  cgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagat
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2730
3'  gcagtagtggctttgcgcgctccgtcgacgccatttcgagtagtcgcaccagcacttcgctaagtgtcta

5'  gtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaag
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2800
3'  cagacggacaagtaggcgcaggtcgagcaactcaaagaggtcttcgcaattacagaccgaagactatttc

5'  cgggccatgttaagggcggtttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttca
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2870
3'  gcccggtacaattcccgccaaaaaaggacaaaccagtgactacggaggcacattccccctaaagacaagt

5'  tgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2940
3'  acccccattactatggctactttgctctctcctacgagtgctatgcccaatgactactacttgtacgggc
```

```
5'   gttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcag
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3010
3'   caatgaccttgcaacactcccatttgttgaccgccatacctacgccgccctggtctctttttagtgagtc
o
5'   ggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgc
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3080
3'   ccagttacggtcgcgaagcaattatgtctacatccacaaggtgtcccatcggtcgtcgtaggacgctacg
o
5'   agatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccgaa
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3150
3'   tctaggccttgtattaccacgtcccgcgactgaaggcgcaaaggtctgaaatgctttgtgcctttggctt
o
                         Bpu10I
o                        |
5'   gaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatc
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3220
3'   ctggtaagtacaacaacgagtccagcgtctgcaaaacgtcgtcgtcagcgaagtgcaagcgagcgcatag
o
5'   ggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacga
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3290
3'   ccactaagtaagacgattggtcattccgttggggcggtcggatcggcccaggagttgctgtcctcgtgct
o
          FspAI
o         |
5'   tcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcggg
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3360
3'   agtacgcgtgggcaccccggcggtacggccgctattaccggacgaagagcggctttgcaaaccaccgccc
o
5'   accagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtc
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3430
3'   tggtcactgcttccgaactcgctcccgcacgttctaaggcttatggcgttcgctgtccggctagtagcag
o
5'   gcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgca
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3500
3'   cgcgaggtcgctttcgccaggagcggcttttactgggtctcgcgacggccgtggacaggatgctcaacgt
o
                                         PshAI
o                                        |
5'   tgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgg
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3570
3'   actatttcttctgtcagtattcacgccgctgctatcagtacggggcgcgggtggccttcctcgactgacc
o
5'   gttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcg
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3640
3'   caacttccgagagttcccgtagccagctctagggccacggattactcactcgattgaatgtaattaacgc
o
5'   ttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcg
o    ----|----|----|----|----|----|----|----|----|----|----|----|----|----|  3710
3'   aacgcgagtgacgggcgaaaggtcagccctttggacagcacggtcgacgtaattacttagccggttgcgc
o            Lac I
o            <----------------------------------------------------------
```

```
5'  cggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaacagct
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3780
3'  gcccctctccgccaaacgcataacccgcggtcccaccaaaaagaaaagtggtcactctgcccgttgtcga
                                   Lac I

5'  gattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3850
3'  ctaacgggaagtggcggaccgggactctctcaacgtcgttcgccaggtgcgaccaaacggggtcgtccgc
                                   Lac I

                          HpaI
5'  aaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccact
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3920
3'  ttttaggacaaactaccaccaattgccgccctatattgtactcgacagaagccatagcagcatagggtga
                                   Lac I

5'  accgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgat
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3990
3'  tggctctacaggcgtggttgcgcgtcgggcctgagccattaccgcgcgtaacgcgggtcgcggtagacta
                                   Lac I

5'  cgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccgga
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  4060
3'  gcaaccgttggtcgtagcgtcacccttgctacgggagtaagtcgtaaacgtaccaaacaacttttggcct
                                   Lac I

5'  catggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccag
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  4130
3'  gtaccgtgaggtcagcggaagggcaaggcgatagccgacttaaactaacgctcactctataaatacggtc
                                   Lac I

                                        PspOMI
                                          ApaI
5'  ccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgaccca
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  4200
3'  ggtcggtctgcgtctgcgcggctctgtcttgaattacccgggcgattgtcgcgctaaacgaccactgggt
                                   Lac I

5'  atgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgt
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  4270
3'  tacgctggtctacgaggtgcgggtcagcgcatggcagaagtaccctcttttattatgacaactacccaca
                                   Lac I

5'  ctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctgg
    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  4340
3'  gaccagtctctgtagttctttattgcggccttgtaatcacgtccgtcgaaggtgtcgttaccgtaggacc
                                   Lac I
```

BclI MluI

5' tcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttac 4410
3' agtaggtcgcctatcaattactagtcgggtgactgcgcaacgcgctcttctaacacgtggcggcgaaatg
Lac I

5' aggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagattt 4480
3' tccgaagctgcggcgaagcaagatggtagctgtggtggtgcgaccgtgggtcaactagccgcgctctaaa
Lac I

5' aatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgac 4550
3' ttagcggcgctgttaaacgctgccgcgcacgtcccggtctgacctccaccgttgcggttagtcgttgctg
Lac I

5' tgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccactt 4620
3' acaaacgggcggtcaacaacacggtgcgccaacccttacattaagtcgaggcggtagcggcgaaggtgaa
Lac I

5' tttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacacc 4690
3' aaagggcgcaaaagcgtctttgcaccgaccggaccaagtggtgcgccctttgccagactattctctgtgg

5' ggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccggg 4760
3' ccgtatgagacgctgtagcatattgcaatgaccaaagtgtaagtggtgggacttaactgagagaaggccc

5' cgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctccctta 4830
3' gcgatagtacggtatggcgctttccaaaacgcggtaagctaccacaggccctagagctgcgagagggaat

EcoNI

5' tgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatg 4900
3' acgctgaggacgtaatccttcgtcgggtcatcatccaactccggcaactcgtggcggcggcgttccttac

5' gtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaaca 4970
3' cacgtacgttcctctaccgcgggttgtcagggggccggtgccccggacggtggtatgggtgcggctttgt

5' agcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagca 5040
3' tcgcgagtactcgggcttcaccgctcgggctagaaggggtagccactacagccgctatatccgcggtcgt

SgrAI

5' accgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatcgatctcgatc 5110
3' tggcgtggacaccgcggccactacggccggtgctacgcaggccgcatctcctagctctagctagagctag

XbaI

5' ccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataatttt 5180
3' ggcgctttaattatgctgagtgatatcccttaacactcgcctattgttaaggggagatctttattaaaa

T7 Promoter — Lac Operator

BarI — BarI'

5' gtttaactttaagaaggagatatacatatgagcgataaaattattcacctgactgacgacagttttgaca 5250
3' caaattgaaattcttcctctatatgtatactcgctattttaataagtggactgactgctgtcaaaactgt

Trx Tag

M S D K I I H L T D D S F D
1 2 3 4 5 6 7 8 9 10 11 12 13 14

RsrII

5' cggatgtactcaaagcggacggggcgatcctcgtcgatttctgggcagagtggtgcggtccgtgcaaaat 5320
3' gcctacatgagtttcgcctgccccgctaggagcagctaaagacccgtctcaccacgccaggcacgtttta

Trx Tag

T D V L K A D G A I L V D F W A E W C G P C K M
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38

5' gatcgccccgattctggatgaaatcgctgacgaatatcagggcaaactgaccgttgcaaaactgaacatc 5390
3' ctagcggggctaagacctactttagcgactgcttatagtcccgtttgactggcaacgttttgacttgtag

Trx Tag

I A P I L D E I A D E Y Q G K L T V A K L N I
39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61

5' gatcaaaaccctggcactgcgccgaaatatggcatccgtggtatcccgactctgctgctgttcaaaaacg 5460
3' ctagttttgggaccgtgacgcggctttataccgtaggcaccatagggctgagacgacgacaagtttttgc

Trx Tag

D Q N P G T A P K Y G I R G I P T L L L F K N
62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84

5' gtgaagtggcggcaaccaaagtgggtgcactgtctaaaggtcagttgaaagagttcctcgacgctaacct 5530
3' cacttcaccgccgttggtttcacccacgtgacagatttccagtcaactttctcaaggagctgcgattgga

Trx Tag

G E V A A T K V G A L S K G Q L K E F L D A N L
85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108

5' ggccggttctggttctggccatatgcaccatcatcatcatcattcttctggtctggtgccacgcggttct 5600
3' ccggccaagaccaagaccggtatacgtggtagtagtagtagtaagaagaccagaccacggtgcgccaaga

T...g | His6-tag | Thrombin

A G S G S G H M H H H H H H S S G L V P R G S
109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131

BglII | Acc65I | KpnI

5' ggtatgaaagaaaccgctgctgctaaattcgaacgccagcacatggacagcccagatctgggtaccgacg 5670
3' ccatactttctttggcgacgacgatttaagcttgcggtcgtgtacctgtcgggtctagacccatggctgc

En...e

G M K E T A A A K F E R Q H M D S P D L G T D
132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154

5' acgacgacaaggccatggcgatATCAACAAGTTTGTACAAAAAAGCAGGCTctgaaaacttgtactttca 5740
3' tgctgctgttccggtaccgctaTAGTTGTTCAAACATGTTTTTTCGTCCGAgacttttgaacatgaaagt

Enterokinase | attB1 | TEV site

BGT25_T02a3-pDONR221

D D D K A M A I S T S L Y K K A G S E N L Y F Q
155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178

5' aggcGTCAATGAGCCGCCGAAACTTGGGGCGTTCGGGACCGAGAGCTTGATCGACCAGCACAGGCTTGTC 5810
3' tccgCAGTTACTCGGCGGCTTTGAACCCCGCAAGCCCTGGCTCTCGAACTAGCTGGTCGTGTCCGAACAG

T... | procollagen-lysine 5-dioxygenase

BGT25_T02a3

BGT25_T02a3-pDONR221

G V N E P P K L G A F G T E S L I D Q H R L V
179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201

5' CGGCAGCGGTCGTTGCGGTCGCGCATCGAGGGCACGCCGCGCATCCTGGTGACAATCCTCGCAAAGCAGA 5880
3' GCCGTCGCCAGCAACGCCAGCGCGTAGCTCCCGTGCGGCGCGTAGGACCACTGTTAGGAGCGTTTCGTCT

procollagen-lysine 5-dioxygenase

BGT25_T02a3

BGT25_T02a3-pDONR221

R Q R S L R S R I E G T P R I L V T I L A K Q
202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224

5' AAGAGCCGGCACTGCCGCTGTACCTCGAGTGCATCGAGGCGCTCGACTACCCCAAGGCGTCGATCGTCCT 5950
3' TTCTCGGCCGTGACGGCGACATGGAGCTCACGTAGCTCCGCGAGCTGATGGGGTTCCGCAGCTAGCAGGA

procollagen-lysine 5-dioxygenase

BGT25_T02a3

BGT25_T02a3-pDONR221

K E P A L P L Y L E C I E A L D Y P K A S I V L
225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248

5' GTACATCAGGACAAACAACAACACCGACAGGACCGAGCACATACTGCGCGAATGGGTGGAGCGCGTTGGT 6020
3' CATGTAGTCCTGTTTGTTGTTGTGGCTGTCCTGGCTCGTGTATGACGCGCTTACCCACCTCGCGCAACCA

procollagen-lysine 5-dioxygenase

BGT25_T02a3

BGT25_T02a3-pDONR221

Y I R T N N N T D R T E H I L R E W V E R V G
249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271

5' CACCTCTATGCCGCGGTTGAATTCGATGCCTCCAACGTGGCCGATAGAGTCGAGCAGTTTGGCGAACACG 6090
3' GTGGAGATACGGCGCCAACTTAAGCTACGGAGGTTGCACCGGCTATCTCAGCTCGTCAAACCGCTTGTGC

procollagen-lysine 5-dioxygenase

BGT25_T02a3

BGT25_T02a3-pDONR221

H L Y A A V E F D A S N V A D R V E Q F G E H
272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294

5' AATGGAACGAGACGCGCTTCAGGGTGCTTGGCCGAATCCGCAACATAAGTTTGCGGAAAACTCTCGAACA 6160
3' TTACCTTGCTCTGCGCGAAGTCCCACGAACCGGCTTAGGCGTTGTATTCAAACGCCTTTTGAGAGCTTGT

procollagen-lysine 5-dioxygenase

BGT25_T02a3

BGT25_T02a3-pDONR221

E W N E T R F R V L G R I R N I S L R K T L E H
295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318

5' CAGCTGCGATTTCTATTTCGTTGCCGACGTCGATAATTTCGTGCGACCGGCCACGCTACGCGAGCTTGTC 6230
3' GTCGACGCTAAAGATAAAGCAACGGCTGCAGCTATTAAAGCACGCTGGCCGGTGCGATGCGCTCGAACAG

procollagen-lysine 5-dioxygenase

BGT25_T02a3

BGT25_T02a3-pDONR221

S C D F Y F V A D V D N F V R P A T L R E L V
319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341

BmgBI XmaI SmaI

5' GCGCTCGACGTGCCGATCGTCGCTCCCCTGCTTCGCTCCATCTCTCCCGGGCAATACTATTCGAACTACC 6300
3' CGCGAGCTGCACGGCTAGCAGCGAGGGGACGAAGCGAGGTAGAGAGGGCCCGTTATGATAAGCTTGATGG

procollagen-lysine 5-dioxygenase
BGT25_T02a3
BGT25_T02a3-pDONR221

A L D V P I V A P L L R S I S P G Q Y Y S N Y
342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364

PmlI

5' ACGCCGAAATCGACGCCAACGGCTACTATATGCAGTGCGATCAATATGGGTGGGTGCTCAACCGCCACGT 6370
3' TGCGGCTTTAGCTGCGGTTGCCGATGATATACGTCACGCTAGTTATACCCACCCACGAGTTGGCGGTGCA

procollagen-lysine 5-dioxygenase
BGT25_T02a3
BGT25_T02a3-pDONR221

H A E I D A N G Y Y M Q C D Q Y G W V L N R H V
365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388

5' GCGCGGCATTATCGAGATGCCGCTTGTGCACTGCACCTATCTTGTCCGCGCCGACGTACTGCCCGAACTG 6440
3' CGCGCCGTAATAGCTCTACGGCGAACACGTGACGTGGATAGAACAGGCGCGGCTGCATGACGGGCTTGAC

procollagen-lysine 5-dioxygenase
BGT25_T02a3
BGT25_T02a3-pDONR221

R G I I E M P L V H C T Y L V R A D V L P E L
389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411

5' ACATACGAGGACGCGACATCGCGCTACGAGTACGTGATCTTTGCCGACAGCGCCCGCAAGGCCGGCATCG 6510
3' TGTATGCTCCTGCGCTGTAGCGCGATGCTCATGCACTAGAAACGGCTGTCGCGGGCGTTCCGGCCGTAGC

procollagen-lysine 5-dioxygenase
BGT25_T02a3
BGT25_T02a3-pDONR221

T Y E D A T S R Y E Y V I F A D S A R K A G I
412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434

5' TCCAGTATATGGATAACCGACAAGTCTATGGCTACATCACCTTCGGCGATGGCCAATATTATGTGAGTGA 6580
3' AGGTCATATACCTATTGGCTGTTCAGATACCGATGTAGTGGAAGCCGCTACCGGTTATAATACACTCACT

procollagen-lysine 5-dioxygenase

BGT25_T02a3

BGT25_T02a3-pDONR221

V Q Y M D N R Q V Y G Y I T F G D G Q Y Y V S D
435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458

5' CGGGATCGAGCGCGCCCGAGCGCTGTTGCACGGTGCCGGCGACACGCCGGTATTCACTGCCGCCACGCCT 6650
3' GCCCTAGCTCGCGCGGGCTCGCGACAACGTGCCACGGCCGCTGTGCGGCCATAAGTGACGGCGGTGCGGA

pr...e

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

G I E R A R A L L H G A G D T P V F T A A T P
459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481

5' CCGGGCGCCGTACCGATGCCAGATCGACATGTCGAGCCGCCGCAAATTCACCTAATAAACTTGGACAGCA 6720
3' GGCCCGCGGCATGGCTACGGTCTAGCTGTACAGCTCGGCGGCGTTTAAGTGGATTATTTGAACCTGTCGT

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

P G A V P M P D R H V E P P Q I H L I N L D S
482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504

5' GCGTCGAGCGCTGGCGCAAATATCAAGATAGGAATCCTCATCTAGCCGCCAATACAATTCGGGTATCGGC 6790
3' CGCAGCTCGCGACCGCGTTTATAGTTCTATCCTTAGGAGTAGATCGGCGGTTATGTTAAGCCCATAGCCG

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

S V E R W R K Y Q D R N P H L A A N T I R V S A
505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528

SgrDI

5' CGTCGACGGAGCATCCCTCGATAGGTCGACCTTGCAAGCCTTGGTGGACGATGGAGTGATTGCCGAGGAT 6860
3' GCAGCTGCCTCGTAGGGAGCTATCCAGCTGGAACGTTCGGAACCACCTGCTACCTCACTAACGGCTCCTA

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

V D G A S L D R S T L Q A L V D D G V I A E D
529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551

5' TGCGGGTATCTGCCGGGCGCATTGGGATGTGCTCTATCGCACATCAACTTGTGGAAGTTAGCGGTATCCG 6930
3' ACGCCCATAGACGGCCCGCGTAACCCTACACGAGATAGCGTGTAGTTGAACACCTTCAATCGCCATAGGC

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

C G Y L P G A L G C A L S H I N L W K L A V S
552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574

5' AGAAGCGGCCAATTACCATTTTCGAGGACGACGTCTACTCGTCGTTCAACTTTCTTGAGGAGTCGACTCG 7000
3' TCTTCGCCGGTTAATGGTAAAAGCTCCTGCTGCAGATGAGCAGCAAGTTGAAAGAACTCCTCAGCTGAGC

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

E K R P I T I F E D D V Y S S F N F L E E S T R
575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598

BsmI
BseRI

5' CATTCTGTCCATGGCTCCAGAGAACTGGGATATGGTAAAGTGGGGATTCAATTTCGACCCTCTATTCCTG 7070
3' GTAAGACAGGTACCGAGGTCTCTTGACCCTATACCATTTCACCCCTAAGTTAAAGCTGGGAGATAAGGAC

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

I L S M A P E N W D M V K W G F N F D P L F L
599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621

AjuI' AjuI

5' TGGTTGAACTTCGATTTCTCCAAAGCCAAGCTGGAGTTCTACAACCGACAACACAAACAGATTCCGGCCC 7140
3' ACCAACTTGAAGCTAAAGAGGTTTCGGTTCGACCTCAAGATGTTGGCTGTTGTGTTTGTCTAAGGCCGGG

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

W L N F D F S K A K L E F Y N R Q H K Q I P A
622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644

SpeI

5' AATTTCAATCTGAAACCTATCCGCGATCTCTTATCCGACTAGTCCATTCCTTTGGTGCAATGGCTTATTC 7210
3' TTAAAGTTAGACTTTGGATAGGCGCTAGAGAATAGGCTGATCAGGTAAGGAAACCACGTTACCGAATAAG

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

Q F Q S E T Y P R S L I R L V H S F G A M A Y S
645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668

5' AGTAACTCCAAGGGGAGCTCAAGTACTTCTTGAAAAATGCCTACCTCTTCGAAAGCGGCTTATTCCGTTT 7280
3' TCATTGAGGTTCCCCTCGAGTTCATGAAGAACTTTTTACGGATGGAGAAGCTTTCGCCGAATAAGGCAAA

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

V T P R G A Q V L L E K C L P L R K R L I P F
669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691

AbsI

5' CCTGGGACTGGAGTGGTCCTCGAGGATGAGGGCATCGACTGCGCAATGTCCGCGGTGTATGACTCGATGC 7350
3' GGACCCTGACCTCACCAGGAGCTCCTACTCCCGTAGCTGACGCGTTACAGGCGCCACATACTGAGCTACG

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

P G T G V V L E D E G I D C A M S A V Y D S M
692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714

NsiI NruI

5' ATGCGTTCGTCTGCATGCCGCCACTGGTCATTCACGACGACGAGCAGGCATCAGATCGGGAAGTCGCGAA 7420
3' TACGCAAGCAGACGTACGGCGGTGACCAGTAAGTGCTGCTGCTCGTCCGTAGTCTAGCCCTTCAGCGCTT

BGT25_T02a3

GT25?

BGT25_T02a3-pDONR221

H A F V C M P P L V I H D D E Q A S D R E V A N
715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738

BamHI HindIII

5' TCGGCAGTAGGACCCAGCTTTCTTGTACAAAGTGGTTGATatcggatccgaattcgagctccgtcgacaa 7490
3' AGCCGTCATCCTGGGTCGAAAGAACATGTTTCACCAACTAtagcctaggcttaagctcgaggcagctgtt

BG...a3 attB2

GT25?

BGT25_T02a3-pDONR221

R Q .
739 740 741

NotI

5' gcttgcggccgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaa 7560
3' cgaacgccggcgtgagctcgtggtggtggtggtggtgactctaggccgacgattgtttcgggctttcctt

BlpI

5' gctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttga 7630
3' cgactcaaccgacgacggtggcgactcgttattgatcgtattggggaaccccggagatttgcccagaact

T7 Terminator

5' ggggttttttgctgaaaggaggaactatatccggat 7666
3' ccccaaaaaacgactttcctccttgatataggccta

T7 Terminator