```
FEATURES             Location/Qualifiers
     misc_feature    complement(4..34)
                     /note=loxP
     misc_feature    complement(244..717)
                     /note=sopC
     misc_feature    complement(790..1761)
                     /note=sopB
     misc_feature    complement(1761..2927)
                     /note=sopA
     misc_feature    complement(3515..4270)
                     /note=repE
     misc_feature    complement(4599..4665)
                     /note=oriS
     CDS             5610..6269
                     /note=Cm
     misc_feature    6272..12252
                     /mobile_element=transposon:Tn1000
     repeat_region   6272..6306
                     /rpt_type=INVERTED
     CDS             complement(6384..8480)
                     /note=unnamed protein product
                     /note=ORF
                     /codon_start=1
                     /transl_table=11
                     /protein_id=CAA42758.1
                     /db_xref=GI:43087
                     /db_xref=GOA:Q00042
                     /db_xref=UniProtKB/Swiss-Prot:Q00042
                     /translation=MELNITSKSNPFGDTTAENDKKMLSNAFIETADFRTLIETDDRT
                     IVVGRRGTGKSALFIQLNEHWKKDKKILILSFSPDDSQIIGFRSMLKPFTGSFNLAI
                     ATRLLWRYAMLMEIASYISSHYKLSSQISSETLLNEHLKKWNSAQGDILRKCRLVA
                     YLDENNPEESIGDLQFNLNISEIENNIVSLLERSDRKVVILMDKLDEAYEPDNIGIG
                     IAGLAYASIELNQKAKCIRPIIFLRDNIFRSLSKEDPDYSRNIEGQVIRLHWDWAQL
                     MLSAKRMKVAFKLDIEKDQRVWDRCTADDLKGRNGFKRCLQFTLYRPRDLLSLLN
                     FSAFRENRETIINTDLEYAAKSISMARLEDLWKEYQKIFPSIQVITSAFRSIEPELTV
                     YTCLKKIEASFELIEENGDPKITSEIQLLKASGILQSLYSVGFVGIRDKNTSSYSFCH
                     DGRTPDKGFESNEKLLIHPCYWLGLNLNRNALAPEEAEEINDEYDINIISDNSAIRM
                     TIGQITTHLDQIPIGNEGATEFEQWCLDALRIVFASHLTDIKSHPNGNAVQRRDIIG
                     NGGKSDFWKRVLEDYKTRQVVFDAKNFEELGPSEYRQLQSYLTGPYGKLGFIINR
                     EVLKSGKDLDWTKEMYQSHNSLIIKLPAKYISKLLQKLRNPEKHDAIDRQMGKLLT
                     ETSYMAIKSTQKKRRK
     misc_feature    complement(8496..9047)
                     /gene=tnpR
     CDS             complement(8496..9047)
                     /gene=tnpR
                     /codon_start=1
                     /transl_table=11
                     /product=resolvase
```

/protein_id=CAA42759.1
/db_xref=GI:43088
/db_xref=GOA:P03012
/db_xref=InterPro:IPR006118
/db_xref=InterPro:IPR006119
/db_xref=InterPro:IPR006120
/db_xref=InterPro:IPR012287
/db_xref=PDB:1GDR
/db_xref=PDB:1GDT
/db_xref=PDB:1GHT
/db_xref=PDB:1HX7
/db_xref=PDB:1RES
/db_xref=PDB:1RET
/db_xref=PDB:1ZR2
/db_xref=PDB:1ZR4
/db_xref=PDB:2GM4
/db_xref=PDB:2GM5
/db_xref=PDB:2RSL
/db_xref=UniProtKB/Swiss-Prot:P03012
/translation=MRLFGYARVSTSQQSLDIQVRALKDAGVKANRIFTDKASGSSSD
RKGLDLLRMKVEEGDVILVKKLDRLGRDTADMIQLIKEFDAQGVSIRFIDDGISTD
MGKMVVTILSAVAQAERQRILERTNEGRQEAMAKGVVFGRKRKIDRDAVLNMWC
ASHISKTMNIARSTVYKVINESN
CDS 9211..12219
/gene=tnpA
/codon_start=1
/transl_table=11
/product=transposase
/protein_id=CAA42760.1
/db_xref=GI:43089
/db_xref=GOA:Q00037
/db_xref=InterPro:IPR002513
/db_xref=UniProtKB/Swiss-Prot:Q00037
/translation=MPVDFLTSDQKQNYGCYAAEPNDVQLARYFHLDERDLAFINQRR
GKHNRLGIALQLTTARFLGTFLTDLTQVLPGVQHFVAVQLNIHRPEVLSRYAERDTT
REHTALIKEYYGYHEFGDFPWSFRLKRLLYTRAWLSNERPGLMFDFATAWLLQNKV
PGATTLVRLISEIRERANQRLWKKLAALPNKWQAAQVMELLVIPEGQRVSALEQLK
PVTVSGPAFNEALERYIRLRSLEFSRLNFSGLPAIQLRNLARYAGMASVKYIARMPC
RKLAVLTAFVKAQEITALDDAVDVLDMLILDIIREAKKTGQKKRLRTLKDLDQAALL
ARACALLLDDNTDVPDLRQVIFKCVPKNRLAESVSKVNELARPQNNNFHDEMVEC
VKRFLPAVLRDLHFRAAPAGEHVLAAIHYLAELNGSKKRILDDAPEHIITGPWKRLV
DAEGRIQRAGYSLCLLERLQDALRRRDIWLENSDRWGDPREKLLQGEEWQTQRI
ALGHPVDGRKGVQQLAIQLDETWKAVASRFEKNAEVHICNEGKYPSLTISCLEKQ
PSLLRLNNRIKQLLPPVDLTELLLEIDAQTGFTHEFAHVSESGARAQDLHISLCAVL
AEACNIGLEPLIKHNIPALTRHRLSWVKQNYLRAETLVSANARLVDFQSTLELAGRV
GGEVASADGMRFVTPVKTINSGSNRKYFGSGRGITWYNFVSDQYSGFHGIVVPG
SIFVLEGLLEQQTGLNPVEIMTDTAGSSDIIFGLFWLLGYQFSPRLADAGEAVFWR
KSANYGVLDKLARGYADLSKAESQWDEMMRTAGSLKLGTIHASELIRSLLKSSRP
AQAIMEVGRVNKTLYLLNYIDDEDYRRRILTQLNRGEGRHAVARAICYGQRGEIRK

```
                REGQEDQLGALGLVTNAVVLWNTLYMEEALSWMRRNGEEIIDEDIARLSPLMHGH
                LGHYTFTLPEDILKGELRALNLNINNELSP
     misc_feature    9211..12219
                /gene=tnpA
     repeat_region   12218..12252
                /rpt_type=INVERTED
     misc_feature    complement(12382..12609)
                /note=BGH pA
     CDS             complement(12631..13665)
                /note=HygB
     misc_feature    complement(12634..15017)
                /note=pL_UGIH derived
     misc_feature    complement(13689..14264)
                /note=IRES
     CDS             complement(14298..15017)
                /note=GFP
     misc_feature    complement(15018..15023)
                /note=Kozak sequence
     misc_feature    complement(15024..16288)
                /note=pTracer-Er/V5-His A derived
     promoter        complement(15102..16288)
                /note=EF-1a Promoter
     misc_feature    complement(16292..16322)
                /note=loxP
BASE COUNT      4292 A      3995 C      3799 G      4241 T        0 OTHER
ORIGIN      ?
        1 ATAACTTCGT ATAGCATACA TTATACGAAG TTATTGCAAC GTGCCGGCAC GGCCTGG
       61 ACCAGGTATT TTGTCCACAT AACCGTGCGC AAAATGTTGT GGATAAGCAG GACACA
      121 CAATCCACAG CAGGCATACA ACCGCACACC GAGGTTACTC CGTTCTACAG GTTACG
      181 CATGTCAATA CTTGCCCTTG ACAGGCATTG ATGGAATCGT AGTCTCACGC TGATAGT
      241 ATCGACAATA CAAGTGGGAC CGTGGTCCCA GACCGATAAT CAGACCGACA ACACG
      301 GATCGTGGTC CCAGACTAAT AATCAGACCG ACGATACGAG TGGGACCGTG GTCCC
      361 AATAATCAGA CCGACGATAC GAGTGGGACC GTGGTTCCAG ACTAATAATC AGACCG
      421 TACGAGTGGG ACCGTGGTCC CAGACTAATA ATCAGACCGA CGATACGAGT GGGAC
      481 TCCCAGACTA ATAATCAGAC CGACGATACG AGTGGGACCG TGGTCCCAGT CTGATT
      541 GACCGACGAT ACGAGTGGGA CCGTGGTCCC AGACTAATAA TCAGACCGAC GATAC
      601 GGACCGTGGT CCCAGACTAA TAATCAGACC GACGATACGA GTGGGACCGT GGTCC
      661 TGATTATCAG ACCGACGATA CAAGTGGAAC AGTGGGCCCA GAGAGAATAT TCAGG
      721 TATGCTTTCT GGCCTGTAAC AAAGGACATT AAGTAAAGAC AGATAAACGT AGACTAA
      781 GTGGTCGCAT CAGGGTGCTG GCTTTTCAAG TTCCTTAAGA ATGGCCTCAA TTTTCTC
      841 ACACTCAGTT GGAACACGAG ACCTGTCCAG GTTAAGCACC ATTTTATCGC CCTTATA
      901 TACTGTCGCT CCAGGAGCAA ACTGATGTCG TGAGCTTAAA CTAGTTCTTG ATGCAG
      961 CGTTTTAAGC ACAGAAGTTA AAAGAGTGAT AACTTCTTCA GCTTCAAATA TCACCCC
     1021 TTTTTTCTGC TCATGAAGGT TAGATGCCTG CTGCTTAAGT AATTCCTCTT TATCTGTA
     1081 GGCTTTTTGA AGTGCATCAC CTGACCGGGC AGATAGTTCA CCGGGGTGAG AAAAA
     1141 AACAACTGAT TTAGGCAATT TGGCGGTGTT GATACAGCGG GTAATAATCT TACGTG
     1201 ATTTTCCGCA TCAGCCAGCG CAGAAATATT TCCAGCAAAT TCATTCTGCA ATCGGC
     1261 ATAACGCTGA CCACGTTCAT AAGCACTTGT TGGGCGATAA TCGTTACCCA ATCTGG
     1321 TGCAGCCATC TGCTCATCAT CCAGCTCGCC AACCAGAACA CGATAATCAC TTTCGG
```

1381 TGCAGCAGCT TTACGACGGC GACTCCCATC GGCAATTTCT ATGACACCAG ATACTC
1441 ACCGAACGCC GGTGTCTGTT GACCAGTCAG TAGAAAAGAA GGGATGAGAT CATCC
1501 GTCCTCAGTA AGCAGCTCCT GGTCACGTTC ATTACCTGAC CATACCCGAG AGGTCT
1561 AACACTATCA CCCCGGAGCA CTTCAAGAGT AAACTTCACA TCCCGACCAC ATACAC
1621 AGTAATGGCA TTACCGCGAG CCATTACTCC TACGCGCGCA ATTAACGAAT CCACCA
1681 GGCAGCTGGT GTCGATAACG AAGTATCTTC AACCGGTTGA GTATTGAGCG TATGTT
1741 AATAACAGGC GCACGCTTCA TTATCTAATC TCCCAGCGTG GTTTAATCAG ACGATC
1801 ATTTCATTGC AGACAGGTTC CCAAATAGAA AGAGCATTTC TCCAGGCACC AGTTGA
1861 CGTTGATCAA TGGCCTGTTC AAAAACAGTT CTCATCCGGA TCTGACCTTT ACCAAC
1921 TCCGTTTCAC GTACAACATT TTTTAGAACC ATGCTTCCCC AGGCATCCCG AATTTGC
1981 TCCATCCACG GGGACTGAGA GCCATTACTA TTGCTGTATT TGGTAAGCAA AATACG
2041 TCAGGCTCGA ACCCTTTAAG ATCAACGTTC TTGAGCAGAT CACGAAGCAT ATCGAA
2101 TGCAGTGCGG AGGTGTAGTC AAACAACTCA GCAGGCGTGG GAACAATCAG CACA
2161 GCACATACGA CATTAATCGT GCCGATACCC AGGTTAGGCG CGCTGTCAAT AACTAT
2221 TCATAGTCAT GAGCAACAGT TTCAATGGCC AGTCGGAGCA TCAGGTGTGG ATCGG
2281 AGTTTACCTT CATCAAATTT GCCCATTAAC TCAGTTTCAA TACGGTGCAG AGCCAGA
2341 GAAGGAATAA TGTCAAGCCC CGGCCAGCAA GTGGGCTTTA TTGCATAAGT GACAT
2401 TTTTCCCCAA GATAGAAAGG CAGGAGAGTG TCTTCTGCAT GAATATGAAG ATCTGG
2461 CATCCGTGAT ACATTGAGGC TGTTCCCTGG GGGTCGTTAC CTTCCACGAG CAAAA
2521 AGCCCCTTCA GAGCCAGATC CTGAGCAAGA TGAACAGAAA CTGAGGTTTT GTAAA
2581 CCTTTATGGG CAGCAACCCC GATCACCGGT GGAAATACGT CTTCAGCACG TCGCA
2641 GTACCAAACA CATCACGCAT ATGATTAATT TGTTCAATTG TATAACCAAC ACGTTGC
2701 ACCCGTCCTC GAATTTCCAT ATCCGGGTGC GGTAGTCGCC CTGCTTTCTC GGCATC
2761 ATAGCCTGAG AAGAAACCCC AACTAAATCC GCTGCTTCAC CTATTCTCCA GCGCCG
2821 ATTTTCCTCG CTTCCGGGCT GTCATCATTA AACTGTGCAA TGGCGATAGC CTTCGTC
2881 TCATGACCAG CGTTTATGCA CTGGTTAAGT GTTTCCATGA GTTTCATTCT GAACATC
2941 TAATCATTGC TTTGCGTTTT TTTATTAAAT CTTGCAATTT ACTGCAAAGC AACAACAA
3001 TCGCAAAGTC ATCAAAAAAC CGCAAAGTTG TTTAAAATAA GAGCAACACT ACAAAA
3061 ATAAGAAGAG CACATACCTC AGTCACTTAT TATCACTAGC GCTCGCCGCA GCCGTG
3121 CGAGCATAGC GAGCGAACTG GCGAGGAAGC AAAGAAGAAC TGTTCTGTCA GATAC
3181 ACGCTCAGCG CAAGAAGAAA TATCCACCGT GGGAAAAACT CCAGGTAGAG GTACA
3241 GGATAGCCAA TTCAGAGTAA TAAACTGTGA TAATCAACCC TCATCAATGA TGACGA
3301 ACCCCCGATA TCAGGTCACA TGACGAAGGG AAAGAGAAGG AAATCAACTG TGACA
3361 CCCTCAAATT TGGCTTCCTT AAAAATTACA GTTCAAAAAG TATGAGAAAA TCCATGC
3421 CTGAAGGAAA CAGCAAAACT GTGACAAATT ACCCTCAGTA GGTCAGAACA AATGT
3481 ACCACCCTCA AATCTGTGAC AGATAACCCT CAGACTATCC TGTCGTCATG GAAGTG
3541 CGCGGAAGGA AAATACGATA TGAGTCGTCT GGCGGCCTTT CTTTTCTCA ATGTAT
3601 GGCGCATTGG AGTTCTGCTG TTGATCTCAT TAACACAGAC CTGCAGGAAG CGGCG
3661 AGTCAGGCAT ACGCTGGTAA CTTTGAGGCA GCTGGTAACG CTCTATGATC CAGTC
3721 TCAGAGAGAC GATGCCTGAG CCATCCGGCT TACGATACTG ACACAGGGAT TCGTA
3781 GCATGGCATA CGGATTGGTG ATTTCTTTTG TTTCACTAAG CCGAAACTGC GTAAAC
3841 TCTGTAACCC GATAAAGAAG GGAATGAGAT ATGGGTTGAT ATGTACACTG TAAAGC
3901 TGGATGGACT GTGCGCACGT TTGATAAACC AAGGAAAAGA TTCATAGCCT TTTTCA
3961 CGGCATCCTC TTCAGGGCGA TAAAAAACCA CTTCCTTCCC CGCGAAACTC TTCAAT
4021 GCCGTATATC CTTACTGGCT TCCGCAGAGG TCAATCCGAA TATTTCAGCA TATTTAG
4081 CATGGATCTC GCAGATACCG TCATGTTCCT GTAGGGTGCC ATCAGATTTT CTGATCT
4141 CAACGAACAG ATACAGCATA CGTTTTTGAT CCCGGGAGAG ACTATATGCC GCCTCA
4201 GGTCGTTTGA CTGGACGATT CGCGGGCTAT TTTTACGTTT CTTGTGATTG ATAACCC
4261 TTTCCGCCAT GACAGATCCA TGTGAAGTGT GACAAGTTTT TAGATTGTCA CACTAA

4321 AAAAGAGTCA ATAAGCAGGG ATAACTTTGT GAAAAAACAG CTTCTTCTGA GGGCA
4381 TCACAGGGTT AAGGGCAATT TGTCACAGAC AGGACTGTCA TTTGAGGGTG ATTTG
4441 CTGAAAGGGC AATTTGTCAC AACACCTTCT CTAGAACCAG CATGGATAAA GGCCTA
4501 GCGCTCTAAA AAAGAAGATC TAAAAACTAT AAAAAAAATA ATTATAAAAA TATCCCC
4561 GATAAGTGGA TAACCCCAAG GGAAGTTTTT TCAGGCATCG TGTGTAAGCA GAATAT
4621 GTGCTGTTCC CTGGTGCTTC CTCGCTCACT CGAGGGCTTC GCCCTGTCGC TCGAC
4681 CGAGCACTAC TGGCTGTAAA AGGACAGACC ACATCATGGT TCTGTGTTCA TTAGGT
4741 CTGTCCATTG CTGACATAAT CCGCTCCACT TCAACGTAAC ACCGCACGAA GATTTC
4801 GTTCCTGAAG GCATATTCAA ATCGTTTTCG TTACCGCTTG CAGGCATCAT GACAGA
4861 TACTTCCTAT AAACGCTACA CAGGCTCCTG AGATTAATAA TGCGGATCTC TACGATA
4921 GGAGATTTTC CCGACTGTTT CGTTCGCTTC TCAGTGGATA ACAGCCAGCT TCTCTG
4981 ACAGACAAAA ACAGCATATC CACTCAGTTC CACATTTCCA TATAAAGGCC AAGGCA
5041 TTCTCAGGAT AATTGTTTCA GCATCGCAAC CGCATCAGAC TCCGGCATCG CAAACT
5101 CCGGTGCCGG GCAGCCACAT CCAGCGCAAA AACCTTCGTG TAGACTTCCG TTGAA
5161 GGACTTATGT CCCATCAGGC TTTGCAGAAC TTTCAGCGGT ATACCGGCAT ACAGCA
5221 CATCGCATAG GAATGGCGGA ACGTATGTGG TGTGACCGGA ACAGAGAACG TCACA
5281 AGCAGCAGCG GCGGCAACCG CCTCCCCAAT CCAGGTCCTG ACCGTTCTGT CCGTC
5341 CCAGATCCGC GCTTTCTCTG TCCTTCCTGT GCGACGGTTA CGCCGCTCCA TGAGCT
5401 GCGAATAAAT ACCTGTGACG GAAGATCACT TCGCAGAATA AATAAATCCT GGTGTC
5461 TTGATACCGG GAAGCCCTGG GCCAACTTTT GGCGAAAATG AGACGTTGAT CGGCA
5521 GAGGTTCCAA CTTTCACCAT AATGAAATAA GATCACTACC GGGCGTATTT TTTGAGT
5581 CGAGATTTTC AGGAGCTAAG GAAGCTAAAA TGGAGAAAAA AATCACTGGA TATACC
5641 TTGATATATC CCAATGGCAT CGTAAAGAAC ATTTTGAGGC ATTTCAGTCA GTTGCTC
5701 GTACCTATAA CCAGACCGTT CAGCTGGATA TTACGGCCTT TTTAAAGACC GTAAAG
5761 ATAAGCACAA GTTTTATCCG GCCTTTATTC ACATTCTTGC CCGCCTGATG AATGCTC
5821 CGGAATTCCG TATGGCAATG AAAGACGGTG AGCTGGTGAT ATGGGATAGT GTTCA
5881 GTTACACCGT TTTCCATGAG CAAACTGAAA CGTTTTCATC GCTCTGGAGT GAATAC
5941 ACGATTTCCG GCAGTTTCTA CACATATATT CGCAAGATGT GGCGTGTTAC GGTGAA
6001 TGGCCTATTT CCCTAAAGGG TTTATTGAGA ATATGTTTTT CGTCTCAGCC AATCCCT
6061 TGAGTTTCAC CAGTTTTGAT TTAAACGTGG CCAATATGGA CAACTTCTTC GCCCCCC
6121 TCACCATGGG CAAATATTAT ACGCAAGGCG ACAAGGTGCT GATGCCGCTG GCGAT
6181 TTCATCATGC CGTTTGTGAT GGCTTCCATG TCGGCAGAAT GCTTAATGAA TTACAAC
6241 ACTGCGATGA GTGGCAGGGC GGGGCGTAAT TGGGGTTTGA GGGCCAATGG AACC
6301 TACGTTTATG GTATAACTTA TTGATTATAA TACAGTATAC AAAGGTCCCT TTTCAGGG
6361 CCTTTCGTAT GTTGTAGCTG ATTTTATTTT CTTCTTTTCT TTTGGGTGCT TTTAATAGC
6421 ATATATGATG TTTCATATAA AGTTAAAAGT TTCCCCATCT GTCTATCTAT AGCGTCATC
6481 TTTTCAGGAT TTCTTAATTT CTGCAGCAGT TTAGATATAT ATTTCGCAGG TAACTTTAT
6541 ATTAATGAAT TATGGGATTG ATACATTTCC TTTGTCCAAT CTAAATCTTT TCCTGATTT
6601 AATACCTCAC TCTCATCTCT ATTAATTATA AAACCGAGCT TTCCATACGG ACCTGTTA
6661 TATGATTGTA ATTGTCTATA CTCAGATGGA CCTAATTCTT CAAAATTTTT AGCATCAA
6721 ACAACTTGTC TTGTTTTATA GTCCTCCAAT ACTCGTTTCC AAAAATCAGA TTTGCCAC
6781 TTGGTGCCTA TAATATCTCG TCTCTGAACT GCGTTACCAT TTGGATGGGA CTTGATG
6841 GTTAGGTGGG ATGCAAATAC TATTCTTAGT GCGTCTAAAC ACCATTGTTC AAATTCA
6901 GCACCTTCAT TTCCTATCGG TATCTGATCT AGATGAGTAG TTATTTGACC TATTGTTT
6961 TTTCTAATGG CTGAATTATC TGAAATAATA TTAATATCAT ATTCGTCATT TATCTCCTC
7021 GCTTCCTCTG GAGCAAGTGC ATTACGATTT AGATTCAAAC CAAGCCAGTA ACATGG
7081 ATTAATAATT TCTCATTACT TTCAAAACCT TTATCTGGAG TTCGCCCGTC ATGGCAAA
7141 GAATAGGATG AGGTGTTTTT ATCACGAATA CCAACAAATC CTACACTATA CAAACTT
7201 AGAATTCCAC TTGCCTTTAA CAACTGTATT TCAGACGTTA TTTTAGGATC TCCATTTT

7261 TCGATTAATT CGAAAGATGC TTCTATTTTT TTTAAGCACG TATAAACTGT TAATTCAG
7321 TCAATGCTAC GAAATGCACT AGTTATAACC TGTATTGAAG GAAAGATCTT CTGATAC
7381 TTCCAGAGAT CTTCAAGTCT GGCCATGGAA ATTGACTTGG CTGCATATTC TAGGTC
7441 TTTATGATAG TTTCTCTATT CTCTCTGAAT GCGGAAAAAA AAGCTTCATT CAACAATG
7501 AGTAAATCCC TGGGCCGGTA AAGGGTAAAT TGCAAACATC GCTTAAAACC ATTCCT
7561 TTAAGATCAT CCGCTGTGCA TCTATCCCAA ACTCGTTGAT CTTTCTCAAT ATCTAGCT
7621 AATGCTACTT TCATTCTTTT AGCTGACAGC ATTAGGAGTT GTGCCCAGTC CCAATGC
7681 CTTATGACTT GACCCTCTAT ATTTCTCGAG TAATCAGGAT CTTCCTTTGA TAGCGAC
7741 AATATATTAT CCCTTAAAAA AATTATTGGA CGAATGCATT TTGCTTTTTG ATTTAATTC
7801 ATAGATGCAT ATGCTAGACC TGCAATGATT CCAATTCCTA TATTATCCGG TTCATACG
7861 TCATCTAGCT TATCCATTAA TATGACAACT TTCCTGTCTG AGCGTTCAAG AAGTGAT
7921 ATATTATTTT CTATTTCTGA GATATTCAAA TTGAATTGAA GATCACCAAT TGATTCTTC
7981 GGGTTATTTT CATCTAAATA CTCCTTTGCG ACAAGCCTAC ACTTTCTTAA AATGTCAC
8041 TGTGCAGAAT TCCATTTTTT CAAATGTTCA TTCAACAATG TTTCTGATGA TATTTGAG
8101 GACAATTTGT AATGAGATGA TATATATGAT GCTATCTCCA TTAGCATAGC GTATCGC
8161 AATAGTCTTG TTGCTGCCCT TGCTAAATTA AATGATCCTG TAAATGGTTT CAACATTC
8221 CTGAAACCAA TAATTTGAGA ATCGTCTGGT GAGAAACTCA GGATTAATAT TTTTTTG
8281 TTCTTCCAAT GCTCATTTAG CTGAATAAAT AAAGCACTTT TACCTGTCCC TCGTCTAC
8341 ACAACAATGG TCCTGTCATC AGTTTCAATT AGAGTCCTAA AGTCAGCAGT TTCAATC
8401 GCATTACTCA ACATCTTTTT ATCATTTTCT GCTGTCGTAT CACCAAACGG ATTAGACT
8461 GAAGTAATAT TCAATTCCAT ATTCAACCTT TTATGTTAGT TGCTTTCATT TATTACTTTA
8521 TATACTGTTG AACGAGCAAT ATTCATTGTT TTTGATATAT GTGAGGCACC TAACCCCT
8581 TGCCACATAT TTAATACTGC ATCTCTATCT ATTTTTCTTT TTCTACCAAA AACAACTCC
8641 TTTGCCATTG CCTCTTGGCG ACCTTCATTG GTACGCTCTA GTATTCTCTG TCGTTCTC
8701 TGGGCCACTG CAGATAGAAT AGTGACAACC ATTTTACCCA TCTCCCCATC GGTACT
8761 CCGTCATCAA TAAACCGAAT GGATACACCT TGGGCGTCAA ACTCTTTTAT TAACTGC
8821 ATGTCAGCAG TATCGCGCCC AAGGCGGTCA AGTTTCTTCA CCAAGATGAC GTCAC
8881 TCCACCTTCA TCCTCAGCAA GTCCAGCCCT TTCCGATCGC TTGAACTGCC CGATGC
8941 TCAGTAAAGA TGCGATTTGC TTTCACGCCT GCGTCTTTGA GTGCCCGAAC CTGAAT
9001 AGAGATTGCT GGCTGGTTGA TACCCGTGCG TAACCAAAAA GTCGCATAAA AATGTA
9061 AAATCAAATA TCGGACAAGC AGTGTCTGTT ATAACAAAAA ATCGATTTAA TAGACAC
9121 AACAGCACTG TTTTTATGTG TGCGATAATT TATAATATTT CGGACGGTTG CAAAATTC
9181 AATGAATAAC TGTCAGGTAG GGAGCCTGAT ATGCCTGTCG ACTTTCTAAC CTCTGA
9241 AAACAGAATT ACGGTTGTTA TGCTGCAGAA CCTAATGACG TGCAACTGGC GCGCTA
9301 CATCTTGATG AACGGGATCT GGCCTTCATT AACCAACGAC GGGGCAAACA TAATAC
9361 GGCATTGCGC TTCAGCTCAC CACAGCCCGT TTTCTGGGAA CATTTCTGAC GGATTT
9421 CAGGTTCTGC CTGGTGTTCA ACATTTTGTC GCGGTACAGC TTAATATCCA CCGTCCA
9481 GTTCTCTCCC GCTATGCTGA ACGGGACACT ACCCTTAGAG AACATACTGC ATTAAT
9541 GAATATTACG GCTATCATGA ATTTGGTGAT TTTCCATGGT CTTTCCGCCT GAAGCGT
9601 CTATATACCC GGGCGTGGCT CAGTAATGAG CGACCGGGTC TGATGTTTGA TTTTGC
9661 GCATGGTTGC TTCAAAATAA GGTATTACTG CCCGGAGCAA CCACACTAGT ACGTCT
9721 AGTGAAATTC GTGAAAGGGC AAATCAGCGG CTGTGGAAAA AGCTGGCCGC ACTG
9781 AAATGGCAGG CAGCTCAAGT GATGGAGCTT CTGGTCATTC CGGAAGGTCA GCGTC
9841 GCACTGGAAC AGTTGAAAAA AGGCCCTGTT ACAGTCAGTG GACCTGCGTT TAATG
9901 CTGGAACGAT ATATCCGATT ACGAAGTCTT GAGTTTTCCC GACTGAACTT TTCCGGT
9961 CCTGCCATTC AACTGCGTAA TCTGGCTCGT TATGCTGGCA TGGCGTCGGT AAAATA
10021 GCTCGAATGC CACAGCAGAG AAAGCTTGCT GTACTTACTG CATTCGTTAA AGCAC
10081 ATAACGGCAT TAGACGATGC CGTTGATGTG CTTGATATGC TAATTCTGGA CATTATC
10141 GAAGCAAAGA AAACCGGGCA AAAAAAAAGA CTCAGGACAC TGAAAGATCT TGAT

10201 GCATTGTTAC TGGCGCGGGC ATGTGCATTG TTGCTGGATG ATAATACAGA TGTCC
10261 CTCAGGCAGG TTATCTTCAA GTGCGTACCC AAAAACAGAC TGGCAGAATC TGTAA
10321 GTTAATGAAC TTGCTCGTCC ACAGAACAAT AATTTCCATG ACGAAATGGT TGAGCA
10381 GGTCGGGTTA AACGTTTTCT TCCGGCGGTG TTGCGGGACC TGCATTTCCG TGCGC
10441 GCAGGTGAAC ATGTACTGGC TGCGATTCAT TATCTGGCAG AACTGAATGG TTCGA
10501 CGCATCCTTG ATGATGCGCC TGAACATATT ATCACCGGTC CTGGAAACG CCTCG
10561 GATGCGGAGG GACGGATACA GCGTGCAGGT TATTCACTAT GTTTGCTGGA ACGC
10621 GATGCACTGC GCCGCCGGGA CATCTGGCTT GAAAACAGTG ATCGCTGGGG AGAT
10681 GAGAAGTTGT TGCAAGGTGA AGAGTGGCAG ACTCAGCGTA TTCCTGTCTG TCGG
10741 GGACATCCTG TCGATGGACG TAAAGGTGTG CAACAACTGG CTATTCAGCT GGATG
10801 TGGAAAGCCG TGGCATCACG ATTTGAAAAG AATGCGGAAG TTCATATCTG TAATG
10861 AAATATCCAT CCCTGACTAT CAGTTGTCTG GAGAAACAGG AAGAGCCACC ATCAT
10921 CGTCTAAATA ATCGGATCAA ACAGCTACTC CCACCGGTAG ATTTAACGGA ACTGTT
10981 GAGATAGATG CCCAGACAGG ATTTACACAT GAGTTTGCGC ATGTCAGTGA ATCTG
11041 CGAGCGCAAG ATTTGCACAT CAGTTTATGT GCGGTATTGA TGGCTGAAGC CTGTA
11101 GGACTGGAAC CGCTGATAAA GCACAATATA CCAGCACTGA CCCGCCATCG GCTCA
11161 GTGAAACAGA ATTACCTTCG TGCAGAAACG CTGGTCAGCG CCAATGCCCG CCTGC
11221 TTTCAGTCCA CACTGGAGCT TGCTGGTCGT TGGGGAGGTG GAGAAGTGGC ATCA
11281 GGCATGCGCT TTGTCACACC AGTGAAGACC ATCAACTCAG GATCTAACAG AAAAT
11341 GGTTCTGGAC GAGGCATCAC CTGGTATAAC TTCGTATCTG ATCAGTACTC TGGGTT
11401 GGCATTGTGG TACCCGGTAC ATTACGGGAT TCGATTTTTG TACTGGAAGG ACTTCT
11461 CAGCAGACAG GGCTGAATCC AGTTGAAATC ATGACAGACA CTGCGGGTAG CAGC
11521 ATTTTCGGTC TGTTCTGGCT ACTGGGGTAT CAGTTTTCCC CCCGGCTTGC CGATGC
11581 GAGGCGGTGT TCTGGCGGGT GAATAAATCG GCAAACTACG GAGTACTGGA TAAG
11641 CGAGGATATG CAGATCTGTC AAAAGCGGAG AGTCAGTGGG ATGAGATGAT GCGA
11701 GGTTCGCTGA AGTTGGGTAC AATTCATGCG TCAGAACTCA TTCGCTCTTT ACTGAA
11761 TCGCGCCCAT CAGGACTGGC ACAGGCAATT ATGGAAGTAG GGCGTGTAAA CAAA
11821 TATCTCCTCA ATTATATTGA TGATGAAGAT TATCGTCGGC GGATCCTGAC GCAGCT
11881 CGAGGAGAAG GCCGCCACGC TGTGGCTCGT GCGATTTGTT ACGGACAACG CGGA
11941 AGAAAGCGTT ATCGTGAAGG GCAGGAAGAT CAACTGGGGG CGCTGGGCCT GGT
12001 GCAGTGGTTC TGTGGAACAC GCTCTATATG GAGGAAGCGT TGAGCTGGAT GCGC
12061 GGCGAAGAAA TTATAGATGA AGATATCGCT CGGCTATCTC CCCTGATGCA CGGGC
12121 AATATGTTGG GCCATTATAC ATTCACGTTG CCAGAGGATA TTTTAAAAGG GGAACT
12181 GCTCTAAATT TAAATATAAA CAACGAATTA TCTCCTTAAC GTACGTTTTC GTTCCATT
12241 CCCTCAAACC CCAATTTTTT TAAGGCAGTT ATTGGTGCCC TTAAACGCCT GGTTGC
12301 CCTGAATAAG TGATAATAAG CGGATGAATG GCAGAAATTC GATGATAAGC TGTCAA
12361 GAGAATTGGT CGACGGCCCA ACCATAGAGC CCACCGCATC CCCAGCATGC CTGC
12421 CTTCCCAATC CTCCCCCTTG CTGTCCTGCC CCACCCCACC CCCCAGAATA GAATGA
12481 TACTCAGACA ATGCGATGCA ATTTCCTCAT TTTATTAGGA AAGGACAGTG GGAGTG
12541 CTTCCAGGGT CAAGGAAGGC ACGGGGGAGG GGCAAACAAC AGATGGCTGG CAA
12601 GCACAGTCGA GGCTGATCAG CGGGTTTAAC CTATTCCTTT GCCCTCGGAC GAGTG
12661 GCGTCGGTTT CCACTATCGG CGAGTACTTC TACACAGCCA TCGGTCCAGA CGGCC
12721 TCTGCGGGCG ATTTGTGTAC GCCCGACAGT CCCGGCTCCG GATCGGACGA TTGC
12781 TCGACCCTGC GCCCAAGCTG CATCATCGAA ATTGCCGTCA ACCAAGCTCT GATAG
12841 GTCAAGACCA ATGCGGAGCA TATACGCCCG GAGCCGCGGC GATCCTGCAA GCTC
12901 CCTCCGCTCG AAGTAGCGCG TCTGCTGCTC CATACAAGCC AACCACGGCC TCCAC
12961 GATGTTGGCG ACCTCGTATT GGGAATCCCC GAACATCGCC TCGCTCCAGT CAATG
13021 TGTTATGCGG CCATTGTCCG TCAGGACATT GTTGGAGCCG AAATCCGCGT GCACG
13081 CCGGACTTCG GGGCAGTCCT CGGCCCAAAG CATCAGCTCA TCGAGAGCCT GCGC

13141 CGCACTGACG GTGTCGTCCA TCACAGTTTG CCAGTGATAC ACATGGGGAT CAGCA
13201 GCATATGAAA TCACGCCATG TAGTGTATTG ACCGATTCCT TGCGGTCCGA ATGGG
13261 CCCGCTCGTC TGGCTAAGAT CGGCCGCAGC GATCGCATCC ATGGCCTCCG CGAC
13321 CAGAACAGCG GGCAGTTCGG TTTCAGGCAG GTCTTGCAAC GTGACACCCT GTGC
13381 GGAGATGCAA TAGGTCAGGC TCTCGCTGAA TTCCCCAATG TCAAGCACTT CCGGA
13441 GAGCGCGGCC GATGCAAAGT GCCGATAAAC ATAACGATCT TTGTAGAAAC CATCG
13501 GCTATTTACC CGCAGGACAT ATCCACGCCC TCCTACATCG AAGCTGAAAG CACGA
13561 TTCGCCCTCC GAGAGCTGCA TCAGGTCGGA GACGCTGTCG AACTTTTCGA TCAGA
13621 CTCGACAGAC GTCGCGGTGA GTTCAGGCTT TCCGGATCTA TCCATCTTTG GTACG
13681 TGGCAAGCTT ATCACCGTGT TTTTCAAAGG AAAACCACGT CCCCGTGGTT CGGGG
13741 AGACGTTTTT TTAACCTCGA CTAAACACAT GTAAAGCATG TGCACCGAGG CCCCA
13801 GATCCCATAC AATGGGGTAC CTTCTGGGCA TCCTTCAGCC CCTTGTTGAA TACGCT
13861 GAGAGCCATT TGACTCTTTC CACAACTATC CAACTCACAA CGTGGCACTG GGGTT
13921 GCCTTTGCAG GTGTATCTTA TACACGTGGC TTTTGGCCGC AGAGGCACCT GTCGC
13981 GGGGGGTTCC GCTGCCTGCA AAGGGTCGCT ACAGACGTTG TTTGTCTTCA AGAA
14041 AGAGGAACTG CTTCCTTCAC GACATTCAAC AGACCTTGCA TTCCTTTGGC GAGAG
14101 AGACCCCTAG GAATGCTCGT CAAGAAGACA GGGCCAGGTT TCCGGGCCCT CACA
14161 AAAGACGGCA ATATGGTGGG AAATAACATA TAGACAAACG CACACCGGCC TTATT
14221 CGGCTTCGGC CAGTAACGTT AGGGGGGGGG GAGGGAGAGG GGCGGAATTC CTC
14281 GCCGAGTCGC GGCCGCTTTA CTTGTACAGC TCGTCCATGC CGAGAGTGAT CCCGC
14341 GTCACGAACT CCAGCAGGAC CATGTGATCG CGCTTCTCGT TGGGGTCTTT GCTCA
14401 GACTGGGTGC TCAGGTAGTG GTTGTCGGGC AGCAGCACGG GGCCGTCGCC GAT
14461 TTCTGCTGGT AGTGGTCGGC GAGCTGCACG CTGCCGTCCT CGATGTTGTG GCGG
14521 AAGTTCACCT TGATGCCGTT CTTCTGCTTG TCGGCCATGA TATAGACGTT GTGGCT
14581 TAGTTGTACT CCAGCTTGTG CCCCAGGATG TTGCCGTCCT CCTTGAAGTC GATGC
14641 AGCTCGATGC GGTTCACCAG GGTGTCGCCC TCGAACTTCA CCTCGGCGCG GGTC
14701 TTGCCGTCGT CCTTGAAGAA GATGGTGCGC TCCTGGACGT AGCCTTCGGG CATGG
14761 TTGAAGAAGT CGTGCTGCTT CATGTGGTCG GGGTAGCGGC TGAAGCACTG CACG
14821 GTCAGGGTGG TCACGAGGGT GGGCCAGGGC ACGGGCAGCT TGCCGGTGGT GCA
14881 TTCAGGGTCA GCTTGCCGTA GGTGGCATCG CCCTCGCCCT CGCCGGACAC GCTG
14941 TGGCCGTTTA CGTCGCCGTC CAGCTCGACC AGGATGGGCA CCACCCCGGT GAAC
15001 TCGCCCTTGC TCACCATGGT GGCGGACTAG TGGATCCGAG CTCGGTACCA AGCT
15061 GCCAGCTTGG GTCTCCCTAT AGTGAGTCGT ATTAGTACCA AGCTAATTCC TCACGA
15121 TGAAATGGAA GAAAAAAACT TTGAACCACT GTCTGAGGCT TGAGAATGAA CCAAG
15181 AACTCAAAAA GGGCAAATTC CAAGGAGAAT TACATCAAGT GCCAAGCTGG CCTAA
15241 GTCTCCACCC ACTCAGTGTG GGGAAACTCC ATCGCATAAA ACCCCTCCCC CCAAC
15301 GACGACGTAC TCCAAAAGCT CGAGAACTAA TCGAGGTGCC TGGACGGCGC CCGC
15361 GTGGAGTCAC ATGAAGCGAC GGCTGAGGAC GGAAAGGCCC TTTTCCTTTG TGTG
15421 TCACCCGCCC GCTCTCCCGA GCGCCGCGTC CTCCATTTTG AGCTCCCTGC AGCAC
15481 GGAAGCGGCC ATCTTTCCGC TCACGCAACT GGTGCCGACC GGGCCAGCCT TGCC
15541 GGCGGGGCGA TACACGGCGG CGCGAGGCCA GGCACCAGAG CAGGCCGGCC AG
15601 TACCCCCGTC CGATTCTCGG TGGCCGCGCT CGCAGGCCCC GCCTCGCCGA ACAT
15661 TGGGACGCAC GGGCCCCGTC GCCGCCCGCG GCCCCAAAAA CCGAAATACC AGTC
15721 TCTTGGCCCG CATTTACAAG ACTATCTTGC CAGAAAAAAA GCGTCGCAGC AGGTC
15781 AAATTTTAAA TGGCTAGAGA CTTATCGAAA GCAGCGAGAC AGGCGCGAAG GTGCC
15841 ATTCGCACGC GGCGGCCCCA GCGCCCAGGC CAGGCCTCAA CTCAAGCACG AGG
15901 GCTCCTTAAG CGCAAGGCCT CGAACTCTCC CACCCACTTC CAACCCGAAG CTCGC
15961 AGAATCACGT ACTGCAGCCA GGTGGAAGTA ATTCAAGGCA CGCAAGGGCC ATAA
16021 AAGAGGCCAG GCCCGCGGGA ACCACACACG GCACTTACCT GTGTTCTGGC GGCA

```
16081 TTGCGAAAAA GAACGTTCAC GGCGACTACT GCACTTATAT ACGGTTCTCC CCCAC
16141 GGAAAAAGGC GGAGCCAGTA CACGACATCA CTTTCCCAGT TTACCCCGCG CCAC
16201 TAGGCACCGG TTCAATTGCC GACCCCTCCC CCCAACTTCT CGGGGACTGT GGGC
16261 CGCTCTGCCC ACTGACGGGC ACCGGAGCAT AACTTCGTAT AGCATACATT ATACG
16321 ATATCGAT
//
```

RA
KE
I
L
NEAF

H
NK
T
DES
TLY

)GE
QQGLG

TL
VLL
KKG
QQ
LL
QYGR
VY
IPVCR
EEP
M
WG
STLRD
LVN
PSGL
KRY

HINM

GTA
GCAG
GACGA
TCTG
AGTGG
AGACT
GACGA
CATGG
TATCA
CGAGTG
CCAGTC
CCAGT
AAAC
CTAT
ACAA
ATGA
AGC
AA
AAGAGC
AAAT
TTGC
ATAA
GTAAG

CTTCG
CAGTGC
TTCTC
GGCAA
ATCGG
TTTGG
GAAA
AGAG
TTCA
CTCC
TACA
AAAC
TCAGCA
GACA
TGGGC
ACAG
CGTCC
GTACC
CACGT
CGCCA
ATCGC
TCA
CTCTG
GGTT
CATT
CTT
AA
AGGAG
GTAAC
GCTCTT
ACACGC
ACTA
AAACTG
CAGG
GACGA
GATAT
GAGA
GGCGGA
GATTT
TAAAC
CGGT
CCCTC
TCGC
GCCT
GCAA
TGGT
AGTGA
GCTG
ATAA

ATTTG
TCACA
ACAAG
GTG
TATAA
TGCGG
TTGTT
TATT
ACAC
ATG
TTTA
TTTA
GCAC
ACTGAT
ATGTG
ACCGTC
CACTTC
TTATC
CCCTG
ACGTAA
TTAT
CACCG
AAT
AAAA
ATC
CCCTT
CACG
AACC
GGG
GTTT
TCAGG
CAGT
GAAAACG
GAC
C
GT
TT
C
AAA
AA
CCA
GTCT
GTG
TA
T
GATGA
AAT
TGG
CT

GT
CTCT
AGTG
GAT
CCCT
TTA
CAAC
CTA
A
GCC
ACT
CT
CCT
AT
CAT
GAT
TCT
CCA
GAAA
TT
A
TGT
CT
GCC
GATT
GATC
CTTCC
CCTTG
ATCG
ATCCT
CATT
GTT
CCAA
ATTTT
GGCTG
AACT
AGAA
TAAG
CTG
CCACT
TCATC
CCGAAC
GTATCA
AAGCG
TCTG
ATATC
AGGAA
CCGC
CAGGCC

CAGAT
GCAAG
AGTAC
GCACCG
AAAAG
TATAC
CTTCAG
TCCTCGC
GCACTG
GAGACC
AAGGT
TGCTT
TACTT
GTGCT
ATATC
AGTTGG
GTTGAT
GCTGAC
TATTTT
TCCAT
TTGAG
CGATATT
CAGGT
GTTGGCC
ACCGCT
AAAGC
ACGCTA
TAAT
AGAGATC
TACTAAT
CGTAAT
CATATC
GAGA
TGG
CTACG
AACAT
TATTGT
ACACC
GGCAC
ACTAGAAG
GCTGGG
CGCGCT
CGTCGCA
GAGTTG
CCGGATG
GAAGAA
GACCGC
GAGGTG
CGACGGA

AATCGC
CCGAA
CGGCTG
CACGGCG
AATCGG
GGCGCA
GATTC
AAACTT
GGTTG
GGGCCT
GATCA
TTGAG
GTGCC
CAGGT
GCTTCC
GGGAA
ATTGCCA
CCAAG
CTAGTGCG
GCGGCG
AGGGCG
GGGGGTG
GATCTTG
TGTTG
CCTTC
CTTGTAG
GCGGAC
GCCGTAG
AGATGAAC
GAACTTG
CAGCTCC
TACCTA
ACACC
GATCCA
ACTTCA
CCTAAA
GTACTCC
GGTGAC
GGGCCG
CGCCCAG
GCTTGAGAC
TGTGCGC
GTGCAGA
CATCAA
CACCAG
CGAAGGG
GGATCA
CCCGTA
AAACCCG

CCTCG
CTTCTC
GATGTG
AAGTT