SEQUENCE

```
  1  GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC
     CGCGGGTTAT GCGTTTGGCG GAGAGGGGCG CGCAACCGGC TAAGTAATTA CGTCGACCGT GCTGTCCAAA GGGCTGACCT TTCGCCCGTC ACTCGCGTTG

101  GCAATTAATG TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA TAACAATTTC
     CGTTAATTAC ACTCAATCGA GTGAGTAATC CGTGGGGTCC GAAATGTGAA ATACGAAGGC CGAGCATACA ACACACCTTA ACACTCGCCT ATTGTTAAAG
```

```
                                                                      BamHI      KpnI
                                                                      ~~~~~~     ~~~~~~
                                                                          SmaI
                                                                          ~~~~~~~
                                                                          XmaI
                                                                          ~~~~~~~
                          HindIII          PstI                          AvaI
                          ~~~~~~          ~~~~~~~                        ~~~~~~~
```

```
201  ACACAGGAAA CAGCTATGAC CATGATTACG CCAAGCTTGC ATGCCTGCAG GTCGACTCTA GAGGATCCCC GGGTACCCCG GCGCGCCTCA CTGATGATTC
     TGTGTCCTTT GTCGATACTG GTACTAATGC GGTTCGAACG TACGGACGTC CAGCTGAGAT CTCCTAGGGG CCCATGGGGC CGCGCGGAGT GACTACTAAG

301  GGGTCTACTT TCGGCGCCTG AGCATCATTT AGCTTTTTAG CTTCTGCTAA AAGGTTAGCG CTTTGGCTTG GGTCATCTTT TAAACTTTGG ATGAAGGCGT
     CCCAGATGAA AGCCGCGGAC TCGTAGTAAA TCGAAAAATC GAAGACGATT TTCCAATCGC GAAACCGAAC CCAGTAGAAA ATTTGAAACC TACTTCCGCA

401  TTCGTTGTTC TTCGTTTAAG TTAGGTAAAT GTAAGATCTC ATAGAACGCG TTTTGTTGTT CTTTGTTGAA TTTGTTGTCT ACTTTCGGCG CCTGAGCATC
     AAGCAACAAG AAGCAAATTC AATCCATTTA CATTCTAGAG TATCTTGCGC AAAACAACAA GAAACAACTT AAACAACAGA TGAAAGCCGC GGACTCGTAG

501  ATTTAGCTTT TTAGCTTCTG CTAAAAGGTT AGCGCTTTGG CTTGGGTCAT CTTTTAAACT TTGGATGAAG GCGTTTCGTT GTTCTTCGTT TAAGTTAGGT
     TAAATCGAAA AATCGAAGAC GATTTTCCAA TCGCGAAACC GAACCCAGTA GAAAATTTGA AACCTACTTC CGCAAAGCAA CAAGAAGCAA ATTCAATCCA

601  AAATGTAAGA TCTCATAGAA CGCGTTTTGT TGTTCTTTGT TGAATTTGTT GTCCACGGCC GCATGGTGAT GGTGGTGGTG AGAACCGGCG TAATCTGGAA
     TTTACATTCT AGAGTATCTT GCGCAAAACA ACAAGAAACA ACTTAAACAA CAGGTGCCGG CGTACCACTA CCACCACCAC TCTTGGCCGC ATTAGACCTT
```

```
                          HindIII          PmlI          HindIII              PacI
                          ~~~~~~          ~~~~~~~        ~~~~~~              ~~~~~~~~
```

```
701  CATCGTATGG ATAAAAGATT CTACCTGGAC CTGGACCTTG GAACAAAGCT TCCAAACCAC GTGCTTGACA AGCTTCCGTA CCACTATTAA TTAAGGTTGA
     GTAGCATACC TATTTTCTAA GATGGACCTG GACCTGGAAC CTTGTTTCGA AGGTTTGGTG CACGAACTGT TCGAAGGCAT GGTGATAATT AATTCCAACT
```

```
       EcoRI
       ~~~~~~
```

```
801  TGGAATTCCT TCTCCTTGAC GTTAAAGTAT AGAGGTATAT TAACAATTTT TTGTTGATAC TTTTATGACA TTTGAATAAG AAGTAATACA AACTGAAAAT
     ACCTTAAGGA AGAGGAACTG CAATTTCATA TCTCCATATA ATTGTTAAAA AACAACTATG AAAATACTGT AAACTTATTC TTCATTATGT TTGACTTTTA

901  GTTGAAAGTA TTAGTTAAAG TGGTTATGCA GCTTTTCCAT TTATATATCT GTTAATAGAT CAAAAATCAT CGCTTCGCTG ATTAATTACC CCAGAAATAA
     CAACTTTCAT AATCAATTTC ACCAATACGT CGAAAAGGTA AATATATAGA CAATTATCTA GTTTTTAGTA GCGAAGCGAC TAATTAATGG GGTCTTTATT

1001 GGCTAAAAAA CTAATCGCAT TATCATCCTA TGGTTGTTAA TTTGATTCGT TAATTTGAAG GTTTGTGGGG CCAGGTTACT GCCAATTTTT CCTCTTCATA
     CCGATTTTTT GATTAGCGTA ATAGTAGGAT ACCAACAATT AAACTAAGCA ATTAAACTTC CAAACACCCC GGTCCAATGA CGGTTAAAAA GGAGAAGTAT

1101 ACCATAAAAG CTAGTATTGT AGAATCTTTA TTGTTCGGAG CAGTGCGGCG CGAGGCACAT CTGCGTTTCA GGAACGCGAC CGGTGAAGAC GAGGACGCAC
     TGGTATTTTC GATCATAACA TCTTAGAAAT AACAAGCCTC GTCACGCCGC GCTCCGTGTA GACGCAAAGT CCTTGCGCTG GCCACTTCTG CTCCTGCGTG
```

```
                                                                    SacI
                                                                   ~~~~~~~
                                                   SacII            EcoRI
                                                   ~~~~~~~          ~~~~~~
1201 GGGAGGAGAG TCTTCCGTCG GAGGGCTGTC GCCCGCTCGG CGGCTTCTAA TCCGTTCCCG CGGTGGAGCT CGAATTCACT GGCCGTCGTT TTACAACGTC
     CCCTCCTCTC AGAAGGCAGC CTCCCGACAG CGGGCGAGCC GCCGAAGATT AGGCAAGGGC GCCACCTCGA GCTTAAGTGA CCGGCAGCAA AATGTTGCAG

1301 GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
     CACTGACCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA GGGGGAAAGC GGTCGACCGC ATTATCGCTT CTCCGGGCGT GGCTAGCGGG

1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGCGCCTG ATGCGGTATT TTCTCCTTAC GCATCTGTGC GGTATTTCAC ACCGCATATA TCGGATCGTA
     AAGGGTTGTC AACGCGTCGG ACTTACCGCT TACCGCGGAC TACGCCATAA AAGAGGAATG CGTAGACACG CCATAAAGTG TGGCGTATAT AGCCTAGCAT

1501 CTTGTTACCC ATCATTGAAT TTTGAACATC CGAACCTGGG AGTTTTCCCT GAAACAGATA GTATATTTGA ACCTGTATAA TAATATATAG TCTAGCGCTT
     GAACAATGGG TAGTAACTTA AAACTTGTAG GCTTGGACCC TCAAAAGGGA CTTTGTCTAT CATATAAACT TGGACATATT ATTATATATC AGATCGCGAA

1601 TACGGAAGAC AATGTATGTA TTTCGGTTCC TGGAGAAACT ATTGCATCTA TTGCATAGGT AATCTTGCAC GTCGCATCCC CGGTTCATTT TCTGCGTTTC
     ATGCCTTCTG TTACATACAT AAAGCCAAGG ACCTCTTTGA TAACGTAGAT AACGTATCCA TTAGAACGTG CAGCGTAGGG GCCAAGTAAA AGACGCAAAG

1701 CATCTTGCAC TTCAATAGCA TATCTTTGTT AACGAAGCAT CTGTGCTTCA TTTTGTAGAA CAAAAATGCA ACGCGAGAGC GCTAATTTTT CAAACAAAGA
     GTAGAACGTG AAGTTATCGT ATAGAAACAA TTGCTTCGTA GACACGAAGT AAAACATCTT GTTTTTACGT TGCGCTCTCG CGATTAAAAA GTTTGTTTCT

1801 ATCTGAGCTG CATTTTTACA GAACAGAAAT GCAACGCGAA AGCGCTATTT TACCAACGAA GAATCTGTGC TTCATTTTTG TAAAACAAAA ATGCAACGCG
     TAGACTCGAC GTAAAAATGT CTTGTCTTTA CGTTGCGCTT TCGCGATAAA ATGGTTGCTT CTTAGACACG AAGTAAAAAC ATTTTGTTTT TACGTTGCGC

1901 AGAGCGCTAA TTTTTCAAAC AAAGAATCTG AGCTGCATTT TTACAGAACA GAAATGCAAC GCGAGAGCGC TATTTTACCA ACAAAGAATC TATACTTCTT
     TCTCGCGATT AAAAAGTTTG TTTCTTAGAC TCGACGTAAA AATGTCTTGT CTTTACGTTG CGCTCTCGCG ATAAATGGT TGTTTCTTAG ATATGAAGAA

                           AvaI
                           ~~~~~~
2001 TTTTGTTCTA CAAAAATGCA TCCCGAGAGC GCTATTTTTC TAACAAAGCA TCTTAGATTA CTTTTTTTCT CCTTTGTGCG CTCTATAATG CAGTCTCTTG
     AAAACAAGAT GTTTTTACGT AGGGCTCTCG CGATAAAAAG ATTGTTTCGT AGAATCTAAT GAAAAAAAGA GGAAACACGC GAGATATTAC GTCAGAGAAC

2101 ATAACTTTTT GCACTGTAGG TCCGTTAAGG TTAGAAGAAG GCTACTTTGG TGTCTATTTT CTCTTCCATA AAAAAAGCCT GACTCCACTT CCCGCGTTTA
     TATTGAAAAA CGTGACATCC AGGCAATTCC AATCTTCTTC CGATGAAACC ACAGATAAAA GAGAAGGTAT TTTTTTCGGA CTGAGGTGAA GGGCGCAAAT

2201 CTGATTACTA GCGAAGCTGC GGGTGCATTT TTTCAAGATA AAGGCATCCC CGATTATATT CTATACCGAT GTGGATTGCG CATACTTTGT GAACAGAAAG
     GACTAATGAT CGCTTCGACG CCCACGTAAA AAAGTTCTAT TTCCGTAGGG GCTAATATAA GATATGGCTA CACCTAACGC GTATGAAACA CTTGTCTTTC

2301 TGATAGCGTT GATGATTCTT CATTGGTCAG AAAATTATGA ACGGTTTCTT CTATTTTGTC TCTATATACT ACGTATAGGA AATGTTTACA TTTTTCGTATT
     ACTATCGCAA CTACTAAGAA GTAACCAGTC TTTTAATACT TGCCAAAGAA GATAAAACAG AGATATATGA TGCATATCCT TTACAAATGT AAAAGCATAA
2401 GTTTTCGATT CACTCTATGA ATAGTTCTTA CTACAATTTT TTTGTCTAAA GAGTAATACT AGAGATAAAC ATAAAAAATG TAGAGGTCGA GTTTAGATGC
     CAAAAGCTAA GTGAGATACT TATCAAGAAT GATGTTAAAA AAACAGATTT CTCATTATGA TCTCTATTTG TATTTTTTAC ATCTCCAGCT CAAATCTACG

2501 AAGTTCAAGG AGCGAAAGGT GGATGGGTAG GTTATATAGG GATATAGCAC AGAGATATAT AGCAAAGAGA TACTTTTGAG CAATGTTTGT GGAAGCGGTA
     TTCAAGTTCC TCGCTTTCCA CCTACCCATC CAATATATCC CTATATCGTG TCTCTATATA TCGTTTCTCT ATGAAAACTC GTTACAAACA CCTTCGCCAT

2601 TTCGCAATAT TTTAGTAGCT CGTTACAGTC CGGTGCGTTT TTGGTTTTTT GAAAGTGCGT CTTCAGAGCG CTTTTGGTTT TCAAAGCGC TCTGAAGTTC
     AAGCGTTATA AAATCATCGA GCAATGTCAG GCCACGCAAA AACCAAAAAA CTTTCACGCA GAAGTCTCGC GAAAACCAAA AGTTTTCGCG AGACTTCAAG
```

```
2701  CTATACTTTC TAGCTAGAGA ATAGGAACTT CGGAATAGGA ACTTCAAAGC GTTTCCGAAA ACGAGCGCTT CCGAAAATGC AACGCGAGCT GCGCACATAC
      GATATGAAAG ATCGATCTCT TATCCTTGAA GCCTTATCCT TGAAGTTTCG CAAAGGCTTT TGCTCGCGAA GGCTTTTACG TTGCGCTCGA CGCGTGTATG

2801  AGCTCACTGT TCACGTCGCA CCTATATCTG CGTGTTGCCT GTATATATAT ATACATGAGA AGAACGGCAT AGTGCGTGTT TATGCTTAAA TGCGTACTTA
      TCGAGTGACA AGTGCAGCGT GGATATAGAC GCACAACGGA CATATATATA TATGTACTCT TCTTGCCGTA TCACGCACAA ATACGAATTT ACGCATGAAT

2901  TATGCGTCTA TTTATGTAGG ATGAAAGGTA GTCTAGTACC TCCTGTGATA TTATCCCATT CCATGCGGGG TATCGTATGC TTCCTTCAGC ACTACCCTTT
      ATACGCAGAT AAATACATCC TACTTTCCAT CAGATCATGG AGGACACTAT AATAGGGTAA GGTACGCCCC ATAGCATACG AAGGAAGTCG TGATGGGAAA

                                                                                        ClaI
                                                                                      ~~~~~~~
3001  AGCTGTTCTA TATGCTGCCA CTCCTCAATT GGATTAGTCT CATCCTTCAA TGCTATCATT TCCTTTGATA TTGGATCGAT CCGATGATAA GCTGTCAAAC
      TCGACAAGAT ATACGACGGT GAGGAGTTAA CCTAATCAGA GTAGGAAGTT ACGATAGTAA AGGAAACTAT AACCTAGCTA GGCTACTATT CGACAGTTTG

3101  ATGAGAATTA ATTCTACCCT ATGAACATAT TCCATTTTGT AATTTCGTGT CGTTTCTATT ATGAATTTCA TTTATAAAGT TTATGTACAA ATATCATAAA
      TACTCTTAAT TAAGATGGGA TACTTGTATA AGGTAAAACA TTAAAGCACA GCAAAGATAA TACTTAAAGT AAATATTTCA AATACATGTT TATAGTATTT

3201  AAAAGAGAAT CTTTTTAAGC AAGGATTTTC TTAACTTCTT CGGCGACAGC ATCACCGACT TCGGTGGTAC TGTTGGAACC ACCTAAATCA CCAGTTCTGA
      TTTTCTCTTA GAAAAATTCG TTCCTAAAAG AATTGAAGAA GCCGCTGTCG TAGTGGCTGA AGCCACCATG ACAACCTTGG TGGATTTAGT GGTCAAGACT

3301  TACCTGCATC CAAAACCTTT TTAACTGCAT CTTCAATGGC CTTACCTTCT TCAGGCAAGT TCAATGACAA TTTCAACATC ATTGCAGCAG ACAAGATAGT
      ATGGACGTAG GTTTTGGAAA AATTGACGTA GAAGTTACCG GAATGGAAGA AGTCCGTTCA AGTTACTGTT AAAGTTGTAG TAACGTCGTC TGTTCTATCA

3401  GGCGATAGGG TCAACCTTAT TCTTTGGCAA ATCTGGAGCA GAACCGTGGC ATGGTTCGTA CAAACCAAAT GCGGTGTTCT TGTCTGGCAA AGAGGCCAAG
      CCGCTATCCC AGTTGGAATA AGAAACCGTT TAGACCTCGT CTTGGCACCG TACCAAGCAT GTTTGGTTTA CGCCACAAGA ACAGACCGTT TCTCCGGTTC

3501  GACGCAGATG GCAACAAACC CAAGGAACCT GGGATAACGG AGGCTTCATC GGAGATGATA TCACCAAACA TGTTGCTGGT GATTATAATA CCATTTAGGT
      CTGCGTCTAC CGTTGTTTGG GTTCCTTGGA CCCTATTGCC TCCGAAGTAG CCTCTACTAT AGTGGTTTGT ACAACGACCA CTAATATTAT GGTAAATCCA

3601  GGGTTGGGTT CTTAACTAGG ATCATGGCGG CAGAATCAAT CAATTGATGT TGAACCTTCA ATGTAGGAAA TTCGTTCTTG ATGGTTTCCT CCACAGTTTT
      CCCAACCCAA GAATTGATCC TAGTACCGCC GTCTTAGTTA GTTAACTACA ACTTGGAAGT TACATCCTTT AAGCAAGAAC TACCAAAGGA GGTGTCAAAA

3701  TCTCCATAAT CTTGAAGAGG CCAAAACATT AGCTTTATCC AAGGACCAAA TAGGCAATGG TGGCTCATGT TGTAGGGCCA TGAAAGCGGC CATTCTTGTG
      AGAGGTATTA GAACTTCTCC GGTTTTGTAA TCGAAATAGG TTCCTGGTTT ATCCGTTACC ACCGAGTACA ACATCCCGGT ACTTTCGCCG GTAAGAACAC

3801  ATTCTTTGCA CTTCTGGAAC GGTGTATTGT TCACTATCCC AAGCGACACC ATCACCATCG TCTTCCTTTC TCTTACCAAA GTAAATACCT CCCACTAATT
      TAAGAAACGT GAAGACCTTG CCACATAACA AGTGATAGGG TTCGCTGTGG TAGTGGTAGC AGAAGGAAAG AGAATGGTTT CATTTATGGA GGGTGATTAA

3901  CTCTGACAAC AACGAAGTCA GTACCTTTAG CAAATTGTGG CTTGATTGGA GATAAGTCTA AAAGAGAGTC GGATGCAAAG TTACATGGTC TTAAGTTGGC
      GAGACTGTTG TTGCTTCAGT CATGGAAATC GTTTAACACC GAACTAACCT CTATTCAGAT TTTCTCTCAG CCTACGTTTC AATGTACCAG AATTCAACCG

4001  GTACAATTGA AGTTCTTTAC GGATTTTTAG TAAACCTTGT TCAGGTCTAA CACTACCTGT ACCCCATTTA GGACCACCCA CAGCACCTAA CAAAACGGCA
      CATGTTAACT TCAAGAAATG CCTAAAAATC ATTTGGAACA AGTCCAGATT GTGATGGACA TGGGGTAAAT CCTGGTGGGT GTCGTGGATT GTTTTGCCGT

                                                                              ClaI
                                                                            ~~~~~~
4101  TCAACCTTCT TGGAGGCTTC CAGCGCCTCA TCTGGAAGTG GGACACCTGT AGCATCGATA GCAGCACCAC CAATTAAATG ATTTTCGAAA TCGAACTTGA
      AGTTGGAAGA ACCTCCGAAG GTCGCGGAGT AGACCTTCAC CCTGTGGACA TCGTAGCTAT CGTCGTGGTG GTTAATTTAC TAAAAGCTTT AGCTTGAACT

4201  CATTGGAACG AACATCAGAA ATAGCTTTAA GAACCTTAAT GGCTTCGGCT GTGATCTTGA CCAACGTGGT CACCTGGCAA AACGACGATC TTCTTAGGGG
```

```
      GTAACCTTGC TTGTAGTCTT TATCGAAATT CTTGGAATTA CCGAAGCCGA CACTAGAACT GGTTGCACCA GTGGACCGTT TTGCTGCTAG AAGAATCCCC

4301  CAGACATAGG GGCAGACATT AGAATGTATA TCCTTGAAAT ATATATATAT ATTGCTGAAA TGTAAAAGGT AAGAAAAGTT AGAATAAGAC GATTGCTAAC
      GTCTGTATCC CCGTCTGTAA TCTTACATAT AGGAACTTTA TATATATATA TAACGACTTT ACATTTTCCA TTCTTTTCAA TCTTATTCTG CTAACGATTG

4401  CACCTATTGG AAAAAACAAT AGGTCCTTAA ATAATATTGT CAACTTCAAG TATTGTGATG CAAGCATTTA GTCATGAACG CTTCTCTATT CTATATGAAA
      GTGGATAACC TTTTTTGTTA TCCAGGAATT TATTATAACA GTTGAAGTTC ATAACACTAC GTTCGTAAAT CAGTACTTGC GAAGAGATAA GATATACTTT

4501  AGCCGGTTCC GGCCTCTCAC CTTTCCTTTT TCTCCCAATT TTTCAGTTGA AAAAGGTATA TGCGCGGCGA CCTCTGAAAT TAACAAAAAA TTTCCAGTCA
      TCGGCCAAGG CCGGAGAGTG GAAAGGAAAA AGAGGGTTAA AAAGTCAACT TTTTCCATAT ACGCGCCGCT GGAGACTTTA ATTGTTTTTT AAAGGTCAGT

4601  TCGAATTTGT CTGTGCGATA GCGCCCCTGT GTGTTCTCGT TATGTTGAGG AAAAAAATAA TGGTTGCTAA GAGATTCGAA CTCTTGCATC TTACGATACC
      AGCTTAAACA GACACGCTAT CGCGGGGACA CACAAGAGCA ATACAACTCC TTTTTTTATT ACCAACGATT CTCTAAGCTT GAGAACGTAG AATGCTATGG

4701  TGAGTATTCC CACAGTTAAT TCTTGAAGAC GAAAGGGCCT CGTGATACGC CTATTTTTAT AGGTTAATGT CATGATAATA ATGGTTTCTT AGACGTCAGG
      ACTCATAAGG GTGTCAATTA AGAACTTCTG CTTTCCCGGA GCACTATGCG GATAAAAATA TCCAATTACA GTACTATTAT TACCAAAGAA TCTGCAGTCC

4801  TGGCACTTTT CGGGGAAATG TGCGCGGAAC CCCTATTTGT TTATTTTTCT AAATACATTC AAATATGTAT CCGCTCATGA GACAATAACC CTGATAAATG
      ACCGTGAAAA GCCCCTTTAC ACGCGCCTTG GGGATAAACA AATAAAAAGA TTTATGTAAG TTTATACATA GGCGAGTACT CTGTTATTGG GACTATTTAC

4901  CTTCAATAAT ATTGAAAAAG GAAGAGTATG AGTATTCAAC ATTTCCGTGT CGCCCTTATT CCCTTTTTTG CGGCATTTTG CCTTCCTGTT TTTGCTCACC
      GAAGTTATTA TAACTTTTTC CTTCTCATAC TCATAAGTTG TAAAGGCACA GCGGGAATAA GGGAAAAAAC GCCGTAAAAC GGAAGGACAA AAACGAGTGG

                                                                 ApaLI
                                                                ~~~~~~
5001  CAGAAACGCT GGTGAAAGTA AAAGATGCTG AAGATCAGTT GGGTGCACGA GTGGGTTACA TCGAACTGGA TCTCAACAGC GGTAAGATCC TTGAGAGTTT
      GTCTTTGCGA CCACTTTCAT TTTCTACGAC TTCTAGTCAA CCCACGTGCT CACCCAATGT AGCTTGACCT AGAGTTGTCG CCATTCTAGG AACTCTCAAA

5101  TCGCCCCGAA GAACGTTTTC CAATGATGAG CACTTTTAAA GTTCTGCTAT GTGGCGCGGT ATTATCCCGT ATTGACGCCG GGCAAGAGCA ACTCGGTCGC
      AGCGGGGCTT CTTGCAAAAG GTTACTACTC GTGAAAATTT CAAGACGATA CACCGCGCCA TAATAGGGCA TAACTGCGGC CCGTTCTCGT TGAGCCAGCG

5201  CGCATACACT ATTCTCAGAA TGACTTGGTT GAGTACTCAC CAGTCACAGA AAAGCATCTT ACGGATGGCA TGACAGTAAG AGAATTATGC AGTGCTGCCA
      GCGTATGTGA TAAGAGTCTT ACTGAACCAA CTCATGAGTG GTCAGTGTCT TTTCGTAGAA TGCCTACCGT ACTGTCATTC TCTTAATACG TCACGACGGT

5301  TAACCATGAG TGATAACACT GCGGCCAACT TACTTCTGAC AACGATCGGA GGACCGAAGG AGCTAACCGC TTTTTTGCAC AACATGGGGG ATCATGTAAC
      ATTGGTACTC ACTATTGTGA CGCCGGTTGA ATGAAGACTG TTGCTAGCCT CCTGGCTTCC TCGATTGGCG AAAAAACGTG TTGTACCCCC TAGTACATTG

5401  TCGCCTTGAT CGTTGGGAAC CGGAGCTGAA TGAAGCCATA CCAAACGACG AGCGTGACAC CACGATGCCT GTAGCAATGG CAACAACGTT GCGCAAACTA
      AGCGGAACTA GCAACCCTTG GCCTCGACTT ACTTCGGTAT GGTTTGCTGC TCGCACTGTG GTGCTACGGA CATCGTTACC GTTGTTGCAA CGCGTTTGAT

5501  TTAACTGGCG AACTACTTAC TCTAGCTTCC CGGCAACAAT TAATAGACTG GATGGAGGCG GATAAAGTTG CAGGACCACT TCTGCGCTCG GCCCTTCCGG
      AATTGACCGC TTGATGAATG AGATCGAAGG GCCGTTGTTA ATTATCTGAC CTACCTCCGC CTATTTCAAC GTCCTGGTGA AGACGCGAGC CGGGAAGGCC

5601  CTGGCTGGTT TATTGCTGAT AAATCTGGAG CCGGTGAGCG TGGGTCTCGC GGTATCATTG CAGCACTGGG GCCAGATGGT AAGCCCTCCC GTATCGTAGT
      GACCGACCAA ATAACGACTA TTTAGACCTC GGCCACTCGC ACCCAGAGCG CCATAGTAAC GTCGTGACCC CGGTCTACCA TTCGGGAGGG CATAGCATCA

5701  TATCTACACG ACGGGGAGTC AGGCAACTAT GGATGAACGA AATAGACAGA TCGCTGAGAT AGGTGCCTCA CTGATTAAGC ATTGGTAACT GTCAGACCAA
      ATAGATGTGC TGCCCCTCAG TCCGTTGATA CCTACTTGCT TTATCTGTCT AGCGACTCTA TCCACGGAGT GACTAATTCG TAACCATTGA CAGTCTGGTT

5801  GTTTACTCAT ATATACTTTA GATTGATTTA AAACTTCATT TTTAATTTAA AAGGATCTAG GTGAAGATCC TTTTTGATAA TCTCATGACC AAAATCCCTT
```

```
          CAAATGAGTA TATATGAAAT CTAACTAAAT TTTGAAGTAA AAATTAAATT TTCCTAGATC CACTTCTAGG AAAAACTATT AGAGTACTGG TTTTAGGGAA

5901      AACGTGAGTT TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA AAAGATCAAA GGATCTTCTT GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC
          TTGCACTCAA AAGCAAGGTG ACTCGCAGTC TGGGGCATCT TTTCTAGTTT CCTAGAAGAA CTCTAGGAAA AAAAGACGCG CATTAGACGA CGAACGTTTG

6001      AAAAAAACCA CCGCTACCAG CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA GCAGAGCGCA GATACCAAAT
          TTTTTTTGGT GGCGATGGTC GCCACCAAAC AAACGGCCTA GTTCTCGATG GTTGAGAAAA AGGCTTCCAT TGACCGAAGT CGTCTCGCGT CTATGGTTTA

6101      ACTGTCCTTC TAGTGTAGCC GTAGTTAGGC CACCACTTCA AGAACTCTGT AGCACCGCCT ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG
          TGACAGGAAG ATCACATCGG CATCAATCCG GTGGTGAAGT TCTTGAGACA TCGTGGCGGA TGTATGGAGC GAGACGATTA GGACAATGGT CACCGACGAC

                                                                                                                ApaLI
                                                                                                                ~~~~~~~
6201      CCAGTGGCGA TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC
          GGTCACCGCT ATTCAGCACA GAATGGCCCA ACCTGAGTTC TGCTATCAAT GGCCTATTCC GCGTCGCCAG CCCGACTTGC CCCCCAAGCA CGTGTGTCGG

6301      CAGCTTGGAG CGAACGACCT ACACCGAACT GAGATACCTA CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA GAAAGGCGGA CAGGTATCCG
          GTCGAACCTC GCTTGCTGGA TGTGGCTTGA CTCTATGGAT GTCGCACTCG ATACTCTTTC GCGGTGCGAA GGGCTTCCCT CTTTCCGCCT GTCCATAGGC

6401      GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC TTCCAGGGGG AAACGCCTGG TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG
          CATTCGCCGT CCCAGCCTTG TCCTCTCGCG TGCTCCCTCG AAGGTCCCCC TTTGCGGACC ATAGAAATAT CAGGACAGCC CAAAGCGGTG GAGACTGAAC

6501      AGCGTCGATT TTTGTGATGC TCGTCAGGGG GGCGGAGCCT ATGGAAAAAC GCCAGCAACG CGGCCTTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC
          TCGCAGCTAA AAACACTACG AGCAGTCCCC CCGCCTCGGA TACCTTTTTG CGGTCGTTGC GCCGGAAAAA TGCCAAGGAC CGGAAAACGA CCGGAAAACG

6601      TCACATGTTC TTTCCTGCGT TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCTTTGA GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC
          AGTGTACAAG AAAGGACGCA ATAGGGGACT AAGACACCTA TTGGCATAAT GGCGGAAACT CACTCGACTA TGGCGAGCGG CGTCGGCTTG CTGGCTCGCG

6701      AGCGAGTCAG TGAGCGAGGA AGCGGAAGA
          TCGCTCAGTC ACTCGCTCCT TCGCCTTCT
```