



	ScaI	ScaI	ScaI	ScaI	
	~~~~~	~~~~~	~~~~~	~~~~~	
1	AGTACTTTAA TCATGAAATT	AGTACTTTAA TCATGAAATT	AGTACTTTAA TCATGAAATT	AGTACTTTGA TCATGAAACT	TCCAACCCCT AGGTTGGGGA
51	CCGCTGCTAT GGCGACGATA	AGTGCAGTCG TCACGTCAGC	GCTTCTGACG CGAAGACTGC	TTCAGTGCAG AAGTCACGTC	CCGTCTTCTG GGCAGAAGAC
101	AAAACGACAT TTTTGCTGTA	GTCGCACAAG CAGCGTGTTT	TCCTAAGTTA AGGATTCAAT	CGCGACAGGC GCGCTGTCCG	TGCCGCCCTG ACGGCGGGAC
151	CCCTTTTCTT GGGAAAAGGA	GGCGTTTTCT CCGCAAAAAGA	TGTCGCGTGT ACAGCGCACA	TTTAGTCGCA AAATCAGCGT	TAAAGTAGAA ATTTTCATCTT
201	TACTTGCGAC ATGAACGCTG	TAGAACCGGA ATCTTGGCCT	GACATTACGC CTGTAATGCG	CATGAACAAG GTAAGTGTTC	AGCGCCGCCG TCGCGGCGGC
251	CTGGCCTGCT GACCGGACGA	GGGCTATGCC CCCGATACGG	CGCGTCAGCA GCGCAGTCGT	CCGACGACCA GGCTGCTGGT	GGACTTGACC CCTGAACTGG
301	AACCAACGGG TTGGTTGCC	CCGAAGTACA GGCTTGACGT	CGCGGCCGGC GCGCCGGCCG	TGCACCAAGC ACGTGGTTCG	TGTTTTCCGA ACAAAAGGCT
351	GAAGATCACC CTTCTAGTGG	GGCACCAGGC CCGTGGTCCG	GCGACCGCCC CGCTGGCGGG	GGAGCTGGCC CCTCGACCGG	AGGATGCTTG TCCTACGAAC
401	ACCACCTACG TGGTGGATGC	CCCTGGCGAC GGGACCGCTG	GTTGTGACAG CAACACTGTC	TGACCAGGCT ACTGGTCCGA	AGACCGCCTG TCTGGCGGAC
451	GCCCGCAGCA CGGGCGTCTG	CCCGCGACCT GGGCGCTGGA	ACTGGACATT TGACCTGTAA	GCCGAGCGCA CGGCTCGCGT	TCCAGGAGGC AGGTCCTCCG
501	CGGCGCGGGC GCCGCGCCCG	CTGCGTAGCC GACGCATCGG	TGGCAGAGCC ACCGTCTCGG	GTGGGCCGAC CACCCGGCTG	ACCACCACGC TGGTGGTGGC
551	CGGCCGGCCG GCCGGCCGGC	CATGGTGTGG GTACCACAAC	ACCGTGTTCG TGCCACAAGC	CCGGCATTGC GGCCGTAACG	CGAGTTCGAG GCTCAAGCTC
601	CGTTCCCTAA GCAAGGGATT	TCATCGACCG AGTAGCTGGC	CACCCGGAGC GTGGGCCTCG	GGGCGCGAGG CCCAGCGCTC	CCGCCAAGGC GGCGGTTCCG
651	CCGAGGCGTG GGCTCCGCAC	AAGTTTGGCC TTCAAACCGG	CCCGCCCTAC GGGCGGGATG	CCTCACCCTG GGAGTGGGGC	GCACAGATCG CGTGTCTAGC
701	CGCAGCCCCG GCGTGCGGGC	CGAGCTGATC GCTCGACTAG	GACCAGGAAG CTGGTCCTTC	GCCGCACCGT CGGCGTGGCA	GAAAGAGGGC CTTTCTCCGC
751	GCTGCACTGC CGACGTGACG	TTGGCGTGCA AACCGCACGT	TCGCTCGACC AGCGAGCTGG	CTGTACCGCG GACATGGCGC	CACTTGAGCG GTGAACTCGC
801	CAGCGAGGAA GTCGCTCCTT	GTGACGCCCA CACTGCGGGT	CCGAGGCCAG GGCTCCGGTC	GCGGCGCGGT CGCCGCGCCA	GCCTTCCGTC CGGAAGGCAC

## NotI

~~~~~

| | | | | | |
|------|------------|-------------|------------|------------|-------------|
| 851 | AGGACGCATT | GACCGAGGCC | GACGCCCTGG | CGGCCGCCGA | GAATGAACGC |
| | TCCTGCGTAA | CTGGCTCCGG | CTGCGGGACC | GCCGGCGGCT | CTTACTTGCG |
| 901 | CAAGAGGAAC | AAGCATGAAA | CCGCACCAGG | ACGGCCAGGA | CGAACCGTTT |
| | GTTCTCCTTG | TTCGTACTTT | GGCGTGGTCC | TGCCGGTCCT | GCTTGGCAAA |
| 951 | TTCATTACCG | AAGAGATCGA | GGCGGAGATG | ATCGCGGCCG | GGTACGTGTT |
| | AAGTAATGGC | TTCTCTAGCT | CCGCCTCTAC | TAGCGCCGGC | CCATGCACAA |
| 1001 | CGAGCCGCCC | GCGCACGTCT | CAACCGTGCG | GCTGCATGAA | ATCCTGGCCG |
| | GCTCGGCGGG | CGCGTGCAGA | GTTGGCACGC | CGACGTACTT | TAGGACCGGC |
| 1051 | GTTTGTCTGA | TGCCAAGCTG | GCGGCCTGGC | CGGCCAGCTT | GGCCGCTGAA |
| | CAAACAGACT | ACGGTTCGAC | CGCCGGACCG | GCCGGTCGAA | CCGGCGACTT |
| 1101 | GAAACCGAGC | GCCGCCGTCT | AAAAAGGTGA | TGTGTATTTG | AGTAAAACAG |
| | CTTTGGCTCG | CGGCGGCAGA | TTTTTCCACT | ACACATAAAC | TCATTTTTGTC |
| 1151 | CTTGCGTCAT | GCGGTCGCTG | CGTATATGAT | GCGATGAGTA | AATAAACAAA |
| | GAACGCAGTA | CGCCAGCGAC | GCATATACTA | CGCTACTCAT | TTATTTGTTT |
| 1201 | TACGCAAGGG | GAACGCATGA | AGGTTATCGC | TGTACTTAAC | CAGAAAGGCG |
| | ATGCGTTCCC | CTTGCGTACT | TCCAATAGCG | ACATGAATTG | GTCTTTCCGC |
| 1251 | GGTCAGGCAA | GACGACCATC | GCAACCCATC | TAGCCCAGCG | CCTGCAACTC |
| | CCAGTCCGTT | CTGCTGGTAG | CGTTGGGTAG | ATCGGGCGCG | GGACGTTGAG |
| 1301 | GCCGGGGCCG | ATGTTCTGTT | AGTCGATTCC | GATCCCCAGG | GCAGTGCCCC |
| | CGGCCCCCGC | TACAAGACAA | TCAGCTAAGG | CTAGGGGTCC | CGTCACGGGC |
| 1351 | CGATTGGGCG | GCCGTGCGGG | AAGATCAACC | GCTAACCGTT | GTCGGCATCG |
| | GCTAACCCGC | CGGCACGCCC | TTCTAGTTGG | CGATTGGCAA | CAGCCGTAGC |
| 1401 | ACCGCCCGAC | GATTGACCGC | GACGTGAAGG | CCATCGGCCG | GCGCGACTTC |
| | TGGCGGGCTG | CTAACTGGCG | CTGCACTTCC | GGTAGCCGGC | CGCGCTGAAG |
| 1451 | GTAGTGATCG | ACGGAGCGCC | CCAGGCGGCG | GACTTGGCTG | TGTCCGCGAT |
| | CATCACTAGC | TGCCTCGCGG | GGTCCGCCGC | CTGAACCGAC | ACAGGCGCTA |
| 1501 | CAAGGCAGCC | GACTTCGTGC | TGATTCCGGT | GCAGCCAAGC | CCTTACGACA |
| | GTTCCGTCGG | CTGAAGCACG | ACTAAGGCCA | CGTCGGTTCG | GGAATGCTGT |
| 1551 | TATGGGCCAC | CGCCGACCTG | GTGGAGCTGG | TTAAGCAGCG | CATTGAGGTC |
| | ATACCCGGTG | GCGGCTGGAC | CACCTCGACC | AATTCGTGCG | GTAACCTCCAG |
| 1601 | ACGGATGGAA | GGCTACAAGC | GGCCTTTGTC | GTGTCGCGGG | CGATCAAAGG |
| | TGCCTACCTT | CCGATGTTTCG | CCGAAACAG | CACAGCGCCC | GCTAGTTTCC |
| 1651 | CACGCGCATC | GGCGGTGAGG | TTGCCGAGGC | GCTGGCCGGG | TACGAGCTGC |
| | GTGCGCGTAG | CCGCCACTCC | AACGGCTCCG | CGACCGGCC | ATGCTCGACG |

| | | | | | |
|------|---------------------------|---------------------------|---------------------------|--------------------------|---------------------------|
| 1701 | CCATTCTTGA
GGTAAGAACT | GTCCCGTATC
CAGGGCATAG | ACGCAGCGCG
TGCCTCGCGC | TGAGCTACCC
ACTCGATGGG | AGGCACTGCC
TCCGTGACGG |
| 1751 | GCCGCCGGCA
CGGCGGCCGT | CAACCGTTCT
GTTGGCAAGA | TGAATCAGAA
ACTTAGTCTT | CCCGAGGGCG
GGGCTCCCGC | ACGCTGCCCCG
TGCGACGGGC |
| 1801 | CGAGGTCCAG
GCTCCAGGTC | GCGCTGGCCG
CGCGACCGGC | CTGAAATTAA
GACTTTAATT | ATCAAACTC
TAGTTTTGAG | ATTTGAGTTA
TAAACTCAAT |
| 1851 | ATGAGGTAAA
TACTCCATTT | GAGAAAATGA
CTCTTTTACT | GCAAAAGCAC
CGTTTTTCGTG | AAACACGCTA
TTTGTGCGAT | AGTGCCGGCC
TCACGGCCGG |
| 1901 | GTCCGAGCGC
CAGGCTCGCG | ACGCAGCAGC
TGCGTCGTCG | AAGGCTGCAA
TTCCGACGTT | CGTTGGCCAG
GCAACCGGTC | CCTGGCAGAC
GGACCGTCTG |
| 1951 | ACGCCAGCCA
TGCGGTCCGT | TGAAGCGGGT
ACTTCGCCCA | CAACTTTCAG
GTTGAAAGTC | TTGCCGGCGG
AACGGCCGCC | AGGATCACAC
TCCTAGTGTG |
| 2001 | CAAGCTGAAG
GTTTCGACTTC | ATGTACGCGG
TACATGCGCC | TACGCCAAGG
ATGCGGTTCC | CAAGACCATT
GTTCTGGTAA | ACCGAGCTGC
TGGCTCGACG |
| 2051 | TATCTGAATA
ATAGACTTAT | CATCGCGCAG
GTAGCGCGTC | CTACCAGAGT
GATGGTCTCA | AAATGAGCAA
TTTACTCGTT | ATGAATAAAT
TACTTATTTA |
| 2101 | GAGTAGATGA
CTCATCTACT | ATTTTAGCGG
TAAAATCGCC | CTAAAGGAGG
GATTTCTCTC | CGGCATGGAA
GCCGTACCTT | AATCAAGAAC
TTAGTTCTTG |
| 2151 | AACCAGGCAC
TTGGTCCGTG | CGACGCCGTG
GCTGCGGCAC | GAATGCCCCA
CTTACGGGGT | TGTGTGGAGG
ACACACCTCC | AACGGGCGGT
TTGCCCGCCA |
| 2201 | TGGCCAGGCG
ACCGGTCCGC | TAAGCGGCTG
ATTTCGCCGAC | GGTTGTCTGC
CCAACAGACG | CGGCCCTGCA
GCCGGGACGT | ATGGCACTGG
TACCGTGACC |
| 2251 | AACCCCAAG
TTGGGGGTTC | CCCGAGGAAT
GGGCTCCTTA | CGGCGTGAGC
GCCGCACTCG | GGTCGCAAAC
CCAGCGTTTG | CATCCGGCCC
GTAGGCCGGG |
| 2301 | GGTACAAATC
CCATGTTTAG | GGCGCGGCGC
CCGCGCCGCG | TGGGTGATGA
ACCCACTACT | CCTGGTGGAG
GGACCACCTC | AAGTTGAAGG
TTCAACTTCC |
| 2351 | CCGCGCAGGC
GGCGCGTCCG | CGCCAGCGG
GCGGGTCGCC | CAACGCATCG
GTTGCGTAGC | AGGCAGAAGC
TCCGTCTTCG | ACGCCCCGGT
TGCGGGGCCA |
| | | Not I
~~~~~ | | | |
| 2401 | GAATCGTGGC
CTTAGCACCG | AAGCGGCCGC
TTCGCCGGCG | TGATCGAATC
ACTAGCTTAG | CGCAAAGAAT
GCGTTTCTTA | CCCGGCAACC
GGGCCGTTGG |
| 2451 | GCCGGCAGCC
CGGCCGTCGG | GGTGCGCCGT
CCACGCGGCA | CGATTAGGAA
GCTAATCCTT | GCCGCCAAG
CGGCGGGTTC | GGCGACGAGC
CCGCTGCTCG |
| 2501 | AACCAGATTT
TTGGTCTAAA | TTTCGTTCCG
AAAGCAAGGC | ATGCTCTATG
TACGAGATAC | ACGTGGGCAC
TGCACCCGTG | CCGCGATAGT
GGCGCTATCA |

2551 CGCAGCATCA TGGACGTGGC CGTTTTCCGT CTGTCTGAAGC GTGACCGACG
 GCGTCGTAGT ACCTGCACCG GCAAAAAGGCA GACAGCTTCG CACTGGCTGC

2601 AGCTGGCGAG GTGATCCGCT ACGAGCTTCC AGACGGGCAC GTAGAGGTTT
 TCGACCGCTC CACTAGGCGA TGCTCGAAGG TCTGCCCGTG CATCTCCAAA

2651 CCGCAGGGCC GGCCGGCATG GCCAGTGTGT GGGATTACGA CCTGGTACTG
 GGGCTCCCGG CCGGCCGTAC CGGTACACACA CCCTAATGCT GGACCATGAC

2701 ATGGCGGTTT CCCATCTAAC CGAATCCATG AACCGATAACC GGAAGGGAA
 TACCGCCAAA GGGTAGATTG GCTTAGGTAC TTGGCTATGG CCCTTCCCTT

2751 GGGAGACAAG CCCGGCCGCG TGTTCCGTCC ACACGTTGCG GACGTACTCA
 CCCTCTGTTC GGGCCGGCGC ACAAGGCAGG TGTGCAACGC CTGCATGAGT

2801 AGTTCTGCCG GCGAGCCGAT GGCGGAAAGC AGAAAGACGA CCTGGTAGAA
 TCAAGACGGC CGCTCGGCTA CCGCCTTTTCG TCTTTCTGCT GGACCATCTT

2851 ACCTGCATTG GGTAAACAC CACGCACGTT GCCATGCAGC GTACGAAGAA
 TGGACGTAAG CCAATTTGTG GTGCGTGCAA CGGTACGTCG CATGCTTCTT

2901 GGCCAAGAAC GGCCGCCTGG TGACGGTATC CGAGGGTGAA GCCTTGATTA
 CCGGTTCTTG CCGGCGGACC ACTGCCATAG GCTCCCACTT CGGAACTAAT

2951 GCCGCTACAA GATCGTAAAG AGCGAAACCG GGCGGCCGGA GTACATCGAG
 CGGCGATGTT CTAGCATTTT TCGCTTTGGC CCGCCGGCCT CATGTAGCTC

3001 ATCGAGCTAG CTGATTGGAT GTACCGCGAG ATCACAGAAG GCAAGAACCC
 TAGCTCGATC GACTAACCTA CATGGCGCTC TAGTGTCTTC CGTTCTTGGG

ClaI
 ~~~~~~

3051 GGACGTGCTG ACGGTTACCC CCGATTACTT TTTGATCGAT CCCGGCATCG  
 CCTGCACGAC TGCCAAGTGG GGCTAATGAA AAAGTAGCTA GGGCCGTAGC

3101 GCCGTTTTCT CTACCGCCTG GCACGCCGCG CCGCAGGCAA GGCAGAAGCC  
 CGGCAAAAAGA GATGGCGGAC CGTGCGGCGC GCGTCCGTT CCGTCTTCGG

3151 AGATGGTTGT TCAAGACGAT CTACGAACGC AGTGGCAGCG CCGGAGAGTT  
 TCTACCAACA AGTTCTGCTA GATGCTTGCG TCACCGTCGC GGCCTCTCAA

3201 CAAGAAGTTC TGTTTCACCG TGCGCAAGCT GATCGGGTCA AATGACCTGC  
 GTTCTTCAAG ACAAAGTGGC ACGCGTTCGA CTAGCCCAGT TTAAGTGGACG

3251 CGGAGTACGA TTTGAAGGAG GAGGCGGGGC AGGCTGGCCC GATCCTAGTC  
 GCCTCATGCT AAAGTTCCTC CTCCGCCCCG TCCGACCGGG CTAGGATCAG

3301 ATGCGCTACC GCAACCTGAT CGAGGGCGAA GCATCCGCCG GTTCCTAATG  
 TACGCGATGG CGTTGGACTA GCTCCCGCTT CGTAGGCGGC CAAGGATTAC

3351 TACGGAGCAG ATGCTAGGGC AAATTGCCCT AGCAGGGGAA AAAGGTCGAA  
 ATGCCTCGTC TACGATCCCG TTTAACGGGA TCGTCCCCTT TTTCCAGCTT

3401	AAGGTCTCTT TTCCAGAGAA	TCCTGTGGAT AGGACACCTA	AGCACGTACA TCGTGCATGT	TTGGGAACCC AACCCCTTGGG	AAAGCCGTAC TTTCGGCATG
3451	ATTGGGAACC TAACCCCTTGG	GGAACCCGTA CCTTGGGCAT	CATTGGGAAC GTAACCCTTG	CCAAAGCCGT GGTTTCGGCA	ACATTGGGAA TGTAACCCTT
3501	CCGGTCACAC GGCCAGTGTG	ATGTAAGTGA TACATTCACT	CTGATATAAA GACTATATTT	AGAGAAAAAA TCTCTTTTTT	GGCGATTTTT CCGCTAAAA
3551	CCGCCTAAAA GGCGGATTTTT	CTCTTTAAAA GAGAAATTTT	CTTATTTAAAA GAATAATTTT	CTCTTAAAAAC GAGAATTTTG	CCGCCTGGCC GGCGGACCGG
3601	TGTGCATAAC ACACGTATTG	TGTCTGGCCA ACAGACCGGT	GCGCACAGCC CGCGTGTCTGG	GAAGAGCTGC CTTCTCGACG	AAAAAGCGCC TTTTTCGCGG
3651	TACCCTTCGG ATGGGAAGCC	TCGCTGCGCT AGCGACGCGA	CCCTACGCCC GGGATGCGGG	CGCCGCTTCG GCGGCGAAGC	CGTCGGCCTA GCAGCCGGAT
	NotI ~~~~~				
3701	TCGCGGCCGC AGCGCCGGCG	TGGCCGCTCA ACCGGCGAGT	AAAATGGCTG TTTTACCGAC	GCCTACGGCC CGGATGCCGG	AGGCAATCTA TCCGTTAGAT
3751	CCAGGGCGCG GGTCCCGCGC	GACAAGCCGC CTGTTCCGGC	GCCGTCGCCA CGGCAGCGGT	CTCGACCGCC GAGCTGGCGG	GGCGCCCACA CCGCGGGTGT
3801	TCAAGGCACC AGTTCCGTGG	CTGCCTCGCG GACGGAGCGC	CGTTTTCGGTG GCAAAGCCAC	ATGACGGTGA TACTGCCACT	AAACCTCTGA TTTGGAGACT
3851	CACATGCAGC GTGTACGTCG	TCCCGGAGAC AGGGCCTCTG	GGTCACAGCT CCAGTGTCTGA	TGTCTGTAAG ACAGACATTC	CGGATGCCGG GCCTACGGCC
3901	GAGCAGACAA CTCGTCTGTT	GCCCCTCAGG CGGGCAGTCC	GCGCGTCAGC CGCGCAGTCG	GGGTGTTGGC CCCACAACCG	GGGTGTCGGG CCCACAGCCC
3951	GCGCAGCCAT CGCGTCGGTA	GACCCAGTCA CTGGGTCAGT	CGTAGCGATA GCATCGCTAT	GCGGAGTGTA CGCCTCACAT	TACTGGCTTA ATGACCGAAT
	ApaLI ~~~~~				
4001	ACTATGCGGC TGATACGCCG	ATCAGAGCAG TAGTCTCGTC	ATTGTAAGTGA TAACATGACT	GAGTGCACCA CTCACGTGGT	TATGCGGTGT ATACGCCACA
4051	GAAATACCGC CTTTATGGCG	ACAGATGCGT TGTCTACGCA	AAGGAGAAAA TTCCTCTTTT	TACCGCATCA ATGGCGTAGT	GGCGCTCTTC CCGCGAGAAG
4101	CGCTTCCTCG GCGAAGGAGC	CTCACTGACT GAGTACTGTA	CGCTGCGCTC GCGACGCGAG	GGTCGTTTCGG CCAGCAAGCC	CTGCGGCGAG GACGCCGCTC
4151	CGGTATCAGC GCCATAGTCG	TCACTCAAAG AGTGAGTTTC	GCGGTAATAC CGCCATTATG	GGTTATCCAC CCAATAGGTG	AGAATCAGGG TCTTAGTCCC
4201	GATAACGCAG CTATTGCGTC	GAAAGAACAT CTTTCTTGTA	GTGAGCAAAA CACTCGTTTT	GGCCAGCAAA CCGGTCGTTT	AGGCCAGGAA TCCGGTCCTT

4251 CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT CCATAGGCTC CGCCCCCTG  
 GGCATTTTTTC CGGCGCAACG ACCGCAAAAA GGTATCCGAG GCGGGGGGAC

4301 ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA  
 TGCTCGTAGT GTTTTTTAGCT GCGAGTTCAG TCTCCACCGC TTTGGGCTGT

4351 GGACTIONATAA GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCGTGCCTC  
 CCTGATATTT CTATGGTCCG CAAAGGGGGA CCTTCGAGGG AGCACGCGAG

4401 TCCTGTTCCG ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT  
 AGGACAAGGC TGGGACGGCG AATGGCCTAT GGACAGGCGG AAAGAGGGAA

4451 CGGGAAGCGT GCGCCTTTCT CATAGCTCAC GCTGTAGGTA TCTCAGTTCG  
 GCCCTTCGCA CCGCGAAAGA GTATCGAGTG CGACATCCAT AGAGTCAAGC

ApaLI  
 ~~~~~

4501 GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT GTGCACGAAC CCCCCGTTCA
 CACATCCAGC AAGCGAGGTT CGACCCGACA CACGTGCTTG GGGGGCAAGT

4551 GCCCGACCGC TGCGCCTTAT CCGGTAACATA TCGTCTTGAG TCCAACCCGG
 CGGGCTGGCG ACGCGGAATA GGCCATTGAT AGCAGAACTC AGGTTGGGCC

4601 TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC
 ATTCTGTGCT GAATAGCGGT GACCGTCGTC GGTGACCATT GTCCTAATCG

4651 AGAGCGAGGT ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCCTAA
 TCTCGCTCCA TACATCCGCC ACGATGTCTC AAGAACTTCA CCACCGGATT

4701 CTACGGCTAC ACTAGAAGGA CAGTATTTGG TATCTGCGCT CTGCTGAAGC
 GATGCCGATG TGATCTTCCT GTCATAAACC ATAGACGCGA GACGACTTCG

4751 CAGTTACCTT CGGAAAAAGA GTTGGTAGCT CTTGATCCGG CAAACAAACC
 GTCAATGGAA GCCTTTTTTCT CAACCATCGA GAACTAGGCC GTTTGTTTGG

4801 ACCGCTGGTA GCGGTGGTTT TTTTGTTCG AAGCAGCAGA TTACGCGCAG
 TGGCGACCAT CGCCACCAA AAAACAAACG TTCGTCTGCT AATGCGCGTC

4851 AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGGTCTGACG
 TTTTTTTCCT AGAGTTCTTC TAGGAAACTA GAAAAGATGC CCCAGACTGC

4901 CTCAGTGGAA CGAAAACTCA CGTTAAGGGA TTTTGGTCAT GCATGATATA
 GAGTCACCTT GCTTTTGTAGT GCAATTCCCT AAAACCAGTA CGTACTATAT

4951 TCTCCCAATT TGTGTAGGGC TTATTATGCA CGCTTAAAAA TAATAAAAGC
 AGAGGGTTAA ACACATCCCG AATAATACGT GCGAATTTTT ATTATTTTCG

5001 AGACTTGACC TGATAGTTTG GCTGTGAGCA ATTATGTGCT TAGTGCATCT
 TCTGAACTGG ACTATCAAAC CGACACTCGT TAATACACGA ATCACGTAGA

5051 AACGCTTGAG TTAAGCCGCG CCGCGAAGCG GCGTCGGCTT GAACGAATTT
 TTGCGAACTC AATTCGGCGC GGCGCTTCGC CGCAGCCGAA CTTGCTTAAA

5101 CTAGCTAGAC ATTATTTGCC GACTACCTTG GTGATCTCGC CTTTCACGTA
 GATCGATCTG TAATAAACGG CTGATGGAAC CACTAGAGCG GAAAGTGCAT
 5151 GTGGACAAAT TCTTCCAAC TATCTGCGCG CGAGGCCAAG CGATCTTCTT
 CACCTGTTTTA AGAAGGTTGA CTAGACGCGC GCTCCGGTTC GCTAGAAGAA
 5201 CTTGTCCAAG ATAAGCCTGT CTAGCTTCAA GTATGACGGG CTGATACTGG
 GAACAGGTTC TATTCGGACA GATCGAAGTT CATACTGCCC GACTATGACC
 5251 GCCGGCAGGC GCTCCATTGC CCAGTCGGCA GCGACATCCT TCGGCGCGAT
 CGGCCGTCCG CGAGGTAACG GGTGAGCCGT CGCTGTAGGA AGCCGCGCTA
 5301 TTTGCCGGTT ACTGCGCTGT ACCAAATGCG GGACAACGTA AGCACTACAT
 AAACGGCCAA TGACGCGACA TGGTTTTACGC CCTGTTGCAT TCGTGATGTA
 5351 TTCGCTCATC GCCAGCCCAG TCGGGCGGCG AGTTCCATAG CGTTAAGGTT
 AAGCGAGTAG CGGTCGGGTC AGCCCCCGCG TCAAGGTATC GCAATTCCAA
 5401 TCATTTAGCG CCTCAAATAG ATCCTGTTCA GGAACCGGAT CAAAGAGTTC
 AGTAAATCGC GGAGTTTTATC TAGGACAAGT CTTTGGCCTA GTTTCTCAAG
 5451 CTCCGCCGCT GGACCTACCA AGGCAACGCT ATGTTCTCTT GCTTTTGTCA
 GAGGCGGCGA CCTGGATGGT TCCGTTGCGA TACAAGAGAA CGAAAACAGT
 5501 GCAAGATAGC CAGATCAATG TCGATCGTGG CTGGCTCGAA GATACCTGCA
 CGTTCTATCG GTCTAGTTAC AGCTAGCACC GACCGAGCTT CTATGGACGT
 5551 AGAATGTCAT TCGCTGCCA TTCTCCAAAT TGCAGTTCGC GCTTAGCTGG
 TCTTACAGTA ACGCGACGGT AAGAGGTTTTA ACGTCAAGCG CGAATCGACC

ApaLI
 ~~~~~

 5601 ATAACGCCAC GGAATGATGT CGTCGTGCAC AACAATGGTG ACTTCTACAG  
 TATTGCGGTG CCTTACTACA GCAGCACGTG TTGTTACCAC TGAAGATGTC  
 5651 CGCGGAGAAT CTCGCTCTCT CCAGGGGAAG CCGAAGTTTC CAAAAGGTCG  
 GCGCCTCTTA GAGCGAGAGA GGTCCCCTTC GGCTTCAAAG GTTTTCCAGC  
  

BclI  
 ~~~~~

 5701 TTGATCAAAG CTCGCCGCGT TGTTTTATCA AGCCTTACGG TCACCGTAAC
 AACTAGTTTT CAGCGGCGCA ACAAAGTAGT TCGGAATGCC AGTGGCATTG
 5751 CAGCAAATCA ATATCACTGT GTGGCTTCAG GCCGCCATCC ACTGCGGAGC
 GTCGTTTTAGT TATAGTGACA CACCGAAGTC CGGCGGTAGG TGACGCCTCG
 5801 CGTACAAATG TACGGCCAGC AACGTCGGTT CGAGATGGCG CTCGATGACG
 GCATGTTTTAC ATGCCGGTCG TTGCAGCCAA GCTCTACCGC GAGCTACTGC
 5851 CCAACTACCT CTGATAGTTG AGTCGATACT TCGGCGATCA CCGCTTCCCC
 GGTTGATGGA GACTATCAAC TCAGCTATGA AGCCGCTAGT GGCGAAGGGG
 5901 CATGATGTTT AACTTTGTTT TAGGGCGACT GCCCTGCTGC GTAACATCGT
 GTACTACAAA TTGAAACAAA ATCCCCTGTA CGGGACGACG CATTGTAGCA

5951 TGCTGCTCCA TAACATCAAA CATCGACCCA CGGCGTAACG CGCTTGCTGC
ACGACGAGGT ATTGTAGTTT GTAGCTGGGT GCCGCATTGC GCGAACGACG

6001 TTGGATGCCC GAGGCATAGA CTGTACCCCA AAAAAACAGT CATAACAAGC
AACCTACGGG CTCCGTATCT GACATGGGGT TTTTTTGTCA GTATTGTTTCG

6051 CATGAAAACC GCCACTGCGC CGTTACCACC GCTGCGTTTCG GTCAAGGTTC
GTACTTTTTGG CGGTGACGCG GCAATGGTGG CGACGCAAGC CAGTTCCAAG

6101 TGGACCAGTT GCGTGACGGC AGTTACGCTA CTTGCATTAC AGCTTACGAA
ACCTGGTCAA CGCACTGCCG TCAATGCGAT GAACGTAATG TCGAATGCTT

BclI

~~~~~

6151 CCGAACGAGG CTTATGTCCA CTGGGTTTCGT GCCCGAATTG ATCACAGGCA  
GGCTTGCTCC GAATACAGGT GACCCAAGCA CGGGCTTAAC TAGTGTCCGT

6201 GCAACGCTCT GTCATCGTTA CAATCAACAT GCTACCCTCC GCGAGATCAT  
CGTTGCGAGA CAGTAGCAAT GTTAGTTGTA CGATGGGAGG CGCTCTAGTA

6251 CCGTGTTTCA AACCCGGCAG CTTAGTTGCC GTTCTTCCGA ATAGCATCGG  
GGCACAAAGT TTGGGCCGTC GAATCAACGG CAAGAAGGCT TATCGTAGCC

6301 TAACATGAGC AAAGTCTGCC GCCTTACAAC GGCTCTCCCG CTGACGCCGT  
ATTGTACTCG TTTCAGACGG CGGAATGTTG CCGAGAGGGC GACTGCGGCA

6351 CCCGGACTGA TGGGCTGCCT GTATCGAGTG GTGATTTTGT GCCGAGCTGC  
GGGCTGACT ACCCGACGGA CATAGCTCAC CACTAAAACA CGGCTCGACG

6401 CGGTCGGGGA GCTGTTGGCT GGCTGGTGGC AGGATATATT GTGGTGTA  
GCCAGCCCCT CGACAACCGA CCGACCACCG TCCTATATAA CACCACATTT

6451 CAAATTGACG CTTAGACAAC TTAATAACAC ATTGCGGACG TTTTAAATGT  
GTTTAACTGC GAATCTGTTG AATTATTGTG TAACGCCTGC AAAAATTACA

6501 ACTGAATTAA CGCCGAATTG CTCTAGCATT CGCCATTAG GCTGCGCAAC  
TGACTTAATT GCGGCTTAAC GAGATCGTAA GCGGTAAGTC CGACGCGTTG

## PvuII

~~~~~

6551 TGTTGGGAAG GCGGATCGGT GCGGGCCTCT TCGCTATTAC GCCAGCTGGC
ACAACCCTTC CCGCTAGCCA CGCCCGGAGA AGCGATAATG CCGTTCGACCG

6601 GAAAGGGGGA TGTGCTGCAA GGCGATTAAG TTGGGTAACG CCAGGGTTTT
CTTTCCCCCT ACACGACGTT CCGCTAATTC AACCCATTGC GGTCCCAAAA

6651 CCCAGTCACG ACGTTGTAAA ACGACGGCCA GTGCCAAGCT AATTCGCTTC
GGGTCACTGC TGCAACATTT TGCTGCCGGT CACGGTTCGA TTAAGCGAAG

6701 AAGACGTGCT CAAATCACTA TTTCCACACC CCTATATTTT TATTGCACTC
TTCTGCACGA GTTTGTAGTAT AAAGGTGTGG GGATATAAAG ATAACGTGAG

6751 CCTTTTAACT GTTTTTTTATT AAAAAAATGC CCTGGAAAAT GCACTCCCTT
 GGAAAATTGA CAAAAAATAA TGTTTTTTACG GGACCTTTTA CGTGAGGGAA

6801 TTTGTGTTTG TTTTTTTGTG AAACGATGTT GTCAGGTAAT TTATTTGTCA
 AAACACAAAC AAAAAAACAC TTTGCTACAA CAGTCCATTA AATAAACAGT

6851 GTCTACTATG GTGGCCCAT TATTAATAG CAACTGTCCG TCCAATAGAC
 CAGATGATAC CACCGGGTAA TATAATTATC GTTGACAGCC AGGTTATCTG

6901 GACGTCGATT TTCTGCATTT GTTTAACCAC GTGGATTTTA TGACATTTTA
 CTGCAGCTAA AAGACGTAAA CAAATTGGTG CACCTAAAAT ACTGTAAAAT

6951 TATTAGTTAA TTTGTAAAAC CTACCCAATT AAAGACCTCA TATGTTCTAA
 ATAATCAATT AAACATTTTG GATGGGTTAA TTTCTGGAGT ATACAAGATT

7001 AGACTAATAC TTAATGATAA CAATTTTCTT TTAGTGAAGA AAGGGATAAT
 TCTGATTATG AATTACTATT GTTAAAAGAA AATCACTTCT TTCCCTATTA

7051 TAGTAAATAT GGAACAAGGG CAGAAGATTT ATTAAGCCG CGGTAAGAGA
 ATCATTTATA CCTTGTTCCC GTCTTCTAAA TAATTTTCGGC GCCATTCTCT

7101 CAACAAGTAG GTACGTGGAG TGTCTTAGGT GACTTACCCA CATAACATAA
 GTTGTTTCATC CATGCACCTC ACAGAATCCA CTGAATGGGT GTATTGTATT

7151 AGTGACATTA ACAAACATAG CTAATGCTCC TATTTGAATA GTGCATATCA
 TCACTGTAAT TGTTTGTATC GATTACGAGG ATAAACTTAT CACGTATAGT

7201 GCATACCTTA TTACATATAG ATAGGAGCAA ACTCTAGCTA GATTGTTGAG
 CGTATGGAAT AATGTATATC TATCCTCGTT TGAGATCGAT CTAACAACCTC

KpnI
~~~~~

7251 CAGATCTCGG TGACGGGCAG GACCGGACGG GGCGGTACCG GCAGGCTGAA  
 GTCTAGAGCC ACTGCCCGTC CTGGCCTGCC CCGCCATGGC CGTCCGACTT

PvuII  
~~~~~

7301 GTCCAGCTGC CAGAAACCCA CGTCATGCCA GTTCCCCTGC TTGAAGCCGG
 CAGGTCGACG GTCTTTGGGT GCAGTACGGT CAAGGGCACG AACTTCGGCC

7351 CCGCCCGCAG CATGCCGCGG GGGGCATATC CGAGCGCCTC GTGCATGCGC
 GGCGGGCGTC GTACGGCGCC CCCCCTATAG GCTCGCGGAG CACGTACGCG

7401 ACGCTCGGGT CGTTGGGCAG CCCGATGACA GCGACCACGC TCTTGAAGCC
 TGCGAGCCCA GCAACCCGTC GGGCTACTGT CGCTGGTGCG AGAACTTCGG

7451 CTGTGCCTCC AGGGACTTCA GCAGGTGGGT GTAGAGCGTG GAGCCCAGTC
 GACACGGAGG TCCCTGAAGT CGTCCACCCA CATCTCGCAC CTCGGGTCAG

SalI
~~~~~

7501 CCGTCCGCTG GTGGCGGGGG GAGACGTACA CGGTCTGACTC GGCCGTCCAG  
 GGCAGGCGAC CACCGCCCCC CTCTGCATGT GCCAGCTGAG CCGGCAGGTC

7551	TCGTAGGCGT	TGCGTGCCTT	CCAGGGGCC	GCGTAGGCGA	TGCCGGCGAC
	AGCATCCGCA	ACGCACGGAA	GGTCCCCGGG	CGCATCCGCT	ACGGCCGCTG
7601	CTCGCCGTCC	ACCTCGGCGA	CGAGCCAGGG	ATAGCGCTCC	CGCAGACGGA
	GAGCGGCAGG	TGGAGCCGCT	GCTCGGTCCC	TATCGCGAGG	GCGTCTGCCT
7651	CGAGGTTCGTC	CGTCCACTCC	TGCGGTTCCT	GCGGCTCGGT	ACGGAAGTTG
	GCTCCAGCAG	GCAGGTGAGG	ACGCCAAGGA	CGCCGAGCCA	TGCCTTCAAC
7701	ACCGTGCTTG	TCTCGATGTA	GTGGTTGACG	ATGGTGCAGA	CCGCCGGCAT
	TGGCACGAAC	AGAGCTACAT	CACCAACTGC	TACCACGTCT	GGCGGCCGTA
7751	GTCCGCCTCG	GTGGCACGGC	GGATGTCGGC	CGGGCGTCGT	TCTGGGCTCA
	CAGGCGGAGC	CACCGTGCCG	CCTACAGCCG	GCCCCGAGCA	AGACCCGAGT
7801	TGGTAGATCC	CCCGTTCGTA	AATGGTGAAA	ATTTTCAGAA	AATTGCTTTT
	ACCATCTAGG	GGGCAAGCAT	TTACCACTTT	TAAAAGTCTT	TTAACGAAAA
7851	GCTTTAAAAG	AAATGATTTA	AATTGCTGCA	ATAGAAGTAG	AATGCTTGAT
	CGAAATTTTC	TTTACTAAAT	TTAACGACGT	TATCTTCATC	TTACGAACTA
				XhoI	
				~~~~	
7901	TGCTTGAGAT	TCGTTTGTTC	TGTATATGTT	GTGTTGAGAA	TTAATTCTCG
	ACGAACTCTA	AGCAAACAAA	ACATATACAA	CACAACTCTT	AATTAAGAGC
	XhoI				
	~~				
7951	AGGTCCTCTC	CAAATGAAAT	GAACTTCCTT	ATATAGAGGA	AGGGTCTTGC
	TCCAGGAGAG	GTTTACTTTA	CTTGAAGGAA	TATATCTCCT	TCCCAGAACG
				EcoRV	
				~~~~~	
8001	GAAGGATAGT	GGGATTGTGC	GTCATCCCTT	ACGTCAGTGG	AGATATCACA
	CTTCCTATCA	CCCTAACACG	CAGTAGGGAA	TGCAGTCACC	TCTATAGTGT
8051	TCAATCCACT	TGCTTTGAAG	ACGTGGTTGG	AACGTCTTCT	TTTTCCACGA
	AGTTAGGTGA	ACGAAACTTC	TGCACCAACC	TTGCAGAAGA	AAAAGGTGCT
8101	TGCTCCTCGT	GGGTGGGGGT	CCATCTTTGG	GACCACTGTC	GGTAGAGGCA
	ACGAGGAGCA	CCCACCCCCA	GGTAGAAACC	CTGGTGACAG	CCATCTCCGT
8151	TCTTGAACGA	TAGCCTTTCC	TTTATCGCAA	TGATGGCATT	TGTAGGAGCC
	AGAACTTGCT	ATCGGAAAGG	AAATAGCGTT	ACTACCGTAA	ACATCCTCGG
8201	ACCTTCCTTT	TCCACTATCT	TCACAATAAA	GTGACAGATA	GCTGGGCAAT
	TGGAAGGAAA	AGGTGATAGA	AGTGTTATTT	CACTGTCTAT	CGACCCGTTA
8251	GGAATCCGAG	GAGGTTTCCG	GATATTACCC	TTTGTTGAAA	AGTCTCAATT
	CCTTAGGCTC	CTCCAAAGGC	CTATAATGGG	AAACAACCTT	TCAGAGTTAA
8301	GCCCTTTGGT	CTTCTGAGAC	TGTATCTTTG	ATATTTTTGG	AGTAGACAAG
	CGGGAAACCA	GAAGACTCTG	ACATAGAAAC	TATAAAAACC	TCATCTGTTC

```

8351 TGTGTCGTGC TCCACCATGT TATCACATCA ATCCACTTGC TTTGAAGACG
      ACACAGCACG AGGTGGTACA ATAGTGTAGT TAGGTGAACG AAACCTTCTGC

8401 TGGTTTGGAAAC GTCTTCTTTTT TCCACGATGC TCCTCGTGGG TGGGGGTCCA
      ACCAACCTTG CAGAAGAAAA AGGTGCTACG AGGAGCACCC ACCCCCAGGT

8451 TCTTTGGGAC CACTGTCGGC AGAGGCATCT TCAACGATGG CCTTTCCTTT
      AGAAACCCTG GTGACAGCCG TCTCCGTAGA AGTTGCTACC GGAAAGGAAA

8501 ATCGCAATGA TGGCATTGTG AGGAGCCACC TTCCTTTTCC ACTATCTTCA
      TAGCGTTACT ACCGTAAACA TCCTCGGTGG AAGGAAAAGG TGATAGAAGT

8551 CAATAAAGTG ACAGATAGCT GGGCAATGGA ATCCGAGGAG GTTTCGGAT
      GTTATTTTCA TGTCTATCGA CCCGTTACCT TAGGCTCCTC CAAAGGCCTA

8601 ATTACCCTTT GTTGAAAAGT CTCAATTGCC CTTTGGTCTT CTGAGACTGT
      TAATGGGAAA CAACTTTTCA GAGTTAACGG GAAACCAGAA GACTCTGACA

8651 ATCTTTGATA TTTTTGGAGT AGACAAGTGT GTCGTGCTCC ACCATGTTGA
      TAGAAACTAT AAAAACCTCA TCTGTTTACA CAGCACGAGG TGGTACAAC

                                                                 XmaI
                                                                 ~~~
 SmaI
                                                                 ~~~

      PstI           HindIII           PstI   SalI   XbaI   BamHI
      ~~~~~
8701 CCTGCAGGCA TGCAAGCTTG CATGCCTGCA GGTCTGACTCT AGAGGATCCC
 GGACGTCCGT ACGTTTGAAC GTACGGACGT CCAGCTGAGA TCTCCTAGGG

 KpnI
      ~~~~~
      XmaI
      ~~~~~
 SmaI SacI EcoRI
      ~~~~~

8751 CGGGTACCGA GCTCGAATTC GTAATCATGT CATAGCTGTT TCCTGTGTGA
      GCCCATGGCT CGAGCTTAAG CATTAGTACA GTATCGACAA AGGACACACT

8801 AATTGTTATC CGCTCACAAT TCCACACAAC ATACGAGCCG GAAGCATAAA
      TTAACAATAG GCGAGTGTTA AGGTGTGTTG TATGCTCGGC CTTCGTATTT

8851 GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAACTCACA TTAATTGCGT
      CACATTTTCGG ACCCCACGGA TTACTIONCTC GATTGAGTGT AATTAACGCA

                                                                 PvuII
                                                                 ~~~~~
8901 TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCTGTG CCAGCTGCAT
 ACGCGAGTGA CGGGCGAAAG GTCAGCCCTT TGGACAGCAC GGTCGACGTA

8951 TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTTCGTA TTGGAGCTTG
 ATTACTTAGC CGGTTGCGCG CCCCTCTCCG CCAAACGCAT AACCTCGAAC

```

```
9001 AGCTTGGATC AGATTGTCGT TTCCCGCCTT CAGTTTAAAC TATCAGTGTT
 TCGAACCTAG TCTAACAGCA AAGGGCGGAA GTCAAATTTG ATAGTCACAA

9051 TGACAGGATA TATTGGCGGG TAAACCTAAG AGAAAAGAGC GTTTATTAGA
 ACTGTCCTAT ATAACCGCCC ATTTGGATTG TCTTTTCTCG CAAATAATCT

9101 ATAATCGGAT ATTTAAAAGG GCGTGAAAAG GTTTATCCGT TCGTCCATTT
 TATTAGCCTA TAAATTTTCC CGCACTTTTC CAAATAGGCA AGCAGGTAAA

9151 GTATGTGCAT GCCAACCACA GGGTTCCCCT CGGGATCAA
 CATAACGTA CGGTTGGTGT CCCAAGGGGA GCCCTAGTT
```