Supplementary materials to manuscript:
"Gene identification in novel eukaryotic genomes by self-training algorithm"

File S2

```
LOCUS          1     1343 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      1
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES             CHR  ---
     CDS             join(500..670,772..843)
                     /translation="MGRKAGNLYINPKKLGNLAKPCMKEMVSFLNCMALNNIKDDKC
                     EKQKQLLSVCMQGQSDHKNKSWGNINYHLQRLTRGRK"
ORIGIN
       1  TTTCTCTCGT CTCGTCGTCT CTCAGAAATC TTTGTAAGTG CTTCGTTCCC CATTTCTCTC
      61  TGATTAGATT TTGTGTTTGT TCGTTTTGCT TCGATTCTGC GATTTTTGAT TTGCCCAATT
     121  AGATTCGATT TCGTTCTATG TTCGGATCTG AGACCATTTG TATGTTTAAT TCTTCTTTGA
     181  TTTTGGCTTT GATGTTGATT CCGATCTTGG GTCTTTAAGT GTGTTCATCA TTTGTTTTAT
     241  CTTCTCATAT GGTTTCTACA AACTTCTCAA CGTCATAGAA ATGTATAATT GTTAGTGTTT
     301  CGTGGGTTAA TCTCAATGTT CTTACTTTCT CTTGTTGTTT AAAGGCCTAG GATTGAATCT
     361  GTCGACGAAA ATGTGATATA TGAGTGTATT CAGGGTAAGG GTAGCTTTGT GGGAAACTTG
     421  ATATGATGCT GCAGATTTGA TTTACTACTG TTTTTTTTTT TTATGGTTTC TGGTTCTTTA
     481  GTTAGTTGAG AGAGCTATAA TGGGGAGAAA GGCTGGAAAT TTGTACATAA ACCCGAAAAA
     541  GCTCGGGAAT TTGGCGAAAC CGTGTATGAA GGAAATGGTA TCGTTCTTGA ACTGTATGGC
     601  TCTTAACAAT ATCAAAGATG ATAAATGTGA GAAACAGAAG CAACTCCTGA GTGTTTGTAT
     661  GCAAGGCCAG GTTAATATCT TTGCATTTTG CTACTGCGTT TTATACTTTT ATTCGTCACT
     721  GCCTTTTGCG AAAACGATGA ACTTTGTTTT GTTTTCTTGC TTGCTTTGCA GTCGGATCAT
     781  AAGAACAAGT CTTGGGGTAA CATTAACTAC CATTTGCAGA GGCTAACCCG TGGAAGAAAG
     841  TAATCGAAAG CAATGGGATC TTGTTATCCA TAGACCGATA GATGAATCCG TTGTTCGATC
     901  TTGTCATTTG TCTCTGTGTT CTTTGAGGTA TGTAGAAGAA GAACTTTGGA CACATTGTTC
     961  ATCGAGAGAA AGGATTATTG GTCTCTTTGT TTTGCAATAA ACATGTCTTC TGAGTCAAGG
    1021  GTCAGGCAGA AGATCCATTA GTATTAATGA TGATGTGCTT AGTCTAGTTT AAATTAGAAA
    1081  TGATATGTAC CAAATTGCAA TTAAAACATT GATTATTAGA AAACTTGAAG CTACAATAAG
    1141  ATTTAAGTTG TAATTTTTTT TTTTTTCTC AACTTAAATT AGGGTTAGTA GTGTCTTTAA
    1201  ATTACATGTT AGCATTACTG GAAAAAAGTC TTAATACAAA TAAGAGTTTC AATAAGGTAC
    1261  AAGTGAATGC AACAGAACTG TTTACATATT TTTTTCTTCA TAAACATTTA TGCATATTCT
    1321  TGAGATAAAT GTTTACTAAA AAA

//

LOCUS          2     1318 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      2
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES             CHR  ---
     CDS             complement(join(500..553,627..818))
                     /translation="MKKVLRVYGGVMRLVRLLPADSRPYYAKYARENFVNYREVDQS
                     ETSLDELFQRAYNHSLWVLKKYSIDESAATKLKEICFE"
ORIGIN
       1  CAAAAACAAT CTCAAAAAGA CTTTGGTTTT GGTTAGGTTG GTTTCATATA AGATTTTCGA
      61  GTTTTCGACA AGGACCTAAA CTCTTAGCTA CTGTTTATTG TCCCTTAGCT ACATATATAT
     121  CACTTTAATT ACCCAAACAA TCCCTAATCT TTCTCTCTCG TTATTACTTA ACTGCCACTT
     181  CTACTTTGCT CTGATGAGTT TCTTTCATTA CCAAAGATGA GATTAAGTAA TGTGATGTTT
     241  TCGAATCTAA GAATGTAACA ATTTTATTCA GGATCATGGA AACAAACAAG AAGCTAGCTA
     301  CTCAAACAAG ACAACGACCA ATTCAGCAGT CAATCACAGA AAGTCAAGAC AACTAAAAAG
     361  GAAAGACAAA GACACATGAA CAACCCCTGT TCTTGATATA GCTATCAAAA GTAGAAAACT
```

```
     421  GATGGTACCC ATTTAACAAA TCGTCCAAAC TTGTTCCAAC ACTCTCAGGG ACATAACGTC
     481  GAACACTTGG TGTGCGTTAT TATTCAAAAC AAATCTCTTT CAACTTCGTC GCTGCTGATT
     541  CGTCAATTGA GTACTGCAAA TGATTGAGAA CAGACAGTTA ACACAAAGGG TACCTAGATT
     601  TAAAAGAAAC CCAGAATCTT TGTTACCTTC TTGAGTACCC AAAGTGAGTG ATTATAGGCT
     661  CGCTGGAAGA GTTCGTCGAG GGATGTTTCG CTTTGATCGA CTTCACGGTA ATTAACGAAA
     721  TTCTCCCTGG CGTACTTGGC GTAATACGGC CTTGAATCCG CCGGCAAAAG TCTCACCAGC
     781  CTCATAACGC CGCCGTATAC TCTCAGCACC TTCTTCATTT TCGTTTACCT TCTGCAACAC
     841  GTTACCCAAT TCTAGGACTC TAATACTGTA TCATAAACAT CCATGGGCTC TTAATGGGCC
     901  TTTAATAATA AGCAATTACA GACCGGTTAA ACGTTCACAT TTGAGGGTTT AAAACCGGTT
     961  CAAACAAGAC TTTTAAAGTT TAAATCTCTT GATCTCTTAA ACCGGTCTCT ATCGATTACA
    1021  GAAACCATAA CTCAAAACAT CTCTCCAGAT GGAAATAAAC AAAAACTTTC ATAATCATAA
    1081  CACAAAGCCT TAACTTCTTT AACTGTTAAA AATGAACCTA TCATCCAATT TCGAAGAAAA
    1141  TCTTTCTTTC TTTCTTTTGC CTCAAATTTC AAGGTCTTCA AAGTACCCCA AAAATGTGAA
    1201  ACAAAGAACT GAACAAACCC GATGCTTTGC AGGGTTTGGT GTCTTCATAT CTCGGGGTTT
    1261  GGGTAAGTAT GTGTTCCAGT ATTGTCCTTG CATTAGATCT AAAACCTTCA TGATTGAA

//

LOCUS          3     1389 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      3
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES            CHR  ---
     CDS          join(500..563,705..889)
                  /translation="MAKIKSLDVITVAIILLLVIADQATAITVQADCIGPCNDDCQQ
                  LCKSKGYTDWTCASFRTKSSCCCKPPRHQIFEQNAQLNN"
ORIGIN
       1  TATCTCTATT AATTAATATT ACATATAGTG AAGTATCATT ACTTTTATTT TTATCACCTA
      61  AAAGGCCATA ACTTAATTAT ATGGTATGAA ATTTTATCTT TATTAATACT TTTGATATTT
     121  TGAAACAAAA ATAAAATAAA TATGGTATTC CATTCCTTAA TGTACAACTT GATTAAATAT
     181  ATTCAGTCCA TCTTTTATTT TCATTTTCAC TTATATAATC AAGTTATTAT TTTGACCAAT
     241  CCATGTTTAA TCTCTTGGTG ATATATTTCC TTACATATAA TTATTTTCTA CCAAATTTAC
     301  TATATGACAT ATTATATTTT ACCAAATTAC TTCTTAGTCA AATTTTCTAT AAGAATAAGT
     361  TGGTTCAGTG TATATATTTT GATTTTGGTT GGATTTTGTG TCACCAATTA TCTATATAAA
     421  AACTCATTAG ATCTCAACAC AAAACATAAG AAAATATTTT ACAACAATAC ATTATACATT
     481  AGAACAATAC TTCAAAAAAA TGGCCAAGAT CAAATCTTTA GATGTTATTA CTGTCGCAAT
     541  AATTTTGTTG TTGGTGATTG CAGGTTAATC AAGCACTCTT ATTTTTCTCA TTTTCAAAAT
     601  ACGTTCTTAG ATATAATCAT GTATATTATA TTTTAAAGCT TACAATTCCT TTGGTTTTAT
     661  ACATTACATT TCTTTTTAAT TAGTACTTTT TATTATTTTT GTAGATCAAG CAACTGCAAT
     721  TACTGTTCAA GCCGATTGTA TTGGACCGTG CAATGACGAT TGCCAGCAGC TTTGCAAGTC
     781  GAAAGGTTAC ACAGACTGGA CTTGTGCATC ATTTAGAACC AAGTCGAGTT GTTGTTGCAA
     841  ACCACCGAGA CATCAAATCT TTGAACAAAA TGCTCAATTG AACAACTAAA ACCGCTTTTA
     901  TATTGTTCTC AATTTGTATG ATCATTAACA AATATTCAAA GATAAGAAAT AATTTAGAGA
     961  AATCAATTGA TATCATTTTT TCTTATCACT TATCATCATT CTACTCTTGA TTTTGGGCTA
    1021  TTTTGTAATA AGATAACCTG AATTACAATA GCACTGTTTG ATCATAGTAT TTTCATTTGA
    1081  TTAAAACGGC ACCGTTTAGG TTAATCTTGG TCTTAGTTTT CGATTCGGGA ACGCCGCCGT
    1141  CCCAATCACA CAACTGGGTA AAAACGATAT GGAAGACGAT GGTCAGAGCT CCGTCGCCGA
    1201  AACCCTAGAT CCGGCATTGC AGAAGAAGCA CCATGATATG CTCGAGCGTC TCTCCGCTCG
    1261  TCACCAAGCT CGAAAATCAG ATTCACCTGA CTCTTCCTCT TCGTCGTCCT CCACTCTTGA
    1321  ATCGACTTCC TCGTTTCTCG CTAAATTCTC CGATTCAAAG CGATCCATCG AGTCTCGGAT
    1381  CGCCGAGTC

//

LOCUS          4     1370 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      4
VERSION      GeneMark.hmm v3.0 self_training
```

```
ORGANISM   A.thaliana
FEATURES             CHR  ---
     CDS             complement(join(500..711,804..870))
                     /translation="MKYGVLFMVSCGVMFLILSHVEEVEAMKKFGCNTTHPFPGKCG
                     NNGKSSWVSDMKKLPSAPKNRDIRCECSDRPSLARGMPGERVCRCDYDC"
ORIGIN
       1  TCTCTTTGAC TCTCGCTATA TATTCCATCA TATGTATATA TACTGAATCC AATTAGGTTG
      61  ATTTGACTTC TATAAGAGAC TTAATCAGAT GTGTAATTCG TTCATAATAA ATGATGTTTA
     121  CATGTTTTGA TACTGTCATT GGCCACATAC GTTATCTCAT AAGAGAAACA AAGTGCTTGT
     181  ACTCCACTTT TACAAAAAAA ATATCATTTT GACACCTCTC ACAAATTAAT AAAGCCACTA
     241  AATAATACAA ATATGTCCTT TATATATATA ATAAGTGATT AATTATACAA TAATGACTAA
     301  AATTCAAAGA AAAATAGGTT TGGATTAAGG GTATATTAGA AAAAAGTCAC ATTAAAATGA
     361  ACATTTTTTT GTAGAATGGA GAGCTATTAA CTCACAATCA TATGTATTTT AGAATAACAA
     421  GCAATCTACT TTGGCAAAAA CTTTTTTATT TTATTTTAAA GCACGTCATG TTCTATTAGT
     481  TATGGAAATA TTATAAGACT TAACAATCAT AATCACATCT ACATACACGC TCGCCAGGCA
     541  TACCCCTAGC TAAACTTGGA CGGTCAGAAC ATTCGCAACG TATATCACGA TTTTTTGGTG
     601  CGCTTGGAAG CTTTTTCATG TCACTTACCC AAGAACTCTT TCCATTATTT CCACATTTTC
     661  CAGGAAAAGG GTGTGTAGTG TTGCACCCAA ACTTCTTCAT GGCTTCAACC TCTGTTGTGT
     721  TATCAACCAC AAAAATAAAA TGTTAAGACT ATGAAAATAT ATATACAAAA CAGTTGATAA
     781  CACAAAATAA GGAAACAAGT TACCTTCAAC ATGGCTTAGA ATTAGAAACA TGACACCACA
     841  AGAAACCATA AACAAAACTC CATATTTCAT GACTTATAGT CTAAACAAAT AGTCTCTTTG
     901  AGCTTCAAAA ATTTGTGTTT TGTTTTACTC TCATAGAAAG TATATTAAGT TTTCTTTGCC
     961  AACAATCTTC ATAATTAATA TACTCAACAA GATAAAAATA TATGTGGCTA TTTGTTAAGA
    1021  TTTAGCAACA AAGAACACAA CATGAAGAAG CTTGAGGAAG AAGAAATGCC AACTCAGCAA
    1081  AGGGGACACG TCAACAGCAG GACCACAAGA CAAAGAAGGA GCCGTTAGCG CCAACCAGGG
    1141  TTATGAGTCG CACGACTAAA GCATTCTTGT CTTGGTTTGG TAAAGTTGCA CGCGGGTTAA
    1201  GCCCAATCCA AATTGGTGGA AGGTGACAAA TTTGGAGGCA GCTTGGAGAC AGCTGTAGAG
    1261  GAGCTTGTAG AGGCAAACCA AGAAGGATCC AGAAGGGAAC CTAGGTTAAA ACCTCTAAGT
    1321  CCATTTAAGC TGAAGGTGGT GCTTAGAGAC AGAGTTAGAA ACGTAGAGAC

//

LOCUS          5     1305 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      5
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES             CHR  ---
     CDS             complement(500..805)
                     /translation="MASGGFRYSSSRSFAPFRRTGHMLAAVGSMVPKVTSVRQFRLN
                     SSLYVSGPTARAGAPRSFFLCGTMPGREGNASRRRTTTRQLHFGFSLHEMSPVHWTD
                     G"
ORIGIN
       1  ACACTGCCGT CTGTTCCCTT CAGTCCCTTA GCGCCCCGGG ACGGGAGTGG GGGACGGCTC
      61  TACGCGCAGG CAACAGCAGC ACCGGCTCCA CGAAGTCTGA ATCGAATCTT TCTGTTGGCT
     121  TTCCCAAATT CATTCGTAGA TCAAAATCAA AGGGCTAGAA GCGAGCGCTT CTAGCTGCTT
     181  CGCCTGCTTC TTCTTATTAT GGCGGCCATG TTGGCGTGGC GAAAATGAAC GAAAAGCGAG
     241  ATGAACGTGC TATTTTCAAA TCGGATTGAT AGATTGATCT GTTCTGATAG ATCTAAAGAG
     301  TAGAAATAGA TAGAGAATAG AGAGGGAATC AATAATAAGG TCTTTGAGCC TATTTCTATC
     361  TATTGATGAG ACAACTATCT ATTCTTGATC CATCAGAAAG AAATTATTTA TCCATCAGAA
     421  AGAAATCGAT ATGTATCTGA TGGAGTCTTC TACATCGTAC GTAGAGCGCC CAAGCGCTTT
     481  TTGGGCCAGC TCAGTTCTCT TATCCATCGG TCCAATGCAC TGGGCTCATC TCATGGAGGG
     541  AAAAGCCAAA ATGTAGTTGT CTTGTTGTTG TTCGCCGCCT CGACGCATTC CCTTCTCTCC
     601  CCGGCATCGT CCCACACAGA AAGAAAGAGC GCGGAGCCCC GGCTCGAGCC GTAGGTCCGC
     661  TAACGTAAAG CGAGGAGTTG AGCCTGAACT GGCGAACCGA AGTCACTTTC GGAACCATAC
     721  TTCCTACAGC TGCTAACATG TGCCCAGTCC TGCGGAAAGG CGCAAACGAA CGTGAGCTGC
     781  TATACCGGAA TCCCCCGCTG GCCATCGGGG ACCGAGTGGT AAGGCCATGA TCTGCAGGGG
     841  AACGGATCAC TCATTCTTCC ATTGGGGACA GGTGCACGAA CGACAACTCC AAACGTCACA
```

```
     901  CATCCGCCGC CTACTTACCG TTTAGGTGGC ACCAGCGAGA TCCAGCTAAG GAAAAAGAGT
     961  GTCGCGGCTG CTCCACTCCG CCGCGGTCTC ATGAACTTCA CTTCGTTGCC TTCCCGCGCG
    1021  CAAAGCGAAT GGGCGCTGTG CTGTGGGTCA GTTTCGGGGC GGGGGGCGCA GAGATACCAG
    1081  AATAAATGAT TCATTTGTTT GGATCGACGA GCTTTTTCAG CCCCAAAACT CAGAATCAAT
    1141  GGAATGTCTG TCCATAAACA TCTATTCTAT CTGTATATAT AGGGGATCTC TCTATACATA
    1201  TCAAAGTCTT TTATGGCATT GATATGATCG CCTAGCTGTA GCCGCATATC ACAGCTGCAC
    1261  TCAATATGCG GATTTCTGTT GGATTAGATC TTTTCTTTCG CTTTG

//

LOCUS          6     1347 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      6
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES             CHR  ---
     CDS             complement(500..847)
                     /translation="MERRNRRTGRTEKARIWEVTDRTVRTWIGEAVAAAAADGVTFS
                     VPVTPHTFRHSYAMHMLYAGIPLKVLQSLMGHKSISSTEVYTKVFALDVAARHRVQF
                     AMPESDAVAMLKQLS"
ORIGIN
       1  TAAAAATATC CCCGTGGATA AGTGGATAAC CCCAAGGGAA GTTTTTTCAG GCATCGTGTG
      61  TAAGCAGAAT ATATAAGTGC TGTTCCCTGG TGCTTCCTCG CTCACTCGAG GGCTTCGCCC
     121  TGTCGCTCGA CTGCGGCGAG CACTACTGGC TGTAAAAGGA CAGACCACAT CATGGTTCTG
     181  TGTTCATTAG GTTGTTCTGT CCATTGCTGA CATAATCCGC TCCACTTCAA CGTAACACCG
     241  CACGAAGATT TCTATTGTTC CTGAAGGCAT ATTCAAATCG TTTTCGTTAC CGCTTGCAGG
     301  CATCATGACA GAACACTACT TCCTATAAAC GCTACACAGG CTCCTGAGAT TAATAATGCG
     361  GATCTCTACG ATAATGGGAG ATTTTCCCGA CTGTTTCGTT CGCTTCTCAG TGGATAACAG
     421  CCAGCTTCTC TGTTTAACAG ACAAAAACAG CATATCCACT CAGTTCCACA TTTCCATATA
     481  AAGGCCAAGG CATTTATTCT CAGGATAATT GTTTCAGCAT CGCAACCGCA TCAGACTCCG
     541  GCATCGCAAA CTGCACCCGG TGCCGGGCAG CCACATCCAG CGCAAAAACC TTCGTGTAGA
     601  CTTCCGTTGA ACTGATGGAC TTATGTCCCA TCAGGCTTTG CAGAACTTTC AGCGGTATAC
     661  CGGCATACAG CATGTGCATC GCATAGGAAT GGCGGAACGT ATGTGGTGTG ACCGGAACAG
     721  AGAACGTCAC ACCGTCAGCA GCAGCGGCGG CAACCGCCTC CCCAATCCAG GTCCTGACCG
     781  TTCTGTCCGT CACTTCCCAG ATCCGCGCTT TCTCTGTCCT TCCTGTGCGA CGGTTACGCC
     841  GCTCCATGAG CTTATCGCGA ATAAATACCT GTGACGGAAG ATCACTTCGC AGAATAAATA
     901  AATCCTGGTG TCCCTGTTGA TACCGGGAAG CCCTGGGCCA ACTTTTGGCG AAAATGAGAC
     961  GTTGATCGGC ACGTAAGAGG TTCCAACTTT CACCATAATG AAATAAGATC ACTACCGGGC
    1021  GTATTTTTTG AGTTATCGAG ATTTTCAGGA GCTAAGGAAG CTAAAATGGA GAAAAAAATC
    1081  ACTGGATATA CCACCGTTGA TATATCCCAA TGGCATCGTA AAGAACATTT TGAGGCATTT
    1141  CAGTCAGTTG CTCAATGTAC CTATAACCAG ACCGTTCAGC TGGATATTAC GGCCTTTTTA
    1201  AAGACCGTAA AGAAAAATAA GCACAAGTTT TATCCGGCCT TTATTCACAT TCTTGCCCGC
    1261  CTGATGAATG CTCATCCGGA ATTCCGTATG GCAATGAAAG ACGGTGAGCT GGTGATATGG
    1321  GATAGTGTTC ACCCTTGTTA CACCGTT

//

LOCUS          7     1659 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      7
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES             CHR  ---
     CDS             500..1159
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITA
                     FLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTE
                     TFSSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFD
                     LNVANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDE
                     WQGGA"
```

```
ORIGIN
       1  GCAGCCACAT CCAGCGCAAA AACCTTCGTG TAGACTTCCG TTGAACTGAT GGACTTATGT
      61  CCCATCAGGC TTTGCAGAAC TTTCAGCGGT ATACCGGCAT ACAGCATGTG CATCGCATAG
     121  GAATGGCGGA ACGTATGTGG TGTGACCGGA ACAGAGAACG TCACACCGTC AGCAGCAGCG
     181  GCGGCAACCG CCTCCCCAAT CCAGGTCCTG ACCGTTCTGT CCGTCACTTC CCAGATCCGC
     241  GCTTTCTCTG TCCTTCCTGT GCGACGGTTA CGCCGCTCCA TGAGCTTATC GCGAATAAAT
     301  ACCTGTGACG GAAGATCACT TCGCAGAATA AATAAATCCT GGTGTCCCTG TTGATACCGG
     361  GAAGCCCTGG GCCAACTTTT GGCGAAAATG AGACGTTGAT CGGCACGTAA GAGGTTCCAA
     421  CTTTCACCAT AATGAAATAA GATCACTACC GGGCGTATTT TTTGAGTTAT CGAGATTTTC
     481  AGGAGCTAAG GAAGCTAAAA TGGAGAAAAA AATCACTGGA TATACCACCG TTGATATATC
     541  CCAATGGCAT CGTAAAGAAC ATTTTGAGGC ATTTCAGTCA GTTGCTCAAT GTACCTATAA
     601  CCAGACCGTT CAGCTGGATA TTACGGCCTT TTTAAAGACC GTAAAGAAAA ATAAGCACAA
     661  GTTTTATCCG GCCTTTATTC ACATTCTTGC CCGCCTGATG AATGCTCATC CGGAATTCCG
     721  TATGGCAATG AAAGACGGTG AGCTGGTGAT ATGGGATAGT GTTCACCCTT GTTACACCGT
     781  TTTCCATGAG CAAACTGAAA CGTTTTCATC GCTCTGGAGT GAATACCACG ACGATTTCCG
     841  GCAGTTTCTA CACATATATT CGCAAGATGT GGCGTGTTAC GGTGAAAACC TGGCCTATTT
     901  CCCTAAAGGG TTTATTGAGA ATATGTTTTT CGTCTCAGCC AATCCCTGGG TGAGTTTCAC
     961  CAGTTTTGAT TTAAACGTGG CCAATATGGA CAACTTCTTC GCCCCCGTTT TCACCATGGG
    1021  CAAATATTAT ACGCAAGGCG ACAAGGTGCT GATGCCGCTG GCGATTCAGG TTCATCATGC
    1081  CGTTTGTGAT GGCTTCCATG TCGGCAGAAT GCTTAATGAA TTACAACAGT ACTGCGATGA
    1141  GTGGCAGGGC GGGGCGTAAT TTTTTTAAGG CAGTTATTGG TGCCCTTAAA CGCCTGGTTG
    1201  CTACGCCTGA ATAAGTGATA ATAAGCGGAT GAATGGCAGA AATTCGATGA TAAGCTGTCA
    1261  AACATGAGAA TTGGTCGACG GCCCGGGCGG CCGCAAGGGG TTCGCGTTGG CCGATTCATT
    1321  AATGCAGCTG GCACGACAGG TTTCCCGACT GGAAAGCGGG CAGTGAGCGC AACGCAATTA
    1381  ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA
    1441  TGTTGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT
    1501  ACGCCAAGCT ATTTAGGTGA CACTATAGAA TACTCAAGCT TTTATGATCC ATTTGCCGTG
    1561  GGTCCCACCA GCTATGTAGG AATACCCTAT CTTCTTTATA AGAGCTTTGA TACATGGAAA
    1621  ATATTGGAAA ACAATGTCCG AACATATAAT TTCCCCCCCC

//

LOCUS          8     1659 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      8
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES             CHR  ---
     CDS             complement(500..1159)
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITA
                     FLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTE
                     TFSSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFD
                     LNVANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDE
                     WQGGA"
ORIGIN
       1  AGGATGAACT GTAACCGGAT TCATATCCTC AAGATGCTCA TTGTATTCTG CATTTAGATT
      61  GACATCACGA ACCACGCGGC TTCCATTAGA TGGTTGCACA AACAAAGGGC TCAACTCAAG
     121  CTTGAGTATT CTATAGTGTC ACCTAAATAG CTTGGCGTAA TCATGGTCAT AGCTGTTTCC
     181  TGTGTGAAAT TGTTATCCGC TCACAATTCC ACACAACATA CGAGCCGGAA GCATAAAGTG
     241  TAAAGCCTGG GGTGCCTAAT GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC
     301  CGCTTTCCAG TCGGGAAACC TGTCGTGCCA GCTGCATTAA TGAATCGGCC AACGCGAACC
     361  CCTTGCGGCC GCCCGGGCCG TCGACCAATT CTCATGTTTG ACAGCTTATC ATCGAATTTC
     421  TGCCATTCAT CCGCTTATTA TCACTTATTC AGGCGTAGCA ACCAGGCGTT TAAGGGCACC
     481  AATAACTGCC TTAAAAAAAT TACGCCCCGC CCTGCCACTC ATCGCAGTAC TGTTGTAATT
     541  CATTAAGCAT TCTGCCGACA TGGAAGCCAT CACAAACGGC ATGATGAACC TGAATCGCCA
     601  GCGGCATCAG CACCTTGTCG CCTTGCGTAT AATATTTGCC CATGGTGAAA ACGGGGGCGA
     661  AGAAGTTGTC CATATTGGCC ACGTTTAAAT CAAAACTGGT GAAACTCACC CAGGGATTGG
     721  CTGAGACGAA AAACATATTC TCAATAAACC CTTTAGGGAA ATAGGCCAGG TTTTCACCGT
```

```
     781  AACACGCCAC ATCTTGCGAA TATATGTGTA GAAACTGCCG GAAATCGTCG TGGTATTCAC
     841  TCCAGAGCGA TGAAAACGTT TCAGTTTGCT CATGGAAAAC GGTGTAACAA GGGTGAACAC
     901  TATCCCATAT CACCAGCTCA CCGTCTTTCA TTGCCATACG GAATTCCGGA TGAGCATTCA
     961  TCAGGCGGGC AAGAATGTGA ATAAAGGCCG GATAAAACTT GTGCTTATTT TTCTTTACGG
    1021  TCTTTAAAAA GGCCGTAATA TCCAGCTGAA CGGTCTGGTT ATAGGTACAT TGAGCAACTG
    1081  ACTGAAATGC CTCAAAATGT TCTTTACGAT GCCATTGGGA TATATCAACG GTGGTATATC
    1141  CAGTGATTTT TTTCTCCATT TTAGCTTCCC TTAGCTCCTG AAATCTCGAC CGCACAGATG
    1201  CGAAGGAGAA AATACCGCAT CAGGAAATTG TAAGCGTTAA TATTTGTTAA ATTCGCGTTA
    1261  AATTTTTGTT AAATCAGCTC ATTTTTTAAC CAATAGGCCG AAATCGGCAA AATCCCTTAT
    1321  AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTTGTTC CAGTTTGGAA CAAGAGTCCA
    1381  CTATTAAAGA ACGTGGACTC CAACGTCAAA GGGCGAAAAA CCGTCTATCA GGGCGATGGC
    1441  CCACTACGTG AACCATCACC CTAATCAAGT TTTTTGGGGT CGAGGTGCCG TAAAGCACTA
    1501  AATCGGAACC CTAAAGGGAG CCCCCGATTT AGAGCTTGAC GGGGAAAGCC GGCGAACGTG
    1561  GCGAGAAAGG AAGGGAAGAA AGCGAAAGGA GCGGGCGCTA GGGCGCTGGC AAGTGTAGCG
    1621  GTCACGCTGC GCGTAACCAC CACACCCGCC GCGCTTAAT

//

LOCUS          9     2157 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      9
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES            CHR  ---
     CDS          complement(500..1657)
                  /translation="METLNQCINAGHEMTKAIAIAQFNDDSPEARKITRRWRIGEAA
                  DLVGVSSQAIRDAEKAGRLPHPDMEIRGRVEQRVGYTIEQINHMRDVFGTRLRRAED
                  VFPPVIGVAAHKGGVYKTSVSVHLAQDLALKGLRVLLVEGNDPQGTASMYHGWVPDL
                  HIHAEDTLLPFYLGEKDDVTYAIKPTCWPGLDIIPSCLALHRIETELMGKFDEGKLP
                  TDPHLMLRLAIETVAHDYDVIVIDSAPNLGIGTINVVCAADVLIVPTPAELFDYTSA
                  LQFFDMLRDLLKNVDLKGFEPDVRILLTKYSNSNGSQSPWMEEQIRDAWGSMVLKNV
                  VRETDEVGKGQIRMRTVFEQAIDQRSSTGAWRNALSIWEPVCNEIFDRLIKPRWEIR
                  "
ORIGIN
       1  TAACGCTGAC CACGTTCATA AGCACTTGTT GGGCGATAAT CGTTACCCAA TCTGGATAAT
      61  GCAGCCATCT GCTCATCATC CAGCTCGCCA ACCAGAACAC GATAATCACT TTCGGTAAGT
     121  GCAGCAGCTT TACGACGGCG ACTCCCATCG GCAATTTCTA TGACACCAGA TACTCTTCGA
     181  CCGAACGCCG GTGTCTGTTG ACCAGTCAGT AGAAAAGAAG GGATGAGATC ATCCAGTGCG
     241  TCCTCAGTAA GCAGCTCCTG GTCACGTTCA TTACCTGACC ATACCCGAGA GGTCTTCTCA
     301  ACACTATCAC CCCGGAGCAC TTCAAGAGTA AACTTCACAT CCCGACCACA TACAGGCAAA
     361  GTAATGGCAT TACCGCGAGC CATTACTCCT ACGCGCGCAA TTAACGAATC CACCATCGGG
     421  GCAGCTGGTG TCGATAACGA AGTATCTTCA ACCGGTTGAG TATTGAGCGT ATGTTTTGGA
     481  ATAACAGGCG CACGCTTCAT TATCTAATCT CCCAGCGTGG TTTAATCAGA CGATCGAAAA
     541  TTTCATTGCA GACAGGTTCC CAAATAGAAA GAGCATTTCT CCAGGCACCA GTTGAAGAGC
     601  GTTGATCAAT GGCCTGTTCA AAAACAGTTC TCATCCGGAT CTGACCTTTA CCAACTTCAT
     661  CCGTTTCACG TACAACATTT TTTAGAACCA TGCTTCCCCA GGCATCCCGA ATTTGCTCCT
     721  CCATCCACGG GGACTGAGAG CCATTACTAT TGCTGTATTT GGTAAGCAAA ATACGTACAT
     781  CAGGCTCGAA CCCTTTAAGA TCAACGTTCT TGAGCAGATC ACGAAGCATA TCGAAAAACT
     841  GCAGTGCGGA GGTGTAGTCA AACAACTCAG CAGGCGTGGG AACAATCAGC ACATCAGCAG
     901  CACATACGAC ATTAATCGTG CCGATACCCA GGTTAGGCGC GCTGTCAATA ACTATGACAT
     961  CATAGTCATG AGCAACAGTT TCAATGGCCA GTCGGAGCAT CAGGTGTGGA TCGGTGGGCA
    1021  GTTTACCTTC ATCAAATTTG CCCATTAACT CAGTTTCAAT ACGGTGCAGA GCCAGACAGG
    1081  AAGGAATAAT GTCAAGCCCC GGCCAGCAAG TGGGCTTTAT TGCATAAGTG ACATCGTCCT
    1141  TTTCCCCAAG ATAGAAAGGC AGGAGAGTGT CTTCTGCATG AATATGAAGA TCTGGTACCC
    1201  ATCCGTGATA CATTGAGGCT GTTCCCTGGG GGTCGTTACC TTCCACGAGC AAAACACGTA
    1261  GCCCCTTCAG AGCCAGATCC TGAGCAAGAT GAACAGAAAC TGAGGTTTTG TAAACGCCAC
    1321  CTTTATGGGC AGCAACCCCG ATCACCGGTG GAAATACGTC TTCAGCACGT CGCAATCGCG
    1381  TACCAAACAC ATCACGCATA TGATTAATTT GTTCAATTGT ATAACCAACA CGTTGCTCAA
```

```
    1441  CCCGTCCTCG AATTTCCATA TCCGGGTGCG GTAGTCGCCC TGCTTTCTCG GCATCTCTGA
    1501  TAGCCTGAGA AGAAACCCCA ACTAAATCCG CTGCTTCACC TATTCTCCAG CGCCGGGTTA
    1561  TTTTCCTCGC TTCCGGGCTG TCATCATTAA ACTGTGCAAT GGCGATAGCC TTCGTCATTT
    1621  CATGACCAGC GTTTATGCAC TGGTTAAGTG TTTCCATGAG TTTCATTCTG AACATCCTTT
    1681  AATCATTGCT TTGCGTTTTT TTATTAAATC TTGCAATTTA CTGCAAAGCA ACAACAAAAT
    1741  CGCAAAGTCA TCAAAAAACC GCAAAGTTGT TTAAAATAAG AGCAACACTA CAAAAGGAGA
    1801  TAAGAAGAGC ACATACCTCA GTCACTTATT ATCACTAGCG CTCGCCGCAG CCGTGTAACC
    1861  GAGCATAGCG AGCGAACTGG CGAGGAAGCA AAGAAGAACT GTTCTGTCAG ATAGCTCTTA
    1921  CGCTCAGCGC AAGAAGAAAT ATCCACCGTG GGAAAAACTC CAGGTAGAGG TACACACGCG
    1981  GATAGCCAAT TCAGAGTAAT AAACTGTGAT AATCAACCCT CATCAATGAT GACGAACTAA
    2041  CCCCCGATAT CAGGTCACAT GACGAAGGGA AAGAGAAGGA AATCAACTGT GACAAACTGC
    2101  CCTCAAATTT GGCTTCCTTA AAAATTACAG TTCAAAAAGT ATGAGAAAAT CCATGCA

//

LOCUS          10     4102 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      10
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.thaliana
FEATURES            CHR  ---
     CDS
join(500..592,852..1119,1204..1343,1420..1555,1631..1879,1941..2026,2100..2196,2
276..2381,2463..2556,2640..2839,2923..3022,3111..3227,3323..3379,3495..3602)
                 /translation="MCGIAIVVCGVRIELSTLSSSLSETETPFERLQVSVEDVKAVL
                 SQRGPDSVGEKTIHLRPNCGQESATLSVVETSEGTYKLEETTSLGELHFIGSTLQLR
                 GTSPIRQPLVDSSGNILAYNGEVFGGIELNSYDNDTEVLLKSLEKAKSLVPDVLSMI
                 KGPWAIIYWQESSRTLWFGKDPFGRRSLLVHSPTVEDPRFLLSSVSPASSVSQGAGL
                 DTENGNSLHRYWEELPCGVYSISFGVSELCIHGEVTKHEWRNTIWKELIEWERKLVV
                 PRPEDLSTSSLSGVQEDKSVSTSLGLALGFAQTVLVVLKESVRRRTSLHSIYQGDKE
                 AVPVAVLFSGGLDSMILAALLHQCLDPKYEVDLLNVSFDGPNAPDRISAKAGIKELK
                 KIAPFRRWKLVEIDADLSKLGFETKRVMSLINPADTYMDLNIGTALWLAARGDGWIH
                 EDNGNPSVEENNQRVKYKSDARILLVGAGADEQCAGYGRHRTKYRNGSWVALDQEMK
                 LDMQRIWKRNLGRDDRCIADNGKEGRFPFLDEDVIKTLLDIPLWEIADLEQPSGTGD
                 KKILRQVASLLGVHEVAKMPKRAIQFGSRIARESNRKNFGSNRAANQASAGSVRFLA
                 PSH"
ORIGIN
       1  TTTAGTTAAT GTATTTTATT TAATTTTAAA AATTTTAAGA ATTAATAAAA ATTACATTTC
      61  GTTGTAAATA TATACAAAGG CCAGGTTTTC AAGTTAATGT ATATATATAT AAAGAAATAT
     121  TTTGTTTTCT TTGCTAAAAA TGTAAAATTA TGTGTGTCTT TATTAGTTTC TAGCTTTTTT
     181  TTATTTTAGT CGAATGTTTA ACATGTTTTT CGCTGTTATG GTATAGATTG GGCCAACTTG
     241  AAAAACATAA AGTCCATAAA ATGCCAATTT GATAAGCCCA TATAAAAAAA TTGAATTTAA
     301  AAAAAGCCTA CGGTGAGGAT CGAACTCACG ACCTTTCGCT TACGAAGCGA ACGCAATACC
     361  ACTATGCTAC GGAGGCAATT TGTTCATTTT ATAGAATCAT TTTAGAAGAA TGTTTATTAC
     421  TTGTTAAACA GGCTTTAAGT AACTTTCATG ACAACAACTT TGACGGAAAC GGAGACCAGA
     481  AACAGACGCC GGCGGAGATA TGTGCGGAAT TGCAATTGTA GTGTGCGGCG TTCGTATCGA
     541  ACTCTCGACA CTTTCGTCGT CTCTCTCTGA AACTGAGACT CCTTTCGAAC GAGTAAATCT
     601  CTCTCTCCAA CGTCAGTTGT GATTTTAGTA TGAATTGCTT TTGGGTTTAG AAGTTTAGGG
     661  TTTTTGTTCT GTCTGAAGTA ATTTAGCATC ATTGGTAAAG CTTGAATCTT TAATCGAGTT
     721  TGTTCTTATA TAGAAAGAAA GACTTACACT TTTGAGTTTT TATCGATGGA TAATGGTAAT
     781  GTATTGAGAT GAGAGTGAAT TTGGTTGTTC ATAGATGAGC TAATTTGGTG TGTTTTCTTG
     841  GTTTAAACCA GTTACAAGTC TCTGTTGAAG ATGTTAAAGC TGTCTTAAGC CAAAGAGGTC
     901  CTGATAGTGT AGGAGAAAAG ACTATTCATC TTCGACCAAA TTGTGGTCAA GAATCTGCTA
     961  CTCTCTCTGT TGTTGAGACT AGTGAAGGAA CTTATAAGCT TGAGGAGACT ACTTCTTTGG
    1021  GTGAATTGCA TTTTATTGGT TCTACATTAC AGCTTAGAGG AACTAGTCCT ATAAGGCAAC
    1081  CATTGGTGGA TTCTTCTGGG AACATTCTAG CTTACAATGG TGTGGTGTTA TTCTTAAGGT
    1141  TTTTTGATTA TCATATGTTT TCTCTTACAC TATACGATAT TGATTTGTAT ATTTTCCGTA
    1201  CAGGCGAAGT ATTTGGTGGA ATTGAACTCA ACAGCTATGA TAATGATACT GAAGTGCTTT
```

```
    1261  TGAAGTCTTT AGAGAAAGCA AAATCTTTAG TTCCAGATGT TCTTTCGATG ATTAAAGGTC
    1321  CCTGGGCCAT TATATATTGG CAGGTAATTT GGTAACCTGC TGCTATTTGT TTATGTAGCA
    1381  TTTTTTTTTA CTTTCTCTAA GAGCATTTTT CTCTATCAGG AAAGCTCGAG AACGTTATGG
    1441  TTTGGTAAAG ATCCGTTTGG TCGGAGGAGC CTTCTTGTTC ACTCGCCAAC TGTGGAAGAC
    1501  CCTCGTTTTC TTCTATCGTC TGTTTCACCA GCTTCTTCAG TTTCTCAAGG TGCTGGTTTG
    1561  TATTACTTAG TAGAGCAATT TTGTCTCTTT ATGTACACTC CAAGCTCAAG TAACATCTGT
    1621  TCTTTTGCAG GATTGGATAC TGAAAATGGT AATAGCCTTC ATAGATATTG GGAAGAGCTT
    1681  CCATGTGGCG TGTACAGTAT TTCTTTTGGG GTTTCAGAAT TATGTATTCA TGGTGAAGTT
    1741  ACAAAACACG AATGGAGAAA TACCATCTGG AAAGAATTGA TTGAATGGGA AAGGAAATTA
    1801  GTTGTACCTA GACCTGAAGA TCTATCAACG TCATCGCTTA GTGGAGTTCA GGAAGACAAA
    1861  TCTGTGTCAA CTTCTCTAGG TACTGCTTAA GAATCGTTTG ATTCCTCCAG TCCTGTCCAA
    1921  ATCTCAACAT GTCTGACAAG GTTTAGCTTT AGGATTTGCG CAGACAGTTT TGGTCGTGTT
    1981  AAAGGAATCA GTGAGGCGGC GAACTTCACT TCATAGCATT TATCAGGTAT GAGCTTAAAC
    2041  TTGTTTTTTC TCTGAAGCTA TGTGTTTTCT TCTGAACACC TTCGTATTGG CATAAACAGG
    2101  GAGATAAAGA AGCTGTCCCA GTAGCTGTTC TTTTCTCTGG TGGATTGGAT TCGATGATTC
    2161  TTGCTGCGCT GTTGCATCAG TGCCTTGATC CAAAATGTAA CCATTCCCCC ACACTGTTTC
    2221  TGTCTCGGCC AACAATTCTG AAACTAATCA AAAGAAGTAA TATCTATTTC TTCAGATGAA
    2281  GTTGATCTCC TTAATGTCAG TTTTGATGGT CCTAATGCTC CTGATAGGAT CTCCGCCAAG
    2341  GCTGGAATAA AGGAACTCAA AAAGATTGCA CCTTTTAGAA GGTTGAACTT CAAAAGCCTT
    2401  TTATGAAAAA TGCATATATC TGTTACTATC TATGTTGATA CAAATGTTTT TGGAATTTTC
    2461  AGGTGGAAAC TTGTTGAGAT TGATGCTGAT CTGTCAAAAC TAGGCTTTGA GACAAAACGT
    2521  GTTATGTCAC TCATAAATCC TGCAGACACT TACATGGTAA ACCTCAAAGT CTTGAGCATT
    2581  GTGATTACAT TCCTCCGAAC TTGATGTGAA TTATTCAATC TTTTTTTGCT ATGTTCTAGG
    2641  ACCTTAACAT AGGAACAGCA TTATGGTTAG CTGCTCGTGG TGATGGTTGG ATTCACGAAG
    2701  ATAATGGAAA TCCAAGTGTA GAGGAAAATA ACCAGCGGGT CAAGTATAAG TCCGATGCTA
    2761  GAATCCTGCT TGTTGGCGCT GGTGCTGATG AGCAATGTGC CGGTTACGGT AGGCACAGAA
    2821  CTAAATATCG AAATGGAAGG TATATGATTC ATTTATGTAT ACATTTCTTA TGGATGTAAA
    2881  CTAATGTTAA CCTCATCGCA CAAGTTTTAT TCATTTTTCC AGTTGGGTTG CGTTGGATCA
    2941  AGAAATGAAA TTAGATATGC AGAGAATTTG GAAAAGAAAT TTGGGTCGAG ATGATCGATG
    3001  CATTGCAGAT AACGGTAAAG AGGTATATGT CATCACGCCA CAACCACAAA CTCAAAAAAA
    3061  AAGTTTATTC TTTGTGCTTT GTTCTTAAAT CTTAAACGTT ACTTGTTCAG GGGCGATTCC
    3121  CGTTCTTAGA TGAAGATGTG ATAAAGACTC TCCTTGACAT TCCCTTGTGG GAAATCGCTG
    3181  ATCTTGAGCA ACCAAGTGGA ACGGGAGACA AGAAGATTCT TAGACAGGTG AGTAATTAAA
    3241  AAGATACCAT TTTTACTGTA AGAATGCCTC AAGAACATTC CAGAAACTTC ATAAACCTGA
    3301  ATTGTTGCGT AAAATCTTGA AGGTTGCAAG CCTGCTTGGT GTACATGAAG TTGCAAAGAT
    3361  GCCTAAGAGA GCAATTCAGG TAAGAAGAAG AGAACTTTCT GAGAGAACTC AGTTTCAAAC
    3421  GAGTAAAGTT GAAATCTTTT TGTTGTGAGC TCTCTCTCTC TCACTGTTGC TTTGGACTTT
    3481  GGTTGTGGTT TTAGTTTGGA TCAAGAATCG CTCGTGAATC AAACCGGAAG AATTTTGGAA
    3541  GTAACCGTGC CGCAAACCAG GCTTCTGCCG GGAGCGTGAG ATTCCTAGCA CCGTCACACT
    3601  AATTCCCTCT GATTCATGTT AGATCAATCA TTGTTATACA TTGTATCCTA TACAAAGAGA
    3661  CATTCTTTCA ATGATCTCAA AACTGGTTTG AACACTGAAC ACAAATGGAC TAATAAGAGG
    3721  ACATGTTAAG ACTTAAACCA CTTGGTTCGG TATTTTGCTC AATTTCAAGG GTAAAATTGG
    3781  AAAAATCACA AGAAAAACAA AAGCCCTAAC ATTATTGTCT AAAATCAATT TTCTCCTCCG
    3841  GTGGTGATCT CTGTGCTCCG GCGCATAAAC CCCTGTAAAG CCAATAACTT GTTGTTCCCC
    3901  AAAGTTTGGT TCTTCATGTT GGAAACCGAA AGCAGTGTCT GCGACAACAT CGCCGGCAAT
    3961  GAGAAATGGC GAGCGGAAGC TGAAATTGGT GGCAACGAGA GAGCTCTGCA AGCGCTTCGA
    4021  GAACTCATTA TCTTCCCTTT TCGTTACCCT CTCGAAGCTC GAACTCTTGG TCTCAAAGTA
    4081  CATATATCCA CACCTCTGTT GG

//

LOCUS          1     1295 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      1
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          join(500..534,609..795)
```

```
                /translation="MVNVMKGVLVECDPAMKQFLLHLDETQSLGKKFIIQDLDEKHL
                FISVEIVEVLQARVDDLMDRISFPLHEKEA"
ORIGIN
       1  TTAATGTTAC CTAACCAACC CGCGAGAGAA TCGAATGAGA ATGGTTTTGA TGGCCCGTTT
      61  TGCGCGGATG TCAGAAGCAG CAGCAGGGGC AGTCGTGAGC AGGCGGGTTG GGTTATGTCA
     121  TGTCGTTTGA CAGTTTTGCG TAGATCGCGC ATTTCATGGG CATGGAAATC TTAGTTTTTT
     181  ACCTTTCTTC AGAGCTTATC ATGCAAGACA TGGCTATTTA GTTAATTGAA AGTTTATTAA
     241  CATCACTTAA ATTTAGATTT TTGTTGTAAT TCGTATAAAA ACATTTATGT CAAGGCTGGA
     301  CATTCAACAT GGCGGTGAAG AAATTCAGAA ACCTAACCTG GCTTTCTGCA CTTTTCCCAT
     361  TTGACGTTTC GTTCCGTCAA CAGCGGTGTG CGAGCGGTTT GTTTACATTG TCTGCGATGC
     421  AATAAATTGG TGAAATAACA ATCTTTCGGC GATTAGTTTA TCGTGCAATC GTTGATAAAA
     481  CAACCCACAC TACTCAGCAA TGGTGAATGT TATGAAAGGC GTACTCGTTG AATGGTAAGA
     541  ATCATCACTT TTCCTCTTGA ATACCTTCCT TCCGTCGTAC TCAATAGCTT CTCACTCAAT
     601  CTCCACAGTG ATCCGGCTAT GAAACAGTTC CTACTCCATC TGGACGAAAC GCAATCACTG
     661  GGGAAGAAGT TTATCATACA AGATTTGGAC GAGAAGCACC TGTTCATTTC GGTCGAAATC
     721  GTCGAAGTAC TGCAAGCCCG CGTGGACGAT TTGATGGACC GCATCAGCTT CCCGCTACAC
     781  GAGAAAGAAG CTTAAAAACT ACCATCCTGC CGAAGCCGTA GCACTCACTA CCAACCTGTA
     841  AACAAACTCT CAAACTCAAC TCCAACAAGA TGGCCGGTCG CGAATCGGGA CGTATCCGGG
     901  AGGCGGACAA AAAGCGAACC CTGGACGATG CCGCACGCCA ACGACGCGCG CGCAAAGCGC
     961  TGGAAGCACT CGAGCAGGAC AACTTCCACG AAGACCCGCA CGCGGACCTG GTCATGTCGA
    1021  AGAAGGTGCC CAAATTCAGC GACAATCTCG ATGGCAGCGG ACCGAGCAAG AAGAAACACC
    1081  GCCGCACGAA AGGGGCCGAA TACTATCGCG CAAAGTATCG TAAAACGTTT CCGCAGCTGC
    1141  TGGATGAGGA ACGGCAGCAG CGACCGGATC CGCCGAACTA TTTCAGTGCC GCTGCACCCG
    1201  CTTCCCGGCT GCCCGAGCGA CACTTTTGTG CGGTGTGTGG ATTTCCCTCC AGCTACACGT
    1261  GCACGGCTTG CGGGACGCGG TACTGCTGCG TACGG

//

LOCUS          2     1322 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      2
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             join(500..666,753..822)
                     /translation="MSEVATRIPKPVLRGLHNATIKRNLIVSGVLCTIAVVAMKVMH
                     NDPKVRDYAEFYKTYDANKAFQRMKDAGLLQSVKD"
ORIGIN
       1  ATTAGTACGA AGCTTTAAAA TAGCAACATT TTTTTATTGA TTTTTATATA GACTTCGAGG
      61  ACATAGCACC ATTGCGCACT TCTGTTATAC AAAGCAAACG GTCCAATCAA ATTTGGTTTC
     121  ACATGTTGAG CGATCTTGCT TGATTGGGTT AATTTTCTTG CAAATAACAT ATTTTATCGT
     181  TTTCATTTCA TCGCTATCTT TGCGTTGACA TTACATCGAG TAGACTTCGA CTGTCATTCT
     241  TGTCTTTCTT CACACTCTTG TGATTTTCGG AACCAGAATT CGGTGCTGTG AAAAAACAGA
     301  ACACACGAGC CAAGGTAAAG AGCAATTATT TAACTCGATT CCAGTTCCAG TGCGCTTAAA
     361  ACACACGTGA TTTACGTGAA GATTTACGAA TTAATCATAC TGGCAGTGGG TGGGCCAAAG
     421  TGTTGGAAAA TGCGCCTCAC GAAGGATGAC CAAATTACGT AATGTAAATT CGGTTTGTTT
     481  TTTTCTTTTC TTCTTGAAGA TGTCTGAAGT CGCTACCAGA ATCCCGAAGC CCGTGCTGCG
     541  CGGCCTGCAC AATGCCACCA TCAAGCGCAA CCTGATCGTG TCGGGCGTCC TCTGCACCAT
     601  TGCCGTCGTT GCGATGAAGG TGATGCACAA CGACCCCAAG GTGCGGGACT ACGCCGAGTT
     661  CTACAAGTGA GTGCAAAATG GTGCTGCGAG GATTGGTTGC GTGGGCCCGT CCGTTCCTGG
     721  TCTAAGTTAA TGTGGTTACA TTGTTTTTGC AGGACCTATG ATGCAAACAA GGCGTTCCAG
     781  CGCATGAAGG ATGCTGGTCT GCTGCAATCG GTCAAGGATT AGATACGGCT GCCGCAGCAT
     841  ACTTTATTTA GTAACAGAGA GAGAGAGAGA AACAAAGAGA GAGATATGTA TGATTCCTTG
     901  GAGACAGTTT GCAGTTGATT GCAATCCAGT ATCTGATGGT AATCGTGCGA TAAAAATAAA
     961  TTCCGGTTAA ACCAAACGAA TGAGTAAGGG TTTCCGCTCT CGTGTAAGTG TAACCGGGAT
    1021  AATGCTAGTT TGTGGTTGAT TTATCATCAT GTCATCATGG GTCAATCAGT TCAATCTCGT
    1081  CTTCTCCACA TCGCTACGAC TATCCATGCA ATTAGGCAAA AGAATCCAGC ACAGCCAAAT
    1141  TTGCTGGTGC GGTCTTTTAT TAATTTAGTA ACTGCTTCCC AGCATAGCAA GGGAGTGCTC
```

```
    1201  TAGAATATGT CGTATCCGAC GAAAAACAGA CGTAAGCTCT TGCAAATCCT ATCAAATATG
    1261  TCAAACATTT TTTAGAACAT CTCAATAGAG TGCCATACGT TTGTCTTCCG CGCCTGCCCT
    1321  AT

//

LOCUS          3     1339 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      3
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          complement(join(500..630,731..839))
                  /translation="MIKLLDRVLSFINYWWFRYLMITELYMVESWERVTIHVFLFAI
                  FLAQWYFNCKVILPFTGNLLGIQPVDQHIASTLPRS"
ORIGIN
       1  ACAGTGACGT TGTCCGGCCA GTTTTCATGT TATCAATATT TCCCACCGTC TCGCACGACA
      61  CGAGCAGTGG AAGATAGCGC GGAACTGAAG CTCTCCTGTC ATAGATACAT TCAATACGGT
     121  TGGTATTGGT GGTGATAATG GTACGGCTTC TTGAGCAATT GGTCATCGAT AGCAGCCAAA
     181  TAGTACGGCC TTTTGGTTGG GCTGTTTTAT TCCATTGCTT TATATTTCCA TTCACATTCA
     241  CCAAGCACAC AATAAACACT GAGTGCGGTA GAAAACACAA ACACAAAATG GAATTTTCTC
     301  TTAAGCGAAT TCAGAATACA GAATACAAAA AGAACAGAAA AAGAACATCA GATCCTTGTA
     361  CAGTAGGATG AAAAGTCGAT GAGATTTAAC AGTGGTTTTC AGTTGTCCGT AAAATACTTC
     421  GTAGTGTGTT TCCCTGCGGG GCTACATTCG CTCCACCGCT TTGGGGCTTT CGGGAGGCAT
     481  CATCCTCTTC GCGCATTGTT CAGGAGCGTG GTAACGTGGA AGCTATATGC TGATCCACCG
     541  GTTGAATGCC CAGCAGGTTC CCGGTGAAGG GAAGTATCAC CTTGCAGTTG AAATACCACT
     601  GGGCGAGGAA TATGGCAAAC AGAAACACGT CTACGGGGAG GGAAAACAAA CATGGAAGCC
     661  GAACGTGAAT GTTAATGTTT TTGCTTTTGC ACGATGACCT AATTGGGTGC ACCCCGTGTG
     721  GCATACTCAC GTATCGTGAC TCGTTCCCAG CTTTCCACCA TGTACAGCTC CGTTATCATC
     781  AAGTACCGAA ACCACCAATA GTTTATGAAG CTCAACACCC GGTCCAGCAG CTTAATCATG
     841  GTACAGCAGG GATGGGCGCT GAAGGTGTAT GTTTACCGGT GGGCACAGTT TCGTTCTCAC
     901  CGGACGGCTA TGCTGGGCGG TTTTATGCTT CTAGAGTTAT CTATGCACTG TCGAAAGCGG
     961  ACTTGATAAA TTGGGTGTTT ATGGAGAAAA AGGAAGGAGA AATGAATACA AACACATTGC
    1021  CCGCGTTTGA CAGCGGCTGA GGCGAGGTTT GTGCACTGTT GTAAGGTTCT GCGGGTTTCC
    1081  GTTTATAGCT GAGGGTTGAC GTAAAAATAG ATTTAAAATC TCTCAATTTT ACACTAATCT
    1141  TCCCATATAA TTACAATCCA GTATTATTTT GTATGATAAA CCTATCTCCA AACAGTTAAA
    1201  ACATATTATG TTTAATTATT TCATTCGCTC CCTGCTCCGA GGCCATTTTT ATTGCAGTTC
    1261  AACCGGCGCA AGTTCATTTC CATTGTTGTT CTGTCCAGTG TGGCCAGATT ATTTTGGCGG
    1321  TTTTCGGCAG GAGCAATGT

//

LOCUS          4     11366 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      4
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          complement(join(500..591,9085..9205,10840..10866))
                  /translation="MDPRKLYGPSHFNWPELAYPSWFERARSMGEEHRALLSSPVPG
                  SPCDTHLFKCAHFQGLPEHRSPMCLLSTSLSSTACG"
ORIGIN
       1  AAATCATATT GTTCCACTTT CCGAATTCGA CCAAACCGAA CCAAATGCCT TCCGGGGGGT
      61  GCGTTACTGC GTTCTGTGGC AGACAGCACA GCCACACCAT CACCACGCAA AGACAGCGCG
     121  AACGCCTGTG ACGCCCGTTT CGGGCACACA CCATTTGGCA AGGACCGGAA CAGTTTTTCC
     181  CTCTTAATTT GCACACCATC TGTCGTCGCC GATGTCGTGC AGAGGAGAAG GGCTCGTGCG
     241  TAGTCGTAGC AGGGAGTGGA AATTATAAAA CTGTTACAAG TTTGCTATTG TCCGTTAATC
     301  TTTCGCATTT TTCCCTCCCA TCCGTGTACG TGCGGGCTCA TCTACCGTCG GCTACAGTAC
     361  AGTGTTCCGA TGCAGTTCTT CACCGCCCCC CAAGGGAGAG CGTGTCCTGG GGTGTGCTTA
```

```
 421  CGGTACGCTG GGAGTTACGG TACACAAGTG TCGGGGCAAG TGGATGACAA ATTTGAGGAA
 481  AAATATAAAC GGATTCTCCT TAGCCGCACG CCGTCGATGA CAACGATGTG CTGAGCAAAC
 541  ACATGGGACT CCGGTGTTCG GGTAGTCCTT GGAAGTGAGC GCACTTGAAC ACTTGTTGAA
 601  GGACGGCGAA CAGTTGCAAA CATATTTTAG TACACGCAAT TGAAAGTGTG TGCTTCCCCA
 661  TTTGCTCAAT AAAATATCCC ATGAATCTCC AGGAACGAGC GTTGGGCTGG GTGAATCGCG
 721  AATTGTGTTG AAAAACAGAA CTCACACAGA TGCTGTTCCG CTGTGCCCGA ATGCCCATCG
 781  GAACACAGCG ACGTTGGATT TTATATTGTC ATTATTTTAC GATTATTTAT CAAACTGTTT
 841  GCCTTCCGAG GAGTGTCCTG CATACACTCG AAGCACACCC TTTCCAAAAA ATGTGTCATC
 901  CGGCACACAG CAGGAACAGC GCGGCGAGGC GGGAAGTCTT CTTTTGCATT CCCCGGGGGA
 961  ACGGGAAAAC GATTCCCTAT AGCGCCTTCC ACAACCAACG ACTCGTGTTG TCATTCGCAA
1021  ACCTGGCAGG ACAAACATAC ATACAAGGAG CTCTGTGATG TGCTCATAAA CGAGCGGTAT
1081  GGTTATAAGA GTATTAGCAA TTATGATTAC CGATTCGCAT ACCCAACGGG ATTGACTTTT
1141  CCATCCATCC AGTGTTGACG TCGCCGGGGT CAGCAAAGCC GAGCTGAGGA CACACTGAGG
1201  ACACAATGGG AGACCCCTTT TTGCTTTTTA TTTGGTGCCG GATTTTTATT GACACCTCGG
1261  TAGCAAAATA TCCGTCACAA GGGAGCACAG TTGTCAGGTC GCTCGCTGGC TCTCTGTCTC
1321  TCTCGGGAGG GAATTCCCTC CTGCCGTGAT AGAAGGATGA TTTGGTGTTG TGATGATGAG
1381  ATTTCAGCAT GTTCAGGTTC GTATGCAAAC ATTTGGGTCG ATTGGTTGTC AACATTTTAT
1441  GACCATTATG ATTACGAGCA ATCGTCCACC GGGTGTTTAA AACGGTCATA TAGTGGCCCA
1501  AGGGTGTGAG CGCTGATTTA AAGGGTTTGG TTGGAGGGAA AACTTTGGTG CATATACAGA
1561  GACTCTGAGT TTACATGTGT TTCTTTCATA TTTACTAAAC GGTTGACCAT TGAAGAATCA
1621  ACGTCCGTCT ATCTAGCTAT TTCTCAAATT CCTGTTTCGT TTAATAAAAT GCGCAGTTTG
1681  CAAAGCCTGG TATGTGGAAT TTGCATAGTA TAAAATAGAA AGGAACGAAA AAAAAGCAAA
1741  GAGAAACAAT ACACAAGCAA TACAAACTCA TAATAAGATA ACAATAATAC AAAACACAGA
1801  ACAACAACAA TAAAATACAG AAAAATCAAA AAAAGAAGAA CAATAATGCA TAACTCTAAA
1861  GCATAAGCCG CTAACAAATA AACGAAAATA ATAGATACAT AAAGAAACAA TACAAAGCTA
1921  AGTAAAATCA ACGCAAAAGT ATTTAGGAAA AAATAGTATA AAATAGCTAT ATAAAGGCTT
1981  TGTAAAAATA AGAAAGTAGT CAATAAGAAG ACATATACAA TATTGAAAGT ATCAAAATTT
2041  TAATTTAGTA AAAGATCAAT CAACAAAAAA TTCAAAAGTG CAGAGAGAGT GTAAGAGCAG
2101  AGAGAATAGA GTGTAAATCA CAAAAAACCC GACAACATAA ACAATAGCAG CTGGTAAAAA
2161  GATTTAACAT TATAATCGCT TACAATTTGT TTACAACTGA CATAAGCAGG CAAGCTAAAT
2221  GACCACAACG AATAGCAAGG TGTGGAAAGG ACCCGCTGTT GAACAGTCCA ATTTATCAAT
2281  GTCCAGGAGG GATATGGCAA CAAAATTGCG AATGTACGAA AGGGTTAGTC TAAAGGGGTT
2341  GGAATCACCG AATATCCGTG CTAATCTCAC TTGATCACTG TTATTGCTCG TGTGCCAATC
2401  CCTATCGTTG GACCAACTGG TATTAATTTA AGTTTGGATT TTTGAACTAG GACACTGTCT
2461  CTTAAATCGC TAGGACCCTC AAATTTAGTG ACAAATGTTT CATTTTTATG TCTAGCTTAG
2521  GTCTCAGAAT AGATACTCTT GCTCTGTTGC TGACATGGTC TCAATATCTC AGCACATCAG
2581  TATCCTTTGT GTAAAGAAAT TGGATAAGAC GTCAATCGTA TGCTATATGG ACCACTTTAC
2641  TTCACTCCAA ACGCTAAAAA ATACATTTTT TCTTGAAAGA ATATCTCATT AAAACCATTT
2701  TGCCTTCAGG CAAAACAAAA CCAAACTACC TCATCAACGC TCAACATCTC TATCCACCTC
2761  CATTCACACT GCGAAGCTTA TTAGAAGCAA AAATGCTACA CATTCCTTCG CCCTTCGGCT
2821  TCAAATTGTT TTTCAAATGT GATTTTCGCC TGCCACACAT TCTACCTTTT TTCCAAGATC
2881  TTGCTACTAA GTCCACCGCC ACCACTGGCT TCGAGCTACA GTGTGGTGTG TTTAATGATA
2941  CTTTGTAACG CTACCCTTGG ACGCGTCTAA CTGACATGCC AATCTACACA GCTCCCCAGC
3001  AGCAGCAGCA GCATCAGCAG CATCGTCAGC CTCGCACTAT CCCACGCCAT ATCGTCCTAT
3061  ACGGTGGTGG AAAAACTAAA ACCCCCAAAC CTATACAGCG CATCTGCGCT TCGGTGTCTA
3121  CTTTCCCATG ATTTCTTCAT CAGCGGTGCA AAACGGTGTG TATGACAAGT CAAAGATGAC
3181  ACGTGTGTGT GTATACACTT CGAGTGCGCC CTTTTGCTCC CCACGTTCAA CACAGGGGCA
3241  GAGTAGTGAG CCGCATGAGC GAAAACAATC TGTCACCTGT CATCTATCGG AACACGACAG
3301  GGAAATGGTA CTCCTCTGCT CTCCTCCCAC CCGCCCCCAC AGGAATGCAC ACAGAGTGGC
3361  CCAAGGGGCA ATTGCTACAA ATGGTCGCTT TTCCGGACGC ACTCCAGGCG GTTGTGTTGT
3421  GGAAAAGTCG TAAAATGTAG CAATGAAATA CACGAAGAAA CCTCAATAGC TCATTGGCTC
3481  AGTAGGCTGT ACATTCCACA ACGAAACACC TTGGCAACAC TGTACACAGC CCGGGCTGAA
3541  GGAGGCCCTA CCAAACGCCC GCTTATTGTT TAATTTTCTC GCTCTCGATT GAATCGAAAC
3601  GAATCGGTAT TCGACATCCT TCAAGCTTCA TTCACATCTC CGCCCGAGCA ACAGATCGTC
3661  TTCTTCTATC GTGCCCCGAG TGAACTCCGG CCATACATAC ACACGCGCGA GCTTAGACAA
3721  TGTGTTTCGC AACCCCTTGT ATTACGCTAC CGCCACGTGG ATTCGTTCCG GGCGAAGTTT
3781  ATTAAACACC AGAGCGCTCT AGTCAAACAC AGGCGAGCGT GGCAGCCGTC CGGAGTCCGG
```

3841 TGCTGGGTGA AGCGAAGGTC TGGCGAAAGG CCCAAAACCG GGGAGGCGCT TACCAAGTGG
3901 TGTGTTTACA AATAAATTCA AATTAACACA CACGCACCCG GTCCCCTTTG GCAGAGCTGA
3961 CCTTGCCACG TTCGAAACGG TCAGTCTCGA AGAGGCGTGA AAAACGGAAA ATGCTTCTCT
4021 GGGGGGGTTT GGGATGCGCT GCTTTCGCTT TCCCATCGAG GGCGCACCGG TTGAAGGATG
4081 CAGAATCGCA CTGTCCGTCG GGTCGTTGAG TAGGTTGCAC ATACCCAGGT GCAGGTTAGG
4141 TAAATACATC TTCCCAGTAC CGTCGCTGAT TCCAGGATCT ATCCGACCTG GGCCACATTA
4201 CACGCGTTGA AACATGGGCG TTGTTTTCCT TGCCTGCTTC TTCTCCATCT TCCACTGACT
4261 CTGTTTCCCT CTTTCATTGC ACGTGTGCGT GTGTGTGGGT GTGGATTGTG TACAGTTCTC
4321 CTGGAAATTC CCCCCCCCCC CCCCTCAAAA CCGGAACGGA ACCTAGAATT CGTGAATTTT
4381 TACGGATGGT CATTGGAAAT CGATTCGCTC TGGCATCCGC ACTTCGATTC CACATCCCTG
4441 CTTCCCGGGG GCGGGGCTGC GTACGCGTCC AATTGGTTCC GGTTTCGGTT TCGGCACCGG
4501 AAAACTCACG TGCGTCAGCA AAAAGTAGCT ATGAAGCCCT CCCTTCTAGC GCGTAGGACT
4561 TCTCGTTCCG AATTCTTGCC ACGGTTTTCG TTCTGGGTTG TAAGAACGCC AGCCAAGCAA
4621 AAGTGCATTC GAGTTCGTGG ATTGAAATGC GTTACTTTCG GGTGGAATTT AGTGGAAGCA
4681 ATGATACGCA TTAGGAACAT CAGTATCATC GCGACGAATG GAAGGAAGAA GCGTAATAAA
4741 GTTCCAAGAA AGTGTGAAGT GGAAATGTGT AAGGATTTCC ATCGTCGAGT ATTCACGTAA
4801 AGATGTTTCT TGTCTTGTCT TGACCTCACC ATCGCTTGTA GATTGGCTTT AGTTTGGCTA
4861 TAAGATACAC CGAAGGCCAC TATATGTATT ATAACATATT GAAGAATCTT TGTCTCGGTA
4921 GCTCCATTGT AAATCGTAGC TGCCTTTGTA GAAATCTTAG ACTTTGAACG ATAAGACATC
4981 ATTAGATTAC GAGATATGTC AACTCCAAGA TGCATCAATC AGCTTGTAAT CTTTAAGTTT
5041 AAAAAAAGAG TAAAGATCAT AATCAGTAGA GCCTTTGGAC GAAGTTGTGA AGGTGTCACA
5101 CACATATCCC AAATAACGAC TTGCTGTAAA CCACCTTCAT AACATTTTTC ATTCACATAA
5161 AATTACATTG AAATTACTCC ATTAACCCAA CGTGTACACG TGCGAACCCT TCGCAACGAG
5221 CAACCTGGTC AGCAATGATT GCCTGTGATT CAAGTCAAAC CCCCTATTTC TAGTAAAACA
5281 ACGACACAAA TCCTTTTTTT AATCGCTATA CAGTCCTGCA CCTGTCCATA CATAACCACG
5341 TCCACCAAGT GCAACCATAC CGTGCCGTAT ATATTCCCTA TGCACCACCC GCACCATTGC
5401 ATGATCCAAA CGCGTAACAA ACTATCGCAT CGAACCTTTC CATCGGGAAG AAAGAAAACA
5461 GGGCCCGCGC GATGTCAGAG TGCACCTCGT GAGCACCTAT CAATCAACAA ATAATTACTT
5521 ATCAATCATA GACGAGGCCG ACACAGACAC TCAAGAGAAA GAGAGATAGA GAGAGAGACA
5581 GCGAAACAGA GAGTTCGAAG GATCGCACCC GGAGTTGTTG TTTGTTGTTT GTTGCGATTG
5641 CGTTTGTTTT TCCCCGTGCG TTGCGTAGCC CAAGGCACAA GCTAAGGGAA AGCTCATGGC
5701 CCTCTGCGTG TGTTTTTTTG CCTAGTTTGC CTTTCCATTC TAGCCGAAAA CGTTCTGGAA
5761 AAAAACGCAG AAACCCTTCA ACAGCAGCGG CGTAAGGTGG GCAAAGCAGG AAACTAATTG
5821 CGCACAATTA ACTTGTCGTC GGTGTCATTT CGGGAAAACA ATTCCGGAAG CGAGCTTGAA
5881 TATAAAAATA CACACAAACA CAAACAGCAC AGAACGATAG TCGACAGTTT CTGTTCTAGA
5941 CAGGCCAGCG TTTTACTTTT CAAGACAGTT CCGATTCCGG GCGAGACACA TACAAAGGGA
6001 GGGGGAGGAG GTAGACAGGA AAGAAAAAGG GAGCACAGTA TAGCACATAA AAGCCCATCG
6061 GGGCTAACGT TCCGTTGACA TGCAGGTGGG CAACCGGATC GGTTCACGCG GCATATAGCA
6121 CCACGCGCCG TGTGCTGGGA CATGATGGCT GCGCACACAT TGATGTTTTG GTCCCCACTC
6181 GCCCGGGTTT TGCGGTTTGC TGTGTGTGTG TGTGGCTGTT CGATGATATA CGACGATGGA
6241 TAACGGCCTG GTCGTCGGAT GATATGTTCC GGACGTGACG GCATCAACCT TTTACCACGA
6301 GCCAGCACAC CCAGGCCAGG CCACCCAGGG AGGGTGGGTT TGGTTTGGTC AGCTGGCCGG
6361 GGCGGTATTG CGGTATTGCT ACCGGGACCA GAGCCTGCGC CTGCGGGTCC CGGCCACAAA
6421 CAGTGTCATG TCTGTCATAC TATTGTTGCT ACCGTTTTTG TTGCTGTGGC ACAGTACAGT
6481 GGTCGCCAGA ACAGTGGGCG CTACATTCTA GGGGTAGACA AAACGTGGTC GGTTATATTA
6541 AAGTGCAACG GAACTTCTAG TTGCACATCT TTAACTACTT CCAGATGAAT GGAAGTTCTA
6601 GATGCACTCT CTTCATCTCA AAAGGTCAAT ACTAACACTC AACACATTGT TACAATTTCC
6661 CAAGGATTTT TTTATACGCC ACAAATAACG CAAAAAATCC CGTCCCGTCC ATCAAAACCT
6721 CGAAAAGAAA TTAATCTTGG TCACAACACA CACACTCAAC GTACAATAGT ACCTCTGAAT
6781 TACATGAACA CAACCACGCA CGCATCCTTC ACATGCCCTG CTGCCACTAT TGTGTCGCTC
6841 AGTGTGTGGT GTCTCCAGTG TCTGCCCCCC CCCCCCCACC CCCCCTTTGC GTTCGGCTCT
6901 ATCAGTCATT CCGTTCGAGC GCTCTTGCTT GGCAGGTGCC ATTGCCGCCT CCTTGCTGCC
6961 AGTCCACGGG CTCGGGACGA ATGAGAAATG GCACCACCGC CTGAAGGGAA ACGCGTCGAT
7021 CATTGAACAT CGATGAGCTT CTGGTCTTAT TTGTCCGCCC GTTGCCCCTG TTCCATGCAT
7081 CGCGCACACG GTAAGAATGT GTTTTCTCTG CGGTCGTCGT TTGCGTTTAT TGCAAAACGT
7141 CGAGGCGCAA CAACAACAAC AACAACAGAC CGCAACAGAC GCAGCCACAT ACGCAACGCC
7201 AACGCCATCT GTCTTTCACA TATGTAGCGT CCAAGAAGGG GAAGAGTTTT GTTTTCCGCA

```
 7261  AAAGAAGAAA ATGATTGATT TTCTTCAACA ATGAGTCGTT CTTTCTTTTT CATTGTCCTT
 7321  CAGCTTCTTC AAACAGAGCT CCGTGTTTGT TGGCACGATG GTTGTGCCGT GTGAATTAGC
 7381  CATCGCCACG CCAAACCGGA GCAAGGAAGG ACATCTTTCT CTCGCTCCGC CGACACGCCG
 7441  CCGAAGGAGA CGGTCTCCGC CCGCCGCCAA AGTAGTCCAT GAATGATGGA CCTCTGTCAC
 7501  GCCACGCCGG ACAGCCGACA GAGAGTGGTG GGGAAAACCT CAAACAACTC ATCATGTAAG
 7561  CTCGTACAAT GTAGGAGAAA GGTGGTGGTC TCGGTAGGCG CTGCTGTGCA GGTAACCCCA
 7621  CCTCCCATTT CGTGACATAG GCCAAGGTAG AGGTTTGTTA GCACGCAATC AACACCTTCG
 7681  CCACTGTCCT GAAGGATGTA GTGCGCGGGA CCCAATACAA AGCGGGACAA TACAAAAGCA
 7741  GCCACGTTTG GACATCCGCG CTAAACACTG TTTTGATTGT CATCGAAATT GTTTGGACAA
 7801  GAACAGAAGG GAAGAACCAT CCCGGACACC AGTCGTAATC CATTCCAGTG GTTGCTTTAA
 7861  TGGGTTGTAG TTTGTACATG TTCCGATTCA TTTGGAATCG ACACCAATCT AAACCTGGTT
 7921  AGATAGGGAT CAGCAGGAAA CTCCACAAAA AACACCTTAA TTAAAAATGT TTGTTTTTTA
 7981  CTCGTTTAAA ACGACATTCC GTGTTAAGCT GCAATAGCGA TTTATTTTCC CTACGAAGGA
 8041  TCATGTCGTG TACGCAATGT AAGTCCAGAT TAGATTGGAC TTAAAATTCA ATGTTTTATT
 8101  TATTCAACCT TTCTAAAACA ATCTCATACT CAAGTTAACA TTCTACAATG ATTTAAAACA
 8161  AACTCAACCC CTAATCGATT GAACCGTCCA ACGCAAAAAC CCACACTGAA GGCTAACTTT
 8221  TGCATTTTTT GCACAAAAAA CCGATTACTG CACCCGACCC CCACCGATTA AAGCGGAGAG
 8281  TTGTTCCGCA AAGAGTTGTA ATGCTCCAAA CTTCTCACAA CTTCTGTGCA CCAACAGCGT
 8341  GTGCAAACGC AACAGGTCAA CCGTTAGCGC TGACCTTTTC AACGGCAAAA CCATACGCGA
 8401  GCATATTAAA CGAGGCAAAG AGGGAGCAAA CTTCTTTATG GCCATCGCGC CCTCGTGTTG
 8461  GGGGGTTTCG CTCTCCAATT AGGCCAACGG GCGCGCGCCA AGGGGTTTTT TTAGTGACAA
 8521  CCGGTTTTCG TGAAACCGGC ATGATACGAT TGCTGCTCGA ACAACTATCA TTAGAACGGC
 8581  ACACCACCAT TTTGCATCGG CAATTTGCTC GCGCTTATTA TGATGTTTAA CTTCAACACC
 8641  TAGAAGCTAA TTGACATCTT TACGTTAGAA CCACTTAATA TACAGGCGCT CCGGGGCCTG
 8701  CCGGATTGTG ACCGTGGTTT TCTCACGGTA TAATTTTTCC ACCCCTCCCA ATCGTTTATT
 8761  TTGTAGCTTT AACACACACA CATGCACCAG CGGGGAAGAT TAAAGGTAAA GCTTTATTAT
 8821  AATTAAATTT TCGCAAAATA TCGAAACGAA AGCGAACGCG GGCGGCACAT TTTAAAGCGC
 8881  GCATAAAGTG TGGCATCAAT CTTTACGAGC GGTACTAATT TTGCTTTACT ACACCAATTT
 8941  CCACCAACGT ACAACAGCCC GTCCGCGCAC ACTACTAAGG CCAGGTCCGG GACAGTGTTC
 9001  TTTATTAGTA TAACGATCAT CATCATCAGT TCATCGTGTT CGACATAAAT GCGCATAGTT
 9061  TTGCAGGGTA ATTAACCGCG TTACGATGCG TATCGCACGG CGAACCAGGG ACCGGCGAAG
 9121  ACAGCAGAGC CCTGTGTTCC TCGCCCATGC TGCGTGCCCG TTCGAACCAG CTCGGGTATG
 9181  CCAGCTCGGG CCAGTTGAAG TGTGACTGCA AAATTTGTAC CGACACTTGT GCCTCTTCCG
 9241  TGACGAAACG AAAGCGGCGA GAGTTTCTTA ATACAACCAG AAAGGGTGCA AAAGAATACC
 9301  CTTCAGTGCT TCAATCCACG GCCAGATTCA CAAGAATGCC ATAAAAAATG CCGCTCCGCT
 9361  TTACCGGTAC CCTGGGCCGT ACCGAAATTC TAACTCCCGA AGGAACCTTT CCCCCGGTAG
 9421  CATTCCCCAT GGGAGTCCTT GCTTTTGTTG CTTTGTTGGA AGGGATTTGG CCTTTGGCGC
 9481  ACAGATAACA TTCAAATACG ACGCATCTGT CATCATTTAA ACGACACTAA ACAAACACTG
 9541  CACCACCTTC TGGGGGAGGC GAATCTCGAC CACGCTCTGC GGGTCCGAGG AGTTTTGGGT
 9601  GTGGGTTGAA TGATTGATAC GCGATACACG GTGCGTGTCG GACGGGCGGT CGGTTGGTCC
 9661  TGTGTCCTGT CCTGTCCCGC GTTCTGTGCC CGTGATAAAG CCTTATCAGG GTACGGCCAC
 9721  CATGCCACCA AACGGCTGCG CTGTCTCCTA TCGACAAGTG GCCCCGACAG AATCGCTCGC
 9781  CTGCAACTGC TTTAATCGTT TTCGCTCGAC CCCCTCGGCC GATGAGAATG CGAACGATTT
 9841  CCACCCCCCT GTCCCTCACG CGTCACCACG CACCACGTCA GAGAGCGAGC AGAGTTTGGG
 9901  TTGCAGGGAA GGAGAATCGT AGTAACCGAT GCGTGGATAA CTTTTTACCG ACGTCAACAG
 9961  CAGTAATCGA AGGATGGGAG AGGATCGGCA GAGGAGGTAA CACAGTACCC AAAGGGGATG
10021  ATGATTCGTG CGTGCGAGGC AGTTTACCAT GGGTTGAGCA AAACATGTGC TTGACTTGTT
10081  GTTACAGTGC GTGCCGGAAA AAGGGGAAAC GTATGATAGG CTGATGAACA AGGATATCAT
10141  AGGTAAGGGA AGGGAAGGTA GTAGTAATTA TAAGGCTTTT ATAATTTAAT ATTAAACCAA
10201  AAATATCTTT TCAAGGATAT TTTATGAAAT GCATCACCAA GTTAACAACT TTCATGTGGA
10261  ACTTTACGGA GCTACAGTGA TATTTCCCCA AGATGTATCT TCAAGGGACC GTCAGCTTAG
10321  AGAGATTTTT ATGTTAAAGA AAACTAATGA AGCTGGTGCT ATGAAGTATT CAAGAAATCA
10381  CGAAAAAAAT CTGACATCCT TTTACCTATA TTTGTGTGAA CTATCTCTCA ATATTCATCT
10441  AGGTTACTAT ACGCTGTCGA ATTACATGTG CAACTTAGAG TATATTCACC TAAGGGGATA
10501  TTCATCTAGA AGAGCCACAA AAAGAATGGA AGGTGATAGG ACCGTCAAAA TGTTAATCTT
10561  AAAGAAACAA ATAATCTTGA TCATACTCCA TCTGAGAGAG ATTTCACTGT AATTAGTTTT
10621  GACAAACTAC AGAACTAAGC AGAAAAAAAT CCGTCTGTGT ATATTATAAC TCCAGGATGC
```

```
   10681  TGAGTTACAG GATAAAAAAA TGAGTTCTCA GGCTTCTTTG CAATCCACAG GCCTGTTTCC
   10741  AAAGCATGAT CCGAAGCAAA ACTGTTCCTA AGCCCATATC GGTTTTGGCA CTGATCCTAA
   10801  ACCCAAACCA ATACTGAAAC TGTTCTTTAA AACTAATACC GGGCCATACA ATTTTCTTGG
   10861  ATCCATGTCG CTTCTTAAAC TGAACCTCAA TAAATATAAA TAAATAAATA AATAAATAAA
   10921  TAAATCTGCA ACCCATACTC GCCCAAGAAC TGATCTTGTT ACCATTATGA CACTGATACT
   10981  GAGCCCATAC CAGCCAAGGA ACTGATACCA AATAAATATA CCATTCCTGG ACCGGATCCT
   11041  GAACTCCAAT ATTCCTTACT CCAATACCTG AATTCTGATT TTTGACAACG AAATTGAGAT
   11101  TTTTCTAACT GGAAATTACG ACGAATTTTA TCTTTCATCA TATACAGTGA ACCCTCTCTT
   11161  ATTTGACACC TCTCCAATTT GAAAAAACCT CTCTTATTTG AACATTTTGC ACAGTCCTTG
   11221  ATTTCCAGTA CATTTTTGTT CCTCTAATTT GGAAAACCTC TGAAATCTGT GTCCCTCTTA
   11281  TTTGTGTATG GTGTTGGCAT GGTTATTGAA TGATGTATGC TCAAATTGGC AATCCTGTTC
   11341  GCAGTTTCCT ATAGCAACAG TTAAGG

//

LOCUS          5     1349 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      5
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..709,808..849))
                     /translation="MSEQRRRVLDLYKRLQYMGREYPGGPEKFRQRCYNAFKRQSTE
                     TDPEKIEKAIGLGEYVVKEIEALYSLRKYRAMKRRYYDEK"
ORIGIN
       1  GCCTGTTTTT CGAATGATTA TTTTGTCAAT GAACAACAGG ATGTTAGAGA CACTATGCAG
      61  GCGGGGATTA ATTAACGTTG TTACCTAATA ATGCCTTTAA AATTAACTAT ACTAAAACCA
     121  TTAAGTTAAT TAGCGTTTTA AACCAATGGT TATAGGCATC CATTTAGATA GGAAAGTTAC
     181  ATTTTGTTGT TTTTATTTGT AGTTTGCTTC TAAATTTTCT ACAAAAACTG AATTAGCTAC
     241  TGCGGTAGAA AAAAGGGAAC CATAAACTTT TCTCGTAAAC TCGTTTGTTG CGTTATTGTA
     301  CCCCACTAAC GACTATTTAC ATCGCCTCCT GTTGTAAGGA CTACGGTCAA GCTTCCACCT
     361  TCGCAAAAAT TGCTGCCGCC CAAACCTCGG ACCCTGCCAT TCTAAGGTCA TACCAATAAT
     421  CTCTCAATTT GTTATCAACA GCAGCAACAT CAGCGGCGAC GGCGGCACCA TCTCCAAGCG
     481  TATGGTGTAC CGTCTGCCCT CACTTCTCGT CGTAATATCG TCGCTTCATC GCTCGATACT
     541  TCCGCAGACT GTACAGTGCC TCTATCTCCT TCACCACGTA CTCGCCCAGC CCGATCGCTT
     601  TCTCGATCTT TTCCGGATCG GTTTCGGTGC TCTGCCGCTT GAACGCGTTG TAGCACCGCT
     661  GGCGAAACTT TTCCGGCCCA CCCGGATACT CGCGGCCCAT GTACTGCAGC TAAGGGATAG
     721  GATTGGATGG GAAGAAAGAG AAGCGCTTAA TGTTACTCTA TTGTTGCCTC CCACAGCACT
     781  GGGTCCTGCA CTCTAGAATG TTGTTACCCG CTTGTAAAGA TCCAAAACGC GTCGCCGTTG
     841  TTCGCTCATC GTGTGCAAAA AAAAAACCTT TGGCACGGTT GTTGCGCAAA ACTGAAACCA
     901  CGTAAAACAA CAAACAAATA GCAGCCAGCC AGCCAGCTCT CGCGCTCTCT GTGTGTAGTC
     961  AGCTGTCAAA AGTGTTTTGC TGATGCGCGA TGGCACGTAC ACTACCGTAC CCCAGGCGCA
    1021  CCAGCTGATT GAAGGGGGCA TGTGTGCGTG TGTGTGTGTG TGTGCTGTGT AGTGTCCGCC
    1081  GCTGGTACCA ATCGTAGAAG GCCGGGTTGT TGTTAAACAA AGCGCGCGCG TTGCCACAGC
    1141  GAAAATGAGG GCGACAAGAG TGCGGTGCAG TGGCGGCAGC AAACCAGCGC AAGTTGTGAA
    1201  ATACACAATT ACAAGTGCAC AGCAGCGGCT CTCGGTCGGG TTCTGGCAGC GCAGCGTCTC
    1261  TTCAAAGATT TGCATTGTGT CGAAGGAGAA AGAGAAACGT TCATCTTTTC ATCCAACATC
    1321  CAACCATCAT GGAGCTGCTG ACGACGATG

//

LOCUS          6     1350 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      6
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             join(500..629,714..850)
                     /translation="MWPLIMQFMRSNAAYITLPIATLVGVIGYNIESLLSDKYTPYS
```

```
                PSIQEKRAERITDDERLKAATDVEKLVYKENVLGRNLSPSLEKGK"
ORIGIN
       1  GTATGAGCAA GTGAGACGGC GGAGCGCAGA TGCTTATTTG TAGGATGTAG TACTCCGTGA
      61  GGTCAATGAG GGGCGGGGAG GTAGTGGAGT GTATTAATTT TTTTGGCCAT TTAATTTCGT
     121  CCACCAACAC AATGGCACAC AAGCGCGAGC ATTTTTGGCT GCCATCTTAA ACCTTCAAAA
     181  ATCCTCGCTC AAAACATCAA AAACCGTCAC AACTACTTAC AAAAAGCATC GCCAGCGAGG
     241  GAAATTCGCA CCGAATTGGA GCGTTGGCGA GCTCATGACA ACCGGTTTTT GGCCTGCGGG
     301  CCCACCAGCT GATTGTCAAC AAACAAACGT CAGCCACAAA CCGGATTGAA ATCAAAACAA
     361  ACCAGAAAAC AAAAAAAGGC TATTTCATCG TAGGTTCGTG TTCGCAAAAG TGTGCGAAAA
     421  TATTTAATTA ACTATTCGAT GTGCTTCTTT TCAGTGCTTA ATTATTTCTA AACCATCAAG
     481  CTAGCATTAG ACGCGCACCA TGTGGCCACT GATTATGCAA TTCATGCGGT CCAACGCTGC
     541  CTACATCACG CTCCCGATAG CCACATTAGT TGGCGTGATA GGGTACAACA TCGAGAGTCT
     601  CCTTTCGGAC AAATATACTC CATACAGCCG TAAGATAGAA ACGCTCCTGA ACACCTACAT
     661  GAATTTCAGT ATGATGCAAA TGTTTAGGCT AAATCTACAA ACCCCATTTT CAGCTTCCAT
     721  CCAGGAGAAA CGAGCAGAAC GAATTACGGA CGACGAAAGG CTGAAAGCGG CGACCGATGT
     781  GGAAAAGCTA GTCTACAAGG AGAACGTACT TGGGCGAAAC CTGTCGCCAT CGCTGGAGAA
     841  AGGAAAGTAG AATTAGAAAG AGCTGCTTGT TAAGTTTGTA GTCGTACGAG GAAAATATGC
     901  AATAAACCTG AGCGTCCATA ACAAGGCAAA GGTATGTATG TGTGCGTTGT ATATGTAAAA
     961  ATTATTTATG AAAATACAGG GTTTCCCACG ATTTATTGGT TGGTTCCCAC GATTTATTGG
    1021  TGCGTTCCCA CGATTTTTTG GTCATTTCCC ATAATTTTTT GGTAGCAACC CACGATTTAT
    1081  TGGTCGGTTC CCAAATATTA TTGGTCGGTT CCCAAATATT ATTGGTCGTT TCCCAAATAT
    1141  TATTGGTCGG TATTATATTA TATNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    1201  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNGGG AGAATGAGAA
    1261  ATTATCTCTT GAATATATGT TCTTCAATTT CACAATTTGC TGAGAAGATT TTAGAAACTG
    1321  TCATACAACA CGCTGTATGA CAGAGTGGCC

//

LOCUS          7     1350 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      7
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..604,686..850))
                     /translation="MSATLVNIAEEDFFGTVIRYGLYVGAIFQMVCLAACIVLPDSS
                     GGSGEPGSWSPKNNDSEDSSSEHSSPQNTPRRPYHRSRKQDKKKRR"
ORIGIN
       1  CTGTTTTGTT ATCTAACGTA CTTGCCCGTT CGTGGGGAGT ATGACGTTGA CCTGCCAGAA
      61  AAGGGTGCAA TGTGTTATGT GTAAAGTTAA TGAAAGAAAC TGTTTTCTTT TGCAATATAA
     121  AGTCTTGTAC TGGTAGACGG TGGATTTGTA AATGCCTTTT AACATGTTTG ATTTTAAATT
     181  TCGAATATTG AAAGAACTAA ACGCCTTTGG TTTGTGAAGT TAAATGTTTT ATAAGTTGAC
     241  AATAAGTACA CTGTTACATT TAGTCTCGGT TGGCTAATGA GTATTTTGTT TCACTACGGG
     301  ATGGTAGTAG ACGAAGATCT CGTTGAGCTT CCTATCGCTA AGTACGCAGT GGTATCGGTC
     361  ATTGCCGAAG TCGCCCGATA AAATATCGAT TAAAAGCTAG AGAAGGGACA AAACATCTAA
     421  CACATCTCAG ACTAGCGTTG GAATGCTATT AAAGTAGATT TTGCGTGGAA AGTGATCAGC
     481  AAGAGTGAGT GCCAGCAGCT TAGCGGCGCT TCTTTTTGTC CTGCTTGCGG CTACGGTGGT
     541  ACGGTCTTCT GGGAGTGTTC TGTGGCGAGG AGTGTTCCGA GCTAGAGTCT TCGCTGTCGT
     601  TGTTCTGAAA GCAGAGGAAC GAACCCAATT TGTTAAGAAT CAAACTGTCC ACTGTACTGC
     661  AAAAACACAC ACACACACAG CTTACCTTCG GTGACCAACT TCCCGGTTCA CCACTGCCAC
     721  CGGACGAGTC GGGCAGGACG ATGCACGCTG CCAGGCACAC CATTTGGAAA ATCGCCCCAA
     781  CGTACAAACC GTACCGAATG ACCGTGCCGA AGAAGTCTTC CTCGGCTATA TTCACGAGCG
     841  TAGCGCTCAT TTTGGTGTCG ACAAACTGGA GCGGCCAAAG ATTTGCTTTT ACCCTTTGGT
     901  GAAAGAATCG ATATCGATTT GTTTTGTTTT GCGCACTGTT TACGCCAAAT CGATTCCCAA
     961  AGAGCAATGA TGTCAGCAGT GTCTGCTGGA TGGCAGCACT GCTCGAGGAA TGTCGGACAA
    1021  AGTTGCCGTT TGAGCAGCTT ATCCGACACG CGACATCCGA CATTGAGCAA AGCCGCAACA
    1081  TGCTTGTGAA TTGATGATTT GTTTTTACAA TGTATAAATA GAGAAACGGA AAACAAAGCA
    1141  CTGAAAAGTA TAATTTTACC GAATGAAACG GTCAAATTTT TGCTCGGCCA ATGTAAGCAA
```

```
    1201  AACGCGAGAT TTTCTACAAG ACCACACATC CTAAGCGCCA CAGATTATCC AAGTGCCACA
    1261  AATCCGCTAA ATGAACACAA AATGGATTAT ACACGGCTCA AGCTCTCAAC CTAAACGGAA
    1321  TATAGAAAAA ACTTCGTTTA GCAGACGGCA

//

LOCUS          8     1477 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      8
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          join(500..545,618..706,816..977)
                  /translation="MADEDKGLVAGHPPAVKAGGMRIVQHKTPNSERPAKDPVEVIG
                  LSNPAPNVNTGEVAQSGTSTKHTDHSVEASQVAHAQKPPAPVHSKPVNHIQQPRK"
ORIGIN
       1  ACATTAAATT CTGTGCTAAT CCTTTCTAAT TATACGATAC TGTAATTTAC TCCAATCGAA
      61  AGCAAAAAAA ACATCTACGC CTCCATTTTC TGTTCTCGGG TTCTGTTTTT CTTCCCCCTC
     121  CCACAAACAC ACACGCACAC ACGTATACAC ACACATACAG TGCGCTGTAT TTATTTTGAT
     181  TTCTATTCCG CGCTTCGCAC CGTTTCTTGC CTTCGTGGAA CGGTGCTAGT TGTCAAAAAG
     241  AGCCCCCACC CCTCTCACAT CGCATCAACG CCATTAGCGC TCGCTTCTAA TCTATACGCA
     301  GCAAAATTGT AGGCCCTGTG TTTGTGTGAT CGCTCATATA TTTTCGGAAC CACGACGATA
     361  GATAAGAGAT AAGCGCCCGT GATCGTGCGT TGGCAATTCG AATTCAGTTT TGTGGGGTGC
     421  ATTTGAGAAG CGCAGCAGCA TTCGCCGCAA GGTAAACGCG TAGTAGTAGT TGACGCAAAA
     481  GACGCGCAGC ACAGAAACGA TGGCCGATGA AGACAAAGGA CTGGTCGCTG GACACCCCCC
     541  GGCAGGTAAG GCGCTACATT TGCTACCTTT GCTGGGTGGG AAATCTAATG CACACGTTGC
     601  ATCTTTTTTT TCTACAGTGA AAGCGGGTGG AATGCGCATT GTGCAGCATA AGACTCCGAA
     661  CAGCGAACGG CCTGCCAAGG ATCCGGTCGA GGTCATTGGG CTGTCGGTAA GTGGTGCAGC
     721  ATCATCCGTG TGGGGGGAGG GGACCGTGTC AGTAGGTTGG GAAACTGGTT TTGTGCAAAC
     781  AATTAATTCT AACTCACCCT CACTCCCTTC CGCAGAACCC AGCGCCCAAC GTGAACACGG
     841  GTGAGGTTGC CCAGTCCGGG ACGAGCACCA AGCATACCGA CCATTCCGTG GAGGCGTCGC
     901  AGGTGGCCCA TGCCCAGAAG CCGCCGGCAC CGGTCCACAG CAAGCCCGTC AATCACATCC
     961  AGCAGCCTCG GAAGTGAATT ACTCGCGGTG CATGCTATTG AACCAGTGGG GCGGTGCGGA
    1021  GTGGAAGAGG GACGGAAATG AACCAGCAGC AGCAAACGAC GACGCGGTGG GGGGTCAAGG
    1081  GGACAAGGGT GTCCGCCGGA TGCACGTGCA CATTGGCAGG TCCAATTCTT CTGCGCCCCC
    1141  TCAGCCGATC AATTTGCCCT GCTGGAACGT TTTCGGAGTC CGTGTTCGGC GAGTGATGAT
    1201  GGCCTTATCG ATAATCAGTA TCAGGGCAGA GTGTTTTTTC AAGTTTCTTC TTCTTTTTTT
    1261  CTTCTCACCA TCATCGGTAG ATTGATAATT GTCTGAATTT AGAGCGTTGC TTAGCGCGCA
    1321  CACGTTGTTT TTCTTACCTT CAATTTACAC GGAATTTACA CGGGAGAGCG GTGTTGTGCC
    1381  TCATCGAATT GAAGGCGTTG TAACATTTAT AGTGGTCTTG GGTTAGTTTT CCTTTTCAAA
    1441  TTTATAATTC TACTAATATT TTGTAATACT GTTTGTG

//

LOCUS          9     1532 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      9
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          complement(join(500..541,624..772,845..890,961..1032))
                  /translation="MADEFDACECFWSHELAMRRLMSLLRQGQSYCNDNECTDLPSL
                  PNVTGGANFFLIVMALIFVAVMYVLRPSSMRRSNDLNKSLPPPSNDGPNNDGGAPPS
                  IS"
ORIGIN
       1  AGTAATCACA ATATGCATTA TCTGCTCGCG CACAGCCTCT AGGAGACCGG GGCGAAGCGA
      61  TGTGGGGTGT TGGGGTGGGT TGGAGAGAAG GAATTTACTT TGAATTGCTT AACACAGCTG
     121  TTGGCCGGTG GTTGAACAAA GAAGTGGGAA GGTGCAACCA AAACGTTACA TTTGAACTAT
     181  ATTAGTGATG TATTGGAAGA AAGAATACAG ATTGGATCAT TGAAACGCTT CCGGGAGGGG
```

```
     241  TGTGTGTGCG CACGGGTTGT CGTTTGTGTT TTTTGCGGTA AAAAAGGAAC GTGCCCCCGA
     301  ACGTGCGTTT TCCATGTGTT GTATAATTGT CTAGAAGAGA AGCGGAGAGC GGGAAATGCG
     361  ATGCAACATT CGTCCGGAAC CGTACATACA CCTCCAGCCA CCACCGCGCA GTGTCATTTG
     421  GGATGGGGGA GTGGAAACAG GCAGCAAATG CTCAGGTTCG GGGTTGTGCT TTTCGTTGGC
     481  CGCTGAACGT GCGCTCCCTT CAGGATATCG ACGGTGGTGC ACCACCATCA TTATTCGGAC
     541  CCTGGGGTGG GGGGAAAGGG CACGAGTAAT TAGAACAAGC ACGAGAGGGG GAACCAGCGC
     601  CACCAGGAGC GCCAAATACT TACATCGTTC GATGGAGGCG GCAGCGACTT GTTCAGGTCG
     661  TTCGATCGTC GCATGGACGA GGGTCTCAGG ACGTACATAA CGGCCACAAA AATCAGTGCC
     721  ATGACGATTA GGAAAAAGTT TGCGCCACCG GTTACATTCG GGAGACTGGG AACTGGTTGA
     781  AAGAGCGGGG TTTTAAATAG TTAATTATCA ATTTAAGAAG TTATCATTTC CTCACTATAC
     841  TTACGATCGG TGCATTCGTT ATCGTTGCAG TACGACTGTC CTTGACGGAG CTGAAAATGA
     901  TGCAACACAG GCTGTAGATA ATGCTTCACT AACGGCTAAA CTCTACAATT GTATACATAC
     961  CAGCGACATC AACCGTCTCA TAGCGAGCTC ATGGCTCCAG AAGCATTCGC ACGCATCGAA
    1021  CTCGTCTGCC ATTTTTGCTG CGCCTTTGCT TAGTTACTTT GAAATTCGTT CAACCGTTTG
    1081  CAACTAACAA AACATCAACG TCAAAAACAC TTCCAGGCCA AACGCGTGTC TATTCCTCCT
    1141  CGTACACTGG ACTGTTTAGA ACTGGACAGA TTATAAAACG GCTTCTGGTT CCTTGCTAGT
    1201  AGTATCCCTA CAATGACACG AATGAAACAA AACATATTTA ACATAAATCA TTTTCATCTC
    1261  TTTCGCTTGC AGATCCGCGG CGCGATTACT TGGATTGGAG AAAACTGTTC TAACTAGCCA
    1321  CAAACTACCA CTTCGCGCAG TACCCGAAAC AAAGCGCGTT AACGTGTGCC TCCCACAGCC
    1381  ACTCGTATTG GGCCGTTGGG AAAATGCAAC ACAGCAGCCG CGGGTTTTTT CCACCCACTA
    1441  TTCCCCCCTC AGAGACCACC TGTGTATACG ATTCCGTACC GAATTGTAAA CACTTACGTG
    1501  GTCTTCTCAG ATTTGCGGCG GATGCGCTGT AC

//

LOCUS          10     26339 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      10
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS
complement(join(500..526,616..705,7922..7953,17779..17876,24276..24317,25802..25
839))
                 /translation="MSLTPSPMRAPATYPVLDVKSLNLIIRHIPTNRSAHFDFVRQN
                 HQISPMPNIARKIHIDDVHHIVLSRDHLCPTETMLVHQSLPPTLLCYAHEDEQQYLK
                 TVASSPFG"
ORIGIN
       1  CAACGTAACG AAAGCTACAA CCATCTTTTT GTCATACACC CTCATTCATT CTCATCCAAC
      61  CTTTGGAGCT CTATCAACTT TTGAAATTTT CACCGGGATT TATTTTTTTG ACATCGCTTT
     121  TGACAGAAGC TTTAGACTTC TTGAACATAA TTGTTCAGGT TGGAAACCAG TGGTTTATAT
     181  TAGTTCTTGT ATTTAGTTGT TTACCCGCCA TCAAAAATAA GCGACGCGTA GCTAACTCTT
     241  GAGAAATAAG CGCTTATCGC ATTTACGATT ATACTGAAGA TGGAATAATC ATAGGCAGTG
     301  AAATCGGATA GAGTATACAA AACTTAATTG CCTGCGTGTC GCATTTCAGA ATAGACCGGC
     361  CTACACTACA AAAACTGTGC AGCCAAGGCT CGTGAATTGT CATGGACCAA ACGTCCATTG
     421  TCCAGGGGAC CCTGATCGCT AGTACTCTGC TAACTGTTAT ACATACTAGA TACAATACAC
     481  TACCGATTTA CGGGGCCCTC TAACCGAAGG GGCTCGAGGC GACCGTCTAG AGCGGTTGTC
     541  GTTGGATCCG CCACTGTTAG TAGGACTGCG AGTGAAAAAG GCTGCCGGAT GGACGCTACT
     601  GATTGCAAGC ATTACCTTCA GATATTGCTG TTCATCCTCG TGCGCGTAGC AAAGTAGTGT
     661  CGGAGGCAAC GACTGATGAA CTAGCATTGT TTCGGTTGGG CACAGCTGAA CGCAAAGGCA
     721  GAAAGACACG CGTGCTGATT TGCTTCGAAT TCTGAAATAC TCCAAACGTT ATGTTGAAAG
     781  AAGGTTGACA GAAGTCGAGC TCAAAAGACC GGTTGGTAGA AATATTAGTG ATAAGAGTGG
     841  CAGATGGAAA GGAACTAATA AACCGCTGCT GGAAAAACAC ATCTAGAATG TACGGATCAA
     901  CAATATCATC TAGTGGTTTT AGTGCCATTT ATCGAGGAAC ATCGGTGATT AGGATGCGCT
     961  CTTTTTCGAG ATACATACGC AACTTGAATT CGAATGAGCA ATTAAGATAA ACCAATCACT
    1021  GGGATCTAAA GTTGAATGAG AAACAAAGAT AATTAAATAA CTGAGACAAT TTCTAGAACA
    1081  AGATCAAGAA ATGGACTGAA CCAGAAGAAA GAAAGTCTTT TGGGCACACT CGGAAGTGAA
    1141  CATAACACCC ACACAATTAT ACTGCGTAGG CTTATAATCG GACCAATCTA ACAATATGTA
```

1201 ACAAAGCATA TCAATGCAAA AACCACGAAG CGTTTGTTTA ATAGTTTAAC ATGCCGCACT
1261 GGTTTGCTGT ATTATCAGCT CTAAGATTAG CGAAACCACA AACCGCAATA TCCAGCACAA
1321 AGTAGAAAGC CACCGTGAGA ATAACACATA AAAATATACT ACTTCAACTA GCGTATAAAT
1381 AGAGAGCCGT TGGGTCACAT TTTAATCATT GTGCTACAGT GTTCGCGAAG TGTCACATCG
1441 AGGAAAGTGT TTTTGCTTTA TCAAAAATCT TACAGATTGC ATATAGTTTT TTCATGTAAA
1501 TTTGTTCCAC AAATAATCAA TTTTTCTGAG TGCTTCTGAT ACGTTTATCA AAGTGTTTAC
1561 GGTGCTCGTT CCGTTCCGCG AGGAATCGGT GATTACGAAA AGGGTTCATC GTTTGGAATC
1621 TTGTTTGTCA TTTGCCCAAT AGAAAAGTTT GGTTGGGTAC TGTTCTTAAC ATCAGCAGAT
1681 TTATCGTGAA ATTGCGGACA AGTTTGATAA GTTGACACGT AACGGTTCTG TTTACTTTCT
1741 TACCATTGTT CTAGCTTAGC GTTTCTGGCG AAGGTTGCTA TATTAGTGGT TTCGTTCGGT
1801 TGACTCAACA CAGCAAAAAA AACATTATCA TTATCATACA AGATCGTTTA TCAATTTCGA
1861 GGCACGCATT AGTTGGAATA TTTCTTTCAC AAAACCTGTG TAGTCCCTTT TTATCATAAC
1921 TGTAGTTTCG ACAGTCGTTA GAATTGAGCC TTGTGACAAA TGTGCATCAT TTTAAACAGC
1981 TATTGTTGAA TGTGTACTGG TCATTCTGTC TAAGATATCA ACTGCGCAAA CTGGGTATTC
2041 CCTTTTTTCT TATCAAAGCG CCAAGAATAT CAATGCTTGA TGCATAACGC ACAACAGTCA
2101 GAATGATAAA AAGTGAATAA AACATAGCAC GAGAATAAAG CGCGAGATAA AAATACGTTA
2161 AGCAACCGAC CTTACTGACT TGTGGGTTTT GCCTAACAGC TTCCCTGACA AACAAATCTA
2221 TTTGTTGTGT TTGAATATTT TAAGGAGTTG ATATATTTGG GACCTATTAC ATAAACTAAT
2281 TCTAGCACAA ACAATAGTGT TCTTATTACT ATTATTGGTT GTTATGTAGT TTGGGTATAA
2341 TTTTTATCGT ATATAAATTT ATCTACAGTT CGCGAAGCAT ATTGTTGATG ATTTGATTAT
2401 TTTAAATATC ATCTATATTG CATGTTATAA ATTTTTAAAT AATCACATTT TACTTTTCTT
2461 GTGTAACAAT CAGCCTTAGT GGCAATTGCT AACTGATTTA TTATTTTATA CTTTATGCTT
2521 TAAAAACCGC CTCAACCCTA TAATACATAC TAACTGTCAT CATCGGTACA GTCTTTACTA
2581 ATTTAATTAC TAATATTCAG TGCTATCGAA ATCCCTACGC GTCTCATAAC TTGAGTAGCA
2641 AAACAATAGC TTTTGTGTAT ATTCGTATCG TACCACAAAA TCTCGTTGAT ATACCGCAAT
2701 AGCGGGTTCT GTGGGTTTCC TTTATTTTGC GTGGTTATTA TTAACATTTT AACATTAACA
2761 GCCCTTTTAT CGAGTATACA CTTAATCTAC TGATTTTGTT TTACCGCACG ACGATTTAGT
2821 TTTTTTCACC CGTTTTTGCT TTACCTTGCT GTGCAAGTGT TTGTCAGCGT CTGATAACAA
2881 TTTTACTGAT GGTACATATT TACATCTGAT AGGATTTGCG AAGCAAGCGA CAACAAACCC
2941 GGACAACTTG CTGATACATC TCCAGTCTTC TTCTTCTTCT TCTTTGTCGT TTGATCTAAT
3001 ATAAGCCATA TTTACAGGTT ATACCATACG ACGGCGTTTG TTAGTAAATA GATTCATCTA
3061 TTCATTCATT TTCTTCTGTA ACATCGTACC ACTTTTCCAA AAACAGAACT GCTACAACAA
3121 GTGTTGTTCA ATTGGGAATC CGATAATTGA ATAAAAGCCT AAAATTTCAA TACCATCATG
3181 GAGCAGCCGG AAGATGTAGC GATACACGTG AAGAAAGAAG GTAGCCGAAC GGTTATCGAA
3241 GTGAAAGGCA AGCACATCGT GCTTAGCGGA GTTAAGCCCC GCATAGAGCA GCTTTTATCT
3301 AGCTCCACTG CCCAAGAGGT TCATATTCTT GCCGGCACAA GTCTTAGCAT TGATGCGGAT
3361 CTTTCTCGCG ATGTATGGCA TGGACTTAAC TTGGTTGTTC TAGCAAACGA AATCACAGTG
3421 CCCAGCACGA TTACCTGGGA TGTGTCTGGG AACGATAGCA ATCATAACTA CGCGGCCAAC
3481 GCTGGTACTA GTGAGGATGG GCACGGAAAA GATGGGAACG ATGGATATCC GGGCGAAAGT
3541 GGAGGGAATG TTTTGCTACT GGCGAACGCT ATACAAAATC CGGAACAATT AACGATCATT
3601 TCAAACGGCG GTAACGGTAC TTGCGGTCAG GACGGCGGAG ACGGAGAGGA TGGAGTCGAT
3661 GGTAATGGTA TTTCCAGGGC AGATTTTGAT AGAAATTTCC AACACTGTTT GTTTTGGAAA
3721 GATATTGAAC GAAGAGCACG TGTACATAAA ATTGTAAAAA ATATGATTGA TTTGGCCCAC
3781 AGCGTAAAGA CACAGTGGGC AAAAGTAACG TATACGTTTG GTAATAGTGA TACTTACACA
3841 AGCAATGTAC TGGAGCTACC AATAAAAACG TTCCATGATT ATTTGGACAT GTACGTAGAA
3901 GTTGTAACGC AAAATGGAAG TGAAATAATA TTTTCTTTGT ATGAAGCTGT AGGAAAAGCT
3961 CACTCATTCT TTGTATACAA AGGATCACCA GGTGTAAAAG GCAAACCGGG CGGAATGCAC
4021 GGATTAGGAG GTCAGGGAGG TTTTCCCGGA GAGATAATCG TGAAACATTT ACACAATGAA
4081 CGTGCGTTTG AAATTAAAAC GTGCAAAAAG GGGGGCAAAT ATGGCGATAA TGGTAAGGGT
4141 GGCCTGTGTG GAAAGCACGG CAAGAATGGT TGGGATATGG GCTATTTGGG CTGGTTACTC
4201 TCCGAACCAA CAATATACGG ACGAGATGAA AATAGCAAAC TTCAGGTGTC TTATTACAAA
4261 GAAGATAATC GAGCATCAAG CCGAATCAAG TGTCCGTATA AGGCAATTCA CAGCAGCTAC
4321 TTGTACGTTG AGATTAACGA ATCGAAAATC GATCATCTAT CGCAGATTGA ATTTAGGGAA
4381 AGGAACAACA TCCAGGCACG TGTCGAAAGA CAACATCATT CGCAAGTGGT AGGAAAGAAA
4441 TCCATTCTGG AAGTGAAAAT TCGTGCTGAA TTTGCTGCAT ACTTGCAATT GATCGCAACA
4501 CAAAAGCAGC GATTAGCACA ACAGCAAATA GAACAAGAGG AATCAGCAAA AGAAACAGAA
4561 AGCTCACAAC TAGTGAATCT CTCAAAAAGT GAGCAGCAAT TGGCGCAAAA GTTAAAAGAA

4621 AATGGGGAAT ACGATGCCCT TCAGCAAACA GTCGAAGGCC AACACACAGA TCTTGAACGG
4681 GCCACGGATC CATTCATTAA ACGCTTCGTC AATTTTACTC AGGACACATC TATCCACAAA
4741 CAATTTTCTT TTGTTGAGCT TCATACCATG ACGAAAGATA CGGAACAAAA GAAAATGATT
4801 GATATTCTGA GCAATAACTT GAATCTAGAC GACTGGCTAC AGCTACAGGC CCAAGTTATC
4861 AGTGACCAGC ATGAGCTGGA GTCCTTAATC AATCTTTTGG TGACTTGTTT TCCTCATTTT
4921 AATACGTTGG TGGATGCAAA TGACAAAGCA AAGGTACAAA TCATTGTGAA AACCTTAAAA
4981 CTGAAGCATT TATTTGCAGC CTTAAACAAA AAAACACACG ATCTGGCGAC CTATCAAACA
5041 AATGTTACTA ATTTCGAATT AACAGCAGAG AGTATAATTA GATATCTGGA AGAAACGGAG
5101 GAAAAGAATG ATACTTTTCA TCATCCCATG CTTGGCAGTT TTAGGAAATA CATTTTTGAG
5161 AACTCCCTAG CACAGCGAGA AAACATTGCC CAATTTGTAG TAAAACTATC GAATAGTAGC
5221 GACGAAGCGA TCAAACGTGC TGTTGAAACA TTCATCTTGG AAACAGGCAG CACAAAAACA
5281 GAAAATGAAG TGGTGAAACA GTTTTACAAT GAGTACCATG AATTTCCAAA GAAGCAGCAA
5341 AACAATATTG ACGCTTTAAC TGAATCATTA GCTACAGAAC TAAAGAAATC AATGCATCAG
5401 GAAGCATTCC GTTTGTGGAA TGAAAGCGTT GGTGGCGATA CTTTGCCTGC AGATCTAGTA
5461 AAGATGATCA GACAAGATGA AAAGTTAAGC TCACTGTACG ATAAACTATT ACTTGAGAAA
5521 AATACTGAAT ATGACTGGGA AAGATGTAGC TCAGATAAAA AAATACAAGA AATATTCAAG
5581 ATCAGCATAG AACAGGAAGG AATCACGGGT AAAAGTTACA GAGTACTACT GGCGTATACG
5641 TATGATGTGA ACATTCGGGT GTATGCTAAA AATGAGGACA ATGAAATCGT TCTGATAGAA
5701 GATCATAATC CAAAATCCGC AACAGTGATA CATATCTTAC AAAAAGAAAA CGAATTCATT
5761 CAACTGAAAG ACAATGAAAG TTACCGCCAA CTAGAAGAAG AGCGTGTGCA CAAAGATAAG
5821 ATTTATCGGC ATATGTTGAC GGAGATTGAT GCTATGCAGA ACATGAAGGA TTACTTTTAT
5881 CAAAAAATAT CCTATCTATA CGGTAACAGT GATGATAATG CAATCGAGAA ATTAGATGAA
5941 AATATAGCAA TTGAAAAGAT TGCAGCATTT TTTTCAGATG AGGAAAAAGA CCAAATTGCT
6001 CAACGTCTGA ATACCATTTC ATCTCAGTAT ACTAACCAGT ACAGAGTGTT CGATTACATG
6061 TTCCAGCGTT TTTCTGTTGA AGGAAGACAT TTGTCATCAA ACGAACTGTG TTTCTTAATC
6121 GACTCTATAC TTGCTCACAC TGTTGAAGCA AGGCAAGAAA GGTACACCTT TTCATGGATT
6181 GTTGCAGCAT ACAATCAGTG CAATTGGACA GATGAGCTGG TGTTGCTGCA GCTGGAGAAC
6241 TGCTACAAGA GACAGCTGCA GGAAAAGCCC AAATGGCGCA AGTACATGTC AAAAATCGAA
6301 AATAAGGCAG TTTTTGTTAG TGTTTTCTGT TCAGTTTACT CGTAAGCCTT GTCCTCCGCA
6361 GTGTGTTGAA GATATCTTGC ACTTGCTGAG TAACATTCCA ATAGAATCGG TTTGTCTAGA
6421 TGAACTGGAG CTAACAGAGT GGCCTTATGT GCTTAAGGAG TGCTATTGGA CACATAAGTT
6481 GAAATCCCTC CTAACATGGG ATAATGAGTC AGAAGGATTA ATTTCTGCTT CGTTCTTCGT
6541 TCTGTCCATA GAGAACACGT ATGGAAACAA GTTGATGCAG AATTTATTTG AATCGTTGGT
6601 GGAAAAGCGC AGTGAGCTTA CGCATTCTAT GTTATTAAAC GTGCTATCTT GCTTTCATAA
6661 ACAAAAGTGG AACTTGTGTG TTGATGATCT TAATGTTTTA AAACAGAGTA ACTTGTTTGA
6721 ATGGATTCAA CACATGGAAA CAAAGTACAA GAAAGATTTA AAAGAACTCA ATATTTCTCA
6781 GATAGTAACA ATCATTAGAG GCAATGATAA TACGTCAAGT AATATAGTGA AGCATTTGTC
6841 AGAAATTGAA CAATCGGTAG CAAATATTCA ACACAATAGC AAAACCATTA ATGCGAGCTT
6901 GATATCTAAC TTTTTATCCA ATGATGTTGA GAAATGGGCT GAAGAGTTCC AAAGCAGTCA
6961 AAATATGCGA TGTTTGGAAA AGTACGAGGA AATGCTCGCT GTAATTGATA GGGCAATTGA
7021 GCTTAAGAGA GGCTTTCAGT TACGAAACAC TCAGCGACTA ACGGTGTTGG CATTGCTGAC
7081 GAACCATAAC AGCACTTTGG CCCAAGTATC GACCGGTGAA GGGAAAACGG TTATTGTGGT
7141 AGCGATCGCA ATTATAAAAG TACTCTTTGG GGAGAAGGTT GACATAATCA CCAGCTCGTC
7201 GGTACTTGCC AAACGCGATG CAGACGACAA TAAGGACATC TACGGATTGT TTGGAATTAG
7261 CGTGTCGCAT AACTGTAGCG AAAAAATACA AGTACGAAAG GAAGCTTACT CTTCGAATCA
7321 GATAGTTTAT GGAGATTTGG GTAACTTTCA ACGAGACTAT CTGTTGGACA GATTCTATGG
7381 AAAACGTATA TTGGGAGATC GTAGCTTTGA AAATGTGATA GTCGATGAGG TGGACAGCAT
7441 GTTGCTCGAT AAGGGAAACA ATATATTATA TTTATCGCAT GACATTGCTG GTATGGATAA
7501 GCTGGAATCG GTATACTTGT ATATTTGGCA AATGGTCAAT ACTCCGACCG ATCTTTCGGA
7561 AGATGAACTC GTGAAAGAAA TCAAACAGTC GTTGTTGAGT GACATGTATA CACTCATCCA
7621 GAAAGCGGAT ATCCAGAAGC TTAGCAAAAG TTTGGGAAAA TCAAATGTAT TGGAGCTATG
7681 GGAGAGATTA ATCAAGTCTG GAATAATAAA TGATTCAGGA AGATTACAGA CTCAAGCTAT
7741 TGATTCCAAC AAACTAGAAA AGCTACTTTC CCCAGATTTC ACACAATACA AAGATCGTCT
7801 AAGCTTCTTC ATTTGGTCAC GTATAGATCG TGAAACGAAT GTTAAGGTAC CCAATTTTCT
7861 CAAAGAATTT GTGGAGAGAC ACTTGGACAG TTGGATATCC AGTGCCTTAA CAGCGCTGTA
7921 CATGGTCCCG GGACAGTACT ATGTGGTGGA CGTCTACAGA TCGGGAACAA GTGCCGACCG
7981 TATTGCAAAT GTAATTATTC TCGACAGGGA TACGGGTACT GATCAAGTTA GCTCGCAGTG

```
 8041  GGATGAAGCG TTGCACCAAT TTTTGCAGCT TAAGCACGGA TGTAAGCTGT CAACCCAAAG
 8101  CCTGAAGGCA GTTTTCATTT CCAACGTTAC ATATTTGAAG AAATACAAAC TGCTGTATGG
 8161  ATTGACAGGC ACGTTAGGAT CACAAAGAGA AAGACATCTG CTTGAGGAGA TACACCAGGT
 8221  AGACTTTGTC ACTGTCCCCA CTTTCAAATC GAAACAATTT GAAGAATATG AACCCATCAT
 8281  TTGTTCTGGC AAAGATGACT GGTTAAAACA AATACACCAA GAGGTAAGTA AACTGACCAA
 8341  GACTGATAAA AGATCGGTCC TGATCATTTG TGAAACAATC CAAGATGTAG AAACATTGCA
 8401  AAAGAATATA GCTTCTAAGG GGCACACAAA AGTTCGCATT TATGTACGCG ATTATGAGAA
 8461  ATTAGATGTT ACAAAGGGTA AACTGATACC GGGAGAAGTT ATCATTGCCA CCAACTTGGC
 8521  AGGTCGAGGC ACAGATATAA AAATATCCGA CGCACTTGAA GAGGCCGGAG GATTACACGT
 8581  GATATTATCA TACCTTCCGG CGAACATTCG AATTGAAGAG CAAGGCTTTG GAAGAGCGGC
 8641  TAGAAAGGGC GAAAGGGGAT CGGGACAGTT GATAATAATG GCAGATGGTA GAAAAGCGAA
 8701  TCTTTCTATG ATATTTACCC TGAAACAAAA ACGTAATGTC GATGAGCTTT TCCGCTTATC
 8761  AGCTATAAAA GCTTATTACA AAACAACGAT TAAAACAGAA GAGAATTGTT TTGAAGCATT
 8821  CAACGAGCTG TACGAGAAAA AAATGAAACT TTTAGCTAAT GCTGAAACGC CAGAGCCGGT
 8881  GAAGGATATT CTGCTCCAAA GTTGTTTGGA CAGATGGTCG TTTTGGTTAG ACGAAAACAA
 8941  TGTACAGATA GGAAAAGCAT CAGATGAGGC TTCCAAGCGG GAAGTAAAAC AGCGACTGAA
 9001  ACACTTTCTA TTGCAATTTG ATGATCTGAA AGCAACGAGC GCAAACAATT GGGCAGCTTG
 9061  GGTTGAAGAA AACCCAATGC AAATGATAAA ATTGGGAAAA TATATTTCTG AACACGAATC
 9121  AGAGCATCGC GATACAGCAT TGCAACTGTT TGGGAAAGTT ATCTCGAGTG AGCCGTGTTT
 9181  TTGCGAAGCC GCCTATTACT ATCGTGCATA TACACGATCA CTTGACAACA TGAACATAGG
 9241  CGACTTCAAA AAGGATTTGC GTTTGGCTGG TAAGCTCTTT GACGATCATA TACAGTGCAA
 9301  AGTTTATGCA TTAGGCATCG TAAGTAAACA AAAGGATACC GCTATATTGA AGAATACCAC
 9361  TACAGTAAGG AATACCATGA GTGAGGACTA TGAAAAGCAA AATAATAGCC TTATTAATCT
 9421  ATACTCACTG TTCATACGAT CAATAGACGA TATGTTGGGC CATTCTATAA GTCCCGAATC
 9481  ATTTGTAAAT TATGAAATAA AGGAGGATGT GGCTGGAATA ATCTTTGAGG ATCTTGTACG
 9541  AATCGGCTTG ATAAAAAAGG CTGAAGTAGT CAGTAACGTA CCAAATGATC AGATGCACAG
 9601  AATTCACAAC GATTACGGGA TTTCTGTAGA TCAATTGAAT AAGTTCCTAG ACCAGCATAA
 9661  AGGACCTATT GATGAGATAG TATTCCTCAA AGCTATTCGA GAATCATTCT CATTTCCAAG
 9721  CAGAATCGAA TTTTGGAAAG AGCTCATCAA TCAAAAGGTG CTACATAGTG TTGAAAGGTA
 9781  TGTTGTAGTA AATGAAGAAC GGCTCAAAAT AATGGATCCT GCTCTAATGA ACGTTCTAAA
 9841  GAAAAAGATA GACAAAAGAG ATCTGTCGAA AGAAGTTGTT AATGTAGGCA AGTGTGAGAT
 9901  ATTGTTGATT GCCGAAGCAC AGTCACAGTT ATACTTTGAA AAAGAGAAAT TCATTGATAT
 9961  GATCGGTAGA AAGAAGTATA ACTCTCTCCA AGAAAATGGT TTACTGTCAT TCAATGAAAG
10021  GGCTACCATC GAAAAAGGTA GAATTGATGG CTGTGTGTTC TCTGCATTTG ATTCTATAAC
10081  AGTTCATGAT ATTACCACCA AGACACGAAT TACCACTCGT GAAGCTATGA TCATCATAGA
10141  CGAGCTACTT GCGCAAAATG TTCTAGAAAA GAGAGCCGAG AGTTTTGCTC TAACAGTTAA
10201  TAATGACGAC ACCCAGGATG TCTTACTTCC TTCTTTTCCA GTGTATGAGG ATGTCGTTAG
10261  AAACGTCCTT GCTTCATGTT TTGCTTACAG ACTAGCTCTG AGAAACATTG CACGACAGCT
10321  AGAAGAGGAG GAGAGTTCTG ACATTCGCAT TGAATTAAAA CTTGATCCTT ATCGGATGTT
10381  TTTCAACGAT CTATTACAGC AACAGATAAT CCAACCTCCT ATGCTTTCCC ACGAGGAGAT
10441  TGATGGAAGC AAGTTGAAAG AGCTGTACGG TAAATCATTG TCGAAAGGGG CATGGAAAAG
10501  CGTACTTTCT AAGGAGATTG GTTTGAGCGT GGAATATACC GAGAGTTTGT TTGGCGTGTT
10561  GCTTGAGAAA CGATGGATTT TTTCGCTAAA AGATGATAAA ACATCAGTTT TATTCTATAA
10621  ACCATGTTCT AAAAGCAATC GGGCAAGTAT TGCTGATCTA AATTCCGGCA TTGCTAAAAT
10681  TCAGGAGGAA CATGGCCTCG ATGAGTCGTA CGACGTTGGT TTCGCAGTTG AAAAATGGTT
10741  TAACAATCAA TTACTGTTAA GTGAAAGTAC AACTGCTGAA CATATTGAGA ATGTGTTAAG
10801  AAAGTCAAAA TCCTCCATTT TAGGATTAAA TGTTCCGCAA GGCACGCTGA AATCGATCAT
10861  GGACTTTTAT GGTAAAGGAA ACTTTGGAAG CATCGATGAG ATGCGCATCA TTATCAGAAA
10921  TGGATTAGAG CATCTGTTAG AGCTAGGAGA GAGAAAATGG ACCATGGAAA TGATACTAAA
10981  TACCTCACTG GTTTTAGCCA TTGGTGTAGC ACAGATAGCA ATTGGCGCTA TAATGGAAGT
11041  GTACTCGGTT GGACTTATGA CTCACGCAGC TAGTTCATTG ATTAGTGAAG GAGTCAACGA
11101  CATCATGTAC GCTGTCGATG CTGCCGTATC TGGTAATTTC AGTTGGCAAG ATTACGCAAA
11161  TTATAAATGG AAAAGTGTAG CCTTTTCGAT TTGCACGGCA GGTGTGGGAA TGTATGTGTC
11221  TAAAGGTGTT AAACTGAGCC GAATAGGATT GAAAATAGTT GGACCAAGTC TCGAGTACGT
11281  TGGCAAACGG GTTGCCGGCA AGCAGCTTAT CAAAGAAATC GGCAGTCGAA CGGTGGTAAA
11341  AGAGATCGCC AAACGTGTGG CGTTGAAAAC GGCGAAAGGT GTCTCGTTTG CTTTGGCAAA
11401  TACTGCTAAA GATATGATTG TTGCCGACTA TCTGCAAACC AGACTAAACT CTGTAGCATC
```

```
11461  GGATATTTTG TCAGGAATCC GGAAAGCGGT GGAAAGTCAT ATAGTTGCTC AAACCTTACA
11521  GAAAGCTTAT GATCAGTTAG GTGANNNNNN NNNNNNNNNN NNNNATGATC AGTTAGGTGA
11581  GAACGACGCT AAAGGCCGGG CTACATTGAT CGTACTCGCA AGCGTAATTT TGATTTTCAC
11641  TAGCGCATCT GGCGGCGACT GGTGGAAGCT AATTTTGGTC ATCCAGGGAA TGGTCATGCC
11701  GGATCTCAAA ATTAAGTAGT TTTTCCATCG TTTTCCATTG CAAAAATGGA TGCAATGCGT
11761  GCAAGACATG TGTATCTACG TTTTACAAAA CTTTTCTCCT CCGTTTTTAG TAAAAAACAT
11821  GGAAAAATAA CGTATTTTCT GCAGCGGCTC CAAATTGTTT GGCAAAATGC TTCGGTCAGC
11881  CGCCGCCAGA TGCGCTAGTG AAAATCGAAA TTACGCTTGC GAGTACGATC AATGTAGCCC
11941  GGCCTTAAAG CTCTTGTGGA CGACTTGACA AATGCATACC TGGTGGAACA GACTATGACT
12001  GATACATTTG AGCTCTCCGC TAAGGAAATA GCCAATAGTG TGATAAAAGG AATAAAAAAA
12061  GGATTTAAAA AAAAACATAC AAATAGCAGC AATTTCATTT CCATGCTAGC TTTAGTACCG
12121  GATGTATTTG ATAACTCTTT ACAAATTGTC GAAATACTTT CCCAAACTAA TAAACTGTTG
12181  GATTACATTA ATGCGGAGAT GGTACTGAAG CTCTCAAAAA TTGAGCTTAA AACGCCAGCA
12241  AAAGAAACAC AAGTAAATCT ACTAAACGGA AAAGATTTTG AGCGCTTTCA AACGGAAGTG
12301  ATCGGTCAAT GGGAAAATAT ATTAAATAAG CACGTTGGAA AATTGATCGC ACAACATGTT
12361  GTGCAGCCAG TGCTTAAATT CGGCGTGCAA TTGGCCTTGT CATATGTCGC CGAAGGTTTA
12421  ATGCATATAA GAGACAGCCC AAACACTACC CTGGACAGAC ATGTTGACAA ACTAAAGAAC
12481  AAGTATGAGA AGAAAATCAC TCGACCTGAT ATGACGGACG ATACGAAGCG TCAGCTTATA
12541  GACCAGTATC ATGGGCGGTT GCTGAAAATG CTGCACAAAA CTCGTAGTCC ATCTCTGTTG
12601  GGCACTATCA TCCGGGAGAA CGTTCCAATG GATTTGAAGT GCGTCGGGGC CTGTACTAAC
12661  ATAGTACATA GAATGTTACA GGAACAAAAT TCTAATATCA CGGGTCTTAC CATAACTGTG
12721  CGTGGAAAAA AGGGAATAGA GCAATCGTTT ACTTCTGGTA AGAATGGTCG CAAAGTATAT
12781  CTGGAACTGT TAGAAAACCA TTTTACTGTA TCACAGGTGA AAGGTTCAAA CATTGACACG
12841  ACTGAGAACA ATAGTCTGTA TGGAGCGCTA ACGATTGCTC TCCCGACTTT GCAATCGATA
12901  GAGCCAACCG TATTCCGATC TAAGTTAAGC AAATACATTG AAAATAACAA AAGCCTCCAT
12961  CATCTCATTC AGCAGAGCTT ACAACACGCT CCGTTAAAAC AAGGGGTGAT TGTAAGTGCT
13021  AAAATTAAGA AGGTAGACTA AGAACGACCA AGCCACATTA CGACGACGTT CTCTTTTGTG
13081  AGTAGGCGAA CGAGAGAGAG AGATGTAGAG AGAGATACAA AGAGAGAGAA AGAGAGAAAG
13141  AGAGAGAGAT GGATACTAAG AGCATAAAGC ATTTTAACGA CAAATCGATC ATAAACCAAG
13201  TAGAGATCAT CATTTTTGTT GTATTAAAAA GGTAAAACAG TTTATTAGTT TTGAAATTAA
13261  AATGATGTAA AAAACCATGT ACTATTGTTT ATCAACTGCT TAGTCTGTTA TCATTGTAAT
13321  GTAGAAAAAC GTTCAACACC TAATTCGTTA AACACATAAG TTAAACAAAA CTGGTCAACG
13381  TTTAGATATC ATTACATTTA TTCGTTGGAT AGCAAAATCT AGAGAATCGA ATATCGTTCG
13441  GTTTTAATTT TCGCGGAACG TAAGGTCCGC AGCAAAAACT ACTGTAGAAT GTGTACAAAT
13501  TGTAATTTAT AACGCATTTT GGTCCAGATG ATCGTATTGG AAGTCAGTTA TCTACAGGTA
13561  GGTAAGATTT CACAAACTAC AGTTTGTTTC TTTGGAATTT AGTAGTAACA CTTGTTTTTA
13621  CAATAATCAT ATTTTCTTCA ATGCAATATT ATGCCCTTGT CTCATCCTTC CACATCAGAA
13681  CGCAGCTAAG CGCACAAGGA TCGAGAGATT ATCCATTGAC ATTGGAACCA GAATCTTGCA
13741  GATAATCATA GTTGTTCAGT CAGAACCGTT ATTGCACAAA TTCTACTTAC TTACGTACTG
13801  ATACATATAG CCTGATATAT GATGTCAATC AGCCTTCGTT CTTCCTATAT GACACGATGA
13861  TACGCTTACT GTTCCTGTGT AGTTCCTGTG ATTTATAAAT TATTATTACG CTGATCATCA
13921  CTTCTAAAGA TAATAACTGG ACGATGATTT TGTAAATATT GAGATATAAG AAGCGGATTG
13981  ATTAGAAATA ACTCATGATA TTATGTTTTC GTTTCAGTAT TTCCGCACCA TAACAACAAG
14041  AAGTTTCTTT CTACGTCAAT AATAACGTTC TGATCAGTTA GCAACGCCTC GTTATCTCAC
14101  TGTTTTTAGA TATTATTATT ATTATTATTA TTATTATTAA ACAGGGACGA AGTAAAATTG
14161  TATTTTTGGC ATAAAAAAGG TCATAGACTC AGTAAATGTT TAGAATTAAG ATTAGTATGG
14221  GATTATATAA CACCGTTAAA ATGTAAACAT ATGTTGTCAT ATGAAGTCTC GTCACAAAAC
14281  TCAAAAAAGT CATTTGCATA AAAATTCAAG ATTTTTAACG TATGCGTCAT TCGCCAAACA
14341  TATGAATGGA TTATGGTTTT ATTTTTAAAT AATTTTAAAT ATCAAGAAAC CATCGAGCAT
14401  GTTAAACGGT GTACCAAACT CAGATAAAAT ATTTGATCTA AATAAATAAT ATATAAAGTT
14461  TTTTTTTATT TATGAAAAGA GAGGGTTTGC AAAAAACACA CTCATACAAT AGGACATTTT
14521  GAAAAAGTGT AAAAAGTCTA AACATCATAT AGGAAATGAT TACCTCTGGG TTCTACTTGT
14581  ATGTAACAAG ATTATATTGT ATGCTCGTTT TAGTATTTTT GCTGTAAGAA ATATGATATT
14641  TCACCAGAAA AACCTAAATA ATCAGTGAAA ATGTTTATGA GTAGATAAAA TAATTATGAA
14701  GATCTAAGCA AACTTTAGGA TTTAAAGAGA TGTAATACCA TGTAATTTAA AAACTATTCC
14761  TTCAAACGCT CAGTGGGTGT TGGATTCGCC TTTTTGCATG GAGACTTCTA AGTGAGTAAG
14821  TTACTTTTCG TATTAAAATT TCTCAAAATT TATATAATTT TCTTAAGTTT TATTAATATT
```

14881 ATTTTATTAT TAGGCCTAAT GAAATTGATA AAGTGCCACA AAACTGCCAT GTGAATGAAT
14941 GTTCAATGAA TGAATCAGCA TATACCATAG CAATGTTTCA ATGAAACAGC TAGCACTCTC
15001 TCTCTTTCTC TTGTTGGGCT GCTCACGATA GAGCACGACG GTGTGACATG GCCCGGTCCA
15061 CAACAATTCT CCAGGATGCT CGGTCCATGG CTGCAGCCTC CAATCCATGT AGACACCCGA
15121 TCTCCGACAG GTCTCGCTAG CAAGATAGCA ATTACCACAA CTGACTAGAA ACATAAAAAC
15181 TCTGTATGGA TAAATGTCTG CCATTCCATC TGCCCTAATA TGCATTGAGC TGATTTTGTT
15241 AATTGTGATT GTGTTAAATT ATGTTAATAC ATAATGGGAT TGGAATCGGT ACTATCTATT
15301 GGCAACGTGA ACGTGAGGTA ACATTATTAA TTTTTTTTAA TGCTGCACCT GGAGGAGGCT
15361 TATCTAACGG TTATCAACAT ATTTGTCTCC TTTCCAGACC TTCGACCGGT TTCGTTCAGC
15421 TCTGCATTAG GACAAATTGC CCAGATGGCA AAGAGCCATC TCAACACTAA ACTGCAACTA
15481 TAGTGGGTCT AAGAACATGT TGTAAAATAG CCGTCTCATG GGGCAGCCTG GTCATCTGAT
15541 AGATAGGGGC ACGTTAGAGC TCCACAACGT CAGCTAATGG GAAAGTCTCA TCCAGCTCAT
15601 CCACCACACC GATTATGTTG ATAAGGATAA GATAATGCAG ACCTTGTGCT GCAATCAAAA
15661 TTAAAATGCA AAATGCAAAA TGCAATTACC TAAATGTGAA AATGTTATAA GTAAGATGAT
15721 TACTATTCAG TTAGTAGTAT AGTAGTTGAA TAATAGATCA CTAGAAATTA GGACTAGAAG
15781 AGCAGAGTTT GTAATGTCGT ATTTATAGTT TGATTGCAGT TTATATATTA AATACATCTT
15841 CCGGAGTAAG TAGGTTGATT CAATTGATTC AATTTATGTA TTAATGTTTG TTTGTTGATT
15901 TCATTTTCAA AAAACGCCTC CATATCAGTT TTATAAACGT ATTTGTGTAC TTGTATAGCG
15961 GTCCCGTGGT GCAGTCGTCA GCTCGCACGA CTTAACAACA TGCCCGTCGT ACGTTCAAAC
16021 TTAGAGGGGG CCGCCCGGCC CCCCGGTAGC AATATCTGGC TATCAAGCTG CATGGTACTA
16081 AAGTCTCGAA ATCCTGTATA GGGCTGCATG TCCCCATAGG AAGACAGAAG AAGAAGGCCA
16141 AATAGAAGAA TAAGTACTTG TATCGTAAAC GTACGCTATT GCCTTTGTGC ACTGTTAATA
16201 TAAATTCATG CCTGAACAAT TTTTTTTTTT TTGCTTATAC CACACTTACT TACATATCTG
16261 GCGCTACAAC CGCTTTGCGG TCTTGACCTC AGGAATGTCG GAAACTGCTC AAGGTTTCGT
16321 GCCTCCTCGT CTTTATACTA GCATTTTTAT TACAATACAT GCTACTAATT GGCAAAAAAG
16381 AATTATTACA ATTTTTTAAC ATACATTTGT AGGTGAAATA CGGTTTTTGT AATACGTTAA
16441 AATATTTATA TGGTATCATG TGTTGAAATA CAAAAGAATG TTAAACGGTT GCTAGCAATA
16501 TTTACATGGT AACCAGTAAC AAGTAGATGG TTTTTGCTAG TTCATTACCC CTTCGCGAAC
16561 CCTGCTAGAT TGCCCTGTCC ATGAGGAAAA AGATGGAAGT TTTTCAAGCA GAAAGCGTAT
16621 CATCTGAGGG CACCTGAAAG GTATGCATCA ACCCCTTTCT TATCTTGGAT CGTGCTTTGA
16681 ACATCGTGCA AAGCAAACGC CGTATCACTT CGTCGCCAGT GCTGTTAGCG GTAACACACG
16741 GAGTGAGTTT GTGTGGCGTG TTCCGTATTG TGAATCGAGG CAAGCTTTTA TCAGCTATAA
16801 TTTCTGTTTA AGAATACAAT TTATCACAGT TATCGTACCA GTGTTTTGAA AAAAGCAATA
16861 TCAAACGAGT TAAACAGCTC GCAACAGATG AGTTAATGCC ACTACCCTTC TCTAATCGTG
16921 TGCGTCGAGC TAAGCGTATC TTGCAAAGGA TGCGTCATCA CATGTTCGTG TCAATACCGG
16981 TAAGCACCGG CCGGCGTCAT GAAGTGCGCA TGACCGAGTG CATATAAATT GAGAGTGAGA
17041 ACGTAAAGTG TTTAAATTGG TTCATTAAGA AAGAAAGTAA CGTAGCAAAA ACGTTAAGTG
17101 ATTTGTGCGA CTTCAAACGA AATGATAAAC ATGTGGTGAT GGGAATTCCC CAATGTAAAG
17161 CATTTTGCTT GTGTTTACTT GCCGATTCCG GCAGACGGAG ACGGTTTCAG TTATTCTGCA
17221 GTTGTGTGAA CCAACGTGAC CTTGACTGTT TCGGAATTTC TACTCCATAG CAGTACAGTA
17281 GTGTGATTGA CAATGTTGCA ATTAAATTGT GAAATTATAC AAATGTGGTT GAATGTAAGT
17341 GAAGTCTGAT GGTATGTAGC GTCAAACGAA GCAAGACAGT GTAACGTGCT GATGCCCTGC
17401 GTGAAGGATT GATTCTTCGA AACGGCTGTA ATCTATTAAG CAATTAACGT ATAAAACCAA
17461 CAAGATACGG GATATATGGT CGTCGATAAG GAAAATGTAT AAAGTGATAA GCGGTTGTGG
17521 CTGGTGCTTA CCATTAAAGT GAATCAGTGA ATAAAAAACA ACGTGCAAAT AAGCAATAAC
17581 ATAGTCCAAT AACAATAACG AAAGACATCC GATCATGGAA TTGCCGGTGA CGATACGGGA
17641 GAGGAGTGTG GACAGCCGAA CGGTGATTGA AGTGAAAGGG AAATATATCG TGCTAAGTGA
17701 AATTAAATCA CGCATAGAAA CCCTCTTGGC CAACTCTACT GTTGAAGAGG TTCGTATCTT
17761 TGCTGGCATC ATTCTTACCA TCGATGTGGA TCTTTCTCGC GATGTTTGGC ATGGGCGAAA
17821 TCTGGTGGTT TTGGCGAACG AAATCAAAGT GTGCACTACG GTTCGTTGGA ATGTGTCTGG
17881 GAAAGATAAC GACCATATCT ACACGTCGAA TGCCGGTACG GATGGGAACG GTGTTGGCAA
17941 ACAAGGAGCC GATGGATATC CGGGCGAAAG TGGAGGTAAC GTTTTATTGC AGGCAAACAC
18001 AATTCAACAC CCCGAACGAT TGACGATCAT TTCGAACGGT GCAAACGGAA GCAACGGTCA
18061 GGATGGTGGC AATGGAAAGG CGGGCGTCGA CGGTAAAGGT ATCTCTAGGG CAGACTTTGA
18121 GGACAAGTTT CCGCCGACTG CCAAGTTTTT GGAGGGGCCT CGTGCTAGCC GCGTAGAAAC
18181 CACTGTCAAA AACATTAAAA ACTTAACTCG AAACATTAAA ACACAGTGGT TGGTTACAAC
18241 AATGCCCTTA TTTCCCGTCA ATACCACTCA CAGAGAAAAC GTACTTTATC ACACGGATTT

```
18301  TCGTATCGAT TACTTTAACA GTCTTTACAT CGAGGTGGAA ACGCACGATG GCTGCGAGAT
18361  AACTTTTTCT TTTCAGCAAG GCACAACGTT TTACAACTGC CAATCGTTTC TGATGTACAA
18421  GGGAACATCT GGCCAGCGTG GACAGGCTGG CGGAGGATAT GGGCTGGGAG GTCAAGGTGG
18481  ATTTGCTGGC GAAGTGACAA TTAAACGTTT ATACGACGGT CGTACGTTTG ACATTAAGAC
18541  ATACTGCAAC CGGGGCAATG ATGGAGATAA CGGAAATGGT GGATTGTATG GTGCGCACGG
18601  TAAAAATGGT TGGGACATGG GTTATGTGGA TTACATAACG CACGAACCAA AATTGTATGG
18661  ACAGGATGAA AAGAGCAAAC TTCAGCTGTC GTTCTATGAA GAAAATTCAT CCAGCAGAAT
18721  GTATTGTCCA TATAAGGCTT ACATCGACAG CAGCTACAAA TATGCTGGAT TCCAGGTATC
18781  AAGAATTGAC CATCAATCTA GTAACGAATT CAGGGAAAGG AACAACACCC AGTCACGTAC
18841  CGACAGACAG CATCATGCGA AAGCAATAGG GAAGAAATCC ATTCCTGGGC TGCGCATTGC
18901  TGCTGCATTT TCCGAATACT TGCAAGAAAT TGAACAAGAA AGGCAGCAAT TGGAAAAACA
18961  AAGAATCGAG AGAGAAAAGC AAGAATTAGA ACGGCTGCGA ATCGAGAAAG AAAAACAGGA
19021  ATTAAAGAAA CAACGGGAAG AACTTGAAAA ACAAAATTTG GAAAAACAAC GGATCGAGAG
19081  AGAAAGGAAG CAGGAATTGG AGAGACAACG GATTGAAAAA GAAAAGGCAC GATTGGAAAA
19141  ACAACGGATT GAGCAAGAAA AGCAGGAATT GGAGAGGCAA CGAATCGAAA AAGAGAAGCA
19201  AGAGCTGGAA AAACAGCGGA TCGCACAAGA AAGGAAAAAC ATTTTACAAC AACAAATCGA
19261  ACAAGAAAAG CAACGTTTAG CACGACAGCA GATTGAAAAA GAAAGGCAAC AAGAAATGGC
19321  TGAACAGGTG AAAGAATATT TGAATTGTGA TGCATTGCAG CAAGTTGTTG AAAGCCAATA
19381  CGCAGATCTT GATCGAGTCA CGGATCTGGA CATAAAACGG TTTGTTAATT TTACACAAGA
19441  TACTTCTGAT CAGGATCGAA ACGAGTTTGT TAAACTTGGT GAAATGTATA AAGATCCTGG
19501  TCGAGAGGAA ATATTCAGCA AATTGTACAA TAACTTGAAT CTAGACGATT GGCTACAGCT
19561  TCGAACACGT CCTGCTGATC GCCCTGAGCT GGTGTCTTTG GTGGAACATT TTAAATCATT
19621  AAAATCACAG TTCGAACGTT TGCTTGGAAG GGATTGCAGT GCTAAGATAA CGAATGTTAC
19681  AGAAACGTTG AAAATCAAGT ATAAAATAGC AATTCTGAAA GAAATTACAC AGGAACTGCC
19741  GACCTGTTTC ACAACAGTGA ATCGTTTCGT ATTGACAGCA AAGACTGTAG GGAGATATCT
19801  GGAAGAAACG GATGAAAAGA ATGACACATT TTATCATCCC GATCTTGGCA CTTTGAAGCA
19861  ATACATTTTT GAAAACAATC CAGCACAACG AGATAACATT AACAAATTCA TGGCAAAAGC
19921  ATCAGAAAGT AACGACAATG CGATCAAACG TGCCGTAGAA ACATTCATTA TGGAAACAGA
19981  TAGCTCAAAA GATGAAAATA AACGTGTGAA ACTGTTTTAT GATGAGTATC AGAATTATTT
20041  GGAGAAACAG CAAAGAAATA TTGATGCTAT AACGCAATCA TTAGCAGCAG AACTAAAGAA
20101  ATCTAAACAT CAAGAAGCAT TCCGTCTGTG GAATGAAAGC ATTGATGACG ATGGTTTGCT
20161  GGAGGATTTT GAAGAATTGA TTAGAAAAGA TAAGGTGTTG AGCTCACTGT ATGATGAACT
20221  GAAGCGTGCG GAAAATGCTG AATATGACTG GGAAAGGTGT TGTACCGATG AAAAAATACA
20281  AGAAATACTC CAGATCAACT TAAAGCAGGA AGGAATCATG TGCAAAAGTT ACAGAGTACT
20341  GCTGGCGTAT ACGTATGATG TGAACATTCG GGTGTATGCC AAGAATGAAG ACAATGAAAT
20401  CGTTCTGATA GAAGATCATA ATCCAAAATC CATTAAAATT ACACACATTT TACACAAATC
20461  GAACGAATTC ATTCAACTGA AAATCAATGA AAGCTTCCGC CAGTTAGAAG AAGAGCGTCT
20521  GCACAAGGAT CGAATTTTCA GACGAATTCT GACGGAGATA AACGGAATGC AGGACATACA
20581  GACCATCGAG AATTACTTCG AACGAGAAGT ATTCTATCGA AACAGTTCTA ATTTCGCTAA
20641  AATGTTTGAA GATTTAGAAG ACGAGATAAC AGTTGAAAAG ATTGCAGCAT TTTTTTCAGC
20701  CGATGAGAAG GACCAAATTG CTCATCGTCT GAAGAACATT TCTCCCAAAT ACACCGGCCA
20761  ACATGGAGTG TTCGAAAGCA TCTTACTGCG TTTTTCTTTC GAAGGCAGCC ATGTGTCATC
20821  AAACGAACTG TGTTTCTTAA TCAACTCTAT ACTTGCTCAC ACTGTTGAAG CAAGGCAAGA
20881  AAGGTACACC TTTTCATGGA TTGTTGCAGC ATACAATCAG TGCAGTTGGA CAGATGAGCT
20941  GGTGTTGCTG CAGCTGGAGA ACTGCTACAA GAGACAGCTG CAGGAAAAAC CCAAATGGCG
21001  CAAGTACATG TCAAAAATCG AAAATAAGGC AGTTTTGTTA GTATTTTCTG TTCAGTTTAC
21061  TCGTAAGCCT TGTCCTGCGC AGTGTGTTGA AGATATCTTG CACTTGCTGA GTAACATTCC
21121  AATAGAATCG ATTTGTCTAG ATGAACTGGA GCTAACAGAG TGGCCTTATG TGCTTAAGGA
21181  ATGCTATTGG ACACATAAGT TGAAAGCTCT TATAACGTGG GGAAATGACT CGGAAGCCTT
21241  AACTTCTGCT TCGTTCTTCG TTCTGTCCAT AGAGAACACG TATGGAAACA AGTTGGTGGA
21301  GAATTTATTT GAATCGTTGG TGGAAAAGCG CAGTGAGCTT ACGCATTCTA TGTTATTAAA
21361  CGTGCTATCT TGCTTTCATA AACAAAAGTG GAATTTATGT GTTGATGATA TTAACCTTTT
21421  AAGTCGATGT AGTATCAAGA AATGGATTAA ACATATGGAA GAAAAGTTCA AGATAAACTT
21481  AAAGGAATTA AATATTTCAC AGATAGCAAC TATCATTAGA GGCAATGCTA ACACATCAAA
21541  TAATATAGTG AAACATTTAC CTCAAATTGA AAATTCAGTA ACAAGCATTT ATAATAACAA
21601  GAGAACGATA CACACGAGTT TGATATCCTC TTTTTTGCTC AACGACGTAG AAAAATGGGT
21661  TCAAGAGTTC CATGTTAAGA ACCAGAAAAC GAATACAAAC ATGTACGAAG AAATGCTTGC
```

```
21721  TGTAATAGGT AGGGCAATTG AGCTTAAGAG AGGGTTTAAG TTACGAGACA CCCAGCTATT
21781  GACAATTTTG GCCTTACTAG TCAACGATAA CAGCACTTTG GCACAAGTAT CTACAGGCGA
21841  AGGTAAATCA TTGATCGTAG TAGCGACCGC GATCATTAAA GCGCTCCTAG GGGAGAAGGT
21901  AGATATTATC ACCAGCTCCT CAGTACTGGC TCAACGAGAT GCTGTCACCA ACAATGATAT
21961  CTACGAATTG TTTGGTCTTA GTGCCTCCCA CAACTGTAGC GAACAGATAG AGGAGCGAAA
22021  GGAAGCTTAC TCATGTAATC AAATAATTTA CGGAGATTTG GCAAACTTTC AACGTGACTA
22081  TTTATTAGAC AGATTTTATG GAAAGCATGT ATTGGGTGAT CGTAGCTTTG AAAATGTGAT
22141  CGTCGATGAG GTGGACAGCA TGTTGCTCGA TAAGGGAAAC AATATGATAT ACCTGTCGCA
22201  TGACATTGCT GGTATGGATA AGCTGGAATC GGTATACTTG TTCATTTGGC AAATGGTCAA
22261  CAGTCCAACC GATCTTCCGG AAGATGTACA TATGGAAGCG GTTAAAAGAT CAGTTTTAAG
22321  TGACATGTAT GCAATCATCC AGAAAGAGGA TATCCAAAGT TTGGACAATG CTTTGGGAAA
22381  ATCGAATGTA TTGGTGCTAT GGGAGAGATT AGTTAAGTCT GGAATAATAA ATGACGCAGG
22441  AAGATTACAA ACGAAAGCTA TTGATTCCAA CAAACTAGAA AAAGTACTTA CACCAGATTT
22501  CACACAATAC AAAGATCGTC TAAGCTTCTT ACTTACGTCT TGTGTCGAAC GCGAAACGTG
22561  CATCAAAATA CCAAATTATC TAAAAGCATT CGTGGAGAGA CACTTGGACA GTTGGATTTC
22621  AAGCGCCATA ACAGCACTGT ACATGGTCCC GGGACAGTAC TACGTGGTGG ATGTCGACAG
22681  ATCTGGAACA AGTGCCGACC TTAATGCGAA TGTAATTATT CTCGACAGGG ATACGGGTAC
22741  TGATCAGGTG AGCTCACAGT GGGATGAGGC GTTGCACCAA TTTTTGCAGC TTAAGCACGG
22801  TTGTAAGCTG TCAACCCAAA GCCTGAAAGC AGTTTTCATT TCCAACGTTA CATATTTGAA
22861  AATGTATGAA CTGCTTTATG GTTTAACAGG AACGTTAGGA TCACAGAGAG AAAGACATCT
22921  GCTTGAGGAG ATACATCGGG TTGACTTTGT TACTGTTCCT ACCGCCAAGT CGAAGCAGTT
22981  TGAAGAGTAT GAACCTATCA TTTGTTCGAG CAGTGAGGAT TGGAAAAAAC AAATAGGACA
23041  AGAGGTGAAT AAACTAACAA AAATTGATAA AAGATCGGTC CTGATCATTT GTGAGACGAT
23101  CAGGGATGTC GAATCATTCC AGATGGTGTT TGATTCTGAG GAGCACTCAA ACGTTCACAC
23161  TTATGTACGT GATTATGAAA CGTTTACTGT TGCGGAAGGG GACAGAAAGT TGGAACCGGG
23221  ACACGTTATC ATTGCTACCA ATTTGGCAGG TCGTGGCACA GATATAAAAA TATCCGACGA
23281  CCTTCGGAAG GCCGGAGGGT TACACGTGAT ACTAACATAC CTTCCGGCGA ACATTCGAAT
23341  TGAAGAGCAA GCATTCGGAA GAGCAGCTAG AAGCGGAGAT AAGGGATCGG GTCGGATGAT
23401  AATAATGATA ACCGATGGAA AAATGCACAA TCGTTCTATG ATATTTACCC TCAAGCAAAA
23461  ACGCAATGTC GATGAGCTTT TCCGCATATC TGCTATAAAA ACTTATTACG AAACCACTAT
23521  TAAAACAGAA GAGAGTTGTT TTGAAGCATT CAACGAGTTG TATGAGAAAA AAATGAAACT
23581  TTTAGCTAAT GCGAAAGAGC CAGTGAAGGA TATTCTGCTC CAATGTTGTT TGGACAGATG
23641  GTCGTTTTGG TTAGACGAAA ACAATGTACA GATAGGAAAA GCATCAGATG AGGCTTCCAA
23701  GCGGGAAGTA AAACAACTGT TGGAGCACTT TCTATCACAA TTTGATGATC TGAAAGCAAC
23761  GAGCGCAAAC TATTGGCCAG CTTGGGTTGA AGAAAACCCA ATGCAAATGA TCAAATTGGG
23821  AAAATATATT TCACAACACG ATTCGAAGCA TCGCGATACA GCAATGCAAC TGTTTGGGAA
23881  AGTTATCTCG AATGAACCGT GGTTTTGCGA AGCCGCTTAT TACTATCGTG CATACACACG
23941  ATCGCTTGAC AACATGAACG TTGGCGACTT CAAAAAGGAT TTGCGTATGG CAGAAAAACT
24001  CTTAAACGAT CATATAAAAT ATGCAATGCA TGCATCAGGT ATGGTGTGTA AGCTAAAGAA
24061  TAACACACCA CCGATCGAGG GCTATGAAGA GCAAAAGAAG TGTCTCATTA ATCTATACTC
24121  ATTGTTCATT CAATCCATAG ACGACATGTT GGGTCATGCT GTAAGTCCTG AATCGTTTGT
24181  TAATTATGAA ATAAAGGAGG ATGTTGCTGG AATAATCTTT GACGATCTTG TACGAATCGG
24241  CATGATAAAG AAGGCTCAAG TAATCGATAA CATACCTAAT GATCAGATTC AGAGACTTCA
24301  CGTCGAGTAC GGGATACCTG TAGATCAGTT GAAGAAGTTC TTAGCTCAAC ATAGGGGCCC
24361  TATCGATGAC AAAGAATTCC TTAAAGCTAT TCGAGAATCA TTTTCATTTC CAAGCAGAAG
24421  CGAATTTTGG GAAACACTTA TTAACCAAAA GGCTCTGCTC AGTGTTGAAA GGTATGTTGC
24481  GATCAATGAA GAACGGCTCA AAATAATGAA TCCTGCTCTA ATGAACGTTC TGAAGGAAAA
24541  GGTAGACAAA GGACAGCTGG TGAAAGCAGT TGTTAAGCTA GATAAGCACC AGATCTTGTT
24601  GATCCACGAA CAACAGTCAC AGTTTTACTA TGAGAAAGAG AAGCTCATTG ATATGATCGG
24661  TAGAACGAAG TATACCTCTC TTGAAGAAAA TGGTCTACTA TCATGCAACG AAAAGGCAAC
24721  CATAGACCCA AGTAGAATTA CTGGTTGTGT GTTCTCTGCC TTTGATTCTA TAACAGTTCA
24781  GGATATTACA AACAAGACAA GAATTACAAC GAGTGAAGCG AAAAATATCA TCGCAGAACT
24841  ACTTAAGCGT AATGTTCTAC AAAAAAGGGG TGAAACGTTT GTCCTACCAA TGAACTATGA
24901  CGATAGCCAG GATGTCTTGC TTCCCACCTA TCCAGTGTAT GAGAGTGTCG TTAGAAACGT
24961  CCTTGCTTCA TGTTTTGCCT ACAGATTAGC TCTGAGAAAC ATTGCACGAC AACTAGAAGA
25021  GGAGAACTGT TCTGACATTC GTATAGAATT GAAGCCCAAT CCCCATCAAA TGTTTTTTAC
25081  CGACTTACTG GAGCATAAGA TAATACTCCC TCCTATGGCC ACGCATGAGA TCGATGAAAG
```

```
   25141  TAAAATCAAG GCAATGTATA AAACATCATG GTCGAAAGGG GCATGGATAA GCATACTTTC
   25201  TAAGGATATT GGTTTGTCTC AAGAATACAC CAAGCTCCTG GTGGATGTGC TGCTTAAAAA
   25261  CCGTTTTATT GTTCCAAACC CCTTAATGGA ATTGTTGTTA TTTTCTCCTT CAAATAACAT
   25321  GTCTGAAGAA ATAAAGAATG CATTGACAAG CTACAGGTTC AATTCGAAAG ATGACCGAAC
   25381  TAGCGTAGCC TATTTAAAAT CATGCATTGC ATCAATCAAT CAAAAGCATG GACTTCTTGA
   25441  ATCCTGCGAC GTTGGAGTTG CGATTGAAAA GTGGTTTAAC AATCAATTGC TGTTAATTGA
   25501  AAACGGAACT GCTAAACATA TTGAGAATGT CTTAAGTAAG TCGAAATCGT CTCTAGCAGC
   25561  ATTAAATGTT CCGCATGGCG CGCTGAAATC GATCATGGAC TTTTATGGTA AAGGAAACTT
   25621  TGGAAGCATC GATGAGATGC GGATCTTCAT CACAAATGGA TTAGAGCATC TGTTAGAGCT
   25681  AGGAGAGAGA AAATGGACCA TGGAAATGAT ATTGAAAACC TCACTGGTGT TAGCGTTTGG
   25741  TGTAGCACAG ATAACAATTG GCGCTATAAT GGAGGTGTAT TCGGGTGGAA TGATGACTCA
   25801  CGTTGCCGGA GCTCTCATGG GCGAAGGTGT CAGCGACATC ATGTTTGCTG TGGGTTCTTT
   25861  TAAGTCAGGT CACTTCAGCT GGAAAGATTA CGGAAAGCAT AAGTTACAAA GTCTGCTTTT
   25921  TACGATCGGT ACGGCCGGTG TGGGAATGTA TCTCTCCCAG GGCGCCAAGA TGTCTCGCAT
   25981  AGGTTATAAA GTTGCTGGAC CAAGTCTCGA GTACGGTGGC AAACGGGTTG CCGAATCATC
   26041  CGGAAAGCAG TTAATCAAAG CAGTTGGTGG ACGGACGGTC GTTAAGGAGA TAGCGAAGTG
   26101  TGTGGCGCTC AAAACAGCAG AAAGTGTCGC ATTTGCTGTG GCAAATACGG CCGTGGATAC
   26161  GGTTGTTGAA AATCATCTGC AATCCATGTT GAACTCCTTA GCATCGGATA TTTTATCTGG
   26221  AGTTAAGGGT GAAGTAGCAA ATCATAAGGT TGCTCACAAT TTGCAGAAAG CGTACGATCG
   26281  TTTAGGTGAA AATGATGCTA GAGCCCTCGT GAACGACTTG ACAAATGCAT GCATGTTGG

//

LOCUS          11     3071 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      11
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             join(500..536,2282..2571)
                     /translation="MMMATNSATKAHAPCTSTMTSMTFGQKKFIPTAPEKGSFPLDH
                     GGQCRKLMLFYMRCLRENADDNSACREESKAYLQCRMDHDLMAREDFSKLGYSSEKS
                     SENKPEQS"
ORIGIN
       1  ACACACTTTC CATACACTTG CGGATAGTCG TTTTATCATT CACACACAGT TCCCCTTATC
      61  GCAACCGAAA TTCCCGTAAC CGTGGTCACT GGTGTAAACA AAAGAACACA TACACACGCA
     121  CACAACTCTA CTTGGGCTCT AATATGCGTA CGAAAGGCAC GATCGAACGG TACCGCCTCG
     181  CTTGCTCTCC GGTACGGCGG CCACCAAGAA ACAAGAACCA CATACACACA CACACAAACC
     241  CGCGCGCACA AGCAGCAACC GTGTCAATGC GTTTTCATCC CCCCCTTCAG ACTGTTCCGA
     301  TACAAAGCGG CCCCTCCGCA CACACACACA CAGCCACACA GATTCGGATC GTAACGGGAG
     361  GGTAATTAAA CCGGAAACGC ACCGAAAAAC GGATGTATAT ACGTTGCCCC AAAATAACAC
     421  TCTAAAAGCG CACACACCGC ACACTATTGC ATTCGCGCTC AATGCACTCT ACGATCGTGT
     481  ACCTCACCCT CTTGCCACGA TGATGATGGC CACGAACAGT GCCACAAAGG CTCACGGTAG
     541  CGGTAGTTCT CGACGTCCTC CCTCTCTCGG GCGTTTTTGT ACAATTTCTA TCTCATGCAA
     601  CGTTGCATTC CGATACTCAC CGCCGAAAAG CGCAATAGAT GCCGGACAGT GTTTTAGCAC
     661  GCACGAAACT GAGACAGAAC AAAAAAAAGG TCGAGAATAA ACACGCGCAC AACAAACGAC
     721  TCCGTTCCGT ACAGGCGGAA TAAAACATAC GCACTGACAG CCAGAGTGGC CAGATTATTT
     781  TTGCGGTTTT CGGTTGGCGC ATCAAAATTT TATCGGTAGT TTTCGGTAGG AGTTTAAAAT
     841  TTTTCGGTAG TTTTCGGTAG GCTTTGAAGT GTTTAGCGGG GGGGGGGGGG GGGTTGTCGG
     901  AGATGGAAGC TAATCTGTCC CAGATGGAGC AGCACGAGAA AATAACATCT TTCATTTATT
     961  TAAAGGTGAT AGTTTAGATT TGGTTCATAG CGGTCACATG AGGCCTTTCT GAAGTGTCGA
    1021  TCTGAAAATT GTACTAGGGA ATTTAACCCA ACAAATGATT GGGACGCTCT GGCAGCAATC
    1081  GAACGCGACA TCAAACATGA AAAATATATG AGATTTTACA TGTCCGATGT GATAACCCAC
    1141  AAATCGAACA TGTGAAGTCC TATACATTTT TCATGTTCGA TGTCGTGTTC TATTGCTGCT
    1201  AGAGCGCCGG GTTATATTAT CTGTCAAATT GCATATAAAA TTTTATCAGC GCGTTCTACT
    1261  TCATGAAGTG CAGACACGGA CCTTAGTTGT CTCTGTTGAA CTTCTGGTTC TGGTTGCCAG
    1321  TGTTTTCCAT ACATTAAATG ATTTGGGAAA TGGTTATATG ACTTTTTGGT TAGTATTACG
    1381  AAAAAATCTG TAATTTTCGG TAGGATAAAT GAAAAATCGG TATTTTTAGG GGGCTCATCT
```

```
    1441  GTGAATCGGT AGGCATATCA CAAATCGGTA GGAATACAGA TAAATCGGTA TTTCTGGTCA
    1501  CTCTGCTGAC AGCTGACAGT TTTGGACGCG TTGACAGCGT CGCTCTGTCT CGCTCTCTTG
    1561  GCCAGCCTTT CTCGCTCTTG GCTTTTATTT GGATTGAATT GCGTAGTTTT CGCAGAAATT
    1621  CTACTATCGT GCTCAGTGAT GTTACAAAAT GCAAGGAACC ACACGATATT TCCATTTTTT
    1681  CTTAAATTTT GATTCTCTTT TTCCGTTATT TTGCTTCCTT TCAAATCATT CCGCCCAACG
    1741  GACCAAAGGG GATATTTTTG AGCTGTCAAA ACCTATCAAC AAAAAATATC CCAAATAGCT
    1801  CCAGCTCGGA TGGAGGTGCG GCGAAACACG CACGATTAAT GGAAAAGTTT ACGAGTTTTC
    1861  CTGCATAAAT TATTGTGTGC GAATGTTTTA CATACTGGTG GAAGTGTGTG CATTGTAGTT
    1921  CAAGCCCTAT TTGCTGTAAT AAATTCGATT GAACCTCAAC GGAACAGAAG ACATGAGGTT
    1981  TATGTAACGT CGGTAGCGGC AAAACAACGA GGTGAGTTAT CGTTTGTTGA AGCTGTCTAA
    2041  TTAGTTTGTT TGACCAATTT TGTGTTAAAT TACATTAACT ACTCTCGTTT CAAATGGTTT
    2101  CATCATATAG CTTATGCCAC ACTTGCTTTT AACAGCGTAA GGAAGGCTCT TCCTCCTCCC
    2161  AAAAATAACC CAAAACAAAC CATGCACAGG GTGTCTCGGA ACGCATCCCA AGAACCGTCC
    2221  CATTTTGCTC TGTTTGCTGC AATGTGCATT AAAATTTCAT TTCCATAAAT TCCAATTCCA
    2281  GCTCCCTGCA CCAGTACGAT GACGTCTATG ACATTTGGCC AGAAAAAGTT TATCCCCACC
    2341  GCACCGGAAA AGGGTAGCTT CCCGCTGGAC CACGGCGGCC AGTGCCGAAA GCTGATGCTG
    2401  TTCTATATGC GCTGCCTGCG AGAGAACGCC GACGATAATT CTGCCTGCCG GGAGGAATCG
    2461  AAAGCCTACC TGCAGTGCCG CATGGATCAC GATCTGATGG CACGGGAAGA CTTCTCCAAG
    2521  CTAGGCTACA GCAGTGAGAA AAGCAGTGAA AACAAACCGG AACAATCTTA ACATCCCTAA
    2581  ACACCAAAAA CCAGAAATAG AGTTTAAGCC CCCCGGGTCT CGTGTAGTGC ATTTTGTCAT
    2641  TGAACGTTGT TAGAAGAAGC TCTGTGAAGA ATACCCTCGA TAATAGTAAA CCAATAGCAC
    2701  CGAATCGAAA TGAAGGTGAA GAAAAACATC CTCGTCGGCT GTACCGGAAG TGTGGCCACG
    2761  ATCAAGCTTC CCCTGCTGGT GGAAAAACTG CTCCAGCTGA CCGAGTTCGA GGTGGAAGTA
    2821  CACGTGATCG TAACCGAGCA TGCCCGGCAT TTCTTCTCGC CGCAGGATCT GCCGGCCGCG
    2881  GTCACACTCC ACACGGACGC GGAAGAGTGG ACGAGCTGGC AGAAGCGGGG CGACCCGGTC
    2941  CTGCACATCG AGCTGGGCAA ATGGGCCGAC CTGCTCGTGA TTGCACCGCT CGATGCTAAC
    3001  AGCCTCGCCA AGATGGCCAA CGGGCTGTGC GATAATTTGC TGCTCTGTAC GACACGCGCC
    3061  TGGGATCCGA C

//

LOCUS          12     1347 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      12
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             500..847
                     /translation="MFRKAKQTIAPLPKPPTEAQMLEDLQLFHETRSTVSQPACKEN
                     LPTLTEESTMDEWWKVYDASVQQHEQFFDQKTNALELKQSLQEMRSELQAVCDKLMA
                     EIEQDLGRIAATMSE"
ORIGIN
       1  CCCGGTTCAC ATTCCCCACT TTACCGATTG CACCAAGTTC TACAAGTGTT TCAACGGCAA
      61  AAAGTACGAG ATGGACTGCC CGGCCGGTTT GCACTGGAAT ATCGAGAAAG ATTTTTGTGA
     121  CTTTCCCGAG GAAGCTAGCT GCGAACGGCC GCTGCAAATC GACCCGTAAT GGTTCGATCA
     181  ATTAAGTGAA TAAAATCTTT TATTATCAAA ATTAAAATGT GGTTAAAATA GTTTTCTGTG
     241  TGCGATATAT TTGACAGTCA CTGTATATAC AAATTTGAAA TTCAAATTTC AAACCATGTT
     301  GGCATGGATT TGGCTAGGAA CAGCTATACG TGCTCAAATC TCAACTCAAC CCAACTGACG
     361  CGTAAACGCG TTCGCGTTGA CAGCGCAAAG TAAACACAGG CCCAAAAGAG AAATCGTACA
     421  AATCATAACA ACGTACCCAA GTTCTGTGAT AAATTTGTCC TGCCCTCGTG AAACGTGTAG
     481  TGCGTGAATA GATAAAACTA TGTTTCGTAA AGCCAAACAA ACCATTGCTC CTCTCCCGAA
     541  ACCACCGACC GAAGCACAGA TGCTGGAAGA TTTGCAGCTG TTCCACGAAA CGAGATCAAC
     601  AGTATCGCAA CCCGCCTGCA AAGAGAACCT ACCAACCCTC ACGGAAGAGT CGACCATGGA
     661  CGAATGGTGG AAGGTGTACG ATGCCTCGGT GCAGCAACAC GAGCAGTTCT TTGACCAAAA
     721  GACAAATGCA CTGGAGCTGA AACAGTCGCT GCAGGAAATG CGAAGTGAAT TGCAAGCAGT
     781  GTGCGATAAA TTAATGGCCG AAATCGAGCA AGATCTGGGA CGGATTGCTG CCACGATGTC
     841  GGAATAGATT GCTGCCTTTG GAATGGTTTG AATGGATGGC TAAAACGTTG AATTAAAATA
     901  TCTACTGTGA TGAGCCTAAC GGAATAAAAC TACTACGGGT TAACTGTAAT CCTGAATCCT
```

```
     961  GTGGTTTTAT TGCTTTACTC TTTGTATGCC ACCAGTGCCG CGCTGGTAAG TACGTGATGT
    1021  GTGTATCCTG TCTCTGTGTA AGCGCGCATT CCGCGCAAGT CGTTCAGTTG TTCGGGTACA
    1081  CGTGGCCACA GACCGACGAC AGCCACTTCC ACGGGTAGCC TATGGTATGG AATGGGGGGA
    1141  TGGAAAATGG GGTTCAGGGT TGGAAGTCGC ATGTTTAAGG GTTGTCTTCG GTTCTTTGCC
    1201  AAAAACACAG TCGGCAGTCG CCTATGCACT ACACACACCC ACCCAACTCC CCGTTACACA
    1261  GACACACATT CCATGCTAAA GCCCTTAGTC CTAGTTCCAT TAGTACGCGT TGCTGTGTCC
    1321  GCATTACAGT GCCAGTAATA TTCACAC

//

LOCUS          13     2598 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      13
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             join(500..538,1678..1851,1961..2098)
                     /translation="MGEQGKMVQEGEDVNVPKQRRTYCKKCKVHRVHKVTQYKKSKE
                     RQAAQGRRRYDRKQKGFGGQTKPIFRKKAKTTKKIVLRMECVECKYRKQTPLKRCKH
                     FELGGDKKRKGQMIQF"
ORIGIN
       1  GAAGAAGAAG AAGAAGAAGA AGAAGAAGAA GAAGAAGAAG AAGAAGAAGA AGAAGAAGAA
      61  GAAGAATAAG AAGAGGAAAA GGTAAAATTT AATGTAATCC AACATTCAAC GGAGCGTTTG
     121  GAGAGCTACC ACCACCACCA CTACACAGGG ACACCCAGGT TAATGGGGGA AAGAAGCCAG
     181  GGATGGGCAA CATGACACCA AGGGGGGAAA GAGAGTGAGC GTGAGATTGG AGAATGTCCA
     241  GCACAAACAA GCCAGGAAAA TATGGAAGGA ATTGAGGGGG CAATGGAGGG TTGTGATCTC
     301  ACACCGGGGT GGTGATTGCG GTGTGTTTGG AGTAAATAAA AAATTTCTAC TCCAATCAAA
     361  ACCATAAAAT GACAATTTAT GTGCGAAGAA GACGCTCTCC AGTTTTGCGT ACCGCAACCG
     421  ATCGGCCCCG AAGGAGTGCG CGCATGGTTG TCCTCCAAAA AGGAAACGGA ATGGAAACGG
     481  ACGAACGGTA CAGCAAACGA TGGGCGAGCA GGGAAAGATG GTGCAAGAGG GAGAGGATGT
     541  AAGCGAGGCG CCAGCGACTT TATATTCAGG CCAGATGTAA AAGTGTGAGT ATATTTGTCT
     601  CCGTGTGTGT GTCGGTGTAT GTGTGGTTCA AATTATGGAA AGAAGTGGAC GATAAGAAGC
     661  GCAGGACACA CGGACCGATC CAAGCAGTGA TTGCTTACCG ACGGGTTGAC GATTACCACC
     721  AACGGAGGAC ATTACGGTAA TGAGTCGGCA GTAAAGGTGG TCTTATGGAT ATTTGATGCG
     781  GTGTCTCGCG GTGACATTCA CTCTCCTCCC GAGTGCTAAA TGGCCAAGTG TTCCCGTGTA
     841  CGGGTCGGGG TAAAATCGGA CAAATCGGGA AAGCAATTCC CGAGCTACAA GTGTTTCGCT
     901  TTAGCTGATG GGAACAGTAA ATTGGTGTCC AATTTCCACT ATTCAGCGCG CTGCGATTGT
     961  TGGTCTAGCA TGTTTCATTA CAAAGGACAA CACAGTGTGC GTTTGAAAGT TTTGTTGGGG
    1021  TGTTTTATCT ATTCATTTAA AATCATTTTT AAAATACGCT CCACAAACGC AACAGAATAT
    1081  TCTGTAAATA CATATTTGCC ATTATGTAGA AACAAAATTA AATTGCTTGT TTTATCGTTA
    1141  GATTGCGTAA AAGAAGCATA AAGTGAACAA TTTGAGGAAA ATCATAACAT ATTCCCCTAT
    1201  CACATCTCTC CGAACGTTCA TCTTGAATCA TGTTCGAATT CATGCTCGAA TTCATCTTCG
    1261  TACGCTTGCT TGAATGTTCA TCTCATATCT TGAATTTTCT TTCAAATTGT CAATCCCGGA
    1321  AACAGGCGTT CTGTCATCTG TCGTTGGACG ACTGACGTTT ACCGCTTGCT CTGTGCTTCC
    1381  TCTTTTCGTT TGTGACTGAA TCGACGAGAA CATCCAAAAT GGTACGTGAT GGCTAAATTT
    1441  CGCGATATTT TTGCCCCTTT GCGCGTTATT GAAGGTGCAA TTGCTATGCA ATGGTCAAGC
    1501  CAATCATTCC GCAAGTGCCG GAAATGAAGG CGTTCGGTGA AATGGACGTT CGAACGATGA
    1561  ATTAGAAACG ATGTTTACTG GCACCGATTT GTGTCTGTGC GGGATTCTAA GGTTATGTTG
    1621  TGCGCCTATA CTAATCGGTT TGTGGTCCAC CGTGCCTCTC TCCTCTGTTC TCTGCAGGTA
    1681  AACGTACCGA AGCAGCGCCG TACGTACTGC AAAAAGTGCA AGGTTCACCG CGTGCACAAG
    1741  GTGACGCAGT ACAAGAAGTC CAAGGAACGA CAGGCTGCCC AGGGTCGTCG TCGTTACGAT
    1801  CGTAAACAGA AGGGTTTCGG TGGTCAAACC AAGCCCATCT TCAGGAAGAA GGTAAGCAGA
    1861  CCGGCGAACG GTTGTGGGGT TGTTCCGCGC CGAGGACACG TTCGCTCTCT CACACTCTCA
    1921  CACTCACCGT TCGTGCTTTT CCTCTTCCAA ACCCTTCCAG GCCAAGACCA CCAAGAAAAT
    1981  CGTGCTGCGT ATGGAATGTG TCGAGTGCAA GTACCGCAAA CAGACCCCAC TGAAGCGCTG
    2041  CAAACACTTC GAGCTCGGCG GTGACAAGAA GCGCAAGGGA CAGATGATCC AGTTCTAAGC
    2101  TGTTATCTAT CGCCTCCGGA ACGTGCGTTG GCGCATTCGG TTGTACTAAC GCATCTTTCT
    2161  GAAAAAAAAA GCATCGGGAA TAAAAGATCA GCCCCTGTCC CCGGCGTCCT CGACGAGCAA
```

```
    2221  GGGATGATGT GAGGACAAAA CAACTCAAAC GCCTGTACTG TAGTGATGTG TGCGCTATTT
    2281  ATTTAAGATA TATCACCGGT TGGCTGTGCT TTGCTCGTCG GAACTACCGC TGCTGGTGGT
    2341  GCCGTTGTTG GTGGCCTCAT TTTCCAGCAA CGATTGCCGT ATCAGCTCCG CCATCAGGTT
    2401  GCATCGCATC GGTTTGTCAT TCTGGCAGCA GAACTCCTTG CCCGCCGAGA GGTTGCTGTT
    2461  CACCCAGCGG GGCGCACAGT TTTGGTAGAA CAGGTAGGCT CGCTCCCGGA AGCAGTCCCG
    2521  ATCAATGGTG GCGCAGATGA CGTTCTGCGA GTTGTGCAGA TGCTTCACGA TCGAGTCCAA
    2581  ACATTTGTTC GCACACTG

//

LOCUS          14     1386 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      14
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          complement(500..886)
                  /translation="MADDHSDSSLTSIDDERTDNTENLCENPTRDILTEGFFRLFKP
                  VIDDLENNIRGASKTQSELRQQLEELSGEMKNLNFENDPVLHEYFKRMVRVKQKVVV
                  IMNILQGAQERLVTLCQQQELQTPVPGT"
ORIGIN
       1  CACTGAACTA CATCAAGGAA GAGAAGGAAA ACGAGAAGGA TGAGGATGAG AAGAGCACAC
      61  TGCATACGGC ACTCGGAATC GTGAGTGATT ATCTTTCGCT CGAGTTCGGA CGGAAACTTT
     121  CGGTGGCGCT GGGATTCCCC GAGGATGAAA ACATCTCCAA GAAGCGCAAA TCCGTGGTCG
     181  ATCTGGAATC GGCCGTGGTG AAGCGGATTA AAAAGGAGGA AAATCACGAT ACAACCCCCA
     241  TAAAGCTACC GGCGGCTGAA AAGAAGGTCT CTGCCAAGGC GAAAGCGCTA GCCAAGGCGG
     301  CGAGTGGGTC GAAGAGTATT TCTAGCTTCT TTAAAAAATA AAATCCCCCC TCCGTGTGTT
     361  AGGTGTGTTA CGTTAAAAGC ATTATTATTA TTATTACGTT ACATTAGTGC GGAAAGCAAT
     421  GGCCTCTTGG TGGATGGAGG AAAGCTGATG GCTACGAGGA ATAGTAAATA CATAATTCTT
     481  TGGTTCGACG TGTGCCTCTT TACGTGCCGG GTACGGGTGT CTGCAGTTCC TGCTGCTGGC
     541  ACAGCGTCAC TAGCCGCTCC TGCGCACCCT GCAGGATGTT CATTATCACA ACGACCTTCT
     601  GCTTTACGCG CACCATACGC TTGAAGTATT CGTGCAGCAC GGGATCGTTC TCGAAGTTGA
     661  GATTCTTCAT CTCACCGGAC AGTTCTTCCA GCTGCTGGCG CAGTTCGCTC TGTGTTTTGC
     721  TTGCCCCGCG GATGTTGTTT TCCAGATCGT CAATCACTGG TTTGAAGAGC CGGAAGAAGC
     781  CCTCCGTCAG GATGTCCCGG GTCGGATTTT CGCACAGATT CTCCGTGTTG TCCGTCCGCT
     841  CGTCGTCGAT CGATGTGAGC GAACTGTCCG AATGATCGTC GGCCATCGTG GGTGGGAATG
     901  TTTCTGCACA AGGAAGTTTA TTTGTAGATG AATAATAGAG CAAATAAGCA AGGATTCATT
     961  CGGCAACGGC GATCGAAATT CACCTTTTTG GAAAAAGGGG CGATTTTTTC CTCGGTTTGT
    1021  TGTTATTTGT TATCAGTTGT TTGACATTTC TGAACCATTA GCATAAGGTG TATACAATAA
    1081  GCTGTGTTGA AATGTGACAG CAACGCAGTT GACGTTACAG GTGTGCCCAA ATCGCCGTTT
    1141  GTTTTTACTT TTTTTAATGT TTAGCAACAT AATTATAATG TAACAAACAC TTTAATAAAT
    1201  ATTATTCCCC CAATAGTCCT CCTCCCATGA GAAATCTTTG GTTTTGGGCG TGACAGAACA
    1261  CGTGCGAAAG AGATAGAGGG CGCCACTGGC ACTCGCTGTG TCCCAATCTG CCAAAACACC
    1321  GTGCTGAAAC CACGCGGTTT CCATTCGCTT GTTATTCATC TAGCGTTTGC GGCCAATGAA
    1381  GCTAGC

//

LOCUS          15     1407 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      15
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          500..907
                  /translation="MSLFTSSNSRPRSLMEALLAKKIEAASQQTGPGGSRLIRTDSM
                  DSTSSIGSHGSLILGDDVCRCDDCLLGIVDRYTIGPKETASGIKKVGENVLATGSEV
                  VVIDLRQNDAFAKSCYFCSSSTNIVGKLCQSVILA"
ORIGIN
```

```
       1  CCTTAGCGAT TGTTGTGTTA TCAAGGAGGG GACCCACACC GAGGGCCGAT CCCGACCGTT
      61  AATACCCATT CGCATCCAAT CTTTTTCCCC CTCCCGTCGC TAGGTTTACG GTGTGCAATT
     121  AGGTGTGATC ATCGATTAAC CCTTGTGTGA GAGTGAGCAG AGCCGGGACG TTAATTACAC
     181  AACGACGGTT AACGATTACA ATTTAATTAA TTGCTTGTGT GATGTGTGTT TCGTGGAGTT
     241  ACTGTTTAGC GTGCACGCAC TGGCTGTACA CTTCCGGTGC GTCGTAACGC CGCCCGGAGA
     301  AAACTGCACC ACGCTCCCGC CCAAGGTTTC CTCCGGCGGT GCCCCACCAG ATAAAGCATA
     361  TCGAAATAAT TCAGCCATTG GGTGCATTTG TGCGAGTGTG CGTGTGTTTG TGTGCGTATG
     421  CAAAGAGAAG ATAAAGGGAA GGAGTGCCGC CGTTGAGAGC AGAGTAGTGT GTTGAAACTG
     481  TGATTGCCAC CGTCGCACAA TGTCGCTGTT TACGTCGAGC AACAGCAGGC CGCGATCCCT
     541  GATGGAGGCG CTGCTGGCGA AGAAGATCGA GGCAGCGTCA CAGCAGACGG GCCCGGGCGG
     601  ATCGCGGCTC ATCCGGACGG ATTCGATGGA TTCGACCAGC AGCATCGGCT CGCACGGATC
     661  GCTCATACTC GGTGACGATG TGTGCCGGTG CGACGACTGC CTGCTCGGGA TCGTCGATCG
     721  CTATACCATC GGGCCGAAGG AAACGGCGAG CGGTATTAAA AAGGTAGGCG AAAACGTTTT
     781  GGCAACCGGG AGCGAGGTGG TTGTGATTGA TCTGCGACAG AACGATGCGT TTGCGAAGTC
     841  TTGTTACTTT TGCAGTAGTA GTACAAACAT TGTGGGCAAA TTGTGCCAGT CGGTAATACT
     901  GGCATGATTG ATCATATCCA TTTGAGCAAC ATTAAAGTAC ACGATGTACG ATCGCAGTCA
     961  GCATAATGTT TGGAACGATT TCGGTCAGAT CTTTTACAAA TCTTTTTCTC ATTTAGGCTA
    1021  TATCCAGTAT GCCAAGGTTA CGACCGGTCG TTGTTGCTAA TTAATTATGG CGTTGATATG
    1081  GGATCTATTC ATTTTGAGTG GCCATTTTTA TGAAATTGAA ACCTCTGCAT TTGGTAGCGT
    1141  AAAAAATGTC AAAACATTGT TCAAAACTAA TAACAACGAA AAAATATGCC TGCTTGGAAA
    1201  AGGTCAAGAT GGGGAAGAGC GAAAATAAAT TAACAGTTTT ATCTGGTGTG AGGTAGATAT
    1261  TAACCTGTTG TTGTTTAAAG CATCAACTTG CATTGATGTT TCATTTTTAT ATTGACTCAA
    1321  TTTATTTATT ATAATTAAAA TAATCCAGAA ACAACGGAGT ATGGCATTCA TCATAAATTG
    1381  AATTTGAGGA GCGTCTTAGT TTTGCCT

//

LOCUS          16     1506 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      16
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..883,965..1006))
                     /translation="MKFLLLALSVFMLVTASTAQSSKPAAVVQMQMTVGKLLMLVRD
                     LSVANNAFAKDTEDQTALNTLYTTSEDLYQLLPVFGSSSTSTLPLVTRERVNRVITN
                     FKDALTKWESAMDERSAPNLVSTFKAVENAFLSLGGVVFSL"
ORIGIN
       1  AGGTTGCGCT TTTGGAAATT GCTTTAAAAG AGCCAATTGG ACGTACATTT CCAATTATTA
      61  TTCGGAGTTA CTTGGAATTA CGTAGTAGCA GTACGATCCA TTTCATCTTG TGGGTTTGCA
     121  CGTGTTTAAA CAATCTATTT GCAACAATTG TGTACACATT GCGAAAATGG ATGTATGAAT
     181  TCCAAGACAT TGGTGTGCGT TTGCCAATTC TTTCGCCGAA CTCTTCTCAC TATTCCATTG
     241  TGCACTGTGT AGCGATCTCC CTAGATCCCC TGGAAGATCT TCAGCTATAC GCCGACACTG
     301  TATCACTGTG GTGGAACAGT CTGTTGACAC AGGTTGTTCC ACATGCTCTG TTTGCAGCAT
     361  CGGTTGATGC TATGTACTAC ACGCACCGTT GTTCCAAAGA ATGTGTACCA ATATATGCCT
     421  ATAACATAGT CTTTTTACAA CTGCACTAAA CTTAATCTAG TGCATTTATT ATAATTTTTA
     481  CACATAGTCA TCTGTGATCT TACAGCGAGA AAACTACTCC ACCCAAACTC AGGAACGCAT
     541  TCTCGACCGC CTTGAACGTG CTCACTAGAT TCGGGGCTGA GCGCTCATCC ATGGCTGATT
     601  CCCACTTAGT GAGAGCGTCC TTAAAGTTGG TTATCACTCT GTTCACCCTT TCCCGCGTGA
     661  CCAGCGGCAG CGTGGAGGTG CTCGACGAGC CAAACACTGG CAGCAGTTGG TACAGATCTT
     721  CCGAGGTGGT GTAGAGAGTG TTTAGCGCCG TCTGATCTTC GGTGTCCTTG GCGAAGGCAT
     781  TGTTGGCGAC GGAAAGATCT CGCACCAGCA TCAGCAGCTT ACCGACCGTC ATCTGCATCT
     841  GTACGACAGC GGCCGGTTTG GAGGACTGTG CCGTGCTGGC AGTCTAGATG GGATAGGAAG
     901  AATAATCGAT TACAGTCACT GTGTTGCTAC TATAAAGGTT TGTTCAACGG TATCTGCAGC
     961  TTACCACCAG CATGAACACG GACAGGGCAA GCAGCAGGAA CTTCATCACG GAGACACACG
    1021  CGGCACAAAA CTGTGAGGTT GGATGATCGT TCGACAAAGC ACGGCACTGG TTTTATCAGA
    1081  ATCGCCTAGA GGCCGTTTTA TTCATCCATT CCGCACCCCG GGTTCTGTTC GATCATGGCA
    1141  TTGCGCACAG CACTGGTTGT AAAATACAGA TGGTTTTTGT TTGCACTTGA GAATTGTGTG
```

```
    1201  AGATATTGAA CATGAGAGGG GGGTTTGTTA GAACTGTGCC TGAAGAGTAG AGGGGTGGTG
    1261  GTGATGTCTT TTTTTGGGTT TATTTGGAGA TGTTTTTTTT TAAATATTGG GCGGTCCCGA
    1321  TCAATCACAC GCACGACACT TTAATTTTGT TGCAAAATGT TTTTTTGTTT CTAGTTACTT
    1381  AGTATAAAAG GTGAAACCCA TTCGCCGTTG AAAAAACACA CTCTACACAC AGCTGAAGTT
    1441  GTTCAGTTAT GTCCTAAAAC TCCCAAGCAA ATGGTCTGAA AACACGCAAC AATTCCATCG
    1501  GGAATG

//

LOCUS          17     2320 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      17
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..678,753..880,1675..1820))
                     /translation="MSIKVGGAFRRSGGIVSAIGKQLKLVNLKGVQRITATFDPFDE
                     SAVPTREFVHHLSAPKISQTNPTCLLKTEVVCDRRPPSIVFQLIPSVQAEAKLKKIE
                     LSSANLSTLELLKLCNKHVSALAPKEIATSVVKTKSEKKAGGAGGGGKRR"
ORIGIN
       1  ATTTAAACAA AAAAAAAGCA AAAACGACTG TTAAAAAAGT AGAAAACTCT ACAAAACCCT
      61  TGTTTGGACA ATTTCTCAAA GAAATCTACA CCCCTTTTTT AAAAACAGGC ATTCGAGCGC
     121  TCGTTCGGTC CATTATTTTC CAGCTCTTTC GAACCGGGTG CCTTTATCAT TGCTACTATC
     181  TCTACTTACA TCGTTTCTTG CTTCCGCCGC TCCGCCTCTT CAGTCATCAC CTTACACCTT
     241  ATGTCTTCAT TAAAAAAAAA AACCTTAGCA AACAATCGAC TAATCGACCG GCTTCAGCGT
     301  GATAAACCGG GACAGCACCA ACACGACGGC AATGGTGGCC AGCACGAAGC AGACGGCAAA
     361  CATGTAGAAG CTCTCCTCGG TAAACACGTC GGTAAAGGCC GGTATCTGGT GGTACCAGTT
     421  CTCGATGATC GCCATGTGGC CGCCCTTTAG CTTCTTGTGC GGCATGGTTT GATGAGCTGG
     481  TTTGCTGACG TCCTTTTTTT TAGCGCCGTT TGCCACCACC ACCCGCTCCG CCCGCCTTCT
     541  TTTCCGACTT GGTCTTGACC ACGCTGGTGG CGATTTCCTT CGGCGCCAGT GCGCTCACGT
     601  GCTTGTTGCA GAGCTTGAGC AGCTCGAGCG TGGACAGGTT GGCGCTGCTC AGCTCAATCT
     661  TTTTCAGCTT TGCTTCAGCT AGGGATGGTA GTAGAGGAAA AGGGAGCATT AGGCTTGTTT
     721  TGAAAGGGTA GAACAGTACG CTTAACACTT ACCCTGCACC GATGGTATGA GCTGGAAGAC
     781  GATGGATGGG GGCCTTCGAT CGCACACGAC CTCTGTCTTG AGCAGGCAGG TTGGATTGGT
     841  CTGCGATATC TTCGGGGCGG AAAGATGGTG CACAAATTCT CTGCAAAAAG TAATGTGGAA
     901  AAGAATTCAG TTTAAGGAAA TGTTTTTGTT CGTCAACATC GTTCTGGGTC GCGCATTCTT
     961  CTCGAACGTT CTGGATTCGC TCTCACAAGC TGCAAAACCC TGCCGTTAAT GCTCTTTCTC
    1021  TCACACACGT GCTTGTTTTT TTTTTCTCCC TCTGCCTTTA CAGACACGGC GCACGATTGA
    1081  GGCGAGGAGG CGGCAATGTC CACCGAACGG GAACTGACTC CCGGAACAAC GCAGCAGCTG
    1141  CACCCACTGT CCTACTCGGA TATGGCCACA CACATTCAAC TGAACGGAGT AATAATGGGC
    1201  AAATCATTCA ATGGTAAGCT GCCGTCCTTC CTTCCTATTA CCTCCGGCCG TGGCCTACTA
    1261  AATACTTTCC CGTCGTCCTG TCTCTTCCCC TGAGAATGGC TACAATTGTC CGAATCTACT
    1321  GTTGCGCCCG TTCTGCTCTT CCGATCGCGA TTCCTTACGT TGTGTAATGC TGCTGCTGCT
    1381  CACTTCTAAT CGGGGGTTTT TCACACTTTA ATCAGCCTCA AAAACGTACC CTAAAATTAT
    1441  ATGGAAAACT CGGTGCCTCG GTCGTACCTT GTTCCTCTGC CTTAGCTGTT CGTTCTCAAA
    1501  ATACACATCA CCCACCGACG ACGGCGTTTC GCGTTTCTCT CGCCCTGTGA CAGACGCGTG
    1561  CGCATTGGCC ACTTTCCGGG AGCGGGCCAT TTTGCATTTG GGGCGCGTTT GCGAGATGTA
    1621  GGATGGAACC GGCGGGGCGG GATGGGCAGC TGGATCACTG GTTACGGCTC TTACCTCGTT
    1681  GGAACGGCAC TCTCGTCGAA CGGGTCGAAG GTGGCCGTGA TGCGCTGCAC ACCCTTCAGG
    1741  TTGACCAACT TTAGCTGCTT GCCGATGGCC GATACGATGC CACCCGACCG CCGGAATGCT
    1801  CCACCGACCT TGATGGACAT CGTTACGCTG GAGGAAGTTT TATTGTTGCG CTTCCTGCAG
    1861  AAATGGTTGT TGTTTGCCGA CTGCGTGTGC GGCAGCGGCA GGTTTCATAA CAATTCTCGT
    1921  TGTGCAAGGG TTGACGTTTT GGAGATCGGA CTCGTGTTCG AAAATGATGT TCGAACTTGG
    1981  GTTAGGCAGT GCTCTGGCAA TTTGTCGCAT GAGACAAATG GTGCGTGACG AATGAATGTT
    2041  GCATAGGTTT TTTCTTGTTT TGTATAATAA AAAATCAAAA ATACTAAATT TTTTGAGGTG
    2101  GTCATTTAGA TACAGAAAAT TTATTTTAAG AAAAGAGTTC AAATACAAAT AGTCTCGAAG
    2161  ATACGCCGTT CCACTTAAAT GCGGATTCGG AGATATGCGG CTTTCTAAAT TTGACAGGGT
    2221  TGAGCACATT GTACTGATTT GACTCATCAA TTTTCAAATG TCAAATACTT TGCCTTTCGA
```

```
    2281  TCGAACGTTA AAGACAGTTT TAAAAATGTG AAAACTGTTA

//

LOCUS          18     1556 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      18
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..898,976..1056))
                     /translation="MSTIYEAAATIKPKTPDEEEVLKNLWSIRPCYLYNEEYEDCTS
                     VKARFHQYFIHGESIDCNQWKRDFDNCVRFEKNATDTKSALELIRSEQKRREERIRA
                     HYGNNVWTKRTAPPQDWAKPLPEHMQKEYEGSFLASKAKELREGPQTPEVESSSSCA
                     IM"
ORIGIN
       1  GGGCGTTTAA ATTTAAGTAA GTTGGCCATG CTTTCAATCG CTTGGTGCAT CATTTCGTTT
      61  CCTTGCTTTC TCCTCCAGGG GCGATCTCAC GCTGCACATT CGCAAGCACA CGGGCGAAAA
     121  GCCGTTCAAG TGTGACATCT GTGGTAAAGC GTTCAACAAA AGCTACAACG TGGTGATACA
     181  CAAAAAGTCG CACCGCAATC TACCGCCGCC GGAACAGTCG ACGCTGCTGG CTGCGGAAAC
     241  AGCGATGCTA CCTCAAGGCG AGGTAAAACC GGAACATACG GCAATATAGC CGGCGCGGCA
     301  AACAGTGCAC TGTGACGATA TTACGCGACC GTTTTTTTGT GTTAGCTTTT AGTTCATCAT
     361  CCCAGATATA TATCATTTTT AATCTTATTA AACATAACTT TATTCACTCA CGATGTAAAT
     421  GACGGATGTT GGGGTTATTG CAACTGAAAA AAGTGATTTA AAAAATGCTA TTTAAACTTT
     481  CGTTCGCTGC GCCTTCTTGT CACATGATAG CGCACGACGA GCTGCTTTCC ACTTCGGGCG
     541  TTTGCGGTCC TTCCCGCAGC TCTTTGGCTT TCGACGCTAG GAACGACCCT TCGTACTCCT
     601  TCTGCATATG CTCCGGCAGT GGTTTGGCCC AATCCTGCGG CGGTGCCGTC CGCTTCGTCC
     661  ACACGTTGTT GCCGTAGTGG GCGCGTATCC GCTCTTCGCG CCGCTTCTGC TCGCTGCGTA
     721  TTAGCTCCAG CGCGCTTTTC GTGTCCGTTG CGTTCTTCTC GAACCGCACG CAGTTGTCGA
     781  AATCGCGCTT CCACTGGTTG CAGTCGATCG ATTCGCCGTG TATGAAGTAC TGGTGGAACC
     841  GGGCCTTAAC GCTGGTGCAG TCCTCGTACT CTTCATTATA TAAGTAACAC GGCCGGATCT
     901  GGACGGGGGC GGAATGCCGG TGAGCCAAAA TTGGATTAAA TGCGGAACGA AAGCAAACCG
     961  CGCGGGTACA CTTACCGACC AAAGGTTTTT CAGCACTTCC TCTTCGTCGG GTGTTTTTGG
    1021  TTTAATTGTA GCGGCCGCTT CGTATATTGT GGACATGGTT GGTTATTTGT TGTTCAGCTG
    1081  CGGCAGTGTG GGCTGGAAAC CCCGTTGCAA TCAATTTTGT TTATGTCCTT TGACAGTTGT
    1141  TGGACAAAAA CCAGGGCTAC ATTCACGCAC ATCGCTAGTT GGTTTACGGT GTAGGATTGA
    1201  GGTTAAAGGC TGAGAGGAAC GGGCAGCTTT TTGAACGCAA AATCAGTTGG CTGACTGAAG
    1261  GTTGAGGCCA CAGTTCGCTT TTCTTCGGTG GCAAATAACA ATAAGATATT AAATATGAAC
    1321  CAAAAATCAT GTTTTACAGC TTCATACTTG GTTACAACTA TTATATTAGT TTTGTTTAAC
    1381  TTTATTGACC TTTGCAAAAT GTTGCGTCTT GAAAAAATAT CTTAATACAA ATGAGTGACC
    1441  CAAAGCGTAA AACTGATCGA CGGATTAACT AAACCAAAAA GTAAACCATA CAATCTATGC
    1501  TTGTGCTTTC GCTTCCTTCA ACGCTTTCCG CGCTTCTTCG TTTCGGTGCG ATTTCT

//

LOCUS          19     1568 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      19
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..670,745..1068))
                     /translation="MFSNCQDVKLAYNEMPISLETVRFDRDSEGVCDTLHAEYEVRE
                     SSDDVEWELIITTYQCEQQTAKICLDNPKEYIEPMHCDRFHSDNSGPWFFIASSMTN
                     GDRCGRMTGRYNLDAAVLKIKYLEQYIAMGKGTYRVRMLFHIPGTNLDTLNVRGCCE
                     MDFDVID"
ORIGIN
       1  CTTAATTGTA ACAGCCCCAA ATGCGCTCAT TTGAATGTGT CTTTCGTTTT CGTCGTCCCC
      61  CCAGTGTGTT GTGTTGTTCA CCGAGACAAT AAAGCACACA CCACACGCAC CAGGCAGTTG
```

```
     121  GTATTTGAAT CTTGAGCAAA ACAATTATTG TGATCGTACA GAATCGAAAA TTTTAGTTTA
     181  AAGCAAAGCA CCACCACCAG TCAGTCAACC AGGCGTATGA AATATGTACT AGAAAATGCA
     241  AGATACAGGA TACAAATACA ACAACCTATC GGAGCTTACT TCGGAATTGT ATCTCTTCTG
     301  CGGCAAAGGA ATCTTATTCC AATCAGAAAC TCAACGCAAA GAAAACAGAT CAATGTGTGA
     361  ACCAGGTGGA AGAATAACGA TCGGCGTTTT TGCAACCCCA TTTCCCCGTC TCGCTTAGCA
     421  AGATGTGGCT TATTTAATGT TTATAGTTTG TACAGTTTCA GCATTTACGT GTTTGAGCAA
     481  CTCGACCACG ATCGCACGTT CAATCGATCA CGTCAAAGTC CATCTCGCAG CAGCCGCGCA
     541  CGTTCAGCGT GTCCAGATTG GTGCCAGGGA TGTGGAACAG CATCCGTACG CGATACGTGC
     601  CCTTGCCCAT CGCGATGTAC TGCTCCAGGT ACTTGATCTT CAGCACCGCA GCATCGAGAT
     661  TGTAACGACC CTGTCGATGA AAGAGAGAGA AAGATTGCAA TGACCATCCA AAACAGCTGC
     721  TCTGTAACCG ACTTAATAAC ACACCGTCAT GCGGCCACAC CGATCACCGT TGGTCATTGA
     781  GCTGGCAATG AAGAACCAGG GCCCACTGTT GTCGGAATGG AACCGATCGC AGTGCATCGG
     841  CTCAATGTAC TCCTTCGGAT TGTCGAGACA AATTTTCGCC GTCTGCTGTT CACACTGGTA
     901  GGTCGTGATG ATCAGTTCCC ACTCGACGTC GTCGGACGAT TCCCGCACCT CGTACTCCGC
     961  GTGCAGCGTG TCGCAAACGC CCTCGGAATC GCGATCAAAG CGTACCGTTT CCAGCGAGAT
    1021  GGGCATTTCG TTGTAGGCAA GCTTCACGTC TTGGCAGTTG GAAAACATGG TCATGTTGAC
    1081  ACGCTGCAGG TTGGCGGGGA AGTTAGGCAA ACGTTTACAC TCTCCACACT TACACTGATT
    1141  TTCGCGCTAC TTACCAATCC ATGAGAAAGT TCAAAGGTAC CACAAATGAT AAGCACATAC
    1201  AGCAAGATGC ACCTTTTGAT AGCCATTTTT CAAAGAAGGA AGTGAATTTA CCCTCCAGCT
    1261  TCGTGGTGAG TAGTGAAATT GAGCTTCACT GCAGTTGTGC CGATCGCTTT ATAGTTATCA
    1321  GAAGGTATCG GAAGGGCTTG GAATTCGTTT CAACAGCCGC TCCTCGGATA TAGTCATTAT
    1381  TATGACGGTG CAAAATTTGC GTAAACGCTA TGGCTTACCG CATTTGCGTA GAGCTGTAAT
    1441  TTATGCGTAG ACTATCGAAT TAGGCGCACG TGCTCGCATT GTTGAAACGG GACAGAGCAG
    1501  TTTCGGTTAC ATTGATTAAT ATATTAACCA ATAATTGATG AGTGTCTGTA TGGTTTACCT
    1561  CAGGAGGG

//

LOCUS          20     1698 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      20
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..643,715..1014,1106..1198))
                     /translation="MGVNDAVNYVLFNNPVSSTVWSMAAQGARAAGLKPKAVTQTAA
                     THTKAVVEQALPPMWKLLGHNRHIVRLAGLSGAAAVILGAYGAHYHFTPKDEDIQER
                     DPKQIFEMTNRYHFIHSLALLAAPFARRPYLTTALMASGMTLFCGTCYYVAFTNDRS
                     VVRLTPMGGFLLIFGWLSFII"
ORIGIN
       1  GGATCAACTG CACCGGCTGG ATTGTGAGCG GCTGGCGTGG TGGCTTTGCG AGCGACAGCT
      61  GCCGAGCGGT GGCCTAAACG GACGGCCGGA GAAGCTGCCC GACGTTTGCT ACTCCTGGTG
     121  GGTGTTAGCG TCCCTCACCA TCATCGGCCG GTTGCACTGG ATCAGTTCGG AGAAGCTGGA
     181  GAACTTTATC CTCTCGTGCC AGGACGCGGA GACGGGCGGA TTTGCGGACC GTACCGGCAA
     241  CATGCCGGAC ATATTTCACA CCCTGTTCGG GCTCGGCGCG CTGTCCCTGC TGGGCGACAA
     301  GCGGCTGAAA CCGGTCAATC CAACCTTCTG CATGCCACAG TACGTGATCG AGCGGTGCAA
     361  TTTGAAGCCG AAGGTAATCG TGCTGTAAGC GTCGCTTTAT TTTATCAATT CATACATCAA
     421  ACGCGACTGT ACACTAAAGT GGATCATGAA ACACGTTGAA AATGCTCTGC TCTTGCTTTC
     481  CCCTCTGTTG TTAGTAAGCT TAGATAATAA AAGACAGCCA TCCAAAGATC AAAAGGAAGC
     541  CACCCATCGG CGTGAGCCGG ACCACGCTGC GATCGTTGGT GAACGCCACG TAGTAGCATG
     601  TGCCACAGAA AAGCGTCATT CCCGAAGCCA TTAGTGCCGT CGTCTGGTCG TGCGGGAAGA
     661  GAGAAAGTAA GAGAGAATGT TCTGTGCATA GTCCTTCAAA TCGAAATGAC TTACCAGATA
     721  CGGACGTCGT GCAAAGGGGG CAGCCAGAAG GGCCAACGAA TGGATGAAGT GATACCGGTT
     781  GGTCATTTCG AATATCTGCT TCGGGTCCCG TTCCTGTATG TCCTCGTCCT TCGGCGTGAA
     841  ATGGTAGTGA GCTCCATAGG CCCCAAGAAT GACTGCTGCC GCACCGCTCA GTCCAGCCAG
     901  TCTCACGATG TGCCGGTTGT GGCCCAGCAG CTTCCACATC GGTGGCAACG CTTGTTCGAC
     961  CACCGCTTTG GTGTGAGTTG CCGCCGTTTG CGTTACGGCT TTCGGTTTCA GGCCCTAGCG
    1021  GGGAAGGACG ACATCCAGGG ACAAAGTCAC AATGGTCGCA GGAAGCGATG GACAAACAGT
```

```
    1081  TTGTTCAACT GCTGATTATA CTTACCGCAG CTCTGGCTCC TTGCGCTGCC ATGGACCATA
    1141  CGGTGCTGGA GACGGGATTG TTGAAGAGGA CATAGTTTAC GGCATCATTT ACACCCATTT
    1201  TGCGGCAGAA TTATACGTTT GCTGTTGTGA AATGAGCTAT TTTTGCACCG TTTGTTGTGG
    1261  ATAAACGAAT TGTCATTGTT GGTTGGGCGA GTTTAGAGGT TGTAGTTGAT TTCTTCGTGT
    1321  GTGAATCTCC AGCAACGGAC CGAATCAAAG TGTCCTGGCG AAGTGGTTCC TGCTCTTGGA
    1381  AATTGATAAA AACGGTACGA TATCATCTCG ATGCTCCTTT TTACATGCTG GCAACCACTT
    1441  CTGCTAAAAT GTCTTCTAAT TTTTGTAGAG AATAGTGTGC GTTTCAGGAA CTACAGCTGA
    1501  GCTTACTACT GGGTTGGAAG TACATCCCAC AGATCCTTTA TCGTGTAACC TTCCTTCAGC
    1561  GATTCCATGG CGAAACCCAG CTCCTTGCAG TAGACCGGTT TCCTTTTCGA ACCATCACCC
    1621  GTAGCCGTAT CCGCCAGGTA AGCAGTAAAC TTGGAATTTA TGTCACCACC GGGCGTGGAA
    1681  TCGAGCTCCT CAATGTCG

//

LOCUS          21     5552 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      21
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          join(500..521,3976..4250,4744..5052)
                  /translation="MEKPINSDMLNLETGEIEYGTVGGYDRERIYLDKETGEVSSDP
                  TEQAGSRRRRSKFHIPKQYLRNQCGIIGRRPAAPAAPSVSCALMACKQKMIPMMKEE
                  ITVGYNSTAHLLFLLPLLFVTPKLFALVISFVEVMLHMWAHRKNSTNGNPDVYYRSP
                  MHVVTRQFCEVCRHERLMGRVGKLQDVRMKQMQNYFRKVTRNIA"
ORIGIN
       1  CTTTTCTGTG CCGTGTACAG AACCAAGAAG AAGACAAGTA CCTCTCTTCC ACGTCCCCAA
      61  AAGCCCACGA TCGGGTTTTT TTGTGTGCTG CTAGCGATTG TTGGTGCGGT CCGTTGCGTG
     121  TGCCATACAA TTGAAGGGTT GTGGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGCAAAAA
     181  AGGCCTAACG TTTCAGCCAA CAATCAAGCA TCTTGTGCGT TCTCTTCTGT GGATAAAAAA
     241  GAAGCAAAAC GAACCAGGTA AGTCGGTGCG CGAAGGCTTG AAAAGGTAAT AGATTAATAG
     301  CCGCACACGG CCTGTGACTT CTGCGTGATG GCATCTAATT GCTGTGCAGC AACCAAATCT
     361  TGTGGCTTTT GTCAATGAAG CATTTACATG TAGGTTTGAT AATTTGTGGT TTGAGTTTGC
     421  TTAATGCATA TGAAGACGGG GGATACATAA ACAGTACATC ATTTGTCATT GGACGGTAGC
     481  ACAAAGGACA GGACACCAAA TGGAGAAACC CATTAATAGT GGTATGTCCC TTGAACTTGA
     541  CTTGGCAGAC CTTCAACCTT CAAAAGTGGT TGCTATCACA CGCTGATGGA ACAATTGTAG
     601  GCTACAAATC TTCATCTCAA ACGTTGTGCG AAGGGGAAGG CGGCTCTTGT TTCTAAGGTG
     661  CGCAAACGTA TGGCTCGGTA TCTTTGGCAT ACGAAGATAA GCATGCGTGG CATTTAATGT
     721  TGCGCCCGGT ACCGCAACAC ATTGATTATG AGCGAAAACA GGGAAAAAAC ACAAATAGCA
     781  GGTATCGCAC ACGAAGGGAT CACCAGGTCT CTTATCGTCT CGGTTCAGTG CCGTTGGTAC
     841  GCTGGTGATA CTCTTCAGCA AAATGACCAC GATTATGGGG GCTTAGCCCG TTATAGCTAC
     901  TGCTGCTGCT GCTGCTGCTG CTAGGCTGTT TGTGAATCGC ATTTTGATCG ATCGGAACAG
     961  ATAATAATTG GGCTTCATTT TCACTTTCAT GCTGGTTTGG TTGGCAAACA GCGTTAGCGA
    1021  TTAATTTGAG ATTCCCCATC AAGTGAGGTG TTGATAATAG GAGGATTATT AGCACTGCTC
    1081  GCAACGGAAA TCCCTGCTCC AAGATGGTTT TAATTGGTTT CATTTTCTTT TAATGGTGAT
    1141  CCAATGGCCA AGAAGCACAC ACCAACATAA GAAGAAATTC TACTTGGTAA AGAAACATGG
    1201  AAAATATATA ATGTGTGCTT TTTTCTTTCC TTTCGCTTTG ATAAGTTCTT ATCACCTGTG
    1261  TTAACACACC TGACCTAAAA ATGTATTGCC GTCTGGGTGC CTGGTGCTTG TACAGCAGCA
    1321  TGTTGAGAAA TCATTTGGAA AAGGAAATCA TTATTAATTT CCTCATTATA CTAGAATGTT
    1381  ATGTGCTAAT AACTATATTC CGATGGTTTA ATGGCTCAAT AGGATTTCCT AGAAACTTGA
    1441  GCTCTCTAGA ATAAGCGGAG AATGTTTAAG TATCAATACT ATGAACTCTC CCGTCTCCCA
    1501  CGCGGTGGTA CAGTCGTCAA CTCGAACGAC TCAATAACAT GCCCGTCATG GGTTCAAGGC
    1561  TAGAATGGAC CGCGCCCCCC TTAGCAAGGA TTGACTATCC GGCTGCTTGT TAATGAATAA
    1621  AGTCTCGAAA GCCTGTATAG GCCGCGGCAT GTCCGCCTAG GACGTTACGT CAAACAGAAG
    1681  GAGAACTATG AACTATGAAC TCTTAATAAA AAAAAAAAAA AAAAACCAAA AAAAAAAAAA
    1741  AACAAAAGAA GACTCTGACC AACATCGAGC TTGACACTGT CACTTGTTTT AAACCGTATT
    1801  CCTCATTACA AGTATTTCAA ACTAATGTCC AAAATTGTCC AAACTTGTAA TTGATGAACA
    1861  GATGTTCTCG ACATCAGTTT TTTTTGGAGA ACTTTTAGCT TTTTTGTTCA TTTTTTTTAC
```

1921 ATAATAGTAA AGAAAATAAT GACAAAGTGA TGACGAATTT CTCTATTCTG AAATGCCACT
1981 TTGCAACATT TGCAACAATA TTTGTACAAA CAAGATCCTA CAAAGATCTG AATGAATTCT
2041 TCAATATAGT TCACTCTTTT CTCCAGTCAG ACTTCAATTT TAGCAGCATT AATCAACACT
2101 CGCGAGATAT TTTTCAAAGG GTCCAGCTGA CAAAAACACA GCCTGACAGC TGTCGAAAAT
2161 CATCTCGCGA GCGTTGAAAG AATGCTGTTT ATATTGAGAA CTGACTCGAA AATCAGTGTA
2221 ACATTGTACA ATTTTTGTTT TCCAATAAAA TGGATTTTTT GACGGCGCTA AGAATATGCA
2281 AACGTATGCA CTTTTGCGAC AGTTTGTTTT TTACCTATAT GTCATTTATT AATGAGAGTT
2341 GACGAACGAT CGTTGAAGCG CCCACAAATC GTTAACATTT CACACAATGC TAATCATAAT
2401 TAGAAAGAAA CAAAAAACAT CGTTGCAAAC CCTCTTATAC AATCAGTATT GTGTTTTGTG
2461 TGACAATTTT ATTCCCCGGA CACATTTCGC GTCGCTTATA TGGGCAAATA AAAAGCCCTA
2521 CATTTAGCTT AATGTTCTGA CTTACCTCCA CTCGCTTGAC AAATGATGAT GAGCGCTCGG
2581 GATGGCTTAA AAAGTGGCAC TTTATACCCC CCATAGACGG TTCTCGGAGT AGTTTAGGCG
2641 GCACGAGCTA CAACCAGCGT CCAGAGTTAA CTCATTTCGC TCTTCCCCAT TCTGTAGTAA
2701 CCGATTTAAC AGCTTCCATC GTTACCGCGC CAACGTGACA CTGGTAAATG AGCGTTAATC
2761 GATGTTTACT AAGAGCGATT CCATTGCATG GGTGATGAGA TTTTGCTTCC TCGTATATCG
2821 TAGTACGCGA ATGGAGCGTG GCGTGTGGGT TAATTGCTAC GATCGTTAGG GTAGGGGCAC
2881 ACACATGATC GCTTCAAGTG AGCGTGCGTT ACAGCCCATT CCGGCTGATT CGAACTTACC
2941 TAAAGGGAAA ATATTAATAA ATTGGACAAA GTTGGACCAT GGTCAAATTG TACCGTACCA
3001 GGTGAAAAAA AAATCTGGGG ACTCCAATAT CCCTTACCGA TCTAACAATA ATTGCCCATT
3061 TTCATTACGC ACCATTATTA GCGTAACGAA TTGTCCATCA GCAAGGCCAC CGGATCCAGC
3121 GCGTCCGTGC GGACCTCGCG CTGCAGGGTT CAAGGGTCCA GCAAATGCTT GCCCATTGCT
3181 TCATTCTTCT TCTACCTCCT CCTCAGTCTC CTCGCGTGCA AATGTCGCAT GCTGATTAGC
3241 GGTCCCGAGG AGACTAAATG AAGACGAGCA GTTCGCATCG TTTCGGTGAT TCTCTCTCCC
3301 CCCCTCAAAA CACCATCGAT AGCGATCCAG ATGTCTCACC AAAAAACGTG AAAAAGACAT
3361 CACACACACA CACACACACG CACAAACATA CAGGCAAGTT TGCACATAAA TGTATGCTTT
3421 ACGGGTCAAT CACGACCGCG CGGGTTGCGA AGGTTTCTCC TAGCTTGCGC CGTAGGAGCG
3481 CCTTTGCACC TCAGACTTGC TCGGGGGAAC GTGCACCCCC CCCCCCCACA TTGACCCCAC
3541 GCCTCGTGGA GAGAATTGAT GGAAATTTAT TGGATTGTTT TTGCTTGCCA CCGGTGTGGT
3601 GTCTCGTGTC TGGCCGAACT CTAGTGGACA CACTGCTGGT TGTGTCCACG TTTTACGGTC
3661 GTCGATCGTC GCGTTGGTAA TTGACGCCAT TTGCATAATT TACTACCAAC TTAGAATTTC
3721 GGCTGTGCAC ACACACACAC ATGCATTTGG GGAACAAAAA CCGGGAAGGG AGAATAAAAA
3781 ATAATGTGTG AAAAGCGTTT TCGTCCAGCA GATTGGATTG TTCGTGCAAT CAATCATTTT
3841 AAAGTTTGCA CCGCTTCTTC TCCGCGCGAA GACGCAATTT AGTGCATATT ACACACACAC
3901 ACTCACACTA GCAGCTCAAT CAATCAAGAT CATCCAAATC ACACAATTTT TCGTCTTTCT
3961 TCCGCTTCTT TGCAGACATG CTGAACCTCG AGACGGGCGA GATCGAGTAC GGCACGGTCG
4021 GCGGGTACGA TCGTGAGCGT ATCTATCTGG ACAAGGAGAC GGGCGAGGTG AGCAGCGACC
4081 CGACCGAGCA GGCCGGGTCG CGGCGGCGGC GCTCCAAGTT CCACATACCG AAGCAGTACC
4141 TGCGCAATCA GTGCGGCATC ATCGGCCGGC GGCCCGCGGC ACCGGCCGCC CCGAGCGTCA
4201 GCTGCGCGCT GATGGCGTGC AAGCAAAAGA TGATACCGAT GATGAAGGAG GTACCTGTCA
4261 AATGACATCC ATACACTCTA GGTACTCTAG GCTATTACTC TAGGCTACGC TCTGCTACTA
4321 CTACTGCTAC CGAACGCGAT GGACCAAGGC ACGCCTGAAG GGGCGCGTGC GCCCGCGGAC
4381 GAGGTACAGA AGACGAACGT TTCGTGTGTT ATCCTATTTA ACCTTTGTTT TTTTGCCCCC
4441 CCCCCCTGTA GTACTAGCCT TAGTAAGCTT AGTGATAAGA GCTACCGTGT AGTGTACGGG
4501 AGTGAATGTT TGTGGGTGTG TGTGTGTGTG TGGGGATGGA TGTCTGTATG TGTGACTGCG
4561 TGCGTCTGGT TAAGATCTTA ATTCGTAGGG CTCAGATAGT ATTAGTAGCA TCGTTTAACT
4621 GCATCCAGAA ACAACAGCTA ACACTCAAAT CTACTACTAT CTCTTCTCTC TCTCTCTCTC
4681 TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TTTGGACCCA
4741 CAGGAGATTA CGGTCGGCTA TAACAGTACC GCACATCTGC TCTTTCTGCT GCCGCTGCTG
4801 TTCGTGACGC CGAAGCTGTT CGCGCTCGTC ATCTCGTTCG TCGAGGTGAT GCTGCACATG
4861 TGGGCCCACC GGAAGAACAG CACGAACGGC AACCCGGACG TCTACTACCG CAGCCCGATG
4921 CACGTGGTGA CGCGCCAGTT CTGCGAGGTG TGTCGCCACG AGCGGCTGAT GGGCCGGGTC
4981 GGGAAGCTGC AGGACGTGCG GATGAAGCAG ATGCAGAACT ACTTCCGGAA GGTTACGCGG
5041 AACATTGCTT AGCGAAAAAG AAGAAGAAAA AACAACGAAA ACAACCCCGG AGGCGCTTGC
5101 TGTCTACTGC CCTTCCTACC GCTACGTACT ACAACTACTG GTGTGCCCTG ACACTCTGTG
5161 TTTTTGAAGC TTACGCTCGC TCCGTCAGCC GTGGAGGTCT GGGGATGATG ACCGATAATG
5221 CTGCTGCGCG CTGCCAATGT GTGCGTGCGT TTTTGTGGCT AGCCCCTCCC CAGATTATAT
5281 TGTTTTGTTA TTTTTTTCTT TCTTTCTTCT GTGCTACTTT CACAAAACAC TGCACTGTAA

```
    5341  AGTGATAGCG ATTTACGGAT TTACACATAT TAATAAAGCA ATCGAAACGC GTATTTTGGC
    5401  ATTTTGGGGA GTCTTCTTTC TTCACCAATT ACAATTTATG GTGGACATCC CAGAGTTACC
    5461  CGAAATCAGT GGAACAAATT AAACTCAAGA GTTCGAGTTC TCAGTCGAAC TGAGAACTTG
    5521  TACAGTCTGT TCCTGAGTTA CGCAATTATA GA

//

LOCUS          22     2689 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      22
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             join(500..847,1588..1816,2158..2189)
                     /translation="MGLPDLPGQPEDDSAFTMDENNTVHLEPLSEDESDNESMDANG
                     YAGYQPLNMDELNHPNRRPSLMEDEAESGVGEGETVGLDVANSINADFLNVDVWNAP
                     RPNELNIELDSSKAAEIVNVMAGIKLPTTAVPEWARGVPEEEWKENLLQRIRQRQQT
                     DVDDGCGTSFTVGSSCPTIPNSSSSSTSSSAGAGGGGGGTSYSVA"
ORIGIN
       1  ACTGGGCTAT GTCGTCGCTC AGCATCTGCA GCTTGTTGTT GAAGTGATAG TTTTCGAAGC
      61  GAAAATAGTT AACACCGAGC TTAAACGTTC GGCGCCGGTT ATCGCCCACG ATGAACCCGA
     121  ACTCGATACC CTGGTCCAGT GCGCTAAAGA TTGCTCCGGA CCGGGAACAA TCGTCCTTAG
     181  ACTTCCATTT TGCTGTGCCA AGCTCGTCCA GTCGGGTGCA AATATCCTGC AGGGAAACTG
     241  CAATGCGATT GGGTGGATCA TTTGCTGGGT TGTCTAACAA ATTAAACGTT GGCACACTTA
     301  CCGGGTCTGC CCACTTCCTG TAGTATGTAC ATTATTTCCG GTAGCATTAT GTTCGGACAC
     361  AGCAATTGTG CTTTGCTTCC TGCTTTCTGA TCCGCCAACA GACAAATGAA CCTTGATGAA
     421  GGCAGCCTCA AAGCCAACTA GATGGTGCAA CTTACAGGAA AACGTTTTCA CTGCTTACTG
     481  TTGTTTTCGT TTGCAGAAAA TGGGACTTCC CGATTTACCG GGCCAGCCGG AAGACGATAG
     541  CGCCTTTACG ATGGACGAAA ACAATACCGT ACACCTGGAA CCCTTGAGCG AAGATGAAAG
     601  TGACAACGAA TCCATGGATG CGAACGGCTA TGCCGGCTAT CAACCATTGA ACATGGACGA
     661  ACTGAATCAC CCAAATCGGA GGCCGTCCCT AATGGAGGAT GAAGCGGAAT CGGGTGTCGG
     721  GGAGGGAGAA ACGGTCGGGT TGGATGTAGC AAACAGCATC AATGCAGATT TTCTCAACGT
     781  AGACGTCTGG AATGCACCAC GACCCAATGA GTTAAACATC GAGCTGGATT CTTCCAAAGC
     841  CGCTGAGGTG AGACACAGAA TTAATATGCG TTTGAATGAA TTCGAGAGGG ATTAATTTCG
     901  TAAGAAAAAA ATGAATTTTG GACTATCAAA ATCAATCAAT AGTTATGCAT ATGGTGGATG
     961  ATTACATGAG CGACATTACA GACACGTCCT CCAAAGGTCT GTTACGTGCT CCCCGAAAGA
    1021  GAACGAAAGT GATTTGAAGC AAACTTAAAT GTATTACACA ACCCCGCGTT TCGGTTGACA
    1081  CCGGTTTAAT TGTAATTAGA ATGGCCCCGG CCATTGAATA GGAACCTACC ACAGAACGTG
    1141  GGTTATTGAT CACATACATG TGTCAAGAAA GGACGTGCCT CGTGAAGGTT GCAGGTCACT
    1201  TACATTGTTG TTTCGTTCCT TTGCTTTCAA ATGTGTATTT CAAGTCCAAC CCTCCCGAGA
    1261  GAAGCGAGCG TGTATTATTC ATTTTCATTT CCATCAACCC ATGGGACACT TGGTCGTCGT
    1321  CCCGTACGGA CCGGACCTGT GTTTTAGTTT CCTTTTCTTT TCTTTTTTCG GGACAAGAAT
    1381  TTATTGCATC CCACCCAAAA CAAAACATTA CCCTTCCGGT TCCGGCACAC AAGGCCAAAG
    1441  TGCTCCCGCA GGAATGGTCA TCAAATAGCT CGAAGTCGAC ACCGCCACCG TCTTTCATTA
    1501  CACTGTTTTC TGTATAGCGA TGAGCGTGTG TTTTTAATAC ATACCTGATT TGTACTTCTC
    1561  CACACACACA CCCCCGTCTC TTTGCAGATA GTGAACGTGA TGGCCGGCAT AAAACTGCCC
    1621  ACCACAGCCG TGCCGGAGTG GGCCCGCGGA GTGCCGGAAG AAGAATGGAA GGAAAATTTA
    1681  CTACAGCGAA TTCGCCAACG ACAGCAAACG GATGTTGACG ATGGATGTGG CACCTCCTTC
    1741  ACGGTGGGAA GCAGTTGCCC GACGATACCC AATAGCAGTA GTAGTAGTAC TAGCAGTAGT
    1801  GCTGGTGCTG GTGGTGGTAG TTCAAGTACT AGGACGTGAT TTATACGCCA TACATACGAT
    1861  TCATGCCAAT AATAGCGTCA GCTTTCATTC CCACATTCCC CCCACTCCCC CAATTCCATG
    1921  CCACGGATCT CACACAGCAA CGGCATTGAT GCTTGATGGT AATATTTAAA AAAATGGTGA
    1981  ATTGAATTGG CTCAACCACT TTCTTCTAGT TCCACGTTGC ATCCTTGGAA TCCTTCTTTG
    2041  AATACAATAA AGCGCACAGC GGACTCACAT TGGCCCGGCA AGCCGGCAAA GGAGTTGTTG
    2101  TTCTATTTAA CGATTTATTT TATGCGAAAC ATGAATGCCA CAACTTCCAA TTCCAAGGGG
    2161  GTGGCGGGAC GTCGTACAGT GTAGCATAAT GGCACGCGGT TCACTGCGCG CTGGAACTCG
    2221  GCATATGCAC CCGTGCGATC CGGTTCAGGC CTTTGCCACA TTTGTACCGT TTAGCCGGCA
    2281  GGCATGCCGT GCATTTGCAT TTCCGTTGCA CTTCCGCCTT GCGCGTGTCC GCTACTGGGA
```

```
    2341  AGGGGGATAG AAACAAATTC CATTCACACG AATCCCAAAG GAAAAAGGTG AATGTCGTCG
    2401  TACAGTGTGG GCACGTTTCA TAAAGTCTGA CCCGACCCGA GTGACCGGTA ACGGTAATGC
    2461  CAAATTTTCC CTAATGCTAA CCCTTTATCC GGCATGCATG CAGGGGAAAA TGAACTTTTA
    2521  AAATAACACA AAAATGTTCA CACGTAATTT AACATATCCG CGCGCGTGCC CGCCGTTGGA
    2581  ATTCGCCCGC CGCCGTATGC CGGCTCTATA TGTGCAAGTC ACTCGATAAG GGGGGGGGGG
    2641  GAGAGAGGAC TGGGATTAAG TGCGACCGCG CTGATTGCTA TGCTCGATA

//

LOCUS          23     1614 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      23
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          500..1114
                  /translation="MEPGALDYEYAPATEADRAEVRQALATYFYPEEPLTLAHRDGP
                  FVTEDDMENALSFLAQGTTILARTAANGELVGVSIGGPSAHEITTRKLADIVAFLEM
                  LDVRAYGFARSSSAYHVYALAVHPAHRAHSIGRRLMEEQIALVRARWPAFAAVTVET
                  TGPGSGRLMGRIGMRETGRLSFADYRDAAGDQVFLGAGEVVAWEMKL"
ORIGIN
       1  GAGTTCACGA TGACGATGAC TTCAAGGAGC ACACCTGAGC TATGAACAAC CTGCTCTGCT
      61  CTACGGCAAA ACACCAAATA AGTTTTGGAA GCACCCCAAT GGTTACTACG TCAGCTTTGT
     121  AACGGAGCCG AGCCTCGAAA TGCTGAGGTA GGTGACTTGA AATAGAAATT TCCCACAACT
     181  GAGACTGTCC AAGTAATGGA TTAGTTACAT TTCATCAACG AAACTGCGTA AGCATAACCG
     241  ATTGCCCTCA GATGCTGGCC AGCTTTAGCA AACATCGACA TCTGTTATCG AAACTCGATT
     301  ACATCGCGAG ATACATATCT CCTCCACAAC GCATCTCGCG AGAATGGAAT TGTGATGCTA
     361  ATAATTGCGC CCAGATCCCA CGGTACAACA TTCCCGTGTG AATTTCTCAA AACCGATCGT
     421  AAACGATCAA GCAAAACGTA AGCAATCGCT TACCACATAA TCGAGCCCAG TTGCCAGTTC
     481  AGTCCGTCTA CACATCAGCA TGGAACCAGG TGCTTTGGAT TACGAGTACG CTCCAGCAAC
     541  GGAAGCGGAC AGGGCAGAGG TTCGCCAAGC TTTAGCAACC TACTTCTACC CGGAGGAACC
     601  ACTGACACTC GCCCATCGGG ATGGTCCGTT CGTTACCGAG GACGACATGG AGAACGCACT
     661  GTCCTTCCTG GCGCAAGGCA CCACCATTCT AGCCCGTACC GCGGCCAACG GTGAACTGGT
     721  TGGTGTTTCC ATCGGTGGTC CTTCAGCGCA CGAAATTACC ACCCGAAAGT TGGCCGATAT
     781  TGTAGCGTTT CTGGAGATGC TCGATGTTCG GGCGTACGGG TTTGCGCGTT CGAGTTCAGC
     841  GTACCACGTG TACGCACTAG CCGTCCATCC TGCTCACCGG GCCCACTCGA TCGGACGGCG
     901  TTTGATGGAG GAACAGATCG CGTTGGTTCG TGCCCGGTGG CCAGCGTTTG CTGCCGTGAC
     961  GGTAGAAACA ACTGGTCCCG GTTCGGGTCG GCTTATGGGC CGTATCGGGA TGCGTGAAAC
    1021  CGGACGGCTT AGCTTCGCGG ACTACCGCGA TGCGGCGGGT GATCAAGTGT TCCTTGGAGC
    1081  AGGGGAAGTG GTTGCCTGGG AGATGAAACT TTGAAAATGA AAAACAAAAC CGTTACAAGC
    1141  ACGTGGTTTA GGATGTGTTT TGGAATAAAG GTTAGTGTGA TGAACATTTT AAAAAATCAA
    1201  TTGTTATATT AATTACTTTA ATTAGACAGT GAAGAAAATC GCGCAAGTTT TTGTTTTAAT
    1261  ACTCTTTATT ATTTCCACTT CTGTTTCACA TTCACATTCT ACACCGTTCT CTTTTGCGCC
    1321  GTTGCGCCCT ATTTGCTGCT TACCCTTCGT CCCTTCGGCG CCTCTGTGTG TATCACTCAA
    1381  ACATGTATTA GTGTTTCGTA AATCTAAACA TATGACCAAC AGACCACGCA CCTCCAGCCT
    1441  AATCCCGTTG TTTTTGCTGT GTATTATCCC AAGAGAAAGG TATGTTAATG TTGAATGTTT
    1501  CTTTCCTCCC CCGTTTGCTT GGGTAGAGTG ACATCGAGGT GTGTTTATGC TGTAGATATC
    1561  TCGCGATTTT ACACATTCTT CTCCACGCAA CACGACGTTG CCTTCAAAGA TTAG

//

LOCUS          24     1779 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      24
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          complement(join(500..784,851..1044,1123..1279))
                  /translation="MSDSDSSDDETNAQLLSAVDTSFLNDNLYNPTAKNEEDNKPNK
```

```
                EVKSSPVPAESVPKSNRFLPEEESIFHSELNVPAAVQKFTAEKLSRFISTVIEFDET
                ERTPQSLVANGTTDEVGVKLKKVPIVRRAVEKEAKLPLADKIAASVCDPGTFPSEVQ
                QWKGPRKRSIEFKYKQKQNGTIVEKSDPFANEFTKARNANMWHESKIKKFKKRA"
ORIGIN
       1  TTCCACCCTT GCTGGATGAG TGCATACTGT ACCTGCTGCA GCTCGGACGA ATAGTACAAT
      61  CGCAAACCGC GCTCGGCCGT TCAACATCGC TGCCCTCGCT GTTCGTTGGA CAGGAATGCA
     121  AGCGGCGATC GCAGAGCGGA CAACTGCAGC ACGCCGAAAG CTTGGTCGAT GAGGTGCGGG
     181  AAACGTTCGT CAAGCTGCTC GATGCGGCCG TACTGAGCCC GAAGATCTAT TCACATTCGG
     241  AGACGTCCGG GAGTTAGAAG ATGTAGGTGA TCGCCATGAT AACTTTTCGC AAGTTGTAGA
     301  TTGCTGTAAG ATAAAAAGTG CTGCAACGAT ACGCTCCGGG CGATTGATGC TAGTGTTCGG
     361  ATTGATTCTA CGCCGGCAGC AGTGGCTACC ACTAGAGACC AGAGCAACTA ACTCACGGAA
     421  CTACTTGTAA CGATAGTGTA CCAGAACAAA GTATAAACTA ACTTGGCCTT GATTTTTATC
     481  ACGATTTAAT GGTTTCGTGT TAAGCACGCT TTTTAAACTT TTTAATTTTA CTTTCGTGCC
     541  ACATGTTGGC ATTGCGGGCT TTAGTAAACT CATTCGCAAA CGGGTCCGAC TTCTCCACGA
     601  TGGTGCCATT CTGCTTCTGC TTGTACTTAA ACTCGATCGA TCGCTTCCGT GGACCCTTCC
     661  ACTGCTGCAC TTCGCTCGGA AATGTGCCCG GATCGCAAAC AGAGGCAGCG ATTTTATCGG
     721  CCAAGGGAAG TTTCGCTTCC TTCTCGACCG CTCTACGTAC TATTGGCACT TTTTTCAGCT
     781  TCACCTCAGC CGGGCAAGTG GGATCCAAAT TGATGTCAAT AATTTCATCA CACCCTGGTA
     841  AAAGACGTAC TCCGACCTCG TCGGTTGTTC CGTTAGCCAC TAATGACTGC GGAGTGCGTT
     901  CGGTTTCATC AAACTCTATG ACGGTGCTAA TAAATCTGGA CAACTTTTCA GCCGTAAATT
     961  TCTGCACGGC GGCCGGTACG TTTAGCTCGC TGTGGAAGAT GGATTCCTCC TCGGGCAGAA
    1021  ACCGGTTTGA CTTGGGCACA GATTCTGTAA TGCAAGCAAT GATTAATGTT CTAATGTTTC
    1081  TTTGGTTTAG AAGGGGCTCT ACGTAAGCTT CTTATTACTC ACCAGCCGGG ACAGGCGAAG
    1141  ATTTAACCTC TTTATTCGGT TTATTATCCT CTTCATTCTT GGCTGTTGGA TTGTATAGAT
    1201  TGTCGTTCAG AAATGAGGTA TCTACCGCCG ACAATAATTG GGCATTTGTT TCATCGTCGC
    1261  TACTATCCGA ATCGGACATG GTTGAGCAGG AGCACTTGCG GACAAGCAAA GCCGGTAAAA
    1321  TGTGGTGAAT GGAAATGCCC AAAACAATAA CAATCTCCCC GCGGGGTAGG GAACGAATGA
    1381  CATTTGCTGC CGAGATGTAT ACAACAATTT TCAGATTTTC AACAGAAATT CGGAAAAAAA
    1441  ATATTATTGA GCGCATTATA AGGTTTAATT GAAATAAATG TGCAACAAAT TAATTATTTC
    1501  AGGTAAAAAA AATGATTTAT CGTTCTTGTA CTTCAGCATT AAAACACAAG ATTTGCAGAT
    1561  ATTCTTGCAG CCCTCTATTC GTCCACGTTT CCTCAAGCTA CCTCGGTTGA TGTTTAGTTG
    1621  GCAAAATTTG TTTGACTGAG GGAATGAAAA AAAAGTGTCT GCTCCGATAG CAGTTGTACG
    1681  GGATCGGGCT CATTATTTTG TAACTTATTT CGTGCTAAAG GCAGCTCGTT GGCCGTAGCT
    1741  GAACGCCAAA ATGCTTTCCT CTGGGTCGGC TGCAATCGG

//

LOCUS          25     7936 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      25
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          join(500..517,4126..4385,5345..5472,7174..7436)
                  /translation="MEETSKPFRATGRLVLGLELALLLLLVCGVTHAHTEPELPAAA
                  ASAQVSPGDATGLAADAGQHLPPAIDAKTDDSAGKSVVVNWKLACKQLCSAGLGGPS
                  CGSTWLRTTTSDSPTLPVDKDTLDGVCPSLCANGLGVSKCNCTSFRPKDQFDQNLIC
                  QAFCTVASVQLAGCSRCDGTGPSLGNFGALEIVETTTPNWDELCSMFCKMGDGGTLC
                  NCDLPPFF"
ORIGIN
       1  CGCTTGCACA AGCATCGAAA TCCATCCGAA ATTCCACCCG CCAACGTGTC AACAGGATCG
      61  ATTGCCCCGT GGTGCCCGTG GTGTCACACA CCTTTCTGAC AATGTTACAC AATGCGCTTT
     121  GCGTGGTTGC AACGTGGTAT AAATAATAAA TAGTTTTACC TTGTTTTGTT CTTCTTTATT
     181  TAATGTTGTT CCCCAAAGGG TACGAAATTT ATGCAAAATG GTACCTTGCC CCCCGTTGGT
     241  TGGTGCCATA AACTATGGGA AAGTGTCTTT TTTTTGTTAC AAGTGAGCCT AGTGGGACAG
     301  CTGGTTTGTA CGCAAATGTA CCATTTCTGA GGTTAGCTGT TTTGGAGAAC CACAACCCAA
     361  ATGGCCGGAA ATCCTTTGCG GGTGTCGTGT CCAATTAGGT GTAGCTGATC CTAGCCAAGT
     421  GTCGTCCGAT GACGTCAGAC AAATGGGAAA AGTGTTTCTC GTTCCATTGG ATAACTTTTT
```

```
 481  GCACATGTCG GTAAATGAAA TGGAAGAAAC GTCGAAAGTA CTTCTGATTC AGACAAGCAA
 541  CTGTTACTCT ATTTCCTTTG GTTATGAAAT CGTTTAAAAA CAGTTTAAGC TTTTTTACTT
 601  TCACCATCAA CCTTTAGGTA AAGTTTTCCA GATAATAAAT ATACAACCGT TCCCACACGG
 661  ATCTTTAAAA AGTATCCAAT AAAAAACAAA GCTCCAATGG GCAAAAACAA TACACTCGAA
 721  AATGCCATTT GCTGGTGCGG ACGATAACGG AGACCAAACG CAATGACCGC ATGTAAACGG
 781  ATTACCCATT TTACGATCCA CCGTCCCAGC GTCCTACGTC AAACTGACAG CGTGTGTGAG
 841  TGTTCATTAG TGCCACGAGC ACCATATCCT TAATCCCGGC GACAAATTAA CCCATTTTCC
 901  AGGCCCAGGC TTGTCACGCT TTAGTAGTAG TAGTAAGCGA CATGGAAAAC GTGTCCCATT
 961  AAATCTATCT CACAGGCAGC AGCTTACCAG AGTGTGCCAG ATGTTGGCGT GTTGGCGTCA
1021  GTTGCCTTCC TTCGTCTACT CCTACACAAT TCCACTTGCC AATCTGACTA TACAACATGG
1081  TGTTGAGGTA GACAAGTGTG GGCGATACCT GAGTGGATCC GCTTCACGTG GCAAAGGACA
1141  AACGCAAAGG TTGTAAAATC GATAAAAAAT CATCCATCCA CTTACTGTAA AACTATAAAC
1201  TATGAAGTGC ATCGATATGG ATATCAGTGA CGCAAGTGGA ATAATGTGCA ACTATTTCAT
1261  AACGCTAGTA GAACACTCTC GCAGAAGAGA ATGTGAAAGG CAAATAGTGA TGAAAAAATC
1321  GATAAGTAAA GGTAACTGAT AGGATTGTAC AAATGGTTCC CGGAGCGAAG CAATTTAGGC
1381  TATAAAATGA GCAAAATGTA GTAGGCAGGT GTGCCATTAG CATGAATCAA TATCTTAATG
1441  ATATGATCGG GAACATAAAA TGGAAAATTG AAAAAAGTTA AACGAGTTCT AATTTCATGT
1501  AGAGCTATCA TGTAACTAGG TTATGATGAA CAAGATGTAA TAAATATAAG ATGTTTTCAT
1561  TTAGATACAA CAATGATTTG TATAATAAAT GGCCTGTAAA TGGCGTAATA ACATATGCGC
1621  GATGCGATAA GTTATATCTA AATAGATGAA TATGAAGTAA AAATAGTACT AGAGTATCAA
1681  ATACTGGAAA TCTTTAGAAT CCAGTCAGAA TATTATACAA TTGTCTTTTC GAGATTGATA
1741  CAATGTTCTG TTTAAATTAA TTGTAAGAAC ACCAAACAGC CTGCACACTC TAATGAAATA
1801  AACTTTACCA AATTCAACCT TTTCTGGTAA CACCCTAGAA ATATCAAACT ATAAACCTTA
1861  CTCCTCCTTT ACATCTCCAG ACTAAATCGT CGTTGTAAAT CGTCCTTTCC AGACTATTAA
1921  ATCGGAATTA GGAGACCCTT TAGCGTGCGT TGAACTGTCC AACCATAAGA CTAAACTGCT
1981  ACTTACTATT CTTCTCCAAT AATCTTTTGG CTCTATAACT GTAAATGTTC AATTCAACCT
2041  AAAAATTGTG GTATAATTTG TAAGAAGTAT CTTAGATCCT CTGGCGTCTA GACTTATGAA
2101  TGTTGCAATA ATCTGCTGAA CAATCGATTA TTTGTTAATT GATTTAATTG TTAACAACAA
2161  ATCCGAAACA TATCTAAATT TCATATACCC CATATCCGTT GAGAACCAAA ATGAATGCTT
2221  TTCCCTTTAA CAATCCGTTC CCATTAGCTT TTCTATACAT AATTGTACAG AACTACAGTC
2281  TACAGACATT TTCAAGGTTC AAGTTTCTAG CCCGGAACCA AGCCAGTTAC ATCATATTCT
2341  TTCTACAAAC ACAGAGTCTT CAATTAAAGT ACTTCAGAAT GAACTGCAAT TATACCCGCA
2401  GGCAAACAGC GACGCGGTTC AGATTTATTG TGCCGCGACA AATATTGTGC TAGGTAAGTA
2461  AACTAAAAGC TTACCGCCAG CCCCACAGCT CGATAAGAAA GCTCCCACAC CGATGTTGTT
2521  CTGTACATTG TTTTTCCCTA TTCCCTATGC CATATGCTCC GGCACTTGAC AGCAGCAGGG
2581  CCCCGTCAGC GGACCCAGCG CCCGCGACCT TAGCAAGGAA GATAATCAAC AAACACCTAC
2641  GACACGGTTT TCTGTATCGC AAACGCTTTA CACCGCCTTC GTCTCCCGCT CGCAGGGGCT
2701  GATAATTGAT CCCGGAACGC GATAGATCGA TACCGGAACG CGGGATTTGG CGGGACGCAC
2761  GGTGGCCAAT TATCGACGGT ATCGCTGTAA CACGCTGATA AAGCAATTTG AAGGAGCCAT
2821  CGTAATCCGC AGAGGAGGTG TTCGATAGCG AATCGGGCTA CAAAAATCGG CTCGCTCGCA
2881  AACGGTTCGT GCAGTGATAA TTGACGTGAC TGTTCGGTTG CACGCAGCGA ACGGATTGAT
2941  GCAAACCAAG CCCAATTTTA CACCATTGGA GTATACACAC AAACATATAT ATATAAAAGA
3001  AGACGAAACT GTAGCAAATT GCAAACTGCA GCCAATTCTG AACCGCATTT AGTGCTTTGT
3061  GATACATTTT GCATTTAAAT AACATGAAAG TGAGTTGAAT ACCGTTTCTC ATTGCTTTGT
3121  CTTCTACGAC GCAGTAGTAG TGTCTTAAGT GTATTCATTG TTTTTTTATC TCTTAAACGT
3181  CTGGTGCGGA AATGAAAAAA AAAAACGCTA AATTGCTCCT TTCTGCAATA CAAACACCGA
3241  CCGAGATGAA AGCGCGACCA TAAAAGTGGT CAGCTCCATC CATGCCCAGC CCGTGCGACA
3301  TTGACTGGAC ACAAAGTGGT CGCAGTACTT CAAAACTGAG CTTGTTGTTG AGTGAAACGG
3361  CTTCAAAAAG ACTTTCTCAA GGAGGGATTC TTATAAATCA CTGCTTGACA AATCGTTGTT
3421  AAATCGGATA AAGCTAACCT TAGCTCAAAC AGTAAAACAA TGTAGTATGT TTGTAAAAGA
3481  GACTCGCTCT ATGAGATAAT ACAGAGCTAT CGGAATATTC TAGAACCCTT CGACGAAAAA
3541  CGGGCGCAGA TTGGAGCTTT TCCACTGGTG ACCGAACGGA ACAGTATATT TTCCATTTAT
3601  TGGAGCATGT GATTTGCGGT AATGGAAAAA CAAAGGCCCT TTCCCCCCGG TCCTCAACAA
3661  TAGTCATCTC ATTTAGTTCC GAATTCGATG GACCAAGAAG CACGAGAAAA GTTCTTCGCT
3721  CCAAACAAAC ACAAAAAAGT GATGATTTCC AACAGTTTGA TGATTTTGTT TGTTGGTGGC
3781  AAATGAGAAT TGTTCGCTTT CCCGTTCCGT TGTGTGTGTC GTTTTTGGGA AGTGAGAAGA
3841  AAAAATATCG ATCCTTGTTT TTATGGTGAA TTCATACAAA CTGTGCAAAA ACGTTGGACC
```

3901 AAAAATTTAG TTGGAAGAAA GTTGATCTTT CTCCTTTGCC GTTCACGGCG GACAACAACT
3961 CCATTGCACG TTGAACCTTT ACCAAGTATA TAAGAAACGC ACCGACCATC AGCGACCACT
4021 GCCGTGGTTG TCCTAGATAA TTGAGTTAAA TTATCCACGC CCAATCGTTG TTATGCGTTG
4081 TGTGATTTTT CCCCGCTTTT TACGCCTCCC TTCCCTCGTT CTCAGCCTTT CCGTGCAACC
4141 GGTCGCCTTG TGCTGGGATT GGAGCTGGCA CTACTCCTGC TGCTGGTGTG CGGTGTAACG
4201 CATGCCCACA CCGAACCGGA GCTCCCGGCG GCCGCGGCGA GTGCGCAGGT TTCACCCGGC
4261 GATGCCACGG GACTCGCAGC GGATGCAGGG CAGCATCTTC CTCCGGCGAT AGACGCTAAA
4321 ACGGACGACA GCGCTGGAAA AAGCGTAGTT GTGAACTGGA AGCTGGCCTG CAAACAGCTC
4381 TGCAGGTGAG TGCAGCTCGT GTGGTTGGTA GTGTGCTGCC GTCCATCGTT TTCTCGAGCA
4441 CCATGGGAGT GAAATAGTGA GCGGACATGG TGTCTCGAGT GCAGGCAAAC AGTCTGTGGC
4501 ACGATTGCCG AAACTAGATA ACGCAAAGTT CTGGGTGTTG ATGGGAAAAG TTGTATGCCG
4561 GAGGGATGTG TGTGTGTGTG ATGTGTTCAT GTGTGTGAGT GTGTGGAAAA ATGGAAAATT
4621 TTACATCTCA ACGCTCTCTC TCATCCGTTT TATGTGGCTC CTCAGGCAGC AACTGTGTCA
4681 AGTGTGAAGG AAAAGCTGCA GATGATTGAT GGCTTCCAAT TTTACGGGCG AGGCTTTTGT
4741 GAAGTTTTAC AGAATGCGTG TCAAATAGCC AAACTGATGT GATGGAAGAT TGACTCAGAT
4801 CTATCATTTT GTCCCCTTTT TTTGCTTGGG CACAGATTTG CCGACGGGAT TAGACGTACT
4861 GATGTGTAGA GTTATGCTTG CACGAGTGAG CCTTGTGAGA CGCTCCTCCA ATAATGTTTT
4921 CAATTTTCTA CAGACATTTT TCCAATACTT GATATCCAAG AGCAGTTGAC ATATTTCACA
4981 CTTCTAGACG CGTTTTTCTC TTCACCGTGT GTCACCTATT TTTCGCGTGC CCATGTTATC
5041 AGCGGTCTAT TGCTTTTAAC AGGCGCAATT AGGCGCAATT GCCATCAACC GGTAGCCACC
5101 AAAACCATAA AGCACACACT ACTGCCACAC GATCGCAAAC ACATTCCGCC GATGGCGATA
5161 AACCCTAGGC ATCTGGACGG GAGTAGCAGA CGAACAGATA ATCTTGTGGT GTCCCAATGA
5221 CGGGACACTT TTGGATACAG GGACCTCCCT AATAGTTCGC CGTAGAGATA CTTATTTTGA
5281 GCGGATTTTA TATTTAATAT TTTCAAATAG TTATATATCT CTCTCTCTTC CCCTCTTTTC
5341 ACAGTGCAGG GCTTGGCGGT CCCAGCTGCG GGTCAACATG GCTCCGCACT ACCACTTCCG
5401 ACTCCCCCAC ACTCCCGGTG GATAAGGACA CGCTGGATGG GGTGTGTCCA TCGCTGTGTG
5461 CCAACGGGCT AGGTAAGCTT GGTTGCTTTT CTTTAACATC CAAATGACTA TGCGTCAATA
5521 GGGTATGAGA CCACGGGTCG AAAGGTCACT TCAACGACGT CAGTTAAGCG CCGTAAATCA
5581 ATCATCGTCC GTATGCATCG ATCCGGCAGG TTCCGCTGAA CTTTAGAATG TAGATCGCCG
5641 ATAGACAAGC GCAGCTGTGT GCCGATTGTA ATATCTAAAT CCCTTAATTT TTTGATAATT
5701 AATCAACGCC CCTCCGACTC CGGTGTCAGT TAATTAGGTT GTTTCCCAAT ACGGGGATAA
5761 AGCATGATAC TGCGGGGGTT TTACACTAAG CCGATATTGT TGGGGCTGAC ATCGCTTCGG
5821 TCAAGCAGTA CGTCATAGAT AGGACGCATC TGGGGGGCTC GTCTCTGTGC GGTCATTTGG
5881 AATTTAGTTA TCGCCAGTGA TACCAGTGTG CTGGTTAACA GTATTTAGAA GATTAGAAAC
5941 CGTTCATGGA TGAACACGTT TGAACGAGTT CCAAATAATG AATCACCCTG TTGTGTTGGG
6001 ATGTTCAGAG TATAATAGCT TGGGGAGTTT GGTGTCTTCT AATGCGACTG ACAGCTTCAT
6061 CTATGTACTA GCGTGAATCA GCATGCCACG TTTCATTTGT GGTTTCCTTG CTGAACATTT
6121 ACTGACGAGT CAAGAGGCTT TTAATGGGAA CGGTCTTAAA TGGAGCTTTG GTTGGACTTA
6181 CATTAATGAG AATAGCACAT CTTTGAAATC AGTATTAAAA TTATGTTTTA AACAACATTT
6241 GGTATAAGGT TATTGGTTGA AATATAGCGA AATAGTCATT TGCCGACGGC ACTTGATAGT
6301 GCTGTGTTAC AAAGACTTTT GCTAAGGCTT ATCAATTGGT GAATGCTTAA CAACATGCCC
6361 GTCATGGGTT CAAGCCCCGT ATGGACCGTG TCCCTCAAAC TTTGGACTGC CTATCATATG
6421 GTACTATGGT ACAAGCAGGC CTTGACCGAC GACGGTAAAA TGACCTCTTT TTCAGATAAC
6481 ACCTTCTTAC CTCCGTTTGG TTCGTCAACA TCTTGGCCGG CATTAGACAT GTTCGAAAAT
6541 GGCACATGGT TTTGGGCTTG CGTCTACTAA TTCGCTTATT TCGGTGTTTG TGATAGACTC
6601 TGCGTACAAA AATTATGAAA TTTCAGGGTT GACCCATTAC TCTTCATCTG TCCAATTGAC
6661 AGGCTGTCCC TACAAAGACA AAGGCTGTTT ATATCCATAT TCTAAGTCCT GCCTATGGGA
6721 AGGTTATGTT TAACGACTGT TCCACTGGAC GGACCCATAT ACTTCGGCCA AAAACTCTTC
6781 TGAGAACCTC TTCTCTCGAG CGTGATCATC AGAGTTTTGC CTTACCGGAT CAACTATGAA
6841 TCACCGCTGT CGTCCTTGCT TCGGAGCTTT GCAAACTACA AGGCTACCAA ATTATGCTCC
6901 GTGGTTTTCG AAGTATTCAT TAATGGTAAT TCTGATGGCC GATACTAGAG AATTACCCTA
6961 AAGCTTGTGA ACAAAAGACA TCACTGAACT GACTGACGTA TCTTGTGTAG AACCGTTATG
7021 TAACACTGAC ATTTCATCAC TGGCTACTAA GATGAGGTTT TCACGTTGCA AGAATACCTT
7081 CTCTTGGTTT CCATAGCAAT CGCGAAGCGT GAAATAAAAA ATCTGACCAA ATTTCTTTTC
7141 CGTTCTCTCT CATTCCTCTC TCTCTCTCCT TAGGTGTTTC CAAATGTAAC TGCACCAGTT
7201 TCCGGCCGAA GGACCAGTTC GACCAGAACC TTATCTGCCA GGCGTTCTGT ACCGTCGCCA
7261 GTGTGCAGCT GGCCGGATGC AGCCGTTGCG ACGGAACTGG CCCCAGCCTG GGCAACTTCG

```
    7321  GTGCATTGGA AATCGTCGAA ACGACGACAC CGAACTGGGA CGAGCTGTGT TCGATGTTCT
    7381  GCAAGATGGG CGACGGCGGA ACACTGTGCA ACTGTGATTT GCCGCCGTTC TTCTAATGTG
    7441  CACACAACCG TGGACACACC ACACACACAC ACAAACCCAA CCCAATACAG AAGACATTAC
    7501  GTTCACGCTT ACTTCAATTA ATTTACCGCG GGGTGCACCT ACCCCAATGC GGTACAAAAA
    7561  ACAAAGACGG CGTTGGAATT CATTTCGTGA TTTTTTTCAT TCACAAAAGA CAACAGCACA
    7621  ACATGTGAAA TACGTTTTTG GGAGGGATTT CGTGAAGAAG ACCGGGTGCA AACGTTATCA
    7681  TTACATATCC GCAGGGTTTG CATGACATGT GCGGCATTCC GCACCAATCT CAAGCTCGCA
    7741  CCGTTTATGC CAGTTTGTTT GCGTTTCAGA CGTTTCAGTT CGATCCGTTC CTAAGGTCTT
    7801  ACCTGAGTTT TGAATGACTA GGATAGTTGG AGGAGAAGCA GACGGTAGGA TAAGGCAGTC
    7861  GTCGTTCGAC TTTGAACGAT CAATGACAAA TATCACTCAA AAGGGCTTCC AGGTTGAGGG
    7921  TGAAACATAG CGGGAC

//

LOCUS          26     1668 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      26
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             500..1168
                     /translation="MNGGFSKILYRIAAPNECDRIRQALLAFYFPEELVTRSYLDES
                     QPQIGPTEEHIQYALSFVHQGMAAVAIEQDHGTIVGVTIARCVKPGTADELLAMVPA
                     AGSRRWAETLRLFAHLEHTGDVCGRFRSRRSYHVFVLAVEPHFRRRAIGQKLMDFQL
                     ARGKSLRFRVVSADVTCEVAARICERMDMRCVCAMSLNQYRNQSGEHPFVVTGPNHI
                     VSTYARYV"
ORIGIN
       1  TTGCTTGAAC ATTTTTTGTG AATTGAAAAT GGAATAGATA ACATTATATA TATAGATATA
      61  ATATTTCAAT ATTTTTAGAT TATTTAAAGA CATCTACACG TGCAGTAATA TAACGCGAAA
     121  TTTCAATACC TTTCCCCACA CAGTGTTTAT GACCGAACCA CCCACTACGA CCATCTGGCG
     181  ATAAGAGAAC CTACGCAAAA CACCGTACCG CGCTTGTACA ATTCGATGCA ACAACCTTGC
     241  GAATGAAAAC ATTGCCCTAG CAGCAAACGG TATGCAAGGG TGAAGAGATT AAGCGCCATG
     301  CTATCGAACG AATTCGATTC CGTGCGAAGC TAATGCGACA GTCTCTCAAC CAACCTGCCC
     361  CATACGCCAA GAAAAAACGA CCAGATTGCA TCGCCGATTC GCTCCTGGGA AATGATATTT
     421  CGTTGACGTG GAAGGCGGTA AGACGGTAAG AGAAAACTGT GCCATCGCCT ATACTGTGGG
     481  TGCATTTCGG TGCCGTGAAA TGAACGGAGG GTTCTCCAAA ATTCTGTACC GCATTGCGGC
     541  ACCGAACGAG TGCGATCGCA TCCGGCAGGC ACTGCTGGCG TTCTACTTCC CGGAGGAGCT
     601  GGTCACGCGA TCGTACCTCG ACGAGAGTCA ACCCCAGATC GGCCCCACGG AAGAGCACAT
     661  CCAGTATGCA CTGTCCTTCG TGCACCAGGG CATGGCAGCG GTAGCGATCG AGCAAGATCA
     721  CGGCACGATC GTTGGCGTTA CGATCGCACG GTGTGTGAAA CCGGGCACGG CTGACGAGCT
     781  GCTGGCAATG GTACCGGCGG CCGGATCGCG ACGCTGGGCC GAAACGTTGC GACTGTTTGC
     841  CCATCTCGAG CACACTGGGG ACGTTTGTGG ACGGTTCCGT TCGCGCCGCT CGTACCACGT
     901  GTTCGTGCTG GCGGTGGAGC CTCACTTTAG GCGCCGCGCG ATTGGACAGA AGCTGATGGA
     961  CTTTCAGTTG GCAAGGGGGA AGTCGCTTCG CTTTCGGGTG GTAAGTGCGG ACGTTACGTG
    1021  TGAGGTGGCC GCACGGATAT GCGAGCGGAT GGATATGCGG TGTGTGTGTG CGATGTCGCT
    1081  GAACCAGTAC CGGAACCAGT CCGGCGAGCA TCCGTTTGTG GTAACGGGGC CGAATCACAT
    1141  CGTAAGCACG TACGCGCGGT ACGTTTAAGG TGGAATTTCC AAGAACTAGT GCAAATTAGA
    1201  AACTGATAAA AGTAGTGTAA ACGTATAGCA AAATTGGAAA AATAAATAAA TGTTAATATT
    1261  TTTATTAACA AATGTTGCTT AATATTCAAT GCTCAGTTGA AATCAATAAT TGTGTAATTT
    1321  GCGTAGAAAA ATGCATAAAA GACAATAAAT TATATATCAC ACTGTAGGAT TCTAATTACA
    1381  ATAAATTCAT CCAAAACTTC AATATTTAAT TAATTGTGAA GTAACACAAA AAAATACATA
    1441  AATCATTGAC CGTAAAAATT ATACCTGATA ATTAAAATGT TTTTCCCTGC ATTCGGATTC
    1501  TCCTTAACGT TGTTTACTAC CATCTATCTG CCATAAATCA TTACGTCTTT TCGACCGGCT
    1561  GTGCCGAAGT GGCGTGAAGG ATACTTACTT CTAGTGCCAA AAATTAAATG TCCAAGGCTG
    1621  ACCGTGCCCG ATGGACAAAC TTTGAGTGAG TGTTGGGCGA TAGTAATT

//
```

```
LOCUS          27     1683 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      27
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          complement(500..1183)
                  /translation="MSSLRVKCVFCGLIIDCQREDTAKLLAHLRNEHSEMGNHKPTA
                  FTTSNRSKAHPAQPAEKAELEEGSDASFAEGISESEELLHRERVGTGTSSDTASKSS
                  SSEDAETIPPHRSTRPLHVSLESQRSVLYQTSLSCWRPGGRPVCCPDCGAHRVPIVR
                  SHATGVTWSSTVASCFLFCWPLCCLSWFFSEPVSEYLHCARCDQLLAKHDIKAERVS
                  PNYAMLDRVEDQL"
ORIGIN
       1  GTCCGCCATT CTGTCTGTTT CATCTTTGTT GCTGTAATTC GTGTTTTGTG TAAAGTGATT
      61  TCTGTGCCAA GTTTTGTAGA ACGAGACACA AACGACACCA ATCTAGCTCG GTACGCCCGC
     121  GAGCAGTAGT GTTCGGCATT TTAAAAAAGG GTAGTGTTAT TTTCGTACTG CCGCAGAACC
     181  ACCGTGCGCT TGTGAGGTGA GGCGGTTGTT TTACGCAAAG CATCCAATTT TCGTCTGCCG
     241  CCCGGTGGGT AGTGCTGCTG GCCGTTTCGT GAAAAGTGTG TGTGTTTCGA CCCATCATCC
     301  ACCGTCAGGG CACGCAAAGA TGAGCACCGG TGCACCGAAG CTCAGCACCT CGCTGGACGA
     361  AAGTCAGCTC AGCCAAGTGA AGGAATGCTT GGCCAGCTTC GCCGAACAAA CGTGAGTAAA
     421  CCATTTACAA AGGCACTGTT TATTACAAGA AGACGAATGT TTCCCTGCCC CTCGCGCAGT
     481  AGCCCTTCAT TTCGCTCGAT TACAATTGAT CCTCAACGCG GTCAAGCATT GCATAGTTGG
     541  GTGAGACACG TTCCGCTTTG ATGTCGTGCT TCGCCAACAG CTGATCGCAC CGGGCACAGT
     601  GCAGATATTC GCTGACCGGT TCGGAGAAAA ACCACGACAA ACAGCACAGC GGCCAGCAGA
     661  AGAGAAAGCA GGAAGCGACC GTCGACGACC AGGTTACCCC GGTTGCGTGC GACCGAACGA
     721  TCGGTACCCG ATGTGCGCCA CAATCCGGGC AGCAGACGGG TCTGCCTCCC GGGCGCCAGC
     781  AGCTGAGCGA TGTTTGGTAC AGCACCGAGC GCTGTGACTC GAGCGAAACA TGCAAGGGTC
     841  GTGTACTACG GTGCGGTGGA ATAGTTTCTG CATCTTCCGA GCTGCTACTC TTGCTGGCAG
     901  TATCGCTTGA CGTTCCTGTT CCAACTCGCT CCCGGTGCAG TAGCTCTTCC GATTCGGAAA
     961  TGCCTTCAGC AAACGACGCA TCGCTGCCTT CTTCCAGCTC CGCCTTTTCT GCCGGCTGCG
    1021  CTGGATGTGC TTTCGAGCGG TTGGAAGTGG TAAATGCCGT CGGTTTATGA TTACCCATCT
    1081  CCGAATGTTC GTTCCGCAGG TGTGCCAGCA GCTTGGCCGT GTCTTCACGC TGGCAATCGA
    1141  TTATCAATCC GCAGAAGACA CACTTTACTC TCAAACTGCT CATCTCTTCT CCTAGCCGAC
    1201  GATAGCATTG ACAGGTGCAG TGCGCGGCGC CTGTGTAGCT GTGTTTTGCT TTTTAAACTG
    1261  AGCAACTCCT TTTTCGCTCA GTAGGCCGTG CTTGGCGATG GGAATTGTTG TTGAAAATTG
    1321  GAAAAATCAG TTTCTTTTTT CATTTATTCT GTACATACAA TTCAAGCTGG TTACAAAATG
    1381  TGAGTGATTC GACCTTGTGA ATTCAATTAA ACAATACAAC GGCTGTTCAT CACATGTCCG
    1441  CCACATGCGT AAAATGGTGT ACAAAATAGT CTAATCAGCG GCAAAGAAAC TAGAAGAAGG
    1501  AAAAACATCT GTTGTGCAAA GGAAACGCAT TTCTAAAGGC GGCGAACTTC CTGTTTCCTT
    1561  CTGCTACCTG GAACGGGGAC CCAAACACAA AGCCTCAATG ATCCTTTTCC TTCGCCGCAC
    1621  TCGAGGCTTT TCCGTTCCAT CTCCGAGAGC GCAGCAGCCC TATCGCCCTG AATCTCGCCC
    1681  GGC

//

LOCUS          28     1807 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      28
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          complement(join(500..1101,1220..1307))
                  /translation="MNVVFVLGFVLVASAFATDLSANDCEINVEECSPTYSSFLSRS
                  GKTVENDLKQYTSQLVDKSFHFLMMSSAFNKHSLDRPGFEKLYRKISDKAWADAIEL
                  IKYQSRRGSFGHLVQPSKGENYGKVLDVQELSSLQFALDYEKQMAKEAHAIHRKISH
                  AHSKAGSNGSDDVYHYDPDAAHYLDENIIEYQSGVVRDLAGYVHNLKHFTSAKHAAN
                  DLGNHVFDEFLAKVE"
ORIGIN
       1  AATCGACACT TTGCTGCACC GTGTGCTATG GAGAGCAGAA AAGCCGTCAC CGTTGCAGCG
```

```
      61  GCTCACGCAA AAGGGCAAAT GACTGATAAC AACGTGCACA TGCACAGCAC AAGTGCAGCG
     121  GGGAATCCAG CGTACAGGGT TCGGTAGTCC TCCAAAACCG ATCGTCGATT AGCAAACGGC
     181  TTTACGTGGT GATAATTTCA ACGTAAACAA GTCACGCTCA GACGTTTGTA AGCGAGCGTG
     241  TGAAAGTGCA TCCCAAGCTT TCAACCTAAA AAGGAAAGAC GCAATGACTA CAGATTCTGG
     301  ATTATTATTT TCTTCCATTT TGTGGATCAC ATGTTGGAAT TTTAATGTTT ATTTCAAACT
     361  CTCTGCAAAT ACATTGTTTG TTTTTTTTTT ATTTAGCCAA GGGAGTGGCG CTCTAAGCAA
     421  CACGATCATG CGATCGCTTA CAACTGTTGC CAACCTTTCT CCCTGCAGCA AACGCGCTGA
     481  CCCGGCGGAC CTACCAGTCT TACTCCACCT TAGCCAGGAA CTCGTCGAAC ACATGGTTGC
     541  CCAAATCGTT GGCCGCGTGC TTCGCCGAGG TGAAGTGCTT CAGGTTGTGC ACGTAGCCGG
     601  CCAGATCGCG CACCACACCC GACTGGTACT CGATGATGTT CTCGTCCAGG TAGTGGGCCG
     661  CGTCCGGATC GTAATGGTAG ACGTCGTCCG AGCCGTTCGA GCCGGCCTTC GAGTGGGCGT
     721  GCGAAATCTT GCGATGGATG GCGTGCGCCT CCTTCGCCAT CTGCTTCTCG TAGTCCAGCG
     781  CGAACTGCAG CGAGCTCAGC TCCTGCACGT CCAGCACCTT GCCGTAGTTT TCGCCCTTGC
     841  TCGGCTGCAC CAGATGGCCG AACGAGCCGC GGCGGCTCTG GTACTTGATC AGCTCGATCG
     901  CATCCGCCCA GGCCTTGTCG CTGATCTTGC GGTACAGCTT CTCGAACCCG GGGCGGTCCA
     961  GGCTGTGCTT GTTGAACGCG GACGACATCA TCAGGAAGTG GAACGATTTG TCCACCAGCT
    1021  GGCTAGTGTA CTGCTTCAGA TCGTTCTCCA CCGTCTTGCC GGAACGCGAC AGAAAGCTGG
    1081  AGTAGGTCGG CGAACACTCC TCTGGTAAAG GGATAAAGGG TAATGCAGGG GAAAGGCGTT
    1141  AGGGGGTAAC TGCGCATTGT ATATATGTGT GTGTGTGTTT TTGTTTATCA CTCGCCATGT
    1201  TGAGCGAACG GCCACTTACC CACGTTGATT TCGCAGTCAT TTGCACTAAG GTCCGTCGCA
    1261  AAGGCGGAAG CCACCAGCAC AAAGCCCAAC ACAAACACAA CATTCATCTT TGCCATCATT
    1321  TTTTTCTTCT TTTTTCACTC GTTCACTCAC ACTGCCTGTC ACTCGGATCG ATCCTCAAGG
    1381  GCGATGATGG CCGCGAACAA ACTTGCTTCC AACTACTGGG GCACGGGTAT CAAGAAGGGA
    1441  CAGTATTTGT GAACCTTCTC GATTCGACAA AACTGCTCAC GGTAAATGCG ATTGTTCGGT
    1501  TATGCTTTAG CCAGCAAGCG AAAAGGATTT CGCTGCAAAA TTGACACAAT TTCGTACGGG
    1561  AAAGCCACAC ACAGCACAGA AAGCGAACGC GATTTGCGTC TTACTTCCTC CAATAGTTGC
    1621  CACTGTTATG AATGCTCTGT CCGACAGTCG AGTGACTTTT GTTCCCGTCC GCACGTCAGT
    1681  GTAGTGGTGG TGTGTGCACG CCGTAACTCA CCTACACAGT TATATACGGT CGTTCGCGGC
    1741  GCACCATGGA GACAACGAAA CGGAATGGTG CCTGCGCTGA TTACGAGCAG ACAGTGAAAT
    1801  AAAGATC

//

LOCUS          29     1704 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      29
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          500..1204
                  /translation="MTIGVILAVACALTFTWPVEGMYLQERANCRTEGCATQYGCVP
                  RYSQYKTALYCPQRCMIERCRSVENEPQPRVCHKITSVATGQTLAAMDSYDRRGDQR
                  YYGATSLQNPNTGRGEQRYASLVNPNAGERDRWNLIVGPGNYYYIQNSDTREYLRAS
                  ENNYLYLVSGRPTDESFLFKPDLVQGSWSSVLLQSVCYRGYIFGEQYSGRSASYAMM
                  THDLDRCYRESVWSISSVNC"
ORIGIN
       1  CGATCTATGC TAACGAAGCG ACCCTGGCTG ATGGTTCCAG CACACGACGA TCAAACCCAT
      61  CGATACCGTC GATTTTTGTG TGCAAGGCCA TGTGCATATC ACCGGAACCA TTTCTCCCGC
     121  AACGTGCACT AACTCATCCA GTACGTTCTT ATCTATCGCT TTACATCTCG CTCGGTCTCT
     181  GTTGCTCCCA CGCCGGCGGG TTCGTTGTCT CTTGTTGCGC GTGTGCTAAA AGATAGCCAG
     241  TGGCGGTCAT CTATTTGTAA AACGATCATC AACTCGATCA TTCATGGGTA TTTGCTGTGC
     301  TACAAGCGCA TTCTTGCTGA TAGAGCGACC CACCTTCCGA TGACAGCATC TGTGTGTGCC
     361  GCCCGTCGTG TCGCAATCGA ATGGAATCTC TATAAAAGTG TGCACAGCTG GAGCAGTGGC
     421  CAGCAGTACT CAACCGTGTG TCTTCTCAGT AACCGTGCGT GACGTTTCAC TCATACTCAC
     481  CATTCAACAC TGCCACATCA TGACGATCGG TGTGATCCTT GCCGTCGCGT GTGCTCTGAC
     541  GTTTACCTGG CCAGTGGAGG GGATGTACCT GCAGGAGCGA GCAAACTGTC GCACGGAGGG
     601  CTGTGCCACC CAGTACGGTT GCGTGCCCCG GTACTCGCAG TACAAAACCG CACTCTACTG
     661  TCCACAGCGT TGCATGATCG AACGGTGCCG TTCGGTGGAG AACGAACCTC AGCCCCGTGT
```

```
     721  CTGTCACAAG ATTACCTCGG TAGCTACCGG CCAAACGTTG GCGGCGATGG ACTCGTACGA
     781  TCGTCGTGGC GACCAACGGT ACTACGGTGC TACATCGCTT CAGAATCCTA ACACGGGCCG
     841  TGGTGAGCAG CGGTACGCAT CGCTGGTGAA CCCCAACGCG GGCGAACGAG ACCGCTGGAA
     901  CCTGATCGTC GGTCCGGGCA ATTACTACTA CATCCAGAAC AGTGACACCA GGGAGTATCT
     961  GCGTGCGAGT GAAAACAACT ATCTCTACCT TGTGTCCGGC CGTCCCACTG ACGAGAGCTT
    1021  CCTGTTCAAG CCCGACCTTG TTCAGGGCAG CTGGTCGAGC GTGTTGCTGC AGTCGGTATG
    1081  CTACCGTGGT TACATTTTCG GCGAGCAGTA CAGTGGCAGA TCCGCCAGCT ATGCAATGAT
    1141  GACCCACGAT CTGGATCGTT GCTATCGTGA AAGTGTTTGG AGCATTAGCA GCGTAAACTG
    1201  TTGAGCGTAG ACCTGCGAGG AGCACATAAA TTAACCTTTC CTGCCGTTCG AGCCGCTCGA
    1261  ACGCTACTGC CATATTTGTT TGATCATGTA TCGAAACGAA ATAAATGAAA GATAAATACG
    1321  ATATGAATTG CATTACCTTT TTGTTTTTGT TTCTGACCAG GAGTAATTAC AAAAATTAAG
    1381  AGCATTGCGC TTCAATTATG GTTGCACTTC ACAAAAGGAG AGCAAAGTCA TTAGGCTTCA
    1441  TTCGCCGTCG ATTGGGTTGA GTGAAATCAC TTTCATGTTG ACCCACATAG CCATTGATGT
    1501  TGCAAATTGA CATTGAGCCG TGAGCAATGC AATCGATCGA TTACTAATTT CCCGGAACCA
    1561  TTTGATTGTA ATAAAGAATA CATAATATGA TGCACTTACC ATAAAAATGA TTTGCCTAGT
    1621  TCAGTGTTTT TAGATTTTTG GGCTATAATA ATTAGTTTTG TATAGGATTC TTATTGAAAA
    1681  AACGCCTTGT ACTTAACTCT CTTC

//

LOCUS          30     1794 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      30
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          complement(500..1294)
                  /translation="MSTLFARVFRQAWATFEHTGAGDAKFASNLNALRRLVDQLTLA
                  DIGLEPSLVATETFQQPTKAPCTYVGVFENDRFAMSVFVLRENYTMPLHDHPQMHGL
                  LRVVSGAVQICSYSEIARRDTVEPRAEGGTLRRHVLVVAEPEKIISSAAGDCAVLTP
                  TERNFHEITAIGGPAAFFDILSPPYNTASQPQYYFYRKVPVPRHLAEMEPLGFGGPD
                  KPSAQEDERPRYVLETIPNPEHYYCDTVHYATPEFMYQQAEGGGGGGGGV"
ORIGIN
       1  CAATACTAGA ATAACCGGAA GTCGTTCCGG TAGGCGGAAG TCAGAGCGCT AGCAGTAGAG
      61  TTAAATTTGG TTTTATTGAA CTTTCGTTAG GCTTTGCTTG CTTGCTTGCT TGTGTATTAA
     121  GGGTGAAGTG CCCTTCTACT ATTCAGTAAC GGTACTACTA CGCGTGCAGT ATTGCATGGC
     181  CGTTGCATCA ACGCAACGCC AGCGCCTCCT TCTCCATCTT CCAGGCGCTC CTGTAAAAAA
     241  CGGGGTGAGA AACACACACA CTACGCTAGG CTCGAACGAT GCCTACTGCG GTTTCTTCTA
     301  ATAGTGACGA AGAAGAACAA AAAAATATAA CATAACAAAA ACAAAAATAA CAATCCAACA
     361  AATTAAAAAG AAACCACCAG CATCGAAGCG TTCGAGTGCA AAATGCTAAT TATAATTTGT
     421  AGTTAGTAGT TGACCGACTT TCTTCCAGCC CCCGGGACTC TTTTCGGCTC GTTTCGCAAT
     481  CCTTCCCCTG AAACGACCCT TAAACGCCTC CTCCACCTCC GCCTCCTCCT TCCGCCTGTT
     541  GATACATAAA CTCCGGCGTG GCGTAGTGCA CCGTATCACA GTAGTAGTGT TCCGGGTTGG
     601  GAATCGTCTC CAGCACGTAC CGGGGCCGCT CGTCCTCCTG GGCAGACGGT TTGTCCGGGC
     661  CGCCAAACCC GAGCGGTTCC ATCTCGGCCA GATGCCGCGG GACGGGCACC TTCCGGTAGA
     721  AGTAGTACTG CGGCTGGCTG GCCGTGTTGT ACGGCGGGCT GAGGATGTCG AAGAAGGCGG
     781  CCGGGCCCCC GATCGCCGTT ATCTCGTGAA AGTTCCGCTC GGTCGGCGTC AGCACGGCAC
     841  AGTCGCCGGC CGCCGAGCTG ATAATCTTTT CCGGCTCGGC CACCACCAGC ACGTGCCGGC
     901  GGAGGGTGCC GCCCTCGGCT CTCGGTTCCA CCGTGTCGCG ACGCGCGATC TCACTGTAGC
     961  TGCAGATCTG TACCGCCCCC GACACGACCC GCAGCAGGCC GTGCATCTGG GGGTGGTCGT
    1021  GCAGCGGCAT GGTGTAGTTT TCCCGCAGCA CAAACACCGA CATCGCGAAG CGGTCGTTCT
    1081  CGAACACGCC GACGTACGTG CAGGGCGCCT TGGTCGGCTG CTGGAACGTT TCCGTCGCGA
    1141  CGAGCGACGG TTCCAGCCCA ATGTCCGCCA GCGTAAGCTG GTCGACCAGG CGGCGCAGCG
    1201  CGTTCAGATT GCTGGCGAAC TTTGCGTCCC CCGCTCCGGT GTGCTCGAAC GTGGCCCATG
    1261  CCTGCCGGAA CACCCGGGCA AACAGGGTGC TCATGGTGGT GGTCACGCGC GGCTATTGCG
    1321  CACTCGGTTC GGCAACACTA TTCCGGGCGC TACTGCAGGA CACGGACACA CGCACTACAC
    1381  ATACCCGGCA CACACACACA CAGCAGCGAT CGATGTGGTT CCGATGCAAA CGTTGGCCGT
    1441  GCTAGTTGTG CTGGATGATC TCCGCAATTA CAGAGTCCGG TTTTCCTTCC GCACCGTGCC
```

```
    1501  CGCACAGGCG ATGTTGTGTT GTTCTGCGTG TGTATGATTT CCCCCAAACT GTTCGCACCT
    1561  CACTGACAGC GTCGACTGAC AGTTGTAGAG CAGAGCGATA GAGAGAGCGA GGCGGACGTT
    1621  TGTTCCCTTT TCTTTTGCAG TGTTGCAGTG CAGGCTGCCT CCTTCGGTAA GGATAATTCC
    1681  ATTTGCAGCT GTTGTGGGAT TGCGTTAAAC TGCGCGCCGG GTGCACGAAA ACTTATTTAA
    1741  TCCATGCTAC GAAAGGTGCG ATATGTAACT AATTTCAATG ATTGACACTA CCGC

//

LOCUS          31     2043 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      31
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             join(500..533,613..1410,1503..1543)
                     /translation="MSANKFFFVFKDSLIESDDASMSIVSLRNPATRNATKYLFRRT
                     SKKACNLYEVNCFNEQHRSWFINESVCSNGKIFLPTPIDPLFLVLPYLVEHCSERAV
                     PLDQILIDDKFPCTSQLVEVLPADRLSLIADEKRAGSILALKYNETKALEWLVSKCH
                     RLSKVIGTEEGPTARSLNYIKEEKENEKDEDEKSTLHTALGIVSDYLSLEFGRKLSV
                     ALGFPEDENISKKRKSVVDLESAVVKRIKKEENHDTTPIKLPAAEKKVSAKAKALAK
                     AASGSKMRKAMASWWMEES"
ORIGIN
       1  CTTGAACTCT CTCATCGAGT GGGAAACCTC TAAAGAGAAA TAACAGCGAT TTTTTTAAAG
      61  CTGATAAAGA AAAGTAAGCG TTAAGCAAGT TCTTGCAAAA CAAAGCAGCG ACGCTTGATC
     121  GTCAACAATT TTGAATAGTA ACGGTTCTGC CTGTGCATCA TATACACCTT GGGCAGTAAC
     181  AATGGATGGT AGGCAGTTTG ATTGTGTGAG ATCTTTGCAT CGATTAGAAA GTTTATATTA
     241  AAATAAAGCA TGATTTGCTC GAAAAATTAA AGCAATTCAA ACGATTCATT ACATTATGTT
     301  GGTTTTTACA CATGGAAGAT CCCTGCAGGA CCAATCATAT ATACCTCACC TCAATCATAT
     361  ATATACTCAC CTTGGTCTCG CCGGGCGCTG TATGTTGCTT GGGAAATGTT TACAAAACAA
     421  ACCCGCTCCA TAGCATCACA ATTTCCCGCT CAAAAGTGTT TGTTTGCGTG TTTCGTTTCT
     481  GTTTCTGTCC GGCTGCACGA TGTCCGCGAA TAAATTCTTT TTTGTGTTTA AAGGTAGGCC
     541  CAATTGGAAC CATTTTCGCC ACTACACTGT TGCTAAAACC CATTTTCTTG ATTTTTCTTC
     601  ATTTTTTTGC AGATTCCCTG ATCGAAAGTG ATGACGCCAG CATGAGCATC GTGTCCCTGC
     661  GCAATCCTGC GACAAGAAAC GCAACCAAGT ATCTATTCCG GCGTACCTCG AAGAAGGCGT
     721  GCAACCTTTA CGAAGTGAAC TGCTTCAACG AGCAGCATCG CTCCTGGTTC ATCAACGAAT
     781  CGGTTTGCTC GAACGGCAAA ATCTTCTTAC CCACACCGAT CGATCCGCTG TTTCTCGTGC
     841  TGCCCTATTT GGTGGAGCAC TGTAGCGAAC GGGCGGTGCC GTTGGATCAA ATCCTGATAG
     901  ACGATAAGTT CCCGTGCACC TCACAGCTTG TCGAGGTGCT GCCAGCGGAT CGTTTGAGCT
     961  TGATAGCGGA TGAAAAGCGT GCCGGTTCGA TACTGGCATT GAAGTACAAC GAAACGAAAG
    1021  CACTGGAATG GCTTGTCAGT AAGTGCCATC GGTTGAGCAA AGTGATCGGT ACGGAGGAAG
    1081  GTCCAACGGC ACGATCACTG AACTACATCA AGGAAGAGAA GGAAAACGAG AAGGATGAGG
    1141  ATGAGAAGAG CACACTGCAT ACGGCACTCG GAATCGTGAG TGATTATCTT TCGCTCGAGT
    1201  TCGGACGGAA ACTTTCGGTG GCGCTGGGAT TCCCCGAGGA TGAAAACATC TCCAAGAAGC
    1261  GCAAATCCGT GGTCGATCTG GAATCGGCCG TGGTGAAGCG GATTAAAAAG GAGGAAAATC
    1321  ACGATACAAC CCCCATAAAG CTACCGGCGG CTGAAAAGAA GGTCTCTGCC AAGGCGAAAG
    1381  CGCTAGCCAA GGCGGCGAGT GGGTCGAAGA GTATTTCTAG CTTCTTTAAA AAATAAAATC
    1441  CCCCCTCCGT GTGTTAGGTG TGTTACGTTA AAAGCATTAT TATTATTATT ACGTTACATT
    1501  AGTGCGGAAA GCAATGGCCT CTTGGTGGAT GGAGGAAAGC TGATGGCTAC GAGGAATAGT
    1561  AAATACATAA TTCTTTGGTT CGACGTGTGC CTCTTTACGT GCCGGGTACG GGTGTCTGCA
    1621  GTTCCTGCTG CTGGCACAGC GTCACTAGCC GCTCCTGCGC ACCCTGCAGG ATGTTCATTA
    1681  TCACAACGAC CTTCTGCTTT ACGCGCACCA TACGCTTGAA GTATTCGTGC AGCACGGGAT
    1741  CGTTCTCGAA GTTGAGATTC TTCATCTCAC CGGACAGTTC TTCCAGCTGC TGGCGCAGTT
    1801  CGCTCTGTGT TTTGCTTGCC CCGCGGATGT TGTTTTCCAG ATCGTCAATC ACTGGTTTGA
    1861  AGAGCCGGAA GAAGCCCTCC GTCAGGATGT CCCGGGTCGG ATTTTCGCAC AGATTCTCCG
    1921  TGTTGTCCGT CCGCTCGTCG TCGATCGATG TGAGCGAACT GTCCGAATGA TCGTCGGCCA
    1981  TCGTGGGTGG GAATGTTTCT GCACAAGGAA GTTTATTTGT AGATGAATAA TAGAGCAAAT
    2041  AAG
```

```
//

LOCUS          32     2875 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      32
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..541,620..656,745..1580,2316..2375))
                     /translation="MSEKEVVMASRSTWITSIRLKTAYEAHRLLVQTYGKHALGKTQ
                     CAVWFGKFKRGDFDIQNKLRGRPPKKFDDGDLQALLDEDDCQTQQQLADRLRVTRES
                     ISVRLKAMRKVHRGGTWVPCELTDRQQEKRKTICEMLLARHQRKPFLRRLVNGGEKW
                     IYFENPKRKKSSSEPTDPAALAAGPSRYAQKTRLCVWWDRDGVICYELLGPGETVTP
                     ERYQLQLIALEQALREKRSEGGDRQDGRTIILLHDPVPSHSGKPVEETIAQFRWEAL
                     PHVANSPDLFPSSYHLFPVLGFALSNQSYRNGGKSAYKAMHNVRTMSHNTVQL"
ORIGIN
       1  CCCCGACGAC CAAAAAAAGC TTCCACCAGC CGCCGCCAGA TGCGCTAGTG AAAATCGAAA
      61  TTACACTTGC GAGTACGATC AATGTAGCCC GGCCTTAACC TTTTAACTTG TACTGAAAAC
     121  GCAATTAGCA GTGGATACAA CATTTACAGC CAACTCACTG AACGCAAACT GATGCGCACC
     181  GCAAAACAAT GCCTTGATGT GATGATGTTG TTAGAGAAGG GAAGCAGTTA CAATGGTGGT
     241  GAACTCTCCT TTTGGGGGGG GAAATTAAGG CTGCTATGTA CGATGCTTTG TTAGAACATA
     301  GCAATACACA CACACACAGT AACACTCCAC TCCACTTTAC TAAATGCGCT AATATAGCTC
     361  TTTCTATGTA TGCATGGGTC AACAGGAGAA ATTACCGTAT TATAAGTATA CTAGGCACGT
     421  ACGCGCTGTA TAAATAGAAT AAAGATCGCA CCATAATGCA ATAAAGGATC CATAAAGCGC
     481  TGTAGCAATT CTTGCCTGTT CACAGCTGAA CCGTGTTGTG CGACATTGTC CGAACATTAT
     541  GCTGGGATGT TTAGAAGACA AACACTTTTA ACACATAAGC ATTGCATAGT TTTTTTTAAT
     601  TATTTTTCAA AATAATTACC ATCGCTTTGT ATGCACTTTT CCCACCGTTC CGGTAGCTGT
     661  TGTAAACCGC GCCAAAAAAA CGACCCATCT TTGGTGCTAA ACCACCTGTC CAGCCACTGT
     721  TCGATATGTT CCAGCGAACC GAACCTCTGA TTGGACAGTG CAAAACCAAG CACGGGAAAC
     781  AGATGGTAGC TCGAAGGGAA CAGGTCTGGC GAGTTGGCCA CGTGCGGTAA TGCCTCCCAC
     841  CGGAACTGGG CGATCGTTTC CTCTACCGGT TTACCGCTAT GACTGGGCAC GGGATCGTGC
     901  AGCAAAATGA TTGTCCTACC ATCTTGCCTA TCGCCGCCCT CTGAACGTTT TTCCCGCAAT
     961  GCCTGTTCCA GTGCAATCAG CTGCAGCTGA TACCGTTCCG GCGTGACCGT TTCGCCCGGT
    1021  CCCAGCAGCT CGTAGCAGAT GACACCGTCC CGGTCCCACC AAACGCACAG CCTGGTCTTC
    1081  TGGGCGTAAC GGCTCGGTCC GGCCGCCAGT GCTGCCGGAT CGGTCGGCTC GGAGCTGGAC
    1141  TTTTTACGTT TGGGATTTTC AAAGTAAATC CACTTCTCAC CCCCGTTCAC GAGCCGGCGC
    1201  AGAAAGGGCT TCCGCTGGTG GCGGGCGAGC AACATTTCGC ATATCGTTTT GCGCTTCTCC
    1261  TGCTGCCTGT CGGTCAGCTC GCACGGTACC CATGTGCCAC CGCGGTGAAC CTTGCGCATC
    1321  GCTTTCAGCC GTACGGAAAT GCTTTCGCGC GTCACACGGA GCCGGTCCGC GAGCTGCTGT
    1381  TGCGTCTGGC AGTCGTCCTC GTCGAGCAGG GCCTGCAGAT CGCCGTCGTC GAACTTTTTC
    1441  GGCGGTCTGC CACGGAGCTT GTTCTGGATG TCGAAATCAC CGCGCTTGAA TTTGCCAAAC
    1501  CATACGGCGC ACTGCGTTTT GCCGAGCGCA TGTTTGCCGT AGGTTTGCAC CAGCAGCCGG
    1561  TGCGCTTCGT AGGCTGTTTT CTTTTGATGG TAGAAGAATA ATAGTGCTGC TCGCAGGTGG
    1621  TGTTTTTTCG GTACAAAGTT CGACATTTTC AGGCACAGGA GGCTAATTTT GTCGGGAGAA
    1681  AGAGATCACA TAACCGTTCA ATATGCGAGT AAACAAACGA AGTACCTAAT GAAACGTAAA
    1741  CAAACCGCAG CGTTTGTTTT GATTCACGTC GCCCAAGGTC TGTATCGTTT CATATAATCA
    1801  CCTTGGTATT TAAATGAAAA AAATAAATTG CAATTACAAC TTAAAAAAGA ATGTATTTTA
    1861  TGCACAACCT GCAGTCAGTC GGTCACTAAG CAAAATTATG ATAGTAAAAT ATCGAATAAA
    1921  GCAACCGCGT TGAATAATCG AACATTTATA TTCATCCACC TTTTTATTCA TCTTATCCTA
    1981  CTGTATCCTA CCCCCCGCCG AAAGTCTGTC AACGACGACT ACATCGCAAC TCTATATACC
    2041  GTGTTCGGAT GGCGCTGTCA CGTTGCCATG TCTTCTTCGC CTTTTCTCCT ACTGATCTTT
    2101  GCTCCTAGAG ATGGTACATA TGATTTTGGC CGGATCATAA CGTTCCAGCG TCTTTTTATT
    2161  TATTTAATTT TAGCCCTTAT ACGCCATGAA AAATACACAA ATAATTATTA ATAAGTTTTT
    2221  CTCAATTATT GCAATGATAT TAATATCAGA AATAGTTAAA CGAGATTGAT TTTCCTTAAA
    2281  AATGGTAAAA TTCATACCGG AACGGAATGA CTCACCAGCC GGATCGAGGT GATCCATGTT
    2341  GACCGACTGG CCATCACTAC TTCCTTTTCC GACATCTTGA ACTACGTGAA GCACGCGTTT
    2401  TTCGGGTGTA CGAGTATTTT GTTGTAAGTA GTGAAAAATT GTGCCGCTAT TGGCCCGAAA
    2461  AGCCCTCGAA ACGGTGCCGT ACCGATCGCT AACCAGGAAT CGATTCGCGT CCTCTCGTTG
```

```
    2521  CAGCCTCGGA GAAAGATGAA ATTCAACAAA AATGTTTCGT CGTCCCGGCG CAAGAGCCGC
    2581  AAGCGCCACT TCCAGGCCCC CTCGCACATC CGCAGGAAGA TTATGTCGGC GCCGCTGTCG
    2641  AAGGAGCTGA AACAGAAGTA CAACGTGCGC TCGATGCCGA TCCGCAAGGA CGACGAGGTG
    2701  CAGGTGGTGC GCGGCCACTA CAAGGGCAAC CAGGTCGGCA AGGTGGTGCA GGTGTACCGC
    2761  AAGAAGTACG TCGTGTACAT TGAGCGCATC CAGCGCGAGA AGGCGAACGG CACGAACGTG
    2821  TACGTCGGTG TGCACCCGTC CAAGTGTGTG ATCGTGAAGC TGAAGATGGA CAAGG

//

LOCUS          33     1980 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      33
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(500..1480)
                     /translation="MDSTLDRTYVSDYFANLNEISKKIALDIDEIRGAYEHIWEDLS
                     REEQMEIVNETIIKPELALKYFDNFSTSTDSLQRILGTDDDDPTTGGGGEEDRNANR
                     YDGKRLQTFQYVRTGRKVITDDSVGLYRDEHSAPFSHKTKSQINLTLFPPDRAEPDA
                     KKPHDVSVKAAAGLLAKKLLLKSPKPDDANGGPLAKGRDGAAATKAKELPAYDSEQN
                     IILTEEAYGTKATGAGFLSKFNNLIFNQSTALEMTTAASAASINGSDELDGGALEQD
                     EFDADNDKVNLFTKSAKNECLHRAQSSSVEEDDKNYDELNSLLGNSKGFDFLNNW"
ORIGIN
       1  ACGCACTACG CCACCGCTTC TACATGCCGT GAGGTTACAA ACGTTCTGTT ATGGAGCATG
      61  ATGTGAGATT ACAATCAGTC TGTTCTGGAA TTGTTTGATA TTTGTTCATT TCGTTCTCTC
     121  GTGTTTGAGT ATTTCATAAG GTGAGTGGGT TGAGTACTAA GTAGAGAGGT CCTAATAGCG
     181  AGCAAACAGC TAATGCGTCC TCTGCTCCAG AGAAGCAGTG CTCTAGCTAA CTGCAGACTA
     241  AGAAAAAGGC TTTTTAACCG CCAGTGCTTT GGTGCGATTT GCAAGGAAAC AGGGGAAAGA
     301  ATAAGAATAT TTTACAGCCC TCCACTGTCC TGAAACCTAT CCTTTCACCA TGGGATGGGG
     361  CATTAGCGCG GCAGACAGAA CTAAAGCGAC TAGCAGAGGG CTCATGTGTC ATGAGACTTG
     421  GTGTGTTGTG CCTGTGCTAC CGTTACTGCT AAGGCAAGAG GGAAGCTTTC TTTTGTTTCG
     481  TTTTGTTTTG TCCTTTACTT TACCAGTTGT TGAGAAAATC AAACCCTTTA CTGTTGCCGA
     541  GCAGGCTGTT CAGCTCGTCG TAGTTCTTGT CGTCCTCCTC CACGGACGAG GACTGCGCAC
     601  GGTGAAGGCA CTCGTTTTTG GCGCTCTTGG TGAACAGGTT CACCTTGTCG TTGTCCGCAT
     661  CGAACTCGTC CTGCTCGAGC GCTCCGCCAT CGAGCTCGTC ACTGCCATTG ATCGAAGCAG
     721  CCGATGCGGC CGTCGTCATC TCCAGTGCGG TCGACTGGTT GAAGATGAGG TTGTTAAACT
     781  TGCTCAGGAA GCCGGCCCCG GTTGCTTTGG TACCGTACGC CTCCTCCGTC AGGATGATGT
     841  TCTGCTCGCT GTCGTACGCC GGCAGTTCCT TCGCCTTCGT CGCTGCCGCC CCGTCGCGTC
     901  CCTTCGCCAG CGGCCCTCCG TTCGCGTCGT CCGGTTTCGG TGATTTGAGC AGCAGCTTCT
     961  TGGCCAGCAG CCCGGCCGCC GCCTTCACGC TGACGTCGTG CGGCTTCTTC GCGTCCGGCT
    1021  CGGCCCGGTC CGGTGGGAAG AGGGTGAGAT TGATTTGGCT TTTGGTTTTG TGCGAGAACG
    1081  GCGCCGAATG TTCGTCCCGG TACAGGCCCA CACTGTCGTC CGTGATCACC TTGCGGCCGG
    1141  TGCGCACGTA CTGGAACGTT TGCAGCCGCT TGCCATCGTA CCGGTTGGCG TTGCGGTCTT
    1201  CCTCCCCGCC TCCCCCGGTG GTGGGATCGT CATCATCCGT GCCCAGGATT CGCTGCAGCG
    1261  AATCGGTCGA TGTGGAAAAG TTGTCAAAGT ACTTCAGTGC CAGCTCCGGC TTGATGATGG
    1321  TTTCGTTCAC TATCTCCATC TGCTCCTCGC GGCTCAGATC CTCCCAGATG TGCTCGTACG
    1381  CCCCGCGTAT TTCGTCGATG TCGAGGGCGA TCTTTTTCGA GATTTCGTTC AGATTCGCAA
    1441  AGTAATCCGA CACGTACGTG CGGTCCAGTG TGGAGTCCAT CGTGCGCCAA GGCTAACTAC
    1501  GGAATGGGCG GTGGCGAAAT TGTGTAGCGG GGCTGTGCCT TACATTAGCG CCCGAAAGAA
    1561  AGAGGGCCCT TTCGATGCGG TTTCGCGAGC GATGGGCGCT CGCGTTCACC AGATGCCAAT
    1621  AGCCTCTTAC GCTTATATAG ATGGCAATCA GAGCGTGTCC GACACTGGGC TATGGTGCTG
    1681  CTGCTGCTGC TGGGGAGGAG AGCTGCCCTT TGGTATGTGA TCGGGTGCTC CGTCGATTAC
    1741  CCCCGCTTTA GCCGAACCGT GCGCGAGTGT TCTCGCTCAA TTAATTGACC GTTTCATCGC
    1801  AAATACCCAG CCCAGCCGAT ACAGTATCGA CCCAATCCCC CTACACTCAG GCACACACTA
    1861  CACTCTAGCC TAAATTAAAG AACGCGTTTG AGAAGTGGAA CGCCTTTTTG TTGGGTGAAA
    1921  AACTATTTTG TTATGACACT GTTTGACGTC GCCGTCACGG TAGGGAAGAA AAGGTAAATA
    1981
```

```
//

LOCUS          34     2145 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      34
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          join(500..656,731..1124,1204..1645)
                  /translation="MKCRIFCAVLLGTLLAPFVHSWDLRTIGIDRFRVRHVSLYHSR
                  AFLTIESSNVSLVEASWPETVGYQWSRVLSTDGDERGSCRGLTNVLGTDIDRLARLW
                  VLCGAGENACPPKLVIRSLLSPGTGEIQHRFPTADAEQQRFHAIVVDPVPASDGDTR
                  AFITLFDQDYVLLYSLFKRSIGKLQFQKSDQTSLHPISLSEVTVNLNRLYVADTVSD
                  RLFALPIRNLRQLAFPEDGVRKIIMKTNVTYLGRLLGRPAGLTLDFRNNLLYVLQRD
                  GAIVKWTPGRSLKAENHRVILQQGTNITQIILGIGGKAWAVSGEYISEASRRHCLKI
                  VV"
ORIGIN
        1  TTTGGTCACT CTGTTCAAAT AACTGAATGA GAGCATTGGG TCTCCAATAA TTCAAAATTA
       61  GGGTCTATCG CAGCCATTAG ATCACAACAC ACTGCCCTAC TAACTTGAGC TGTTTATTGA
      121  AGCTGATTGA TCTCCAAAAC ACTTCCAAAA CAACGACCCA AACCATCTCA AAGCACGTAC
      181  TACACCACCG TGTTAACGCC TTTTATCACC TAACACCATC TCAAGTTACG TCCAGCGTGT
      241  ACCTCCTCCC ACTGTAAAAC TGCCCGCATT CCGACCTAGC TTCACCGCTG TACCGGCTGA
      301  ACCCGAACCT CTGCAACGAC CTTGCCATGT TCGCATGTTT GTTACCAACA GTCCACCCTT
      361  TGTCCCATCC AACCCCAAGC ATTTTGCCGA CGTCAGCGGA AACCCATTTC TCGAAAGAGA
      421  GGTATGCGCC ACTCAACGGC ACACGCCACA GGCCGCCCCA CTTCAACAAC AGTGGCTGGC
      481  GCTGACGCTT TAGTTCACGA TGAAGTGTCG CATCTTTTGC GCGGTACTGC TAGGTACGCT
      541  CCTAGCACCT TTCGTCCACT CGTGGGACCT ACGCACGATT GGGATCGATC GGTTCCGCGT
      601  GCGGCACGTT TCACTGTACC ACTCGAGGGC GTTCCTAACG ATTGAGTCAT CGAACGGTGA
      661  GTAGTGGATG AAGCGGCGTG TTACACAATC AAAAAGGGGG CTTAGGTCTA AAATACCTGT
      721  CGCGTTACAG TGTCGCTGGT GGAAGCGAGC TGGCCCGAAA CCGTCGGCTA CCAGTGGTCG
      781  CGCGTGCTCT CAACCGACGG CGATGAACGA GGATCGTGCC GCGGGTTGAC GAACGTGCTG
      841  GGCACTGATA TCGATCGGTT AGCACGGCTG TGGGTGCTGT GTGGCGCTGG GGAGAACGCC
      901  TGCCCCCCGA AGCTGGTCAT CCGTAGTTTG CTGAGTCCGG GCACGGGTGA GATTCAGCAC
      961  CGATTTCCCA CAGCTGACGC TGAGCAACAA CGGTTCCATG CGATCGTGGT CGACCCGGTA
     1021  CCGGCAAGCG ACGGTGACAC ACGTGCGTTC ATCACGCTGT TCGATCAGGA CTACGTGCTG
     1081  CTGTACTCGC TGTTCAAACG CTCCATCGGC AAGCTACAAT TTCAGTAAGA CTCGTGTGAC
     1141  CGTCCCTTCA TCTCCGGAAT TGCATTCCAT GCTTACCACT CTCACTCCAA TCATTTCGTC
     1201  CAGAAAAAGC GACCAAACCT CACTGCATCC GATCTCACTG TCCGAGGTAA CGGTCAACCT
     1261  GAACCGGCTG TACGTGGCCG ACACGGTGAG CGATCGGCTG TTTGCCCTGC CCATCCGGAA
     1321  CCTGCGCCAG CTCGCCTTCC CGGAGGACGG TGTGCGGAAG ATTATCATGA AAACGAACGT
     1381  CACCTATCTC GGGCGGCTGC TCGGGCGGCC CGCCGGGCTG ACTCTGGATT TCCGCAACAA
     1441  TTTGCTGTAT GTGCTGCAAC GTGACGGTGC GATCGTGAAG TGGACGCCGG GCCGCTCGCT
     1501  GAAGGCCGAG AACCATCGCG TCATTCTCCA GCAGGGCACG AACATTACGC AGATCATTCT
     1561  TGGAATTGGT GGCAAAGCGT GGGCCGTATC CGGGGAGTAC ATATCCGAGG CTTCCAGGCG
     1621  GCACTGTCTG AAGATCGTTG TTTAGGAAGA CTGCTGTCGA AAAAAGGGAA AGAAAACTGT
     1681  GATAAAGCGT AAATAAAGTG TGTTTGTAGA TTAAACAGAT GTGCCAGAAG AAATTGGAAC
     1741  CTTCCTGGAG CAACCTGATA TAGAGCAAAA CAAACAAAAC ATTCCATACA CGTGCCCGAC
     1801  CGGGTTCGAT GACGTCTTTT TGGTGGTGAG ACAAGTTTTT TGTTGTTGTT GTTTTGCTGG
     1861  CGAGGCAAAC ATCCTTCCAC AATGATTCCC TGAACTTGAC CATGGCTCGA CGAGGGTAAA
     1921  CACGCAACCA ACGATGTCCG TGATGTCCGT TACGGTGACA TTGAAATTCG CTAACCGTAC
     1981  TCTAAATCCT TACGATTGCC CAGTGAGAAA CGGCAATGGA GAACCGAAGG AGGGGTTGTG
     2041  AAGGTACACG CCGGGTTCGT TCGTGCAGAA AACTTGTTTT AATGCCTTTT CATTAACCAG
     2101  CCGAGGCAAT GCGTCTGGCA CTAGCGGCTC TATCGTATAG AGATG

//

LOCUS          35     2159 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      35
```

```
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..1319,1412..1659))
                     /translation="MASQKSTASTSTTTPALAVTNRILRSPSHESFSTDLSLISISV
                     SSMNSEATSVLNRSSCSLLNRTMEEKLGSLLNPKDSKRNRSSTLTERWNESNDLIRN
                     CATLSAALGIHKSFSTTDIYQLPSDQHQLGRLSSSELALTPFQKVGYLFDHSRLEHA
                     SRSYSTWVAVGELASTSQLPSPHGGQSAGQVPTLTTTNPPAPSFTVTDLISSVNKKI
                     RQNYIRRRLFTTYRALERLSQSEFNLDRLEAAAFAATNPGPTELIVPSTSLSNASAS
                     PVGGKSTSSPAPPGGGGGGEALADHPGKKTSGGVARGVHHGPKKHLTFHDIENERGK
                     PLSKYERHMLIFNWLHTIDDTALEIEN"
ORIGIN
       1  GAGAGAAACA GGAGTTCTCT TGGAACCGAT CAACTAAAAT CCTAAGAGAC ACAGCAAATA
      61  TAACTTAATA CGTTACGTAG AAGACATCAC ATATCGTATA GAAGTAGTTT TTCCTACATT
     121  GGAGGTATTC CCTTGATTGC ACACTGATTA GTAGAAAATA ACACAGTTAA TAAGAAGTAC
     181  TCCAAAAGAG AAGAAGCTTT CCCCTTACAG AAGTGAAGAG TACACGTAAC GACCATTCCA
     241  CCAAACATCA GCTAAACCTC AAGTAAACTA CAACACACGC ACGAACTCCG AATGCACTTT
     301  GTGACACTTT GTAAAAGCAG GACGAGACAA ACTTGCTGGG GACGAGGACT TATTCTGGGC
     361  TCATTTACTT CCTCCTCTCT CTCCCTCTCT TTGTGTGAGT ATGGGTAAAT AAATCTTACT
     421  CTTTTGTGCA GAAGAGACTT GGGTCGCGTC CTCCTCTTCC TCCTCCTCCT CCTCCTCCTC
     481  CCGGGAGCTC CTGCGGCGAT CAGTTCTCGA TCTCGAGCGC CGTATCGTCG ATGGTGTGCA
     541  GCCAGTTGAA GATCAGCATG TGCCGCTCGT ACTTGGACAG CGGTTTGCCC CGCTCGTTCT
     601  CAATGTCGTG GAAGGTGAGA TGCTTCTTCG GACCGTGATG TACACCGCGA GCAACACCTC
     661  CGCTTGTCTT TTTACCGGGA TGGTCCGCGA GTGCCTCACC ACCACCACCA CCACCCGGTG
     721  GTGCCGGTGA GCTGGTCGAT TTTCCACCCA CGGGCGAGGC GGACGCATTG GACAGGCTGG
     781  TCGAGGGTAC GATCAGCTCG GTTGGCCCCG GATTGGTGGC GGCAAAGGCT GCCGCCTCCA
     841  GCCGGTCGAG GTTGAATTCG CTCTGGGAGA GCCGCTCCAG CGCACGGTAG GTCGTGAACA
     901  GCCGGCGGCG GATGTAGTTC TGCCGTATCT TCTTGTTGAC CGAGCTGATC AGGTCCGTTA
     961  CGGTGAAGCT CGGTGCGGGC GGGTTGGTCG TGGTCAGTGT CGGTACCTGG CCGGCCGATT
    1021  GGCCACCGTG CGGGGAGGGA AGTTGCGAGG TGGACGCTAG CTCACCGACC GCTACCCAGG
    1081  TCGAGTAGGA GCGGGACGCA TGCTCCAGCC GGGAGTGATC GAAAAGGTAC CCGACCTTCT
    1141  GGAAGGGCGT GAGCGCGAGC TCCGACGAGG ACAGACGGCC AAGTTGGTGC TGATCGGAGG
    1201  GCAGCTGATA GATGTCGGTC GTGCTGAAGC TCTTGTGGAT GCCCAGTGCG GCCGACAGTG
    1261  TGGCACAGTT CCGGATGAGA TCGTTACTTT CGTTCCAGCG CTCGGTCAGG GTGGAGGACC
    1321  TACTTTCAGA ATTTGAAAGA AAACTTTAGA AATCACAACT CCCAGATATA TGTAGTGCCC
    1381  AAAGCCTCCC CCCCCTCAGT TCCTTTCATA CCGATTCCTT TTGCTATCCT TTGGATTGAG
    1441  CAGACTGCCG AGCTTTTCCT CCATCGTGCG GTTGAGCAGG CTGCAGCTGG ATCGGTTCAG
    1501  TACGCTGGTC GCCTCCGAGT TCATGGACGA TACGGAAATC GATATCAGGG ACAGGTCGGT
    1561  GGAGAAGCTT TCGTGCGACG GCGAGCGCAG GATACGGTTC GTTACCGCTA GCGCCGGTGT
    1621  GGTGGTGGAG GTGGAGGCGG TACTTTTCTG TGACGCCATT CGTGTTCGGG ACGCACCGGG
    1681  GTGGAGTTGT CTGGAATAAA CGCAGAGCGG ATGCCTGATG CTTTATGATC ATTGGAAGTA
    1741  GTTGGTAGGT GTAGGCTCCA TGCTGTTTGT CTAGTAATAT ACCAGCCCTT CCCTAACCTA
    1801  CAACTTCCAA CCAGCAGCGG GTGAATACAA ATGAGCTAGC AACCGGGCTA CCTGAGCCTG
    1861  GGAGTTTAAA TTTAGAATAG CTGCGACAAC GGATTAATGC CAATCAATTT CCCCCGAATG
    1921  ATGCGACACC GGGCACCGGG CTCAATAGTT ATAATCGCTT TCACTGCTAA TCGCGATGTA
    1981  TGGATTATTA AGTAGAGCCA CTGTGTGTCC TTTTTTTCAT CGACACGGTA ATAATCGGAA
    2041  GGTTTCAGTA TTTTGTTCCA GTTCCTATGA AAGGATGTAG TGAGCGGTTG CAGAGCTCTA
    2101  GAGATGGACG GTACGCGATA CCGGTTGATA TGCATGAAAG CCGTGGCAGG CGTTCGTGC

//

LOCUS          36     2259 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      36
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..1259,1325..1533,1634..1759))
```

```
                     /translation="MSEATNEIPLDRKLNEIFDEGYSAMNRLDENAALPSNAPEFQV
                     AVKKTIGHFEDATRLVSLVGMFSTNESYEEVPTENLRYFLLPFFLGKMTLRLCNTNR
                     EEVVEVAEVYFKDFLARCENYKLYETPDRDTGELAIVPGGQSDKIRELQRMGAQRNE
                     KIRKFQEKKELDEKVKQLRYVVEQPEAKIDDETKREFFLSLLKSAVLEAQEELESVA
                     REKQMLQYRAAAKLRGEEELEEASHGAFGGGTNKRPPVQPLKPIIITRDAVQKAVYG
                     RGYPSLPTMTVAEFYEQRVAEGIFPDPERMKEVNKNSLMNRVHMDNAAEQDREDEQQ
                     EQLIERDDEEYLARQRAKDEFKDEHRRGEGNRYNRS"
ORIGIN
        1  CAGGGTAAGG GAGGCGCCGG CGGTGGTACG AAGAAGGTGA AGGAGGAAAT TTCCGACGCG
       61  AGCAGCGCCA CCATCCTGCC CGCGGACAAC ACGACGGAGC TGATCAGTGC GGACGAATCG
      121  CTGCTGGAAA CGAAGCTGGA GGAGGAGGAG GAGATACAGC CGGACGATGA GGCGTGGGTC
      181  AAGATGAAGG ACGGCATAAT ACGCACCGAG TGGAACGCGG GTGACCTGAA CGTGGTGCGC
      241  GAGTACGTCG TCGAGTTTGT GGAGGATTCG AAGGAGATTG AAATATAATA ATACGTACGT
      301  TGTGTAAGCT TAACAATTTA TCCAAAATCA CGAAATACAT TCATATTGCC GCAGCGCATT
      361  CCGCGAACGC ACTTCTCTAT TTTTTTTATG AAGCTACCGC AACATGCCGA GTGAGTTGTT
      421  GTGATCGTGT GTGGTAAAAG AGAAAAGGAA AAATACAAAC TAATTTCGAA CGAACCGCAC
      481  AGTAGGCGTC AAAGCCCCCT TAGCTACGGT TGTACCGGTT GCCCTCGCCA CGGCGATGCT
      541  CATCCTTAAA CTCGTCCTTT GCCCGCTGCC GGGCCAGGTA CTCCTCGTCG TCCCGCTCGA
      601  TCAGCTGCTC CTGCTGCTCG TCCTCGCGAT CCTGCTCGGC CGCATTGTCC ATGTGCACCC
      661  GGTTCATGAG CGAGTTTTTG TTCACCTCCT TCATCCGTTC CGGGTCGGGG AAGATGCCCT
      721  CCGCTACCCG CTGCTCGTAA AACTCAGCCA CCGTCATGGT GGGCAGGCTG GGGTAGCCCC
      781  GCCCGTACAC GGCCTTCTGT ACCGCGTCCC GCGTAATGAT GATCGGCTTC AGCGGTTGCA
      841  CCGGCGGTCG CTTGTTGGTG CCGCCCCCGA ACGCACCGTG GCTTGCTTCC TCCAGCTCTT
      901  CCTCGCCCCG CAGTTTGGCG GCCGCCCGGT ACTGCAGCAT CTGTTTTTCG CGCGCCACAC
      961  TTTCCAGCTC CTCCTGCGCC TCCAGGACGG CCGATTTGAG CAGGGAGAGG AAAAACTCGC
     1021  GCTTCGTCTC GTCGTCGATT TTCGCCTCCG GCTGCTCGAC CACGTACCGC AGCTGCTTGA
     1081  CCTTTTCGTC CAGCTCCTTC TTCTCCTGGA ATTTGCGGAT CTTTTCGTTG CGCTGTGCGC
     1141  CCATTCGCTG CAGCTCGCGT ATCTTGTCGC TCTGCCCACC GGGCACAATG GCCAGCTCGC
     1201  CCGTGTCCCG GTCGGGCGTT TCGTACAGCT TGTAGTTTTC ACAGCGCGCT AGGAAATCCC
     1261  TACAAAAAAA GGAAGCGAAA TGAGAATTGT TTACCTTTTA GCGATGTTCA CGCATCAAAC
     1321  TCACTTAAAG TAAACCTCGG CCACCTCCAC CACCTCCTCG CGGTTTGTGT TGCACAGCCG
     1381  GAGCGTCATT TTGCCGAGGA AAAAGGGCAG CAGAAAGTAG CGCAGGTTTT CCGTTGGCAC
     1441  CTCCTCGTAC GATTCGTTCG TGCTGAACAT CCCGACCAGG CTGACGAGCC GGGTAGCATC
     1501  CTCGAAATGG CCGATCGTTT TCTTTACCGC TACCTGCGGC AGTGGAACAG TTTTCATAAA
     1561  CGTACAGGGT CAGAGACAAA AAATGGGGGA AAACGTGAAC AAATGCGTTA AAGCCACAAA
     1621  TCATTAAACT CACCTGAAAT TCGGGTGCAT TCGAAGGCAG TGCCGCATTT TCATCCAAGC
     1681  GGTTCATGGC GCTGTACCCT TCGTCGAAAA TTTCGTTCAG CTTGCGGTCA AGCGGTATTT
     1741  CGTTCGTCGC CTCCGACATT TCGCTCGATC GCACACTGGT TTTCTTTGCA AGCTAGAAAC
     1801  TATTTCGCAC GTCGTTTGCT AATTTTACCA ATTTAACTAC GGAGGGCGAA CGTAATGAGC
     1861  TGCGGATTGT GTTTTGCTAG CTGCTACCAA AACGTTCCCA GGGTTGTTAC GTGCGTTCAG
     1921  GTGCTCTGAC CTTGTTCTCT ACACACCTTG CACACCTTTG ACAGCTGCAC TGTTTCGGCA
     1981  ATGTTTGTTT ACAAAACAAT GGAGCAGCTG CAGGGCGAAG GTAATTAGTC AAATTAGCGC
     2041  GCATTTATTA CATAGTTTAT CATTCTTTAT TGTTTTATTA CACTTTTTTA CATTTTAAGA
     2101  TGCTGCTACA ACGTCCTGTC CAATGGAGTC GTACTCGACC GCACCGCAAA CAGCCGTCCT
     2161  GCAGCGATTC TGCCGACTGT GCATGCGAGA GTTCCCGTTT CTGCTGCCGT TCAATGCTAC
     2221  ACTCAAAGAA ATCGCGCTAG CCGACATGCT GGAACGACT

//

LOCUS          37     2097 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      37
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          complement(500..1597)
                  /translation="MDCTHSGNGGSGVSGTSTGCHKNDLNLLVKNIVKILLTTSKYL
                  HCPFCADEYLFEFTLKHHLLKNHERDLEKIVRSPDASIRFCESNICPYCGALFYYIS
```

```
                ALPKHIMNSHSRECLMKWQAIEQENSLVRKLANDPSIQCVPCSPGLSDFFDELTTTT
                TAAATGGDAGRPTAASRNIMQLTKPSPLLKTALRNVNRITVAANGGNGGIDPTGTSA
                WSSFGFDGAQTRRSSNLTNQSVRRELNFGDDRSSGRREDGCSPVSGSPSLCNLIAGS
                CSPSPSSESISSARKPKFQRKRFNLRSIKPKLNFRKYVLSKFSKKYHCKIVTSTPNN
                FLDDGLSSGGGCGGGGSSRDTTRFLHNQERWNRIHFE"
ORIGIN
       1  CCTACTACGG CTACGATTGC ATTGGCAGCT CATCGTACCC GATCACAGAT GACGATCGTT
      61  TTCGTTTTCG ATCAAAATCT TCCAATCGCC GAAGAATGGA CCCTCCTGGT GGAAGTATAG
     121  CGGCAATGTG ATACAACGCC TATTTTAAAC GCTCTCTGCC ACATTCTGGA ACTGTTGGAA
     181  AGTTAACAAA CTGCTTCTTT TGCTGCAAAG CAAACTACAC AACTTTCCAC CAGTTCACCA
     241  ATTTGCTCTC TATGGGTTCT CTCTCTCTCT CTCTCTAGGA AAGAAATGAT TGCACAGCCA
     301  AGTTTTGGTA CATCTCGTTG CACGCTATCC GCTTAACGGT TGGTTTGTGT TGGCCATTGG
     361  CAGGGTTGGC CAACGGGTAA AAAAATGGAA TTAGTTGAAA CGAACGAGAG GCAAACGAAC
     421  GGGCTTCGCC TTACACGCAG CAAAACAGCA GCCCGCACAG ATACTCGACG ACGCACGAAA
     481  ACGGCGACCC ATACCTTCGT CATTCGAAGT GAATCCGGTT CCAGCGCTCC TGGTTGTGCA
     541  GGAAGCGGGT CGTGTCGCGG CTGCTGCCGC CGCCGCCGCA GCCGCCACCA CTGCTCAGCC
     601  CATCGTCCAG GAAGTTGTTC GGCGTGCTGG TAACGATCTT GCAGTGGTAC TTTTTCGAAA
     661  ACTTCGACAG CACGTACTTG CGGAAGTTAA GCTTCGGCTT GATGGAGCGC AGATTGAAGC
     721  GCTTGCGCTG GAACTTGGGC TTGCGGGCGG ACGAGATCGA TTCGCTGCTC GGCGAGGGGG
     781  AGCAGCTGCC CGCGATCAGG TTGCAGAGCG ACGGCGAGCC ACTGACGGGC GAGCAACCGT
     841  CCTCCCGCCG GCCGGACGAC CGATCGTCGC CGAAGTTTAG CTCGCGTCGG ACGGATTGGT
     901  TGGTGAGGTT GCTGCTGCGG CGCGTTTGCG CACCGTCGAA CCCGAAGCTC GACCAGGCGG
     961  ACGTGCCGGT TGGATCGATG CCACCGTTGC CACCGTTCGC GGCCACCGTA ATCCTGTTCA
    1021  CGTTGCGCAG GGCCGTCTTG AGCAGGGGCG ACGGTTTGGT AAGCTGCATG ATGTTTCTGC
    1081  TAGCGGCGGT CGGGCGACCT GCGTCGCCAC CCGTAGCCGC CGCCGTCGTC GTCGTCGTAA
    1141  GCTCGTCGAA AAAGTCGCTC AGCCCGGGCG AACAGGGCAC GCACTGGATG CTCGGATCGT
    1201  TCGCCAGCTT GCGCACCAGG CTGTTCTCCT GCTCGATCGC TTGCCACTTC ATCAGACACT
    1261  CGCGGCTGTG GCTGTTCATG ATATGCTTGG GCAGGGCGGA GATGTAGTAG AACAGGGCGC
    1321  CACAGTACGG GCAGATGTTG CTCTCGCAGA ACCGGATGCT GGCGTCCGGC GAGCGGACAA
    1381  TCTTCTCCAG GTCGCGCTCG TGGTTCTTCA GCAGATGGTG CTTGAGCGTG AACTCGAACA
    1441  GATACTCGTC GGCACAGAAC GGGCAGTGCA GGTACTTCGA GGTGGTCAGC AGTATCTTGA
    1501  CGATATTCTT TACCAGCAGG TTCAGGTCGT TCTTGTGGCA TCCGGTGCTG GTGCCGCTTA
    1561  CGCCACTGCC ACCATTACCG CTGTGGGTGC AATCCATCAT CCTATTGGCA GCGGAAGTGT
    1621  GTGTGTGTGA TTTCCTCCTT CCGGGAAAGT GGCACGTTTA TTACACACAG CACAGCAGTT
    1681  CACTGTAACG ACGGTTGGGG CGTACTGTTT CTTTGTATCC CGCTATTACT CGCATGGGAG
    1741  CTGCCGGGCA CGGGTTGATC CAGGGAAGGA GTAATGGCAC AAAAGAAAAT CCCCGAAAAA
    1801  AAGAAACGGA TATCGGTCAA CTCCAAGCAC TGCCTACTTT ACACGCCCGA CGAAACAGTA
    1861  GAAAGTGGTG CAACGAAAGC AGCAGCGCTC GCACGGCGAC CCTCACACGT TAGTGTCCGT
    1921  TGGCCGTCCG CCGATCGATG CGAATGATGT ATGATGCGCT GGTGGTGCTG CTGCGGTGGT
    1981  GCACCGATGC AGGTAAGCGA AGTGCAACGG CATGGAACGG GCACAGCGCG CGCGTCCGGA
    2041  GCAGCCAGAG CGGGATGGAA TGAGCGACAC AGAGAGAGAA AGAGGGGAGA GAAAGAA

//

LOCUS          38     2347 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      38
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES           CHR  ---
     CDS          join(500..544,613..898,978..1500,1580..1847)
                  /translation="MDALSTVRTYEQFQQDFPHWLLNVRNPAELFNAQPSYVVSQAL
                  CIIGGLLSLAHALHRGGRWPFLWMAGALTGVLVEGSMYFSPYGETIWFSPTVIDLFH
                  QRIPLFIFFVYPFFYYQACWAASKLQLKCRWSEHIAAGLLVVLADLPFDMVSIKFLH
                  WTLHDTEQLLSERVYSAPWTLLLFFAVASFAFSYLFHNLRSWMDRSVGTPQPTDRRW
                  VVGTIGAELVAMVGAASVSLSVGTGLFLAFSYPLHTVLGIPHRIIVIGVFLCVATVF
                  WKFDRKSNRRMPMSQSLLDHALNVITVGHFVLYFVLAFVLNPEDTVSSGRHQPIGDC
                  HHTTGTNAPPLCLDTFSRTDYDFHCISKPPNVGTYWYTVCGTPYE"
```

```
ORIGIN
       1  GTGATACAAA ACAGCCTGCT AGAAAATCGA GTTGTGCATT CTCTTTTGCA CATATGACCC
      61  GATTTGGGAC GTCGAACCTT TTGTTAATTA ATTTTAATTA AAACAAAAAA GCTTACCGAA
     121  CAAATATTTT ACCAATCGGA CCTGAACCTA CGTGTAGCAA TTTACGAACA ATTAAAAAAA
     181  AACGAACAAA GAAACTAATA AAACATTATT TAAATCGCTT TACTTTACCT TTACGGTTGT
     241  TTTTTTTTAA AATATTTTTT GAATCATTTT TCCCTCCTGC CGTGTGTCCT TCAAAAAGGC
     301  TATCGGCAAG GTTACATGGT ACGGGTACGT GTGTGTGTGC GTGTGCTCAG TCCAAATTTC
     361  TTCAAAAGAT TGCAGATTAC AATTACAAAG TCCCTCCCCT GCCTGCTGCC ACACCATGAG
     421  CGATAAATGA ACGAGCGCTG ACGTTCTCAG TTGTGTTTTA ATCTTTAAAC GTCGAATGAC
     481  GGACACAAAA ACACTGAACA TGGATGCCCT TTCGACCGTA CGTACTTACG AGCAGTTCCA
     541  GCAGGTAATT TCATAGTTTA AAAACATAAT AAATAAATTC TAATTCATAG TAACATACCT
     601  TTCTGGCTGT AGGATTTCCC ACACTGGCTG CTGAACGTGC GCAATCCGGC GGAACTGTTC
     661  AACGCGCAAC CGAGCTACGT CGTTAGTCAA GCTCTCTGCA TCATTGGTGG ATTGCTGAGC
     721  CTAGCACACG CCTTACACCG GGGCGGCAGA TGGCCCTTCC TGTGGATGGC CGGTGCGCTA
     781  ACCGGGGTGC TGGTCGAAGG TTCCATGTAC TTTTCGCCGT ACGGTGAAAC GATCTGGTTC
     841  AGCCCGACCG TGATCGATCT GTTTCATCAG CGTATACCGT TGTTCATATT TTTCGTCTGT
     901  AAGTAATGGA TCGGATCGGA AGTTCATAGT CCATGAAGTA CATACTTTAA TTGTATTTCT
     961  ACTTCTTTTT CTACCAGATC CCTTTTTCTA CTATCAAGCT TGTTGGGCTG CTTCGAAGCT
    1021  GCAGCTCAAG TGTCGCTGGT CGGAACATAT CGCCGCCGGG CTGTTGGTAG TGCTGGCCGA
    1081  TTTGCCCTTC GACATGGTTA GCATTAAATT TCTCCACTGG ACGCTGCACG ACACCGAGCA
    1141  GCTGCTGAGC GAGCGCGTCT ACTCCGCCCC GTGGACGCTG TTGCTGTTCT TTGCCGTGGC
    1201  CAGCTTCGCC TTTTCCTATC TGTTTCACAA CTTGCGCAGC TGGATGGATC GGTCAGTGGG
    1261  GACCCCCCAA CCGACGGATC GTCGCTGGGT GGTGGGGACA ATCGGAGCGG AGCTAGTGGC
    1321  AATGGTCGGG GCTGCCAGTG TTTCCCTTTC CGTTGGAACG GGTCTGTTCC TTGCGTTCAG
    1381  CTATCCCCTG CACACCGTGC TTGGCATTCC CCATCGGATC ATCGTGATCG GTGTGTTCTT
    1441  GTGTGTGGCG ACCGTTTTCT GGAAGTTTGA TCGTAAAAGC AATCGCCGGA TGCCAATGAG
    1501  GTAAGTGCGA CAAGGAACTT GTCCTTGCGT TCGACCTATT TAAAGATGAA GTTGCTTGCT
    1561  TTTTCGGGTG GGTTTTTAGT CAATCTCTGC TGGATCACGC ATTGAATGTC ATCACAGTGG
    1621  GACACTTTGT GCTGTACTTT GTGCTTGCAT TCGTCCTAAA CCCAGAAGAT ACTGTTTCGA
    1681  GCGGTAGACA TCAACCGATT GGCGATTGTC ACCATACGAC TGGTACCAAC GCTCCTCCAC
    1741  TCTGTCTCGA TACGTTCAGC AGGACCGATT ATGATTTCCA CTGCATTTCG AAGCCACCGA
    1801  ACGTTGGCAC TTATTGGTAC ACGGTCTGTG GCACGCCGTA CGAGTAAGTG GAAACATTTG
    1861  CCTTGATTTT TCAGCACCTT TCTATTTAAA TTGTCTTTTA TACCAGGAAT CGAGCGGAAT
    1921  TTTTGTTCGC AATGGCAGTT ATAACATTTA TTGCCGCGCT CGTACATTGG ACAATTCACT
    1981  ATGACTTTGA TGTGCGGTTT AAAATCTACG ACTTTGTCAA GCGAACTTCA TCGGTCAAGG
    2041  GTGGAAACAA TAAGAAGGTT TTGTAATATG TTTAACTGGT GTAATAATAC AAAAAAACAT
    2101  GTTTAAAATA TTTGGTGCAT TTTCACTGCA AACACACACG TGGGCTTTGT TTATTGTATA
    2161  TTTGCTTTAT CCGTATCAGC TGACTGGATT TGGGCGACAA AAAGAGGGGG TGGGGAAACA
    2221  AAAGTAAAAC TTATAATTTA TTGCACAAAC AGAAAGTACA ATGCTACACC TATCAACACA
    2281  GCCACTACTG GCGCTACGTA AACCGCCCAA TGCGTTCCCT TTTCCTCCTG CTGCACCTGG
    2341  TGCCGCT

//

LOCUS          39     2203 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      39
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          complement(join(500..1488,1565..1703))
                  /translation="MWKLFKDLRLIGGLARRSSTVPHTNPTLCRSLSSTAKSNSEPF
                  LSYEKDPAPAFLRDDVQQLLKSITRLELDKVFRKRSVKDNTVEYRFMTDEQLRRELL
                  HSIEQAQQMLQMPPVVQEQQDSCRVISKDGALKSLSTSKFVFTDITYGLKNSQRSIV
                  VRHPDGTLQEAPYEMRKRLNQIYFPLTGRSIHQPPMFEDEHLQRLLEEGKYEFVLDR
                  ACVQFEPYEKDYHHITAKVYQHINESRSFDALRSTRHFGPFVFFLAWHKLIDDLLLD
                  MIKQDYLRNGIELIVLYSTLHGMPIDEGVKQIALQTQMPSLFDAAEKSVDELKQDEQ
                  YLSIIEQFVNGQATKKVQLNLAVQAYREMAAEKLRLAQGIRRMHGES"
```

```
ORIGIN
       1  GTAAGCCCCC TAAAATTACC GATTTTTCAT TTAGGGCGGT GGCAACGCTG CTCCGATGCG
      61  ATTTTATAAA AGGTTCCTGC AAGGTTTTGT TCCACGGAAG GACCTTTGTT TACATTTTTC
     121  CGAAAAATTG CCAAGCTTTA TCGATCGGCG AAGGACGTTG GTTCTAGCGA CCTTGGCTAG
     181  AATGTCAGCG GCCAGGGCTT TTGCCTTCGG TCGGTGTGTC TCTGCAAATT GAGACGGAAT
     241  AATTCAACCG TTGGCGAATC TTCACGCTAA ATTAAGGGAT TATTTTCAAC AAAATATTTG
     301  CATTTCTATT TTCTATCGAT TTTCGATGCA TTCACAGCCG CTAATTTTCA TTCTTTCGTT
     361  TCTCCTGCAG CAATACCACC GTTCGGTACA GGAAGTCGCA CTTCGCGGAA CGTTATGTAT
     421  TTTCGTTTTA TTTTCTCGTT CTCAACCTAC ACACTACACA TTTAACCTCG ACGTCCCATT
     481  ACGTCGACGT CCCACCCATT CAGCTTTCCC CATGCATCCT CCTGATGCCC TGCGCTAGCC
     541  GAAGCTTCTC CGCTGCCATC TCACGGTACG CCTGTACGGC CAGGTTCAGC TGCACCTTCT
     601  TTGTGGCCTG CCCGTTCACA AACTGCTCTA TGATCGACAG GTACTGTTCG TCCTGCTTCA
     661  ACTCGTCGAC CGACTTTTCG GCCGCATCAA ACAAGCTGGG CATTTGCGTT TGGAGTGCAA
     721  TCTGTTTCAC GCCCTCGTCG ATCGGCATAC CGTGCAGCGT GGAGTAGAGC ACGATCAGCT
     781  CAATTCCGTT GCGCAGATAG TCCTGTTTGA TCATGTCCAG CAGCAGATCG TCGATCAGCT
     841  TGTGCCAGGC AAGGAAGAAC ACAAACGGAC CAAAGTGGCG CGTCGACCGC AGTGCGTCGA
     901  ACGATCGGCT TTCGTTGATG TGCTGGTACA CTTTGGCCGT GATGTGGTGG TAGTCCTTCT
     961  CGTACGGCTC GAACTGTACG CAGGCCCGAT CCAGCACGAA CTCGTACTTG CCCTCCTCGA
    1021  GCAGACGCTG CAGATGCTCG TCCTCGAACA TCGGTGGCTG GTGGATGGAG CGGCCGGTGA
    1081  GCGGGAAGTA GATCTGGTTC AGACGCTTGC GCATCTCGTA CGGCGCTTCC TGCAGCGTGC
    1141  CGTCCGGATG GCGCACCACG ATCGAGCGCT GACTGTTCTT CAACCCGTAC GTGATGTCGG
    1201  TAAAGACGAA CTTCGAGGTG GAAAGTGACT TGAGCGCGCC GTCCTTCGAT ATGACACGGC
    1261  AGCTGTCCTG CTGTTCCTGC ACGACCGGTG GCATCTGCAG CATCTGTTGG GCCTGCTCGA
    1321  TCGAATGCAG CAGCTCGCGT CGCAGCTGCT CGTCGGTCAT GAACCGATAC TCGACCGTGT
    1381  TATCTTTGAC CGAGCGCTTC CGGAACACCT TGTCCAGCTC CAGCCGCGTA ATGGACTTGA
    1441  GCAGCTGCTG CACATCGTCG CGCAGGAATG CAGGTGCCGG ATCCTTCTCT GCAAAGGTAG
    1501  CGAGGAATGG GTTGGATAGA TTTCGTGGCA AAGCACACAC ACGACGGCCG GGACCAACAC
    1561  TCACCGTAGC TAAGGAAAGG TTCACTGTTG GACTTGGCTG TGGAGCTTAG CGAGCGACAC
    1621  AGGGTCGGAT TGGTATGAGG AACTGTACTA CTTCGCCTTG CCAAACCACC GATCAGGCGG
    1681  AGATCTTTAA ACAGCTTCCA CATCCTTTTA GTTGGAGCAG ACTCTGTGGA AAGTATTGTA
    1741  TAATTTGTAG CACGTTTGTT TGGAGTCGGC AAATTTACAC TGCGAAGAAA AACATCACCG
    1801  AACGAAACGT CAAACAAAAC ATGGGTGAGC TGCTGACCAC ACGAAGCCAA ATTGGTTTGT
    1861  TTTTTCGTTG TGCATTTTTT ATAACGAATC GGTTGATATT TTGCCTATAA AAAATACGCA
    1921  AAATTATTTT CTCAAATAAT TCTACATTCC GGATAAAATT GTAAAGAGCA AAAATTAAAT
    1981  TACGCTGCAC TGTGGGACGC TGCACCAGGC TGACAGCAGA ACGCGCTGAC AGTTGAAAGA
    2041  CCGTTCCTTC AACGCGTTTC GAACCCCCCC CCCCCTCCTC CAGTAGGCAG CAGCACTGCC
    2101  CCGATAGTTG CCCGATTCTC GCTGGTGCTA ATCCGCGCTT CATTGGACGA TTCGAGTGCA
    2161  GAAAACGCTC AAGAGGTGAA CACACACATT TTTTGCGTCA ATT

//

LOCUS          40     2172 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      40
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          500..1672
                  /translation="MVGPSSQVSKILLSLLFLLIAFFAWMDVNLYIRIQDYPIRDTD
                  PFLLAANSTLLKPAPAGSHPSTVTSPLLAASSSMPLQLNRHIVRYEDVAWVNCDLNP
                  LCDVTVKAMMLDHTNHYLFAPIATIVDNVAGFSQGDLVTPNMISFFHVFVAIAAGRM
                  IASDSLGYRRIGVVLFQFRTFLDDLDGHVARAKKHIRGERSDIGSAGYYIDGICDGL
                  GCIALMIGVFVFLKNNPPRRGYTQLQSIIPVSESKSGSESGVIYKVKVTTKKVARKV
                  LCYTGMLVLSSTGWNRYIAIYQDMLERENVTPAQYLHQESVFRSTGFFVIAWLWRIF
                  NVHALLHFLLLSVFCDKLWEYLRMIQYAGYVALLVIISFAEMHLLGVQTFIYKSLTV
                  NNTSL"
ORIGIN
       1  TTGTTGTGAG GCACATCGTG GGTGTGCCCG GATCGCGAAG GAAAGGACGC TAGCGCGCAG
```

```
       61  TGACCGAACG AACAAACGGA CGGGAGCAAA ACGAACGGTT CTTAGTAGTA GTTGGTGTCG
      121  TTGTCGTTGT GTTGGTCGTG GTAGTAGTGG TGTTTTAGTC ATTGTCCACC TGTCAGTCAG
      181  GGACCACAGG GGGACCGTGT CAGAGTCGCG CTGGTCAGTT AGTCTTTTGT GGTCGATTGT
      241  CACACACACA ATCCTTGAAG TCTGGGAAGT GTACCCAATA GTTGTTAAGT TAAGTTGTTG
      301  TTATCACCTC CATAGGGAAG TGGGAATATA GTTGTAGGTT AACTACAGAG AGAGAGAGAG
      361  ATAAATAGGG AACGATAAGG TTAAACATAG TGAAATACAT CTAGGAAAGG TTTTTTTTTT
      421  TATTTTAACG TATTGTTTGA AGGACCAAAA AGTAATCCTA TTTATAGCAG AACGGTACAC
      481  AACACGAGTA AAAACAACCA TGGTGGGTCC GAGTTCGCAG GTCAGCAAAA TACTGCTGTC
      541  GCTGCTGTTT CTGCTCATCG CGTTCTTCGC CTGGATGGAC GTGAACCTGT ACATTCGCAT
      601  CCAGGACTAT CCGATCCGGG ACACCGATCC GTTCCTGCTG GCCGCCAACA GCACGCTGCT
      661  GAAGCCCGCC CCAGCCGGTA GCCACCCGAG CACGGTCACC TCACCCCTGC TTGCCGCGTC
      721  CTCCTCGATG CCGCTGCAAC TGAACCGACA CATCGTGCGG TACGAAGACG TGGCCTGGGT
      781  CAACTGTGAC CTGAACCCGC TCTGTGACGT CACCGTCAAA GCGATGATGC TGGACCACAC
      841  GAACCACTAC CTGTTCGCAC CGATCGCAAC GATCGTCGAC AATGTGGCCG GCTTTTCCCA
      901  GGGCGATCTG GTCACCCCGA ACATGATCTC CTTTTTCCAC GTGTTCGTCG CGATCGCGGC
      961  CGGCCGGATG ATTGCGTCCG ATTCGCTCGG CTACCGGCGG ATCGGGGTGG TGCTGTTCCA
     1021  GTTCCGCACC TTTCTGGACG ATCTGGACGG GCATGTGGCG CGGGCGAAGA AACACATCCG
     1081  CGGCGAACGA TCGGACATTG GTTCGGCGGG CTACTACATC GACGGGATCT GCGACGGGCT
     1141  CGGGTGCATT GCGCTGATGA TTGGGGTGTT TGTGTTTTTG AAAAACAATC CACCCCGACG
     1201  GGGATACACC CAGCTGCAGT CGATCATACC GGTGTCGGAG TCAAAGTCCG GGTCCGAGTC
     1261  GGGCGTCATT TACAAGGTCA AGGTAACGAC GAAGAAGGTC GCCCGGAAGG TGCTGTGCTA
     1321  CACCGGCATG CTGGTGCTCA GCTCGACCGG CTGGAACCGG TACATCGCGA TCTACCAGGA
     1381  TATGCTCGAG CGGGAGAATG TGACGCCGGC CCAGTACCTG CACCAGGAGT CGGTGTTCCG
     1441  GTCGACCGGG TTCTTCGTGA TCGCCTGGCT GTGGCGCATC TTTAACGTGC ACGCGCTGCT
     1501  CCACTTTCTG CTGCTGAGCG TGTTCTGTGA TAAGCTGTGG GAGTACCTGC GCATGATCCA
     1561  GTACGCCGGG TACGTGGCGC TGCTCGTGAT CATCAGCTTC GCCGAGATGC ATCTGCTCGG
     1621  CGTGCAGACG TTCATCTACA AGTCGCTCAC GGTCAACAAT ACCTCGTTGT GATTGCGCTA
     1681  AACATTGCGC AAAAAAGCGC CTTGCGGGCT TGTCAACGAT TGTTGTTTTA TTGACTCTTT
     1741  TTATTTCAGT TTCATCATTT ATTGTATTTT GTGTTTGTGT TCCCCATACA CGTGTGTGCC
     1801  CACGTGTATA TGTGTGTTTA TGATTATATT ATACAGGCGT GCGGTGTGTA AGTGTTAGAC
     1861  ATTTTCACCG CAAACCCCCT CCACCCTTTT CCGTCACTGA AACATTACGA AAACCACGAA
     1921  ATTATTATCA ACAAATGCTA AAACAATAAG CAGCCCTCCC TTCCTTATTA AGCCTAAGAA
     1981  GAAGCGTTTA AAGCAGGTGT GAATGGATTA AGTTGCGCGC TCGAACCTAG TAAGAGTATT
     2041  GCAATTGCAT TCGGACGGTT AGGACGTCCG TTATGTATGT GTGTGTGTGT TTTTGTTTTT
     2101  TGTTTTAGAT GAGATCATTG TATATCTACA TATTATGTAT GTGTTGTATA GTGTGATACA
     2161  AGCGCGAACC GG

//

LOCUS          41     2326 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      41
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          complement(join(500..1650,1727..1826))
                  /translation="MITIDRRLAAGVLLLVCLLPFLCHGKEAETQKDLMFLEPGRLV
                  SYTYDEWSITRPGLIVYCYPGQQESLLNAFCTVSIRMHCDHDNYNQYLGSSPAEVKK
                  HYKAEQSLFSFNVFASRKRRQVELEPFNQSCLGVVSSGKYDIEVLLNRFDLWRVGLL
                  VLGVLLFFIALPMSQNPLFYYTGGILIGLTSFALALVYMFSKMMPYRPLMLGVMVSG
                  WTLGFYGMQMMYENLRLIFVLYQHYVFWYLAIVGSISFFVCYRLGPPKDRRSKRVIQ
                  WTLQLIALTMVFFSSYQRQLSTGFVLIVIGAYYHATPLNALHWVRNRFRSRYPGKRR
                  LLTVEEFEEQGRIETDRALKDLREYCRSPECKQWTTMMRLRDPVRFASFVEGSPHVM
                  EEECFEYDTVHSSISNDDSDEVVELTEGEED"
ORIGIN
       1  TTCTTACGTT TATTTACATA TTTTATGCTC ACAAGACACG CGGCAGGCCC TTCGCGGCAC
      61  AAAAGCTGCT GGCACAAGTA GTTACTTTTA ACCTTTTTTA CCGAATCTAC CATCATCTCC
     121  GCTACGATCA AGTCTTATTG TCTCAACTAG AGGCATTTAG TATGGCCGGT TACAACAACA
```

```
     181  AACATTAATT AGGTTTATCA CCACACACGC CTGGATAAGT TTCACTATCA GTATGAACAT
     241  TTGCCATAAT TTTCGAGGCA GCTGCTGATG TATCTACAAT AATGATTACC TAAGCATCGA
     301  AGTAAAGCAT TTGACAATAT TGGTGCTGCC GATATCAGAA ACGATAGGTA AAGTTAAAAT
     361  GTAAGCCGAA AGTTCTATGA ATATTACCTC CAATTTCTAT AGAGTGTTCC GCTTTGCTTT
     421  GAGTGATGAT GTGTATCTCT ATTTGTTCGA GATAGAACCC CTAAACTGGC AAACTATTAC
     481  AGCGCCAACC GCCCCCCGAT CAGTCCTCTT CGCCTTCGGT CAATTCCACC ACCTCGTCCG
     541  AGTCGTCGTT GCTGATGCTC GAGTGTACCG TATCGTACTC GAAGCACTCC TCCTCCATCA
     601  CGTGCGGCGA TCCCTCGACG AACGACGCGA ACCGTACCGG ATCCCTCAAC CTCATCATCG
     661  TCGTCCACTG CTTGCACTCC GGGCTGCGGC AGTACTCGCG CAGATCCTTC AGCGCACGGT
     721  CCGTCTCGAT CCGGCCCTGC TCCTCGAACT CTTCCACCGT CAGCAACCGT CGCTTGCCCG
     781  GATACCGGCT GCGGAACCGG TTGCGTACCC AGTGCAGCGC GTTCAACGGT GTCGCGTGAT
     841  AGTACGCCCC AATCACGATC AGCACAAAGC CGGTCGAAAG CTGGCGCTGG TAGCTGCTGA
     901  AGAACACCAT CGTCAGTGCG ATCAGCTGCA GCGTCCACTG GATGACGCGC TTGCTACGGC
     961  GATCCTTCGG TGGGCCAAGC CGATAGCACA CGAAGAAACT GATCGAACCA ACGATCGCCA
    1021  AATACCAAAA CACGTAGTGC TGGTACAGCA CGAAGATTAG GCGCAAATTT TCGTACATCA
    1081  TCTGCATGCC GTAGAAACCG AGCGTCCAAC CACTGACCAT AACGCCGAGC ATAAGGGGCC
    1141  GATACGGCAT CATCTTGCTG AACATGTACA CCAGCGCAAG GGCGAACGAG GTTAATCCAA
    1201  TAAGAATACC GCCGGTGTAG TAGAACAGTG GGTTTTGGCT CATGGGCAGG GCGATAAAGA
    1261  ACAGCAGCAC GCCCAGCACG AGCAGTCCAA CGCGCCACAG ATCGAACCGG TTCAGCAGCA
    1321  CCTCGATGTC GTACTTGCCG GACGAAACCA CGCCGAGACA GCTCTGGTTG AACGGTTCCA
    1381  GTTCCACCTG CCGGCGCTTG CGGGACGCAA ACACATTGAA GCTGAACAGG CTCTGTTCCG
    1441  CCTTGTAGTG CTTTTTCACC TCCGCCGGTG AGCTACCGAG GTACTGGTTG TAGTTGTCAT
    1501  GATCGCAGTG CATCCGTATC GAGACGGTGC AGAACGCGTT CAGCAAGCTT TCCTGCTGCC
    1561  CCGGGTAGCA GTAAACGATC AGCCCGGGCC GGGTAATGCT CCACTCATCG TACGTGTACG
    1621  ATACGAGGCG TCCCGGTTCG AGGAACATTA CTTTGGGAAG AAAGAAGAGA AGGATTTAAT
    1681  AAAATTGAAA ACAAACCTGT AACATGACTC CTTCACTGCC ACTTACAATC TTTCTGCGTT
    1741  TCCGCTTCCT TGCCATGGCA GAGGAATGGT AACAGACAGA CCAACAGCAG CACACCCGCC
    1801  GCCAGGCGTC GGTCAATCGT GATCATCGCT GGGTATGCGC TTTGCTACCT GCGATTTTCT
    1861  CCTCCTTTGC ACCACCAAAC ACAACCAAAG AAAAGTCCAC GAAATGGTTC ACTAAACAAT
    1921  GCACAGGACA AGCCAATTCT GGGTTTCACT TTCACCTACT GCAGTGCATT TTAAACTAAC
    1981  AAAACCGACT AAAAACCAAG TTTGGCCATC GTCGCGCTGA AAATGATGCC GACCGGTGAG
    2041  GTAAGCGACA CAAACATGTT CTGGCGAGTG CAGCGCCACC GGTAAACAAC GGAAAATAGC
    2101  GGCTTTTTGT TGTGACAGCT GGCAGGGCGA CCAGAAAAAG GGGGTTATTT GTCATTAGAC
    2161  CAGTCGCACA AGGTGTGTAG AGGGTTATTT TGACAGGTGG CTGAGTAAAA TTTGAATCTT
    2221  TTGCCAATCA ACGGAATTTT GTTCTGTGAG ATGCTTTTTA TTCTAAAATG TAGCGAATTT
    2281  TCATAATCGT ATTCAATTGT GTCTGCTTTT AGAAGGTAAG ACTTCT

//

LOCUS          42     2364 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      42
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          complement(500..1864)
                  /translation="MFWGFCEALNFIQQYQKKLSADESLPESLNILLFGSGDPRHIL
                  ATASQLFLHPDLKVNVYLAEGCIELLARHMVLLAVAFEDPELLSLKGKTHLFMDLFG
                  NNLIRPFSSAYLSSKAKELTDIITDAEYAQRQAPMFNYETLRYKERDQLENVFRFWT
                  NAPEHVFNIARYWEDRLRVQLGERYDHRNGAFDWDLQMRLRENGAKQVCPQEYKHWR
                  ETGIAFTFPEYEQSDPNKTFAVGLVRNGKGFLHRGSVGDNMTGPYIAFGHKCAEERL
                  SRSKHGVNDFRSTDVTERNVLQIVYEIQNRKPYCFDPKDIHQYGAHQLDTGKNLNKH
                  EARTESAEAIHYNKPLLRCENLTIHFLSVEDVLRMHEMERFAGKFDVVFVASNYLGL
                  VKDGFSRAWKDNCLVCFETRQLTVFSKEEIKEHTDKIKAFAQKESLAAVTNFSINKN
                  HSVLLYKRAGNK"
ORIGIN
       1  CCCAAAGTTC TAACTTCCGT TCTGAGCCAG GCGAATACGG GAGGCGTCGA AAACACTATG
      61  TACTTTACAC ACGACCACGC ACAATATGGA AAAGGTTCTG CTTGACTTTC CTCTACTTTT
```

```
     121  CCCTCCCTTT TTCTAGCCTG CTCACACACG AAGGAGCACT GCTGGCGTTT TCCGGATTCG
     181  GCGACAAGGA CGCCACGGTC ACAGCGGCGA TCGCGAGCAA CATCTGGTCC GTTTACGAAA
     241  AGAATGGCCG CAACGCATTC CGAGAAGATC GACTCCAGTA CGTGTTTCTG CAGTGTGAGA
     301  ACGGTAACGT CGTCATTACG CAGGTGGCCA ACCTGCTGCT CTGTCTGTAC GCCCGGGACA
     361  GTGTGGGGCT TGGCATACTG AAGGAAAAAG CGCAAGCCCT TGCCACCTAC CTGGAGGGAC
     421  CGCTGAAAGA GATCGCTACA TCGTAGGAGG GCCAAACTTT TGTTATATTC ATATTTATTA
     481  GAGCTTAATA CATTTGCGTT CACTTATTGC CGGCACGCTT GTATAAAAGG ACGGAATGAT
     541  TTTTGTTGAT GGAAAAGTTT GTCACCGCTG CGAGGGATTC TTTCTGCGCG AAGGCCTTAA
     601  TCTTGTCGGT ATGTTCTTTA ATTTCCTCTT TGCTGAATAC GGTCAGCTGA CGCGTTTCGA
     661  AACACACGAG GCAGTTATCC TTCCAGGCGC GGCTAAAACC GTCCTTAACC AAGCCCAGAT
     721  AGTTGCTTGC CACAAATACG ACATCGAACT TGCCAGCAAA CCGCTCCATC TCGTGCATCC
     781  GCAAAACGTC CTCCACAGAA AGGAAGTGAA TGGTTAAATT TTCGCACCTC AATAGAGGTT
     841  TGTTGTAATG TATGGCTTCC GCTGATTCGG TGCGTGCCTC GTGTTTGTTG AGATTTTTCC
     901  CCGTATCGAG CTGGTGCGCA CCGTACTGAT GGATGTCCTT CGGATCGAAG CAGTACGGCT
     961  TCCGGTTTTG AATTTCGTAC ACTATCTGCA GCACATTGCG CTCCGTCACA TCGGTCGAGC
    1021  GGAAATCGTT CACACCATGC TTGGAGCGGG AAAGCCTCTC CTCGGCACAC TTGTGCCCGA
    1081  ACGCGATGTA CGGACCGGTC ATGTTATCAC CAACCGAACC ACGGTGCAGA AACCCTTTCC
    1141  CGTTGCGCAC CAACCCGACG GCAAACGTTT TGTTCGGATC GCTCTGCTCG TACTCGGGAA
    1201  AGGTGAAGGC GATTCCCGTC TCCCGCCAGT GCTTGTACTC CTGGGGACAC ACCTGCTTCG
    1261  CACCGTTTTC CCGCAGTCGC ATCTGCAAAT CCCAATCGAA TGCACCGTTG CGATGATCGT
    1321  ACCGCTCACC GAGCTGCACG CGCAACCGAT CCTCCCAGTA GCGTGCGATA TTGAACACGT
    1381  GTTCCGGCGC ATTCGTCCAG AAGCGAAACA CGTTCTCGAG CTGGTCGCGC TCCTTGTATC
    1441  GTAGCGTTTC GTAATTGAAC ATCGGAGCTT GGCGCTGCGC GTACTCCGCG TCGGTGATGA
    1501  TGTCGGTCAG CTCCTTCGCC TTGCTGCTCA AGTAGGCACT GCTGAACGGG CGGATGAGAT
    1561  TGTTTCCAAA GAGGTCCATA AAGAGATGCG TTTTGCCTTT GAGTGAGAGC AGTTCCGGAT
    1621  CTTCGAAAGC GACCGCAAGC AGCACCATAT GCCGTGCCAG CAGTTCGATG CATCCTTCCG
    1681  CCAGATACAC GTTGACTTTC AGGTCGGGAT GAAGGAACAG CTGCGATGCG GTCGCCAGGA
    1741  TGTGTCGCGG GTCACCGCTT CCGAAAAGTA AAATGTTCAA GCTTTCCGGT AAAGACTCAT
    1801  CGGCTGATAA TTTCTTCTGG TATTGCTGGA TAAAATTAAG AGCTTCACAG AATCCCCAAA
    1861  ACATCGTGAT CGCACGTGAA ATGCACCGGT TGTACTGTGA GAGAGTACTG CTGCGCTAGT
    1921  GTTGCTGCAA ATAACAAGTC ACTCGGCCGT TGCCATCGGA GACCCCGTTT CCATGGTAAC
    1981  CGTCGAGCGC TTCTATGATT GGGTTTCGGT AGCATTTGTT CTGTTTCACA AAAAAAAGAA
    2041  AATGTTTCAA ATTTAAAAGA ATATATTTTT CTTCGCATTC TGCATAATAA TAGTTGTGAA
    2101  ATCAGATTCT GCATTGATCG TATAATATTT GCCCTAATTT TTACTTTTCA CGTGATGTTA
    2161  AACAAAAAAA TAATCTCTCA TCGCTTCGAT TGCGTAGTTC TTACAAGCTA TTCAAACAAC
    2221  GCGTCAGCAG TTAAAAGTCG CAGTTAACGG TGCGCAGTAA ACATGTCTTG TGTGGTGGTG
    2281  GTGGCAGTTG AAGGTAGCAC AGCTTTTGAT TCGATATCGA TTGATTGTGA AGCGCGGAGC
    2341  TTCGTTTTTA GCCTTCCATC TTGA

//

LOCUS          43     2400 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      43
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(500..1900)
                     /translation="MEAPPSVDDSIERGKSTECDGEKVEETQTKAASAATSQMPTES
                     GAGEKSDTPQPSGSEESKDSANGKATADHATSSSSSSSSSSSSSSKGVPPTSRTFDK
                     VPIFVVEDHHEVLTFLYRCLGSKHLPLQGNRIVHFDSHPDMCIPKHMPANYVFNKDD
                     LLDSISIENWLMPTVFAGHVERIVWIKPPWSEQIPKGKYAFHVGEFEGSIRTDSTLE
                     YFVSEGCYQPEDKLENRKRLQLEVCSVEEYPVAEDTDLAQGYILDIDLDYFSTHNPF
                     LKIYDRVQLYEKLKDIFVSPELADDAENDLAKLQRVARDREEKLEFLESIFLYLEEV
                     GNLKHFLVDYQQEISEEYAGLLEKVTVLVRTLKREYKEEEIDWSMIYDAGCTCDVTD
                     LPHHESSREEIETMVGQLERFLEKVPCPPVVITVSRSSEDDYTPAGQVEMIQEMVLA
                     ALTKCLRAELDTPILYYKDQELNL"
ORIGIN
```

```
       1  CGGGTCGGGA ATTGGGGGAA GCACTCTGAA ATGTCCCTTC TGTCCGAGTG TCCGGAAGAT
      61  GCTTCGTGAT TGTTTTTTTA TTACTTTTAT TCCTATACAT TATCGAATAT TTAGCGAGCC
     121  TTAAATGAAG GGATTTGTAT CCAATAATTA ACGAAAAATC GGTTATTTCT TCTCCCAGAA
     181  AATGTATCAC ATGATTGTAG AAGGGTTGAC ACGTAGCTAC TGCTACAAAT GTATTCATGC
     241  TGCAAAGAAA ATGCTTTCCT TTTATCCTTA GAGACAACTA GGCAGCTCTC TATGGGCTTC
     301  TTTTACTGTA ATCTGACTAA GTCCGGTCTT GCACGAGTTA CCCTTTATCT ACATCCACAA
     361  AATACGTCAT CATTATCCTC CCGTGAAATG AAAAAAAAAC ACAACAAAAG TTTCGTTTGT
     421  TAACTTAAAT AGGGAGAAGG GGGCATTGGT GTGTATGGTG GTGGTGTTTT TTTGTTGTCG
     481  TTTTCTTTTA TTATTGTCCT TACAAATTAA GTTCCTGATC CTTATAGTAC AGTATCGGGG
     541  TGTCGAGCTC CGCCCGCAGG CACTTTGTAA GCGCTGCCAG CACCATCTCC TGTATCATCT
     601  CCACCTGCCC GGCCGGTGTG TAGTCGTCTT CGCTCGATCG CGACACCGTG ATGACGACCG
     661  GCGGGCACGG CACCTTCTCC AGGAAGCGCT CCAGCTGGCC GACCATCGTT TCGATCTCCT
     721  CCCGGCTCGA CTCGTGGTGC GGCAGATCGG TCACGTCGCA CGTGCACCCG GCATCGTATA
     781  TCATCGACCA GTCGATCTCC TCCTCCTTGT ACTCGCGCTT CAGCGTCCGC ACCAGGACCG
     841  TCACTTTCTC CAGCAGCCCG GCGTACTCCT CGCTGATCTC CTGCTGGTAG TCCACCAGGA
     901  AGTGTTTCAG GTTGCCCACC TCCTCCAGGT ACAGGAAGAT CGACTCGAGA AACTCCAGCT
     961  TCTCCTCCCG GTCGCGCGCC ACGCGCTGCA GCTTGGCGAG ATCGTTCTCC GCGTCGTCGG
    1021  CCAGCTCGGG CGAGACGAAG ATGTCCTTCA GCTTCTCGTA CAGCTGCACC CGGTCGTAGA
    1081  TCTTCAGGAA GGGATTGTGT GTGCTGAAGT AGTCCAGATC GATATCGAGG ATGTAGCCCT
    1141  GGGCCAGGTC CGTGTCCTCC GCCACCGGGT ACTCTTCGAC CGAGCACACC TCCAGCTGGA
    1201  GGCGCTTGCG GTTCTCCAGC TTGTCCTCCG GCTGGTAGCA ACCCTCCGAC ACGAAGTACT
    1261  CCAGCGTCGA GTCCGTCCGG ATCGAGCCCT CGAACTCGCC CACATGGAAC GCGTACTTGC
    1321  CCTTCGGGAT CTGCTCCGAC CACGGCGGCT TAATCCACAC GATGCGCTCC ACGTGCCCAG
    1381  CGAACACGGT CGGCATCAGC CAGTTCTCGA TGCTGATGCT GTCCAGCAGG TCGTCCTTGT
    1441  TGAACACGTA GTTGGCGGGC ATGTGCTTCG GGATGCACAT GTCCGGGTGC GAGTCAAAGT
    1501  GCACGATCCG GTTGCCCTGC AGGGGTAAAT GCTTCGAGCC GAGGCAGCGG TAAAGGAACG
    1561  TGAGCACTTC GTGATGGTCC TCCACCACAA AGATCGGTAC TTTGTCGAAC GTTCTGCTCG
    1621  TCGGCGGAAC ACCTTTGGAA GATGACGATG ATGATGATGA TGATGATGAG GAAGAGGATG
    1681  TCGCGTGATC TGCGGTCGCT TTGCCGTTGG CGCTGTCCTT CGATTCCTCC GACCCGCTCG
    1741  GCTGCGGTGT GTCACTCTTT TCCCCCGCGC CGGATTCGGT TGGCATTTGG CTGGTCGCTG
    1801  CGGAGGCTGC TTTTGTTTGT GTTTCTTCCA CCTTCTCCCC GTCGCATTCC GTCGACTTCC
    1861  CCCGCTCGAT AGAGTCATCC ACACTCGGGG GAGCTTCCAT CGTGGTTGGT TGTGTATTCA
    1921  ATCGTACAGA GTATAATCCG TAAAGAAAAT AATTCCACTT CACAACACCG CCACTGCGCC
    1981  GCACACTCCG GGATCAAGGT AACCCGTTAC CAGCAGCCAG ACATACAGAC GCACACACAC
    2041  ACGCACGTAC ACACACACTC GGGGAAGGAG GGCCCACACT GCTGCGGCTG CCAATCAATT
    2101  TTCACTCATT AGCCCCTGCT AGAGCGAGCG CGAGCGGATA CTGCTACTAC CCGGGGACAA
    2161  TCGCGCCGTC CACCAGGGCG TCCTGTATCG GTGCGCGTCC GTCCGCTGGC CCACGATTCC
    2221  TTTCCGGGTG GGCGTACCGT AGTCCGGGGT GCGATGTGTG CGTGTGTGTA GAAAATAGTC
    2281  ACGAAACTTA TACGCGCTGG AGCTGCTCCA CACAACTTGC GCTTGCTTCT TTTCGGGGCC
    2341  CTACGCACGA TGACGGTTCA TTACGACACC CTTCTCAAAA TATACGCGCT CTCAACAACT
    2401

//

LOCUS          44     2412 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      44
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          complement(500..1912)
                  /translation="MDPADYRALHESQMENNNGTTAVNVFMHIAPPVFSVFHTVQLI
                  AVASITHVPLRFLVEFCLIVLPFVLLVTVLHRLALNIALTLFVLSAISAVHQIRNRL
                  HLMPFVQVPGKSPQFLTCVRALVNLTTAVCILAVDFHCFPRELAKTETFGYGLMDIG
                  VGLYVFSNGIVYKVRPAEERRLNTARVWKVLISTAPLLLLGFGRFFLTNELDYQLHV
                  SEYGVHWNFFITLACVKFAGTLTMDLVHDPEVIKFVAITLLCCHEMMLHLGVSRYVL
                  SADTQRTTFFAANREGIVSLPGYLALFLASIYIGWVVRPTEEITVAKKFLRRAGKVA
                  IMAAVCWKMIYVCEDMFGISRRLANMGYCFWILAIGASVTALFMFFELFVYFVRFEE
```

```
                PKSPDQIMKTDDFCIPYVPLVLQAINANGLFFFLIANLLTGLVNLVFQTFLLGNGAA
                LTIIGYYMFMLCVAVCFLYVNNVKLKIW"
ORIGIN
       1  GCTTCAGCTC GTTCTATTGA CGAAGCGAAA CGGTTTTATC GTACTTAAAT GCCTTAAGCT
      61  TTACACAAAT GATGAGATTG GGTTGGATGG AGGTGGATCT GATGGCTTTA CGATTAATTA
     121  TACGCTCACG CAGTCTTAAC TTAACGACAA AAGGGGAATG TGAACAGTAT TTTTTATGTC
     181  CAAAAAAACC GCAGAATGTA ATATCAATTT ACCAATAACG AATGTGAGAT TAAGCTTAAC
     241  TTTAATACTA GATTGTTTTA TTTCTCATTT GTTTTCAAGG CTTTTGATGA CGAGATGATG
     301  AGTTTGTTTT CTTGCTGTCT CCCTTCCCCC GAATAGTGTT CAATATTATG TCCTAATGTA
     361  ATTTTTGCTT TAATTTTAGC TTGGTAAACC ACCCATCCTT TTTGCTTAAT TTACAACGAT
     421  GGCTTTCGAT GACTAGAGAT CGATTTTGGG AACGGATCGG AGTGCCTGTT TGCACCTGGA
     481  ATTCAACGCC TTAGTGCGTT TACCATATCT TAAGCTTTAC ATTGTTAACG TACAGAAAGC
     541  AGACGGCCAC ACACAGCATA AACATGTAAT ACCCGATGAT CGTGAGGGCG GCTCCATTCC
     601  CTAGCAGAAA CGTTTGAAAC ACCAAATTCA CCAGCCCCGT AAGCAGATTG GCGATCAGAA
     661  AGAAGAACAG CCCATTCGCG TTGATTGCCT GCAGCACGAG CGGAACGTAC GGGATGCAGA
     721  AATCGTCCGT CTTCATGATC TGGTCGGGCG ATTTCGGCTC CTCGAACCGG ACGAAGTACA
     781  CGAACAGCTC GAAAAACATA AACAGCGCGG TGACGCTAGC CCCGATCGCC AGTATCCAGA
     841  AGCAGTAGCC CATGTTGGCG AGCCGGCGCG AAATCCCAAA CATGTCCTCG CACACGTAGA
     901  TCATCTTCCA GCAGACCGCT GCCATGATGG CCACCTTGCC CGCCCGTCGG AGGAACTTTT
     961  TCGCAACCGT GATCTCCTCC GTCGGCCGCA CAACCCAGCC GATGTAGATG GAGGCGAGGA
    1021  AGAGGGCGAG ATAGCCGGGC AGGGACACAA TTCCTTCCCG ATTGGCGGCG AAGAAGGTAG
    1081  TTCGCTGGGT GTCCGCACTC AGCACGTACC GGGACACGCC CAGATGGAGC ATCATTTCGT
    1141  GGCAGCAGAG CAGCGTAATG GCGACGAACT TGATCACTTC CGGATCGTGC ACCAGATCCA
    1201  TGGTGAGCGT GCCGGCAAAC TTGACACAGG CGAGTGTGAT GAAGAAGTTC CAGTGCACCC
    1261  CGTACTCGGA CACGTGCAGC TGATAGTCCA GCTCGTTGGT GAGGAAGAAG CGGCCAAAGC
    1321  CGAGCAGTAG TAGGGGAGCC GTGCTGATTA GAACCTTCCA CACTCGTGCC GTATTCAGTC
    1381  GTCGCTCCTC TGCAGGTCGC ACCTTGTAAA CGATCCCATT GCTGAACACG TACAGCCCCA
    1441  CTCCGATGTC CATCAGCCCG TAGCCGAAGG TTTCCGTTTT GGCAAGCTCG CGAGGGAAGC
    1501  AGTGGAAATC GACCGCCAGT ATGCAGACGG CCGTTGTTAG GTTTACGAGC GCTCGCACAC
    1561  ACGTCAGGAA CTGCGGTGAC TTCCCGGGCA CCTGCACGAA CGGCATCAGG TGCAACCGGT
    1621  TCCGGATCTG ATGGACGGCC GATATCGCCG ACAGGACGAA CAGGGTGAGC GCGATGTTCA
    1681  GTGCAAGCCG GTGCAGTACG GTCACCAGCA GCACGAACGG CAGCACGATG AGACAAAATT
    1741  CCACCAGAAA TCGGAGCGGA ACGTGCGTAA TACTGGCCAC CGCAATGAGC TGAACCGTGT
    1801  GAAACACACT GAACACGGGC GGTGCGATGT GCATGAACAC GTTCACGGCC GTCGTCCCGT
    1861  TGTTGTTTTC CATCTGGCTT TCGTGCAGGG CACGATAATC GGCTGGATCC ATCGTGTTGG
    1921  GTGCCTTTTT TTTGTTTGGA ATGATGTGAA GTATCCGTGC CGGAAAAAGA GAGGAAGGAA
    1981  AAAAACCGAA TCGCGTACGC GTAATTTTGC AAACGACGAC GCGGATTAAA ACGTAAATAA
    2041  ATAACAAACA TCTGACAAAC GTCAAACCGC TTTCGGATTT GACGTTTGTA GATTGTGGCG
    2101  GGTTTTTGCT GCATGGGCAC CGCTTTTGCT TCGAAAGCTA CGAAACACGT GCGCAATGCG
    2161  TTGTTTACCT TTTTAAAATT CATTTCTCTC ATTTATTATG AAATAAATTT CATTAGAGTA
    2221  TTCTTCTTAA ATTATTTATT TCCTTACACA CTACCAGTTC AGCTTAACCT AACAGAATCA
    2281  CGATGATGAT AAATGAGCAG TTATCAAGTC GCGCACATTT CGACATCGTT GCGCCCGAAA
    2341  AAATGCTTCT CCACCTACTC AAGGGCTCAA CCCTTGACCA CGGTTTTGTA CACGCCCTTG
    2401  CTGACCTTGA CC

//

LOCUS          45     2548 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      45
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          join(500..1933,2013..2048)
                  /translation="MAEEKFVLPSPRFQDEVKKGLDSFSRHLNPHFSTPNGMLLPSS
                  KQPPPMFGPRGSTVPAQFLNNPRKRPASGPHRYPGPGTYRQNQPGGINQNQRNNQGA
                  GFKNKAGGPNHRSPRPPNNGLEKDVKPELLVLDWKLWCEGCDVNCRSEEEFEQHKAN
                  HAPCTVPGCTFVGHPMIMKRHGRQVHSDGGCTKDTIGPSAEEIEQWKEERRKRYPTK
```

```
                QNVILRQQAQEARFNRGERIEEKKDRFPKRNAPVDDDSSMNGPNTTPNGRNANRSKR
                KRRAKKPAFVAPVEEKTLDRIPFKGTSSLKDYKDPPKNALSMLGDYGTDTESSCDEE
                EQASPVVAQKKSPEKSSVKEPTQKPERKLSEGEIEDDGDDAPSTSSAPAAQPVLAPN
                ESQSKDAVMQSTTTASKSVDCAVPGGSSSTVVKRPPKQQNKPPRQNNFKDKRARNPN
                RPLLDYSKLRRSNQNTMLEKLLDADIRHERNVLLQCKKQMYQLYKSP"
ORIGIN
       1  CTTTTCCTTC GCCGGTTCGT CCGCTGGTTT CGTCGCGGGG TTTTCTTCGA TATCGGGCAT
      61  TGCAAGATTA ATCACTTTTA TCACTCGTAC AGCTTAAAAT CTGTGAAAAA TACAGAAATT
     121  TTCACTTTTT CACAGTAAAT TCACGCGTAA ACAAACCGAT TGCCGTCCTT TTCCTCGCGC
     181  TTCTTCTTCC TCCAGTCACA AGTGCTGTCA GATAACTTTG ACAGAAAATG TCATGATTTC
     241  ACTCGATTTC ATCGGGAAAT TACCAGTAAA TTAGATGAAT AATTTGATTT CTGTGTGTTT
     301  TTTAGGTCGA GTAATATTTT ACCGAAATGT GTATTACGCT AATTAAATCA AAATGTAATT
     361  TGTCTTATTT GCAAAAAAAA AGCAAAGGAA AGGATATTCG TTATCGCTAC TGGCAACACT
     421  GCACGGCACG ACAGAACGCA GTTAATTTGT AAACAACTTT CGGTGCATGC AAGCAAAACG
     481  TGCTCCGCCG TACATTGCGA TGGCTGAGGA AAAGTTTGTT TTACCATCGC CCAGGTTTCA
     541  GGACGAGGTG AAAAAGGGTC TGGACAGCTT CAGTCGGCAT TTGAATCCAC ATTTCTCTAC
     601  GCCCAATGGA ATGTTGCTAC CGAGTAGCAA ACAACCGCCT CCAATGTTCG GACCACGTGG
     661  TTCCACCGTT CCGGCCCAGT TTCTGAACAA TCCGCGAAAA CGGCCAGCGT CCGGTCCGCA
     721  CAGGTATCCC GGTCCGGGTA CCTATCGCCA GAACCAACCA GGGGGAATCA ACCAGAACCA
     781  ACGAAACAAT CAGGGGGCAG GGTTCAAAAA CAAAGCCGGA GGGCCAAACC ATCGATCACC
     841  TCGACCACCG AACAATGGAC TGGAGAAGGA CGTGAAACCC GAACTGTTAG TTTTAGATTG
     901  GAAGTTGTGG TGTGAAGGGT GCGACGTGAA CTGCCGATCA GAGGAAGAGT TCGAGCAACA
     961  CAAGGCCAAC CATGCACCGT GCACCGTGCC CGGGTGTACG TTTGTCGGCC ATCCGATGAT
    1021  CATGAAGCGA CACGGTCGAC AAGTGCACAG CGATGGCGGT TGTACGAAGG ACACCATCGG
    1081  ACCGTCAGCG GAGGAGATAG AGCAGTGGAA GGAGGAGCGC CGGAAACGGT ATCCCACGAA
    1141  GCAGAACGTG ATCTTACGCC AGCAGGCACA GGAAGCTCGG TTTAATCGGG GTGAACGTAT
    1201  CGAGGAAAAG AAGGACCGTT TTCCCAAGCG CAATGCGCCG GTGGATGACG ACTCTAGCAT
    1261  GAATGGTCCA AACACGACAC CGAACGGACG GAATGCGAAT CGTTCAAAGC GCAAAAGACG
    1321  CGCAAAGAAA CCTGCCTTCG TCGCACCGGT AGAAGAAAAA ACATTAGACA GAATACCATT
    1381  CAAGGGCACT TCGTCGTTGA AGGATTATAA AGACCCACCC AAAAACGCAC TGTCAATGCT
    1441  GGGAGACTAT GGCACAGATA CAGAATCTTC CTGCGACGAA GAAGAGCAAG CAAGTCCGGT
    1501  AGTTGCGCAA AAGAAGTCGC CAGAAAAATC GTCCGTCAAG GAACCCACTC AGAAGCCGGA
    1561  ACGTAAACTC AGTGAGGGGG AAATTGAAGA CGATGGTGAC GATGCACCCA GCACAAGTTC
    1621  GGCCCCAGCC GCTCAACCAG TGCTTGCTCC TAATGAATCA CAAAGCAAAG ATGCAGTAAT
    1681  GCAGTCGACT ACCACAGCTT CAAAAAGTGT CGACTGCGCC GTTCCGGGTG GGAGCTCTTC
    1741  AACCGTTGTG AAGCGTCCGC CCAAACAGCA GAACAAACCG CCACGACAAA ACAACTTCAA
    1801  AGACAAACGG GCTAGGAACC CGAACAGACC ATTGTTAGAC TATTCGAAGC TCCGACGGTC
    1861  CAATCAGAAT ACCATGTTGG AGAAGCTGCT CGATGCCGAC ATACGCCACG AGCGGAATGT
    1921  GCTGCTGCAG TGCGTGAGGT ACGTTCTGGT GAACAAATTT TTCGGCATTG GACAGCCAAA
    1981  AGAGGTGAAG GAATCCGTTG CAATTGTGAC AGAAGAAGCA AATGTACCAG TTGTACAAGT
    2041  CACCGTAGAA ACAACTGCTC CAGTTGTAGC ATTGAACGAA GAAGCAAATC CTCCAGTTGT
    2101  GACAATTACC GAAGAAGTAA ATGTACCAGT TGTTGCAGTA ACCGACGAAG CAAATGTTCT
    2161  AGTTGTAGGA TTGAACGAAG AAACAAATTC TGCAGTTGAG GCAATCACCC AAAACGCGAA
    2221  TACTCCAGCT GTAGCAGTGA CCTTAGAACC AATTGATCCA GTTGTAACAA TGACCGAAGA
    2281  AGCAAATCCT TCAGTTGCAA CAATGACCGA AGAAACGACT GCTCCAGTTG CAACAATGAC
    2341  TGAAGAAACG ACTGCTCTTG TTTGCACAGT GACCGAAGAA ACGAATGATC CTGTTTCTAC
    2401  AGTGACCGAA GAAGCAAAGG AATGAAATAA AAGTGTTTTA ATGCACACGA TTTATTCAGA
    2461  AAATGTCTTA ATCTCTAGCA TCTATAGCAT AACTCCCGCC GTGCTCGATG ACGTGTTCTG
    2521  TCCGACACGA TTACTTTGCA TCCGGTTC

//

LOCUS          46     3316 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      46
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
```

```
     CDS             join(500..1826,1938..2086,2802..2816)
                     /translation="MQTSQQLSSGHRTANEDEDDEDGESELLDLAIVGAGISGLMAA
                     KTISEKRADIRFRLFEKSTHPGGTLDGLKTRWITPHHYHAMNLCRELQIPLGTVWRQ
                     SEVSEARRSLVSIGPFRKRPSGSQGANSFMSSLRDLLVRLESTRFMAELDCLCTVKY
                     IESNAQNMECFLQRKLLFEASRRFFRFLIKIGTGFYPSELTVAACLRLFRSMSSVRD
                     LYDMLTLQNGHLQPASSPNWDVLIERLVARVGPEHVTYSTNIVQVEITSEQRSDIVS
                     LTDGTGRRWRARFVILAVSCLDLVQISILPEGPLYFHQPDIQLGLWSMANFTVRYPA
                     PYWRDHGYTGSIFCPAQCLICYESGRNQLSGTYFSPLRGVLNDTERHLIRDTILRLL
                     RTNFACRTMQKPLEFGLELHPIPFYFDVFPTFERCIIFSSTNVSCWYRGFINGSIQG
                     GVRAAILALLEIRPQTITFREVTDMQCMHFKYFRQRSSYERVWYSLNLASLKVP"
ORIGIN
        1  TGAGAATGAA GTCAATAAAA CATGCACATA CAATAATTCT TGTAAGACCT TGACGCCCAA
       61  AGTAGCTTAC ACAGGGTCTC TATACAACGC CACTAGATAC TTAGTCCTTG CCACGGGAGA
      121  ACAATCCTAG TGGGAATAGC ACCCTGTACA GCTTTCAAGA ATATGGTGCG TCATGCAACA
      181  TCCCAAGGAT AGAACACATT TCAGCATAAA TTGAATTCAA TTGAATTTGA AGAACTTCTA
      241  CCAATTATTC TGTCACACTT CCCTGAAGAA TACCACATCA AACCAAGAAC AGACGCCACA
      301  TCAACGTCAA CGAACAGCTT GGCAACAGTT TCAAACACGT GCCCAATCAA AGTAACCATA
      361  GACAACAGTG ACATTTTTGT TCTGCTTCCT CCCCGCTGTC AGCTACATTA AATTCTTTCT
      421  GTACACAAAG TTTAGTTCGT CCGTTTTTTT TTGTTATTGT TTTGCTATTG GTTCGGTTCG
      481  GTACAACAAA AGAAAAAACA TGCAAACTTC ACAGCAGCTA TCGTCCGGGC ACAGGACCGC
      541  GAACGAAGAC GAGGACGACG AGGACGGCGA GTCCGAGCTG CTCGATCTGG CCATCGTCGG
      601  TGCGGGCATC TCGGGCCTGA TGGCGGCCAA AACGATCAGC GAGAAGCGGG CGGACATACG
      661  CTTCCGATTG TTCGAAAAGT CCACCCATCC GGGCGGCACG CTGGACGGGC TGAAAACGCG
      721  CTGGATAACA CCGCACCACT ACCACGCGAT GAACCTCTGC CGGGAGCTGC AGATCCCGCT
      781  CGGCACGGTT TGGCGCCAGT CGGAAGTTAG TGAAGCGCGA CGCTCCCTCG TTTCGATCGG
      841  TCCGTTTCGG AAGCGTCCCA GCGGCTCGCA GGGAGCGAAC AGCTTCATGT CCTCCCTTCG
      901  CGACCTGCTG GTACGGCTGG AAAGTACCCG GTTTATGGCG GAGCTGGACT GTCTCTGCAC
      961  GGTAAAGTAC ATCGAAAGCA ACGCCCAGAA CATGGAATGC TTTCTGCAGA GAAAGTTGCT
     1021  CTTCGAAGCG TCGAGGCGAT TTTTTCGTTT TCTCATCAAA ATCGGCACGG GCTTCTACCC
     1081  GTCCGAGCTG ACCGTGGCCG CCTGCCTGCG GCTGTTCCGT TCGATGTCGT CCGTGCGCGA
     1141  TCTGTACGAC ATGCTTACGT TGCAGAACGG CCACCTTCAG CCGGCCAGTT CGCCCAACTG
     1201  GGACGTACTG ATCGAACGGC TCGTGGCACG CGTCGGCCCG GAGCATGTCA CCTACTCGAC
     1261  CAACATTGTA CAGGTGGAAA TTACCAGCGA ACAGCGGTCG GACATCGTCT CGCTAACCGA
     1321  CGGGACCGGC CGGCGGTGGC GTGCGCGGTT CGTCATTTTG GCCGTCTCCT GTCTCGATCT
     1381  GGTCCAGATT AGCATCCTGC CCGAGGGACC GCTCTACTTC CACCAGCCCG ACATACAGCT
     1441  GGGCCTTTGG TCGATGGCCA ACTTTACCGT GCGCTATCCG GCTCCGTACT GGCGCGATCA
     1501  CGGCTACACC GGCAGCATAT TCTGCCCCGC CCAGTGCCTA ATATGCTACG AATCGGGGCG
     1561  CAATCAATTA TCCGGCACCT ACTTTTCGCC GCTCCGGGGC GTCCTGAACG ATACCGAGCG
     1621  GCATCTCATC CGCGACACGA TACTGCGCCT GCTGCGCACG AACTTCGCCT GCCGCACGAT
     1681  GCAGAAGCCG CTGGAATTCG GCCTCGAACT TCACCCGATA CCGTTCTACT TCGACGTGTT
     1741  TCCGACGTTC GAGCGGTGCA TCATCTTCTC CTCGACCAAT GTTAGCTGCT GGTACCGGGG
     1801  GTTCATCAAC GGGTCGATCC AGGGCGGTAA GTTTCGGGGA GTGTGTGGGT TTCGAGCATC
     1861  TACCCAACTT CCCATCAAAC GGGAGGGCAG GAAGTTGTAA AGCTTTGGCT CATCATTACC
     1921  TACCCTTTGC TTTGCAGGCG TGCGGGCCGC AATTTTGGCG CTGCTCGAGA TTCGACCGCA
     1981  AACCATCACC TTCCGGGAGG TGACCGATAT GCAGTGTATG CACTTCAAGT ATTTCCGCCA
     2041  ACGGTCATCT TATGAGCGCG TCTGGTACTC GTTGAACCTT GCGAGCGTGA GCCGGTTTTT
     2101  GCTGGGAGTC GGTGCTGTCC TACTGACTTA CGGCGCTTAT CGCGTGCTGC TGAAGGCTGA
     2161  CATTTGTTCC TGGTGGTCAA ATCGTTGACT GTTTGTCTGT AATACAAAAT TAACTCCGTA
     2221  TCTGATTTCG GTATGATTTC GTGCTTTAAT ACTTGTGTTT TTCGCTTTAA CACAAGACTA
     2281  GTACCGTATT TCTCAAGCAA TTATGTCATC AGCATCTCAC TCCCCAACAG CAATTGTCAC
     2341  ATGTCACACA ATCCCTCTCA CAAACAGCCT TTCTGAGTGC CAAACAAAAA TGGGTCTGCT
     2401  CGCGGGCACT TTCAAACAAA TCGATTTCAA CCGGTAAACG TTCAAACGAC AAAATTGAAA
     2461  CACTCGCTGC AAATTGAGCG AACGTTACAA AGATACGGGG GGACCGTACC CTGTCAATCG
     2521  TAGCTGACAT ATTGGCAGCG TCAACCAGCG TGGAAGACAG GAGAAGATGA AATGGAGTTT
     2581  TCCAGTTCCT GTTGCTTTCT TGCCCACCCC GTAGCGTTAG TTGAGCACTC GCAATATTGT
     2641  TCTGCTCATC GATGAGGTTA GGTTAGGTGC TGCACTAACT ATGGGAACAT GGCGGCTGGA
     2701  AAGCAAAGGG GTGTGCTAGG AAGCCTGCAA AACGATTGGG TAATAGTTGC ACCGATTGCA
```

```
    2761  GGGTGTCATA GCGTTCCCGG GTCGCTCTGA TCACGGCGCA GCTGAAGGTG CCCTAATGAT
    2821  AGCTTGATTG CGACTCCGTA GCAGGTCGGC CATAACCTCC GGTGCAAACG ATGCGCGTCA
    2881  ATTGATGGGC GGTACCTGTA TGATGACTTT GCAGCCAATT TCACCTAATT ATGTCGTATC
    2941  GGTTGGGCTG GATTTAAAGC TTCGGCGAGG AAGCGTTTTT GCACGGGGTC TTTAAAATCG
    3001  ATTGAAAGCG ATTGATGCAT CAGTTGGGAG CATGTTAGGC ATTTCAGATC ATGTTTGTGA
    3061  TCGGTAAATA GGGTCTGTGC AAACATTTTC TGGCAATAGA ATTTTAAAGT CTAAGCGTTG
    3121  ATACAATTCA AGTAAATTGA GTTTTTTTGT TGCTTAATTA GACATTCTCC GGAGGAACTA
    3181  TCAAACAAAG TTACATTATT TTATACCGTT GTTACATCTA AAATCTTTAT TTTGTTCTGT
    3241  TGAATTTAGT AAACATCTGA ACATACCTTT AATGTTAAAC CATAAACCAA ATTCACCAGT
    3301  TATTTCTTGA AACTTA

//

LOCUS          47     2583 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      47
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(500..2083)
                     /translation="MDNSSDITDKPAEAMEHKESPNAEENHPEGEQAKQPESSSEGA
                     ATSEQTNTPQEETKQPASNEPEEEDEAKLQEKLQTLKGDRIGETMYSERFVLSTILK
                     LSKQGNRLLEEDEEFEQDLCNLWDMTIEPDVVRFLLEQEVIELFLALAADSEDFRLI
                     EILFGIIGNMCCVEEMHELFYRQSALLVTFCDFLSISDAPVLVQLMRTLTTLFDRSD
                     DERFHWFGVMKRVDNLTEKLALILATSTNRVLLEHTAETVNAMINRFTEAEVDYAQL
                     NEFYRLFAKSCLLEGLFEAFKQLYPLTGRETEGTSAGFDEETLTLKDIKTMRRFLEI
                     HEHFIVNANELYEAYVSEVDQCVHRVLCSFVKESVLFPVTKNHICIFVSISNIYSGI
                     MYHFRAESFIYMVKIYDQLAEALKNPPQQEEPSNAAENNGTAMDEEDAADDGGEFDV
                     ETACAEVREAIFFLVGPIDEQIIRAAIRMGQLKESTFEQMCRELRETMDGEPLLYQT
                     CAKLMNAVRSTGAEESDIPMQQSNNEEQ"
ORIGIN
       1  GTGCGTTTTT CCACCAGACC TGGGCGGCTC TGCTGCAGGC ACTGGAGCAG TCGGACAATT
      61  TGGTAGATTA CAACGAGTAC AAGCGAAGGG ACTCGCTGCA GGAACAGTTG TGCCTTTCGC
     121  TGGCGCATCT ATTGCGCCTG GCGACGAGGG AAGATGTGGT ATCGATGGCA ACGGTGCTAT
     181  TACCGCTGTA CGATGCCGTA CGGGGGAACT GGGTGCGTGT GATAAGTCGC ATACTGCCGG
     241  AAAAGAGTGC CGTGCTGTTG GAGTCCTACC GTGTGCTGAC GGAGTTGCGC AAGGGCAATG
     301  CTGAAGGTGA ACCGGTGCCC GCATCGTCGT GGGATTTGCT GCTAAAATGT TTCAAAGATT
     361  CGGACGTTTG TTAAGGGGAG AAGAGATCAA CTTTTTGAGA TGTTTTTCCA AATGGCAGTC
     421  ATTACATTAT CGTTTTTTTA TAATCATTCG AAATGCGGCT TAAACATGTG TATTGATGTT
     481  AGTTGCTTAA ATGTATCCTC TACTGTTCCT CATTGTTGGA TTGCTGCATT GGAATGTCGC
     541  TTTCCTCCGC GCCCGTCGAT CGGACCGCAT TCATCAGCTT GGCACAGGTT TGGTACAGCA
     601  ACGGCTCCCC ATCCATCGTT TCTCGCAACT CCCGGCACAT CTGCTCAAAG GTGGACTCCT
     661  TCAACTGCCC CATCCGTATG GCCGCCCGTA TGATTTGCTC GTCGATTGGA CCGACGAGGA
     721  AAAAGATTGC CTCCCGTACC TCCGCGCAGG CCGTTTCCAC ATCAAACTCT CCGCCATCGT
     781  CTGCGGCGTC TTCCTCGTCC ATCGCCGTAC CGTTGTTCTC CGCCGCATTA GACGGCTCTT
     841  CCTGTTGTGG TGGATTTTTC AGTGCCTCCG CCAGTTGGTC ATAGATTTTC ACCATGTAAA
     901  TGAACGATTC CGCCCGGAAG TGGTACATAA TGCCGCTGTA AATATTGCTG ATCGATACAA
     961  AGATGCAGAT GTGATTCTTC GTAACCGGAA ACAGTACGCT TTCCTTCACG AACGAGCACA
    1021  GCACCCGATG CACGCACTGG TCAACCTCGC TAACGTACGC TTCGTACAGT TCGTTCGCAT
    1081  TCACGATAAA GTGCTCGTGA ATCTCGAGAA AGCGACGCAT CGTTTTAATA TCCTTCAGTG
    1141  TTAATGTTTC CTCATCAAAC CCTGCCGAGG TGCCCTCGGT TTCGCGTCCA GTCAAGGGGT
    1201  ACAACTGTTT GAACGCTTCG AACAGTCCCT CCAGCAGGCA GCTCTTTGCA AACAAACGAT
    1261  AAAACTCGTT CAGTTGCGCA TAATCTACTT CCGCCTCGGT GAAACGGTTT ATCATTGCGT
    1321  TCACCGTCTC GGCCGTATGC TCCAGCAGCA CTCGGTTCGT GCTGGTGGCC AGGATGAGGG
    1381  CGAGCTTTTC CGTCAGATTG TCGACCCTTT TCATCACACC GAACCAGTGG AACCGTTCAT
    1441  CGTCCGACCG GTCGAAGAGG GTCGTTAGGG TGCGCATCAG CTGAACTAGC ACGGGCGCAT
    1501  CGGATATGGA GAGAAAGTCA CAGAAGGTTA CCAGCAGGGC GGACTGGCGA TAGAACAGTT
    1561  CGTGCATCTC CTCCACGCAA CACATGTTGC CGATGATGCC AAACAGGATC TCGATCAACC
```

```
    1621  GGAAGTCCTC GGAATCGGCG GCAAGCGCAA GAAACAGTTC AATCACCTCC TGCTCCAGTA
    1681  GAAAACGTAC CACGTCCGGT TCGATCGTCA TGTCCCACAG ATTGCACAGA TCCTGTTCAA
    1741  ACTCTTCGTC CTCTTCCAGC AGACGGTTGC CCTGCTTCGA GAGCTTCAGG ATGGTCGACA
    1801  GGACGAACCG TTCCGAGTAC ATCGTCTCAC CGATCCGATC GCCTTTGAGG GTTTGCAGTT
    1861  TTTCCTGCAG CTTCGCTTCA TCCTCCTCTT CGGGCTCATT GCTGGCAGGT TGTTTGGTTT
    1921  CCTCTTGAGG AGTATTTGTT TGTTCGGAAG TGGCTGCACC TTCTGAGCTA CTTTCTGGTT
    1981  GTTTTGCTTG TTCGCCTTCG GGGTGGTTCT CTTCTGCATT AGGAGATTCC TTATGCTCCA
    2041  TGGCTTCTGC GGGTTTGTCC GTAATATCAC TGCTATTGTC CATTGCACTT ATAAATACTG
    2101  TTATCTATGA AATATTCAAC TATTGGGGTG AAAATGAAGC AAAATTCGTT TGTTTATATT
    2161  TTACCGGTAG AAGCTATTTC GGGACAATTG TTTACAGCGG TAAAAGAGCG TCGATCGAAG
    2221  CACCGCCAGC TGGAGCTAAC ATGCTTTCTC GCTTTGACCA ACATACCGGC GTACCTCGCG
    2281  TGAGAGCGAG AAGGCACGAG CCTTCTGCTG GCTGGGAGCT GTCAACGAAA TGGATTTTTA
    2341  GTACAACACA CACGCTGTCA CTTTGACAGC AGCCAGAAAA AAACGCGTGC AGACGTAAAC
    2401  ACCACCGCAC AACAAAACCT CGTGTTTCTG TGTGGCCGGA CAAAAATCGT TTCCCAGCTT
    2461  TGTAGGTGAA TTTGTGGTTC AAAATGAGTG CGTTGGCACT GGACGAGGTG CCCGGCTTTT
    2521  CGCTGCACAA AGCCGAAGTA GATTACGACG ATGAGGGGAA CGGATTCGGT GGCAAGAAGG
    2581  GCA

//

LOCUS          48     2667 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      48
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS          500..2167
                  /translation="MFDVRTVEAGACCEQERDSADGAMSMTARRRRAFQRQACLLSV
                  LAAFSLGLALGVFVPLIGISQAASKGSIVADGSPEQEPQQPALQALHHPAHQMPHFP
                  TIDDIRSADDGPRGRYASVSFIAEEKLTGEGEDVLFHPKLEVDGFVERKLPSPAASI
                  VQQMLAGSSKMAQAAVPERRESRASKFDDAEREDGEGVIRDGIYWGPSVERALPDGF
                  DQRQSESWARYLHEAEVERLEIGCGRMQNRLLVFRDGTRACARYRQNTDQIQGELFS
                  FYLGQLLNLTNLAPSAASIIDPEARLWAAATEELAGAQWKPTKPVVITKWIGNLEPA
                  GIPQPFRPLERHLNRFDVRNITEGLDRPRDGVRRLLLDRLGPARDGPAATPATPSSS
                  TSTLPAQPPTPEVLERLVELAQWSDLIVFDYLIANLDRVVNNLYNFQWNADIMAAPA
                  HNLARQTDSSLLVFLDNESGLLHGYRLLKKYEAYHGLLLDNLCVFRLSTVRALEALR
                  DASVGQRLNALFERTTTAQIRDVLPPLPEKSVKILVDRIDRVLGQVQKCRETFADR"
ORIGIN
       1  CGTGTTCCCC AAGGAGGGTT CGTCGCTTGG TGATGTGGTT GTTGCTGCAG TGCCGTTTTT
      61  TTTCCGTGCG CTGTTCGCCT GCCCGGATTC TGTGAGGGAG GAAGAAAGAG AGTTTATCGC
     121  ACACAAATAA ATCAATTATT ATAAGCTCCA AACGCGCAGT GGAACGCGCA GTGAAAACAA
     181  CGCAAAGAAG ATCGGCTGAA GCGAGCAAGT GAGTGTGTGT GTATGATGCG ATCGTGTGGG
     241  TGAAAAATCG TGAACAAACC GAAAATTGCT GTTGCTGCAA GTGTTTTAAA AATGTTTACG
     301  CCGTGTTAGC GATACGAAAA AGCATTAAAG TGTCCCGCAG TGTCTAGTAC ATAAAACTCC
     361  TAACGTGTGT GTACGAGTGT CTCTTGTGTA TGTGCCCTCT CGATCAACAG CGGAAAATCA
     421  ATCACAAAAC TTAACCGAGC AGCAGCCAGA AAAACGTCCT CCCCCCCGCA ATCTGGTGTG
     481  GTGTTGCTGC ACGTGGAAAA TGTTTGACGT GCGTACCGTC GAAGCGGGTG CCTGTTGTGA
     541  ACAAGAGCGT GATAGTGCGG ACGGCGCAAT GTCAATGACC GCGCGGCGAC GGCGCGCCTT
     601  CCAGCGGCAG GCCTGTCTGC TGAGCGTGCT GGCCGCGTTT TCGCTCGGCC TGGCGCTCGG
     661  CGTGTTTGTG CCGCTGATCG GCATCTCGCA GGCCGCCTCG AAGGGTTCGA TCGTGGCGGA
     721  CGGCAGTCCG GAGCAGGAGC CACAGCAGCC GGCGCTGCAG GCGCTTCATC ATCCCGCACA
     781  CCAGATGCCA CACTTTCCGA CGATCGACGA CATTCGCTCG GCGGACGATG GACCGAGGGG
     841  TCGGTACGCG TCCGTGTCGT TCATTGCGGA GGAGAAGCTG ACCGGCGAGG GCGAGGATGT
     901  GCTCTTCCAC CCGAAGCTGG AGGTGGACGG GTTCGTGGAG CGAAAGCTTC CGTCGCCGGC
     961  CGCCTCGATA GTGCAGCAAA TGTTGGCCGG CAGTTCGAAG ATGGCCCAGG CAGCGGTTCC
    1021  GGAGCGGCGT GAGTCGCGGG CGAGTAAGTT TGACGATGCG GAGCGTGAAG ATGGCGAGGG
    1081  CGTGATACGG GATGGCATCT ACTGGGGCCC GAGCGTGGAA CGAGCCCTGC CCGATGGCTT
    1141  CGATCAGCGG CAGAGCGAAT CTTGGGCGCG ATATCTGCAC GAAGCGGAGG TGGAGCGGCT
```

```
    1201  CGAGATCGGC TGTGGCCGTA TGCAGAACCG TTTGCTGGTG TTCCGGGACG GTACGAGGGC
    1261  CTGCGCACGG TATCGGCAGA ACACGGACCA GATCCAGGGC GAGCTGTTCT CGTTCTATCT
    1321  CGGCCAGCTG CTTAACCTGA CCAATCTTGC GCCGAGTGCG GCGTCCATCA TCGATCCCGA
    1381  GGCGCGACTA TGGGCTGCGG CAACGGAAGA GCTGGCCGGC GCCCAGTGGA AACCCACCAA
    1441  ACCCGTCGTC ATCACCAAGT GGATCGGGAA CCTGGAGCCT GCCGGCATCC CGCAACCGTT
    1501  CCGGCCGCTC GAGCGACATC TGAACCGATT CGACGTGCGC AACATCACCG AAGGGCTCGA
    1561  TCGACCGCGG GATGGCGTGC GAAGGCTGCT GCTCGACCGG CTCGGCCCGG CCCGAGACGG
    1621  CCCAGCGGCC ACTCCGGCCA CGCCATCGTC GAGCACGAGT ACGCTCCCCG CGCAGCCTCC
    1681  CACGCCCGAG GTGCTCGAGC GGCTGGTCGA GCTGGCCCAG TGGTCCGATC TGATCGTGTT
    1741  CGACTACCTG ATCGCCAACC TGGACCGGGT CGTCAACAAC CTGTACAACT TCCAGTGGAA
    1801  CGCGGACATC ATGGCGGCGC CGGCGCACAA CCTGGCCCGC CAGACCGACT CCTCGCTGCT
    1861  CGTCTTTCTC GACAACGAGT CGGGCCTGCT GCACGGCTAC CGGTTGCTGA AGAAGTACGA
    1921  AGCCTACCAC GGCCTGCTGC TCGACAACCT GTGCGTGTTC CGGCTGTCGA CGGTGCGGGC
    1981  GCTCGAGGCG CTGCGCGACG CCAGCGTCGG CCAGCGACTG AACGCGCTGT TCGAGCGCAC
    2041  GACGACGGCG CAGATACGGG ACGTGCTGCC GCCGCTGCCG GAAAAGTCCG TCAAGATTCT
    2101  AGTCGACCGC ATCGACCGGG TGCTCGGGCA GGTGCAGAAG TGTCGGGAAA CGTTCGCCGA
    2161  CCGATAGCCG CCGGGCCCGC CTGGTGCTGG AGGTGGAGTG GAAGCGGCCA GTCGGAGACA
    2221  GGTGGGTGGG TGGGTGAGCG GCGCACGCGC CATAAGACTG TGAGTCCTTT GTGTGAGCGA
    2281  TGGGAATTTA TTATTATTAT TATATTATTG CTAGTTTGCT AGTTGGTTGT TTTTTTTTTT
    2341  ATTTTGCATT TCCATTTACG GACTGTAAAG GTAGTGGTAA GGTTAGCCCG TACAGTTAGG
    2401  CTAGGGATGC AAGACTAATC GAGGAATGAT GATAGTAGAC AGCGAGATTG CGTGAGAGCG
    2461  AGAATGAATG ATAATGCGTA TAGCTTAACA AAACAGGTAA AAGAAAAGAG TGAAAAACAG
    2521  CAGGAAAGTA ACACAAACAG AACAATTCCA CTTACACAGA CCTACCACTA GTTTGAGGGT
    2581  AACGCCAGAC CACAGTTTTG CTGCTCTTAA GGGGTTGATG CATAAAACAG AGCGATACAA
    2641  GTAAACTATT TCCTCAAAAA CTTCTTT

//

LOCUS          49     3354 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      49
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          complement(join(500..2100,2704..2854))
                  /translation="MSAKTPSEETKAQTKASSGDMATDKRRVQQDDDASGEQCEAGS
                  SSAAREDDTSKPPASKRTKRCDDRNTELVSINGCFFDDEFRERFKAAWQRRETSAGC
                  NYELKQTPFQLAVLQNFLQEDAKRADATKRLEAEFSTVDWKRKQMDLYEFYQTDDLC
                  SLERDYLQAFYTVLKEQMMPLVEELTGIKLTHVSASCSMYNAGDYLLVHDDLLSDRR
                  IAYVFYLSPWDCHRRWQAKDGGALELFRADGNQLPVFPVTDKIYPQNNQLVLFRVCE
                  QSFHQVGEVTTFEYPRLTINGWFHGPAAPLPAGGKGAHAGPQAPSIIIDTHYMSPRK
                  IDAELGEWINDVYLVDEVKQNVQKKVEMFSEVSLEQFLLGPRFTALLESLRTGANLE
                  WRIKGPAHLRKYEVLNFKALPEANPLRALYDLFTSRTMFRLLYDYTELDLYGRKAKT
                  PKCAIELQRWERGCYTVLGDSSTYADSTLDITFYLNAQENVGVITYLVPEGEGKEQE
                  RQGSANQSDAPQASSSSTTSPYTECYEDDPVLLTLLPKDNVLNVVYRAEGTTKFTKY
                  VSHNTHLEKADGDGHAYTYILVCSYKE"
ORIGIN
        1  AGAAAAAAAA CCGACTGTTT ACGTACAATT TTGTTCGAAT TTTTTGCGCA ACTCGAATGA
       61  TAGAATTTCG CGCCGTTCGG CCGACTGGCA GTACAGGCCC AACCGAAACG TCGCCTTTAT
      121  TGGATCCTCA TTCAGGGCAC GAATTTCAGA GTAGGGTAGG GAAACGCTGT AAAGAGCATT
      181  GGTTAGTTAA GGGAAAATCA AGCGGTCCCG TTCAAACAGC CTGCTTACCG GTAGACTTCG
      241  GTCGTTTTCC ACCACATTGG CACACCAATC ACGATGATTA TGACTATGAA AGCAATCGTG
      301  GCGTAGATGT GGATGGCAGC TGCAAATAGA GTAGAAAACG GTAAATGGTT AGCATACATG
      361  AACATAGTTT TAAAAATAAG TTTAAATTCA ATAAACAGTT CAATAAAAAA AACACATCAT
      421  TGTTTCATCA TGTGCCATCG TTGTGCAATG TTTTTATTTG TAATTATTAT CTTTACGCAC
      481  GATACGCTTG TCCCTTTTTC TACTCCTTGT AGCTGCACAC CAGTATGTAG GTGTACGCGT
      541  GACCATCACC GTCCGCCTTT TCGAGATGCG TATTGTGCGA AACGTATTTG GTAAACTTGG
      601  TCGTTCCTTC CGCCCGGTAC ACCACGTTCA GTACATTGTC CTTCGGCAGC AACGTTAGCA
```

```
     661  GCACTGGATC ATCCTCATAG CATTCCGTGT ACGGCGAGGT GGTCGACGAT GACGACGCTT
     721  GCGGCGCATC GCTTTGATTG GCCGATCCTT GCCGCTCCTG CTCCTTTCCC TCGCCCTCCG
     781  GCACGAGATA CGTAATGACG CCGACATTTT CCTGCGCGTT CAGATAGAAG GTTATGTCCA
     841  GCGTGCTGTC GGCGTACGTG GACGAATCGC CCAGCACCGT GTAGCAGCCC CGCTCCCAGC
     901  GCTGCAGCTC GATGGCACAT TTGGGCGTTT TCGCCTTCCG CCCGTACAGG TCCAGCTCGG
     961  TGTAGTCGTA CAGCAGCCGG AACATCGTTC TGCTCGTGAA CAGATCGTAC AACGCTCGCA
    1021  ACGGATTGGC CTCGGGCAGT GCTTTAAAGT TCAGCACCTC GTACTTCCGC AAATGGGCCG
    1081  GTCCCTTGAT TCTCCACTCA AGGTTCGCCC CGGTACGGAG ACTTTCCAGC AGTGCCGTAA
    1141  ACCGGGGCCC GAGCAGAAAC TGTTCCAGCG ACACTTCCGA AAACATTTCC ACCTTCTTCT
    1201  GCACGTTCTG CTTCACCTCG TCCACGAGGT ACACGTCGTT GATCCACTCG CCCAGCTCGG
    1261  CGTCGATTTT GCGCGGGCTC ATGTAGTGCG TATCGATAAT GATCGACGGA GCCTGCGGAC
    1321  CGGCGTGCGC ACCCTTCCCG CCAGCCGGCA GTGGTGCTGC CGGCCCATGA AACCACCCGT
    1381  TGATGGTTAA CCGCGGATAC TCGAAGGTGG TCACTTCGCC CACCTGATGG AACGATTGCT
    1441  CACACACCCG GAACAGTACC AGCTGATTGT TCTGGGGGTA GATCTTGTCC GTCACCGGAA
    1501  AAACGGGCAG CTGATTGCCG TCCGCCCGGA ACAGCTCCAG TGCGCCGCCA TCCTTCGCCT
    1561  GCCACCGTCT GTGGCAGTCC CAGGGGCTAA GATAGAACAC GTACGCTATC CGCCGATCGG
    1621  ACAGCAGGTC GTCGTGCACG AGCAGATAGT CGCCCGCGTT GTACATCGAG CAGGAGGCGG
    1681  ACACATGCGT CAGCTTAATA CCGGTCAGCT CCTCCACCAG CGGCATCATC TGCTCCTTCA
    1741  GCACCGTGTA GAACGCTTGC AGGTAGTCCC GCTCGAGGCT GCACAGATCG TCCGTCTGGT
    1801  AGAACTCGTA CAGATCCATC TGCTTTCGCT TCCAGTCGAC GGTGCTGAAC TCCGCCTCGA
    1861  GCCGCTTCGT GGCGTCCGCC CGCTTCGCAT CCTCCTGGAG GAAGTTTTGC AACACGGCCA
    1921  GCTGAAACGG GGTCTGCTTG AGCTCGTAAT TGCAACCGGC GCTGGTTTCG CGCCGCTGCC
    1981  ACGCCGCCTT GAACCGTTCG CGGAATTCGT CGTCGAAAAA GCATCCATTG ATTGAGACTA
    2041  GCTCGGTGTT GCGGTCGTCG CATCGTTTGG TACGTTTGGA GGCGGGTGGC TTCGATGTGT
    2101  CTGTAAAAAG CGGGGGAACA GAGAGAAAAA CGCGGAACGT CTACGTTAGA TCTCATACAA
    2161  GCGCACTCTA TTTGTGATGG GAAGATGGAA TCGGAACGGG TTTTGTCACG TGTTTTGTGT
    2221  AGGAGTCCAC ATAACAAAAC CAAAAAGGAA ACGAAAGAAA ATCACACTCA CCCATCATTT
    2281  TTTATCCAGA GTCTGGCCCT CCCGACACGG CTTACGATTT TAAAAGCAAA ACCTGATCGT
    2341  ACTGAAAATA TAAAACACTT TCCGGGACGC GCATCGCGTT TCGCGGGTTC GGTTTCGTAG
    2401  AGCGCGGGTG GTTTCGGGAT TACTTGACAG AAAGTTAAAC CAAAAACCGA TCGTTTCAAG
    2461  TGCCTTCATC CGGGTAAAAA TGTTGGATAA ACGTGTGAGC ATGAAGCGAA AAGCGCATGG
    2521  CATTTGGTTG TTCCCTCCGT TTGAGTGGAA AATCGCGCGC GCGTGTTTCT GTGTGTGTAC
    2581  ATGTGCGCCG GGCACGCACA CACACCCTTT CGGATGCTTT TGGCCAATGG AACAGGGGAT
    2641  AAAACACACA GCGCACGGAG GGAAATTGTA AGGTGAAAAT GTAATGAAAA AGTGGTGCTT
    2701  TACCATCCTC GCGCGCTGCT GAAGAAGAAC CGGCTTCGCA TTGTTCTCCG GAGGCATCGT
    2761  CGTCCTGCTG CACCCGCCGC TTGTCTGTGG CCATATCGCC GCTGCTCGCT TTGGTTTGTG
    2821  CTTTTGTTTC TTCACTCGGC GTCTTCGCAC TCATGCTTAC AGTTTACTTG CAATCGTGCT
    2881  GCAACTTGCA CATGCAGCTA CACTTTCCAA ACGGGACTAG AACTGCACCG CAGAGTTATC
    2941  GGTAGACCAG CAACTCCACC TTGTTTTCTT GCGTGTGCGA AGCAAAGCAA AGCAAAAAAA
    3001  AAATACAATG CATTAGCAGG CGTCAAAACG TCAAACCGAT GACAGTTCGT TTTAGTGATG
    3061  GGCAAAACGG GCCTGCAACC GGATTTGGGT TGATCCGAAT CTAGCTGCTT TCGGATTCGA
    3121  ATCCGGATTA ACTCCGGAGT TAACGGGGGA GGAGCTAGTT GACCTAGTAT AATGTTTAGC
    3181  CATTTTTTAA TTTCTTGGGT AAATTTAATT TAATACTAAT TATCCTTAGT AATTGCATAA
    3241  CAAGTAACTT ATAACTACAT TTCCTAATCT GAATTCCTTT GTCTGCACTA TACAGTCTGC
    3301  GAGCGTTTCT ACATACACCG AGAATGCAGC TGAGAAAATA ACTGACAGCA TGCT

//

LOCUS          50     3096 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      50
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(500..2596)
                     /translation="MEVYEFNKSFLTRLLETTNVLNGGGSNIPASAIRSEWINFVEV
                     AVEPTGWQALWKASRPVCEQLGVKYPLVVLGTVEQVLFDELKAVFLIEAILDDDVHL
                     PEEQCTVQLEELWPLREQNNPALNVNITADCIDKLRFFYQHLWMPWDTDTEDEQNWI
```

```
                EKHLESRIRFCHDLKNKTMSRQLSSHALALLAEARYVQRKREFIEQELSEEDGQEEE
                EEDNDTSILKDSRSGELLQLNLRLNTIKNEINVLENPVWRSVFEKVRFGAGKSKGNS
                AMAYIVTQAVPMDEQLTYLQSAKEMMDPKIAVKMCDSLQHALDHCSPDSAIYLPPGT
                RQDIKFLEYLNSGGSFRGVSSASFVDDFEQACKLNPTIVSKNDDSVLLTIDGDFTLE
                NVRLDCSNVRTGVVIKKGNIVFRNCCFTGDPTSSTKQGIVIFGNCTITFDRCLIKEF
                SSGIYSNHDCTINLIGSTISHCMTGLETLDQCQIVFRSASITHCSQYGVLLEDFNED
                VQQDQKAAGDRSTSGSQVYADYNTIEREEFTFEGACEFRDNAKGNFVIRKGFSERFN
                SSCFVDEDESQLHDAEDGLDESICNATNVSFNANYLEASKLSSTKHIEPSVESSERE
                DRSASDTGVSGSLSQSEEDLSTGYCEDSTQIVKDSEYSVREVKRSEQTSESDPEDGS
                EEYDDDDDDDDEGDDQVIEIEDTIIEID"
ORIGIN
        1  ATGTTCGGAG GATCAATACA CTTTCCAAAT ATCGAATAGT TCGAAGCGTT GCTGTTAAAA
       61  TGGAATCTCT TTATTTATAA TTCTAGTGAA CCGATAAAAA GCAAATCGCA AAGATTATTT
      121  TGTTTGTTCC TTTGACAGCT AGCTTCCGTT GCATGTTTTG ACAGCGGCGG AAGAATTGGC
      181  TGCGTGCGTA AAAATATAAT TTTGAACTGG CTGCGGTTAA CGGATTAATC CTCCATCGGT
      241  TGAATGTCCT GAAATGGAAA AATAATAAAT AATTTAAATT AACGCATAAT TTCTAGAAAA
      301  AAAAATGAAT CACACGAAAC AATCTAAAGA ATCGAAAAAA AAAAACAAAT ACAAAACATA
      361  ACCTAACAAC AAACACTAGC TAAAATACAA TCCCAATCTC CAAGAGACTA CATTTTAACA
      421  CTTGGTACAA GACAACGAGA CAAGCACAAG CAGCTTTATC TACATCAATT TCATAAACTA
      481  ACCAGTCCTC GGTACTGCAT CAATCGATCT CAATGATCGT ATCCTCGATT TCTATAACCT
      541  GATCATCCCC CTCATCGTCA TCATCATCAT CATCATCATA CTCCTCCGAT CCATCTTCCG
      601  GGTCGGATTC ACTCGTTTGT TCCGACCGTT TCACTTCGCG CACAGAATAT TCGCTGTCTT
      661  TCACTATCTG CGTGCTGTCT TCGCAGTAAC CCGTACTTAA ATCTTCCTCC GACTGGCTCA
      721  ACGACCCAGA CACGCCCGTA TCGCTGGCCG ATCGGTCCTC CCGCTCGCTC GACTCAACCG
      781  ACGGTTCGAT ATGCTTGGTA GAGGATAGCT TTGATGCTTC CAGATAATTG GCATTGAAAC
      841  TTACATTCGT AGCATTGCAG ATCGATTCGT CCAGACCATC CTCCGCGTCA TGCAGCTGGC
      901  TCTCATCCTC ATCGACAAAG CAGGAGCTGT TAAACCGTTC GCTGAACCCT TTCCGGATGA
      961  CAAAATTGCC CTTGGCATTA TCCCGGAACT CGCACGCACC CTCGAACGTA AACTCTTCCC
     1021  GCTCGATCGT GTTGTAATCG GCGTACACCT GTGACCCGCT CGTGCTTCGA TCACCCGCTG
     1081  CCTTCTGATC CTGCTGGACG TCTTCGTTAA AGTCCTCGAG CAGCACGCCG TACTGTGAGC
     1141  AGTGCGTTAT GCTCGCCGAT CGGAACACTA TTTGACACTG ATCGAGCGTC TCCAGCCCGG
     1201  TCATACAGTG GCTAATGGTG CTGCCGATCA GATTGATCGT ACAATCGTGA TTGCTGTAGA
     1261  TGCCGGAAGA AAACTCCTTG ATCAGACACC GATCGAACGT GATCGTGCAG TTGCCGAAGA
     1321  TGACGATGCC CTGCTTGGTG CTGGAGGTTG GGTCGCCCGT GAAGCAACAG TTGCGGAAGA
     1381  CGATGTTGCC TTTCTTGATA ACCACCCCGG TGCGCACGTT CGAACAGTCG AGGCGGACGT
     1441  TTTCCAGCGT AAAGTCCCCG TCGATGGTCA GCAGCACACT GTCGTCGTTT TTCGACACGA
     1501  TGGTAGGATT CAGCTTGCAC GCTTGCTCAA AATCGTCAAC AAAGGAGGCG CTGCTAACAC
     1561  CCCGGAACGA TCCGCCACTG TTCAGGTACT CCAGGAACTT AATGTCCTGT CTGGTTCCGG
     1621  GTGGAAGATA GATCGCACTA TCCGGACTGC AGTGATCGAG CGCGTGCTGA AGAGAGTCGC
     1681  ACATTTTCAC CGCAATCTTC GGATCCATCA TCTCTTTAGC GCTCTGGAGG TAGGTCAGCT
     1741  GTTCATCCAT CGGTACTGCC TGCGTTACGA TATACGCCAT TGCGGAGTTG CCCTTGCTCT
     1801  TTCCCGCACC GAAGCGTACC TTCTCGAAGA CGCTGCGCCA GACGGGATTC TCCAGCACGT
     1861  TGATTTCATT CTTGATCGTA TTTAAGCGTA AGTTCAGCTG CAACAGCTCT CCCGACCGGC
     1921  TATCCTTCAG GATGGACGTA TCATTGTCCT CCTCTTCCTC CTCCTGCCCA TCCTCCTCGC
     1981  TAAGCTCCTG CTCGATAAAC TCCCGCTTGC GCTGCACGTA CCGCGCTTCC GCGAGCAAGG
     2041  CCAGCGCGTG GGAGGAAAGC TGCCGCGACA TGGTCTTGTT CTTCAGATCG TGACAGAACC
     2101  GGATGCGCGA TTCCAGATGC TTCTCGATCC AGTTCTGCTC ATCCTCCGTA TCGGTGTCCC
     2161  ACGGCATCCA GAGATGCTGG TAGAAAAACC GCAGCTTATC GATACAGTCG GCCGTAATGT
     2221  TGACGTTCAG GGCGGGGTTG TTCTGCTCGC GCAGCGGCCA CAGCTCCTCC AGCTGCACCG
     2281  TACACTGCTC CTCCGGCAGA TGCACATCAT CGTCCAGGAT GGCTTCGATG AGGAAGACGG
     2341  CCTTCAGCTC GTCGAACAGT ACCTGTTCCA CGGTGCCCAG CACGACGAGC GGGTACTTTA
     2401  CGCCGAGCTG CTCGCATACC GGACGCGATG CTTTCCACAG CGCCTGCCAG CCGGTCGGCT
     2461  CTACGGCCAC CTCCACGAAA TTGATCCATT CCGAGCGGAT GGCGGATGCG GGAATGTTGG
     2521  AACCGCCACC GTTCAGCACG TTGGTCGTCT CGAGCAGACG GGTGAGGAAA CTTTTGTTAA
     2581  ACTCGTACAC TTCCATCTTT GTGTTGTGAT CGAATGGGGC GGGTGAAAAT CGTTGGACGG
     2641  TTGCAGAGAA TCGAAATTTT TGAAGCGTTC CAGCTGTCTA GTGAACGTGG GGATGTAAAC
     2701  GGACACAGGT TCGAATGACA GTGCGGTCAT TTTTGCTGCC GCCATGTTTT AACAGAAATG
```

```
    2761  CTGTTTTGCT TGCTTATAAA TTAAATTTTT ATGGTTATTT TAACATTTTT AAAATATCAA
    2821  AATCACTAGA TATGCAATTA ATTTGTTTTA TGTGGCATGA AAGAACAACA GGCATTGGGT
    2881  TAGGAAGGTT TTTTGCAATC TTGATCAAGT CGATTGCTAG AATGAGCCAA ACTGGTTCAC
    2941  TAAATGATAA ACCGACTCCA CACGACGCAC TATGGGCTTT CGGTGCTGGA CAAAATATGA
    3001  ACAGTTGATT GCGTTTGTCA TGATAGTTGT CACTTTTTAT AAGAATCCCG GAGAGGCAGG
    3061  ACCGCAGTCG GTGTCGTCGT AGTCGGTTCG TTGAAA

//

LOCUS          51     3358 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      51
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES             CHR  ---
     CDS             complement(join(500..2475,2568..2757,2841..2858))
                     /translation="MIIPKEVFCCAKCRQKHEDGSHQEDAMLRKICYICNDRPFPLR
                     KDTKITPNNLLIVHILKEHLPLECNRSDFQQINKCFTLCSSDAGKEGECYLESSGTI
                     GHHALPTIEADNKENIAVEQPADDRAMPSRERRKSSLVPGDFQPFGEVDNKIDEVDE
                     ALLTPLTKINLRWKRKSRQSFESMLSMTNNTSEVNASGLLRVDGGSGPRKLVRTTST
                     PMYQWQLMPTGPKQPNESYSATMGQMSSIHCSSGSDTDSRAEDLGASPPVPNELDKV
                     RAIIRNRSKAVATPLRQVMSKSIQRAIAQHTAGVGGVYSKMLQPGTQRKMSFNSTMS
                     SGMNSMLASPTKDPLDLRTTPTLKRLGSGRGGGGSATSAQSIQAQRMVRPGDPAQES
                     VQLAQADAIPNANTTDGSEYYETHRDMESLHGSSALGVSRGVDSLRKLHSAESYPVT
                     PKPGGGGTALKKVSSFQALQDLQLKGEQQSNSAQNDTLEDVWGTPCGGGIPGRSYSC
                     SAAPRLGDTFKAYDASVEDAENDSGTDDVFYPVKGRPNIRQGKSDRDAFADSTSHES
                     KVGKLDGMNATRSGGGGGKLWSIVSNVIRLASRSDVTQEEHSMDAPRATTGAVAGGG
                     GILHRAASFADYLKNRYTTRSDGGISRTSSSGSDEYRIAHESNGKKRRRVASTYARP
                     YAPALTNDGPCAAAKVQQHSPVAKRKRIQSRQPIHRLRHTSSESGGSGSASGPGRYN
                     "
ORIGIN
       1  AAATGTCTAT TTTACGTTAC AGATTTAACG GCACGACTGT CCTGGATCAA TTGGGAAAAC
      61  TTGGGCGTAC ATCCGGGCAG ACGTACCATC CTGACGTCGC TGGCACTCGG ATCGTCAGAG
     121  AGTATGGGTG CGTGTGGACC GCCGCCACAT TTGATCTAGA TTGCAACCAC CGCACTCCTT
     181  TAACGCGGCA CCATTTTTAA GTCTTCCGTT TTGTTTATTC TTTAAATATA AATACTAAAA
     241  CTGCATCTAA TAACTATTTA TCAGCACTAG GCTTACACCT CCCTATTCAA ATAGGAATCA
     301  GCTAGTTCTC TGTCGTACCC CATTTTAAAT GAGAATAAAA ATGAAAAACG CACTCTCGTC
     361  GCCTATTATT GCTTAAAACA ATTCAAAAAT AATGTATTTT TTCGTTCCGG TTCCATTTAA
     421  TGGCACAAAT GTTTGACAAT CTTCCCTAAA TATAAAGCAT ACAAAAGAAA AAAATAATAT
     481  CCAATCACAC TCCCGCACGC TAATTGTACC TTCCGGGGCC ACTAGCGCTT CCGCTGCCAC
     541  CGCTTTCGCT GCTCGTATGG CGCAACCGAT GGATAGGCTG GCGGCTCTGG ATGCGCTTCC
     601  GCTTTGCGAC CGGACTGTGC TGCTGTACCT TTGCGGCCGC ACACGGACCA TCATTCGTTA
     661  GGGCCGGAGC GTACGGCCGC GCGTAGGTTG ACGCTACGCG ACGCCTCTTT TTGCCGTTCG
     721  ATTCGTGCGC GATACGATAC TCGTCCGAGC CGGAGGAAGA CGTGCGCGAA ATTCCACCGT
     781  CCGACCGGGT GGTGTACCGG TTTTTAAGAT AGTCCGCAAA CGATGCGGCT CGGTGTAGGA
     841  TTCCACCACC GCCAGCCACC GCGCCAGTGG TGGCTCTTGG CGCATCCATC GAGTGTTCCT
     901  CCTGCGTGAC GTCGCTGCGC GATGCCAAGC GAATGACGTT GGAAACGATC GACCAGAGTT
     961  TTCCACCACC ACCACCGCTG CGAGTAGCAT TCATTCCATC CAGCTTTCCT ACTTTGCTCT
    1021  CGTGTGATGT ACTGTCGGCG AACGCGTCAC GGTCGCTTTT GCCCTGTCTT ATGTTTGGTC
    1081  GCCCTTTGAC CGGATAAAAC ACATCATCCG TGCCGGAATC GTTTTCCGCA TCCTCCACCG
    1141  ACGCATCGTA CGCCTTAAAC GTGTCGCCGA GCCTAGGAGC GGCCGAACAG CTGTAACTAC
    1201  GGCCAGGTAT ACCACCTCCG CACGGCGTTC CCCAGACGTC TTCCAGCGTG TCGTTCTGGG
    1261  CACTGTTGCT CTGTTGCTCC CCTTTCAGCT GCAGATCCTG CAGCGCTTGG AATGAGCTCA
    1321  CCTTCTTCAG TGCCGTTCCG CCGCCGCCCG GTTTCGGTGT CACCGGGTAG CTTTCTGCCG
    1381  AGTGTAGCTT CCGCAGACTG TCTACCCCTC TCGACACTCC TAACGCGCTA CTCCCATGCA
    1441  GTGATTCCAT ATCGCGATGT GTTTCGTAAT ACTCTGAACC GTCTGTCGTG TTCGCATTCG
    1501  GTATGGCATC GGCCTGTGCT AGCTGTACCG ACTCTTGAGC CGGATCACCT GGCCGTACCA
    1561  TACGTTGCGC TTGAATGGAT TGAGCCGAAG TTGCGGAGCC GCCACCGCCA CGCCCACTAC
```

```
    1621  CGAGGCGCTT CAGTGTCGGT GTGGTACGCA GATCCAGCGG ATCCTTCGTC GGTGACGCCA
    1681  GCATACTGTT CATGCCGCTG CTCATCGTAG AGTTGAAGCT CATCTTACGT TGCGTGCCCG
    1741  GCTGCAGCAT CTTCGAGTAC ACGCCGCCCA CACCGGCCGT ATGCTGGGCG ATGGCACGCT
    1801  GAATGCTTTT GGACATTACC TGCCGCAGCG GCGTGGCGAC CGCCTTCGAC CGGTTGCGTA
    1861  TGATTGCCCG CACCTTGTCC AGCTCGTTCG GCACCGGGGG CGACGCGCCA AGATCTTCCG
    1921  CCCGGCTGTC CGTGTCACTG CCGCTGCTGC AGTGTATGCT CGACATCTGC CCCATCGTCG
    1981  CGGAGTACGA CTCGTTCGGT TGCTTCGGCC CCGTCGGCAT CAGCTGCCAC TGGTACATGG
    2041  GCGTTGAGGT GGTGCGCACT AGCTTGCGCG GCCCGGACCC CCCATCCACG CGCAGCAACC
    2101  CGGATGCGTT CACCTCCGAC GTGTTGTTCG TCATCGAGAG CATGCTTTCG AAGCTTTGAC
    2161  GGCTCTTGCG CTTCCAGCGC AGATTAATCT TTGTCAGCGG CGTCAGCAGC GCTTCGTCCA
    2221  CTTCGTCTAT CTTGTTGTCC ACCTCGCCGA ACGGTTGAAA ATCGCCCGGC ACGAGACTGC
    2281  TCTTTCTCCG TTCACGGGAC GGCATCGCCC GATCGTCTGC AGGTTGCTCA ACCGCAATGT
    2341  TTTCCTTATT GTCTGCTTCA ATGGTGGGCA GCGCATGGTG ACCTATTGTG CCGCTGCTTT
    2401  CGAGATAGCA TTCACCTTCC TTTCCCGCGT CACTCGAGCA GAGCGTGAAG CATTTGTTAA
    2461  TCTGCTGAAA GTCGGCTAAT GAATAGAAAA CCTGTCGCGC ACGGTTTGTG GTTAGTTTTG
    2521  TTGCTAAGAT CAAATAGGAA TCTAATTAAA TGTTTTTACC TTCCTACATC GGTTGCATTC
    2581  CAGGGGCAGA TGTTCTTTCA GTATGTGTAC GATGAGCAAA TTGTTTGGCG TGATTTTAGT
    2641  GTCTTTCCGC AACGGGAATG GTCTGTCGTT GCAGATGTAG CAAATTTTGC GTAACATGGC
    2701  GTCTTCCTGA TGGCTTCCAT CTTCATGCTT TTGCCGGCAC TTGGCACAGC AGAACACCTG
    2761  CAGGCAGGTT CTACACTGCA CCTGATAACG AACATATTAT TGAATAATAG TGCAAACTGG
    2821  CGCGCATCTC ACCAACTTAC CTCCTTAGGT ATAATCATCC TGTAGTTCAC TGTTTTCTCT
    2881  ACTTCACTTC ATGTACTTTC GACTCGCTGG CGATGATGGC CGTACGCTTC CAAAAAGGAT
    2941  GATGGTTGTC CTAGAAGATC AACACCACTA CTGTTACCCG ACCGCAGTGT CTCATCGACC
    3001  CGCATCCGCA CGATTAGGTG TCTTAGTGCT ATTATTGTCC ACCGCCGGCA CAACAGCCAA
    3061  CTGTCGCAGT CCGGCAGCAC CAACACCAGG AAGATTTGAC GTTCGCTGTT TTGAATTGGA
    3121  CGATCGCCGC CCACCGGCGA GTTTCGCACG GAATCGAACC TCCGATCGAC GCTGGTTCAG
    3181  CTTTGCCCGA ACCTCCTCCA CACTTTCGAG CTCCGGTGCG TACGAGAGGT GCAGTATACC
    3241  GCCGTAAAAG TTGCGTGCGT CCAGCAGCTT TTTCGCTTGT CGGGCCTTTT CTAGTTTGAC
    3301  GAAACACGCA TGAAATACGT CCGTAAACGG TTCGGGCAGC AGCTCGTCCG AGCCTTCA

//

LOCUS          52     3201 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      52
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          500..2701
                  /translation="MSGEVSEPIKKCSLILKNAKSDTEKFAALFMVTKLIKGKDCNQ
                  AGKRLLFESIGFDFLRKLLTSKDVPDDCPASVYQSVALSILSCFCDDEQLATHQDML
                  DSIPVFMGIVSTCDDDEYDDNLIIINEAYHCLQSISLYESGRLALRQHDVITKMAQI
                  YTQRSFQIDEALTLIVTLVARFGANSWDNDPKLFHALLQRVSLDFETDHAERKFELA
                  EMISSLLFHCRRDLVARTVQGEIWPECLYKGISDILKSKIGKAQRDPALKLAANAVD
                  VLGIEWTLHDVENPKKFFLLLLQLAAIEVRMQMDNKSFNQCVQQADLITACFIILEL
                  SINFMSTDQLDLDQKDKQQVYTGLKGAFSAVLGVLVKLANDTKKDRLQKTEKAFAYA
                  MVRVLTAWLAQETSAMKNQVSKVLPFLFKLANESFYESRDYRIAHKADNVDDHEQQP
                  PADILRVMLPAICHLVVEEEARQIFLKEKEEQVLYDCLLFHWSIAHYKKPPVPRAER
                  LKRMNEPDPEMTPQQLDDMKDSRTAIVSLCNILMNITVLEAKLVEESTLFAQLLRFI
                  FENLPELKDIPDNLVMHGHLAVLGLLLLKQQASKIKKNDFSICRYIQTTIRFLWDAY
                  NIDESNDPQALVVSMQYKEHWFEIMELWFLGMQTMSGIIKLIPWISEFAIESGWAEG
                  IVETLRKVKIGTLPPNVKLAYEDFLSQLVDANPAVAPVLKKADALKVCRNHRMMDLG
                  KKLFGD"
ORIGIN
       1  CCGTGAGGGT TTCGTGAAGG TCTTCCTAGC AGCTGCACAG CGTGTGTTGC GTTATTGAAC
      61  AGCGGCCACC AGACAGAGCG CAGGTTGGGT GCCGGTGGTA GCACGAAAAG GGAGAGACAT
     121  TTTCCATCAG GATCGTGAGC AAATCTATTT TAACGATTTT CCGACTCTCT CGCAGAGAAC
     181  ACGGAGCACG CGGAGCGTAG AAATGAGCTG CACCACATTG CCGGTGCGAC TAGGAAAATA
```

```
     241  GAGAATGGGT TTCCACCGGT CGTCATCCGA ATCAGTCGAG AAAACGTCTT GGGACGGTAC
     301  GGTTCGGATG GTGAACAGAA GTGCTACTAT TTGAGACACA TCTGTAACTA GCTGATTGTT
     361  GCTAGGCAAC CTATCATCTG CTTTGTGGTG CACGTGTGCT TGTACGTGAA ACAATTGTGT
     421  GCTGCGGCAG AAAACTGAAA ACAGTTCAAT TAAGTAACGC AAAGAAACCA CACGAAAAGT
     481  ATCCCAAGTC GCAACAAAAA TGTCAGGCGA AGTGAGTGAA CCCATTAAAA AATGCTCGTT
     541  GATATTGAAA AATGCCAAGA GTGACACGGA AAAGTTTGCC GCCCTGTTCA TGGTGACGAA
     601  ACTGATCAAG GGGAAGGACT GCAACCAGGC CGGCAAACGG TTGCTGTTCG AGTCGATCGG
     661  GTTCGACTTC TTGCGCAAGC TGCTCACTAG CAAAGACGTG CCGGATGATT GTCCCGCCTC
     721  GGTTTATCAG AGTGTGGCTC TTTCGATTCT GAGCTGCTTC TGCGATGACG AGCAGCTCGC
     781  CACGCACCAG GACATGCTAG ACAGCATACC CGTGTTCATG GGGATCGTAT CGACGTGCGA
     841  CGACGACGAG TACGACGACA ATCTGATCAT TATCAACGAG GCGTACCACT GTCTGCAGAG
     901  CATTTCGCTG TACGAAAGTG GCCGGTTGGC GCTGCGCCAG CATGATGTCA TCACGAAAAT
     961  GGCGCAGATC TACACGCAGC GCAGCTTCCA GATCGATGAA GCACTCACGC TAATCGTAAC
    1021  ACTGGTCGCC CGCTTTGGAG CGAACTCGTG GGACAACGAT CCGAAGTTGT TCCATGCGCT
    1081  GTTGCAGCGC GTTTCGCTCG ACTTTGAAAC CGACCACGCG GAGCGCAAGT TCGAGCTGGC
    1141  GGAGATGATC AGCTCGCTGC TGTTCCACTG TCGCCGAGAT CTGGTGGCCC GCACGGTGCA
    1201  GGGCGAAATC TGGCCCGAAT GCTTGTACAA GGGTATTAGC GATATTTTGA AAAGCAAAAT
    1261  CGGCAAGGCA CAGCGCGATC CGGCCCTAAA GCTGGCAGCG AATGCGGTCG ACGTGCTCGG
    1321  CATCGAATGG ACGTTGCACG ATGTGGAGAA TCCGAAGAAG TTTTTCCTGC TGCTGCTACA
    1381  GCTGGCCGCA ATCGAGGTGC GCATGCAGAT GGACAACAAG AGCTTCAACC AGTGCGTACA
    1441  GCAAGCCGAT CTAATCACGG CCTGCTTCAT CATCCTAGAG CTGTCGATCA ATTTCATGTC
    1501  GACGGACCAG CTAGATCTCG ACCAAAAGGA TAAGCAGCAG GTATACACGG GCCTAAAGGG
    1561  TGCCTTTTCC GCGGTGCTTG GTGTGTTGGT AAAGTTGGCC AACGACACCA AGAAAGACCG
    1621  ACTGCAGAAG ACGGAAAAGG CGTTTGCGTA CGCGATGGTT CGCGTGCTTA CAGCCTGGTT
    1681  GGCGCAGGAA ACAAGTGCCA TGAAGAATCA GGTTTCCAAA GTGTTGCCTT TCCTGTTCAA
    1741  GCTGGCCAAC GAATCGTTTT ACGAGTCACG CGACTACCGC ATTGCTCACA AGGCCGACAA
    1801  TGTGGATGAT CACGAGCAGC AGCCACCGGC AGACATCTTG CGGGTGATGT TGCCCGCCAT
    1861  ATGCCATCTG GTGGTGGAGG AGGAAGCGCG CCAGATTTTC CTGAAGGAGA AGGAAGAGCA
    1921  GGTGTTGTAC GATTGCCTGC TGTTCCACTG GTCAATTGCG CACTACAAGA AACCGCCGGT
    1981  GCCACGCGCC GAACGACTCA AGCGCATGAA CGAACCCGAC CCAGAGATGA CGCCTCAACA
    2041  GCTGGACGAC ATGAAGGACT CGCGCACTGC GATCGTAAGC TTGTGCAACA TCCTGATGAA
    2101  CATCACCGTG CTGGAGGCCA AGCTGGTCGA GGAGAGTACG CTGTTCGCTC AGTTGCTGCG
    2161  GTTCATCTTT GAAAACTTAC CGGAGCTGAA GGACATACCA GACAATCTGG TGATGCACGG
    2221  TCACCTGGCT GTGCTGGGAC TGCTCTTGCT GAAGCAGCAA GCGTCAAAGA TTAAGAAAAA
    2281  CGACTTCTCC ATCTGTCGGT ACATCCAGAC GACGATCCGG TTCCTGTGGG ATGCGTACAA
    2341  CATTGACGAG TCGAACGATC CGCAGGCGCT GGTGGTGTCG ATGCAATACA AAGAACACTG
    2401  GTTCGAGATC ATGGAGCTGT GGTTCCTGGG CATGCAGACG ATGAGCGGCA TCATCAAGCT
    2461  GATTCCCTGG ATCTCTGAGT TTGCAATCGA GAGTGGCTGG GCCGAGGGAA TTGTAGAAAC
    2521  GTTGCGCAAG GTGAAGATTG GCACGCTGCC CCCGAACGTG AAGCTGGCGT ACGAAGACTT
    2581  CCTGTCCCAG CTCGTCGATG CTAATCCGGC CGTAGCGCCC GTGCTCAAGA AGGCCGATGC
    2641  GCTGAAGGTG TGCCGTAATC ATCGGATGAT GGACCTGGGC AAGAAACTGT TCGGAGATTA
    2701  AAGCTAACGT GCGCAGTTTT TCTTTAATTT TACACTCTAA ATTCACCTGT TTCTTTGTCC
    2761  ATGCGTCATG TTGCCGGGCG GCATAGCTCC ATTTCTTATT ACATGTAATA CAATTTGCTG
    2821  CATTTTTGTT TCATCATACT CCATTTACTT AATCTCATCG CATGTTGGTT GTCGCGTTCG
    2881  CTTATTGCGC CGTATGTTCG GTAATATTGA AAAAGAAAAA TACGAAAAAG TAAAAATGTA
    2941  ATGTACACGC TATATGTGCA GAATAAACGT GTTTTCTACG TATATGTATA ACTACACATC
    3001  GGGCTTTTAT AATTCAATAA TCTTCAAGCT AAAATCTTCT ATATGGAATA AACGTAGTTT
    3061  CACGCTCCAG AGTACAAGCT CGAAAAGAAA GGACCCCAAA ATGAATGCCC TACTTTCATT
    3121  TGCCCAATGT CTTGGACAAG AAAAATCTTA GCTAGAAACA CGGAGGAGAC GCGAATAGCA
    3181  GTTTTAGTAG CCGTGTTTAA C

//

LOCUS          53     3482 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      53
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
```

```
FEATURES             CHR  ---
     CDS             join(500..527,598..2259,2373..2830,2905..2982)
                     /translation="MWFLTNNKTQHVYYLVSHTPCHTVGRTGTDLIITGDDSISRNH
                     ASLQPHKDVLKLTDIGSRYGSYVNDNINKTVPISKEHPTELKAGDVIRFGRCGSVWT
                     VGKASFRCLTSTLVMDDRLKAVLEKIGAEMLPLYTPGLTHLIMPTITVTTKFLQCMV
                     GQVPIVKPDYFHAIERDCIGKGKALPAVKEFVPSCTETYIRNEELIFHPNPARSTLF
                     RGKEFIFLNTAQYNQFEHIVKLAGGVCVSAQREKIAKSRFLKPNVVTVKLKSSGSSQ
                     SQSQSFDTLSQYIASRGRRMVPDLEIGLAIMFCSTEKYCNPEYSFAFNVEECYSSAQ
                     GGETLANSTEQQTESTEAGKSASGPEINTIPETEPLTEQKSRAAAATKTPPVGKAVN
                     GSTRSIASFLVPAPVPIAAAPEERRKSKRMQEAEKSKQQTNDDDEHSSNPSKRPRRT
                     GSLVESVSYSPPVDSKLTESERNEIHVPETQPTPDSAMSQALQARGFRAVNRDSMES
                     DSRAAAKDSAKTRRAANLHQKLDDEDMFNFDEIAPRKKPRMDKANEKPITSTNTAEV
                     RQTVAVGDECFNNSNNSHSKTNPSLATSDASITRALAGAQNRSKGSNDSRMDGYKEF
                     IKPVQPSAMGWLSSTMCGLKLTGGGGEDHSLDEINIKSEPLDETDDPYGMQRDCKKW
                     IKSMENAFPVREVSMKLVAHRPMDVSGQEVSCVTADGGKNFKAFVKKRNYPVQQLVL
                     GTKSVCVRDDNQCA"
ORIGIN
        1  TCAGCTGGCC GGAAATTGGT AAAATGTTCA CGAATATTTC GCACTGCTGC CTTCCAGCTT
       61  GTTACTCCTT TCGCCACCTC ATGAAAAACA CTGGCTTTTG ACAGTTGTGA ATTGCTCAAT
      121  AGCAGATGCA GGGTTCGTGT CATCAAGTCC ATCGGGAAAA GAGAATAAGG TGTTGCTTTG
      181  TAATCATTAT CTCAATTTAT TTGTATTTTG AGAAATTTCA AGCTTATAAA AATAAATTTT
      241  TAATAAACCT CACTCCAAAT TGTCAAATTC ATTTTTATTT ATGATATTTT TTATAGAAAC
      301  TTATTCCACA ACAAAGAGGC GAAGACCCCT GTGAATCGAT TTTGAGCAAC TTCCCAGAAC
      361  TGTCATTTTA AAGCGCAAAA CAACAATATT CATTTATTTT CAGTTTTTTT CCACGCTACA
      421  GATAGCCTCA ACCACGGCGA GGTGGTTTTC TTGTGCATCG TCTGATTATA TACAGTGTTT
      481  TCCGTGGTAG AAAGTGAAAA TGTGGTTCCT AACAAATAAT AAAACACGTA AGTGTACTTG
      541  TACCCGTGCG TGATCAAACG TAAATGTATT GATCGATCGC TTGGTTTTCT CTTCCAGAGC
      601  ACGTATACTA TCTTGTATCA CATACCCCCT GCCATACGGT GGGCCGAACC GGTACGGATC
      661  TCATTATAAC CGGCGACGAT TCAATCAGCC GCAACCATGC CTCGCTACAG CCCCACAAGG
      721  ACGTCCTGAA GCTCACCGAT ATAGGATCCC GGTATGGATC GTACGTTAAC GACAACATTA
      781  ACAAAACAGT CCCCATCTCC AAGGAGCATC CGACCGAGCT GAAGGCTGGC GATGTGATAC
      841  GCTTTGGCCG CTGTGGCAGT GTCTGGACGG TCGGCAAAGC TAGCTTTCGG TGTCTCACAT
      901  CCACACTCGT CATGGACGAT CGGTTGAAAG CGGTGCTGGA GAAGATAGGC GCCGAGATGC
      961  TGCCCCTGTA CACACCCGGT CTGACGCATC TCATAATGCC GACCATCACC GTCACCACCA
     1021  AATTCCTGCA GTGCATGGTC GGCCAGGTGC CGATCGTAAA GCCTGACTAT TTCCACGCAA
     1081  TCGAGCGCGA TTGCATCGGT AAAGGGAAAG CACTGCCGGC GGTGAAAGAA TTCGTTCCCT
     1141  CCTGCACGGA AACCTATATC CGGAACGAAG AGCTGATCTT TCACCCTAAC CCGGCCCGGA
     1201  GCACACTTTT CCGTGGGAAG GAGTTTATCT TTCTCAACAC CGCACAGTAC AATCAGTTTG
     1261  AGCACATTGT CAAGCTGGCC GGTGGTGTGT GCGTGAGTGC GCAGCGGGAA AAAATAGCCA
     1321  AATCACGCTT TCTCAAACCG AACGTGGTAA CGGTGAAGCT GAAATCATCC GGCTCATCCC
     1381  AATCGCAGTC CCAGTCGTTT GACACGCTGT CCCAGTACAT TGCGTCACGG GGACGCCGCA
     1441  TGGTGCCGGA TCTGGAAATT GGGCTAGCGA TCATGTTCTG CTCGACGGAA AAGTACTGCA
     1501  ATCCGGAGTA CAGCTTCGCG TTCAACGTCG AGGAGTGCTA CAGTTCGGCG CAGGGCGGAG
     1561  AAACGCTTGC CAACAGCACG GAGCAGCAAA CGGAAAGCAC AGAAGCGGGA AAGAGTGCGA
     1621  GCGGGCCCGA AATAAACACC ATCCCGGAAA CGGAACCGCT AACCGAACAG AAGTCAAGAG
     1681  CAGCTGCGGC AACCAAAACG CCCCCCGTGG GTAAAGCTGT TAACGGTTCT ACACGTTCTA
     1741  TCGCGTCCTT TCTCGTTCCG GCGCCGGTGC CGATCGCTGC TGCTCCCGAA GAACGACGCA
     1801  AATCGAAACG AATGCAGGAG GCAGAGAAAA GCAAGCAGCA AACGAATGAC GACGATGAGC
     1861  ACAGCAGCAA TCCGTCCAAG CGTCCCCGGC GTACTGGTTC TCTAGTGGAA AGCGTGAGCT
     1921  ACTCGCCGCC TGTGGATTCG AAGCTTACAG AATCGGAGCG AAATGAGATA CATGTGCCCG
     1981  AAACGCAACC GACACCCGAC TCCGCGATGT CGCAAGCTCT GCAAGCACGT GGCTTCCGTG
     2041  CGGTTAACCG AGATTCCATG GAGTCGGATT CGCGGGCGGC CGCTAAGGAC TCCGCCAAAA
     2101  CTCGCCGTGC CGCCAACCTT CACCAGAAGC TGGACGATGA GGATATGTTT AACTTCGATG
     2161  AGATTGCTCC AAGAAAGAAA CCTCGCATGG ACAAGGCAAA CGAAAAGCCG ATCACCAGCA
     2221  CGAACACAGC GGAAGTAAGG CAGACGGTTG CGGTTGGAGG TAGTCGTAGC CGACGGAACG
     2281  CTCCGGCAAC AGATCAAGAG GATCTGTTTT GCTTCGAAGA AACTCCAAGC ACCCAGCAGC
     2341  GTCGTGCTGC CGCTGCTGCT AACCAACGCA AGACGAATGC TTCAACAACA GTAATAATAG
```

```
    2401  CCATTCGAAA ACCAACCCAT CGTTGGCCAC ATCCGATGCG TCCATAACCC GCGCGCTGGC
    2461  CGGGGCACAG AATCGAAGCA AGGGAAGCAA CGATTCGAGG ATGGATGGGT ACAAGGAATT
    2521  TATTAAACCA GTTCAACCAT CGGCGATGGG CTGGCTGTCG TCGACGATGT GCGGCTTAAA
    2581  GCTGACCGGT GGCGGTGGTG AGGACCATTC GCTCGACGAA ATCAACATTA AATCGGAACC
    2641  GCTGGACGAA ACCGACGATC CGTACGGTAT GCAGCGGGAT TGCAAAAAGT GGATCAAAAG
    2701  CATGGAGAAT GCGTTTCCGG TGCGGGAGGT AAGCATGAAG CTGGTGGCAC ACCGTCCAAT
    2761  GGATGTCAGC GGCCAGGAGG TGTCGTGCGT TACGGCCGAT GGAGGCAAGA ATTTTAAAGC
    2821  ATTCGTAAAG GTAAGGATTG ATCGGGGTAC AGTGCGTTGC TTTAAAGCTA ACGTGTGTAA
    2881  TGCCTTCTTT TCTTTTACTT TCAGAAACGC AACTATCCGG TACAGCAGCT CGTTCTCGGC
    2941  ACGAAATCTG TTTGCGTGCG GGACGACAAC CAGTGTGCAT GAAATGTGCG AATGTTAATC
    3001  CCTTCTTTTA CATTTCATTC CCAACTTTGT GGCCCTATTA AGACAGCAGG CCCTTTTCTT
    3061  CATCTATTTA ATTACCAGTA ACGACACTAA AATACCCAAA AAGATACAAA TTAACACAGT
    3121  GAAATGAAAT GAAATGAGTA GACCAAATGA ATATAGACCA GCAGACGGAT GCTGTAAGTT
    3181  TCTGCTTTAT TGAAAGCGAT ACACGATTCA CTCACGTGCC CTCTGTAACC AGCTGTAGAA
    3241  ACTGTTCCTT GAAGAACTTT ACAAACTCCA CAGATGAAAG TCGGTTGAAT TCCTTCATCA
    3301  AGCGTACATT GAAATCACCA AACTGCTGCT CCGAAAGCCC TTTTGTCAGC GCTTTCGCCG
    3361  GGCTGTCGGT GCGATTGCTT CGCTCCTGCT TCAGGAACTG CTGCACGTGG AACTTGGAGC
    3421  GTATGTGGTC GAGCTGTTTG GCATCGAATT GCGTCGCAGG TGCATTCCGT CCCTTCCTCT
    3481  CC

//

LOCUS          54     3627 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      54
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
     CDS          complement(500..3127)
                  /translation="MNKLERTLQLLDESTVQSPSGLLRSKDSVRYLQQQAREEQRQL
                  GASGLGESMMLDSVQDVGLYRTVGELEPAGGGLSGSQLSLVAGSTVSYTNANVREAS
                  KQLCEEFLQVLQRYSNASDVFEATEELVKSLDDTLVQLNNASKQLLPTGRERRLHAE
                  EMWLNAERDTWRLMICLYRDRLITQKQDSEMDDLPLVNSEQTIISHLYGGNANLREY
                  QIIVDWLEQMALRQQQTELGHYMNRTVAWENTLHQLLEVGQTAFGSGRALVKGLDPD
                  APIREKLPLHDLDMEDQTQLAKQVFRAIRQGRLKEAQIKCENYGQAWKAALLEGWRL
                  HHDPNYDRENVSLAREPIEGNPRRDLWKKFAWQMAESRMLDPYTKATIGSLCGHLDS
                  MVEVLSEHSWNDMLWAYLKAQIDIRVESEIRSHCVKSYLPMPDRYWNGKMSLEQIFD
                  ALEAHKNARISLAAKDVDKVIQKYIMLDDIPELMRTVDGWLEGTEVVLIPQMLRFLT
                  HFVMFLRQIGKTFQEDIGDRVVKRYVEYLSTLGSAHMVAFYTAALPPAMQLLLYSQF
                  LQTISDSQQRKQALEQAYNFGLDVPTITVYTVEWYRALEATEDELTPIASGRLTELD
                  EQKISSLEWLTFYPDQRGELLWQSNTLIRYFLARRSIEAARKTFAVIPADTMELIFT
                  HYGSKDDIPCREEVSIREYLCYQTYLAAIQGYNDWCQLFYNGKPKPPPAVKLSNFTE
                  RVTSEHREATYRKELSNWELRVAELTNVTRDLLYNVLLFPEAGWLVDTDTTKPPPDD
                  DEDWLHRAVQMENLRKLCIPEIVLLLHQLHTLTERYSENLLLADVLSSETRKLYSVF
                  PKHKLAETLIKIAESSLTLMNARKDPFVGDAAKK"
ORIGIN
        1  CGCACACACT TTACCAATGG CTTACTTTAG CGGCACACGA TGAGAACACG ATAGAATAGC
       61  GCGGTGGACA ATTGTACTGA AGACGGCAGG CAGCAGCAGC AGCAGCAGAA GCAGAACCGA
      121  CACTGTTTTT CTTACCCGCA ATGTTTACGA ACTGGAGAAA TCCGGGTGGA GCGCGAGAGC
      181  TTCAACCCAA CTGTCAAAAT GATAAGCTCG CCCCAAGCAG GCCGCTTGCA CGCACACACT
      241  CGCAGCTGCT CCAGCCCTTG CTGAGCGTGT AGCCCATCTC GCTTTATCTT GCTCTTTCTT
      301  TCTCTTTCTT TCTCCCTTCG CTCGTTTTGG TTTCTTTCAC CGCAAGTGTG TTAGCCTTTT
      361  TGAATTTTGA ATTAAAAAAA TCGCATTTTT ATCTTACACA TACTTTATCA CTTACGATTT
      421  TATTTAGAAC ACGGTAGGAA AATTAAGAAA TTAAGAACTG TAGTAAATGC AGAGCTAGCG
      481  TGGGAAAGAA TGTATGCCTC TACTTCTTCG CTGCATCGCC CACGAACGGG TCTTTCCGGG
      541  CATTCATGAG CGTGAGGGAC GATTCCGCTA TCTTGATCAG CGTTTCCGCC AGCTTGTGCT
      601  TGGGGAAGAC GCTGTACAGC TTGCGCGTTT CCGAGCTGAG CACATCAGCC AGCAGCAGAT
      661  TTTCGCTGTA CCGCTCGGTC AGCGTGTGGA GCTGGTGCAG CAGCAGCACG ATTTCCGGTA
```

```
     721  TGCACAGCTT GCGCAGGTTC TCCATCTGTA CCGCCCGGTG CAGCCAGTCC TCGTCATCGT
     781  CCGGCGGCGG TTTCGTCGTG TCCGTATCGA CCAGCCAGCC AGCCTCCGGG AAGAGCAGCA
     841  CATTGTACAG CAGGTCGCGC GTCACGTTGG TCAGCTCGGC GACACGCAGC TCCCAGTTGC
     901  TCAGCTCCTT CCGGTACGTT GCCTCCCGGT GCTCCGAGGT GACGCGCTCG GTAAAGTTCG
     961  ACAGCTTGAC TGCGGGCGGT GGCTTCGGCT TGCCGTTGTA AAACAGCTGA CACCAGTCGT
    1021  TGTAGCCCTG GATCGCGGCC AGGTACGTCT GGTAGCAGAG GTACTCGCGT ATCGACACCT
    1081  CCTCCCGGCA CGGGATATCA TCCTTCGAGC CGTAGTGCGT GAATATCAGC TCCATCGTGT
    1141  CGGCCGGTAT GACGGCGAAC GTTTTGCGGG CCGCCTCGAT GCTGCGCCGG GCCAGGAAGT
    1201  AGCGGATCAG TGTGTTCGAC TGCCACAGCA GCTCGCCGCG CTGGTCCGGA TAGAAGGTGA
    1261  GCCACTCGAG CGAGGAAATC TTCTGCTCGT CCAGCTCGGT CAACCTGCCC GAAGCGATCG
    1321  GTGTCAGTTC GTCCTCGGTG GCTTCCAGCG CACGGTACCA CTCCACGGTG TACACGGTAA
    1381  TGGTCGGCAC GTCCAGCCCG AAGTTGTACG CCTGCTCCAG GGCCTGCTTG CGCTGCTGGC
    1441  TGTCGCTGAT CGTCTGAAGG AACTGCGAGT ACAGCAGCAG CTGCATGGCC GGTGGCAGGG
    1501  CGGCCGTGTA GAACGCGACC ATGTGCGCGG AACCGAGCGT GCTCAGGTAC TCCACGTACC
    1561  GCTTGACCAC CCGATCGCCT ATGTCCTCCT GGAAGGTTTT GCCAATCTGT CGCAAAAACA
    1621  TGACGAAATG CGTGAGGAAG CGTAGCATCT GCGGTATCAG CACCACCTCC GTGCCCTCCA
    1681  GCCAACCGTC CACGGTGCGC ATCAGCTCGG GAATGTCATC CAGCATAATG TACTTCTGGA
    1741  TCACCTTGTC CACATCCTTG GCGGCCAGAC TGATGCGCGC GTTCTTGTGC GCCTCCAGTG
    1801  CGTCGAAGAT TTGCTCCAGC GACATCTTGC CGTTCCAGTA CCGGTCCGGC ATGGGCAGAT
    1861  AGCTCTTGAC GCAGTGCGAC CGTATCTCGC TCTCCACGCG TATGTCTATC TGCGCCTTCA
    1921  GGTACGCCCA CAGCATATCG TTCCACGAGT GCTCGGACAG CACCTCCACC ATCGAGTCCA
    1981  GATGCCCGCA CAGCGATCCT ATCGTGGCCT TCGTGTACGG ATCCAGCATG CGGCTTTCCG
    2041  CCATCTGCCA CGCGAACTTT TTCCACAGAT CGCGCCTCGG ATTGCCCTCG ATCGGCTCGC
    2101  GGGCCAGTGA CACGTTCTCG CGGTCGTAGT TCGGATCGTG ATGCAGCCGC CATCCCTCGA
    2161  GCAGGGCCGC CTTCCACGCC TGCCCGTAGT TTTCGCACTT GATCTGTGCC TCCTTCAACC
    2221  GGCCCTGGCG AATGGCGCGA AACACCTGCT TGGCCAGCTG CGTCTGATCC TCCATGTCCA
    2281  GATCGTGCAG GGGCAGCTTC TCGCGGATCG GTGCGTCCGG ATCGAGCCCC TTCACCAGTG
    2341  CCCGGCCACT GCCGAACGCC GTCTGGCCTA CCTCTAGCAG CTGGTGGAGC GTGTTTTCCC
    2401  ACGCCACCGT CCGGTTCATG TAGTGGCCCA GCTCGGTCTG TTGCTGCCGC AGCGCCATCT
    2461  GCTCCAGCCA GTCCACAATG ATCTGGTACT CGCGCAGATT CGCATTGCCA CCGTACAGAT
    2521  GGGAGATGAT GGTCTGCTCA CTGTTCACCA GCGGCAAATC GTCCATCTCG CTGTCCTGCT
    2581  TCTGCGTTAT CAACCGATCG CGGTACAGGC AGATCATCAG CCGCCACGTA TCCCGTTCCG
    2641  CGTTCAGCCA CATCTCTTCC GCGTGCAGCC GGCGCTCCCG ACCAGTAGGT AGCAGCTGCT
    2701  TGGACGCATT GTTCAGCTGT ACCAGCGTGT CGTCCAACGA TTTCACCAGC TCCTCGGTCG
    2761  CCTCGAACAC GTCCGAGGCG TTGCTGTAGC GTTGCAGCAC CTGCAGGAAC TCTTCGCACA
    2821  ACTGCTTGCT CGCTTCCCGG ACGTTTGCAT TCGTGTAGCT TACCGTGCTG CCAGCGACCA
    2881  AGCTAAGCTG ACTGCCGGAC AGTCCGCCTC CGGCCGGTTC CAGTTCGCCC ACCGTTCGAT
    2941  ACAGGCCCAC ATCCTGCACG GAATCCAGCA TCATACTTTC GCCCAGCCCG GACGCTCCCA
    3001  GCTGTCGCTG CTCCTCGCGG GCCTGCTGCT GCAAGTAGCG CACCGAATCC TTCGACCGCA
    3061  GCAATCCGCT GGGCGATTGG ACAGTGGATT CGTCCAACAG CTGCAATGTG CGCTCGAGCT
    3121  TGTTCATCTT GATGATGTGA CACCAGCGCC AAAATCCGCA ATCGTTTTCC TTTATCGCAT
    3181  TCAATTCCTT CGCCTTGCCG CAACAACACC GAGCACCATA ACCTAAAACA CAGTCGAGCC
    3241  ATTAAAATCA CATCGTTTCG CTGCACTTCC ACTCGTGTTA TATCTTTACT TACAGGGAAA
    3301  CTTCTCACGT AGAACGAGTA AAGTGTCTTT TATTTCAAAC TATTACAAAA ATAAGCCTAA
    3361  ATGGCTTAAT TTAGTGCGGC CAATGCAAAC GAAACCGGGA GATCGCCAAG CTAGCGCGCA
    3421  AACCAATTGT CAACAAATCA AAACGAGCAT CTGTTTGACA GTTCTGCATG CTGGTCGCGG
    3481  CCCATTCAAC ATTTTTGGAA CAACAGATAT TTTGAACAAA AAAAAAAACA AACCGATGCT
    3541  AAAATATACA AAATTGCTGT TAAGCTTATA CAAAAATTGT CATAATTATA TTTAACTTAC
    3601  TTACTTTTAT CGAAAAACGC AACAATA

//

LOCUS          55     28648 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      55
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   A.gambiae
FEATURES            CHR  ---
```

```
     CDS
join(500..512,1080..1219,8388..8506,9744..9819,17559..17789,24830..26933,27027..
27519,28136..28148)
                 /translation="MVVVERVLENSYHGLIKENETFVEITPLIKVDETKICGFHIIK
                 KNKEIPFQIELVNELGILKAKKTLNCEKRKNYKFDITAVFCDGSHSKSASVHISVID
                 INEYSPTFLQPSYVTECACDRFRPSSSKGRSSIQMALTRFEREAEGAWNPQIFVIVT
                 ERQSASKSIGTAFPQTTPHGQFSPNASKPPLRRRSQVDEGRLYQEIIRVEATDKDCT
                 PLFGDVCKYEILTNDQPFTIDNEGSIRNTEPLSHKISHNHILSVVAYDCAMKQSAPV
                 MVNIRVRRVCEAHVMGVAERIDYVANSMESVSLFPKIHLELCDMQCKGEDMVIQSSV
                 SLKTKHISFGCDRDASQCLRRLSESGSGRERDGDGLVDLLQKNTDWTKDLTYDEGAE
                 AIFHFDGSSGAIVPKNVVQPQDFAAHQFSIVTMFRHNSIVSTDKHTKEHIVCSADDH
                 KMNRHHMALFVRNCRLILLLRKDFNDGDLNIFSPAEWRWKIPQVCDNEWHHYTINVD
                 LPKVDLYIDGIRFESNVEDRHSNPEVIDDWPLHAAHGINTTLAIGACYQGSENRLKH
                 GFSGDIAEIKLSTRKVLTASQIRCGTDCAEKLLPPADHYLEPEQQIQSNTQMNKIII
                 EGSNKTNIEQLLQQIQYINSKESPTIGRRNIQVMTTVSCPNKKAIRLPTIDTYIMVS
                 SAGSSLTGGSMLASATSSSSSSGLYNKILVDKPLATDQKPQIVISGNSNHLVSYPDI
                 KEGVKILAQINISVYTGKDIRPELQKLDSCNVNVFPSLNPDHEEITIDDKDGADESL
                 FKFDIKTQISKEGVEMIGYDTIENYQHVLKSLVYINKKPAYYLNRVFKLTCSQVDDH
                 FRSAEYTLTLTVLHPKQQPTSSIASASAPSSNVLLAISTAPSEQHPAVAVDSHDNAN
                 QYAHVMLHSHDVQESQSKVRSSGGAALLGHEKLASSHSTMLIVVICGSFVLLICGVG
                 IARLRNSNAALAGGSSGHGAAGGAAGGSAFARNLTDKHLPCPKVAPDQQLDWDDSAL
                 TITINPMQHDALSDESSESDNSDSEDEEVLNGRYKNVSQLEWDNSTIASA"
ORIGIN
       1  CTATTAGAAA ACTATTTCGA AATAGAATTG GAAATATTAC AACATTCCCG ATGTTTGTCC
      61  ACAGGAACAT GACGCTTTGC GGATGAGTTT TGCCTCATTA AAAGCTCTGC CCCTTTTCCG
     121  TCCAGGAACT GCTTGACCAT GAACCGGTAT GGAATGAAGT CTGACGCCAA AATTAACGCT
     181  CCTCCACACA CGCCACCCCA AAATGTTTGC CGCTTCCAAT TACATTGGGA AAATATTTGT
     241  GTGGTCGCGT AATGAATATC GATAGCCACC GCCGCCGAAC AGTGTCGCAA CATGCCAACA
     301  TGTGTGTACT TACTACAGCC GTGTTCACCG TGGGACAGTT TCCTGGGAAA AATGTAATAC
     361  CCGTGATAAC AGGCGCATTG AAAATGCATG GTAGTTATCG AGAAAATTAT TCACTTCCAG
     421  TGTACTTCCA GTGTGTTGGA ATGGCGGAAC TATCGCGCAG GAAATGAAAC CAATATCGAT
     481  TGTACAGTAC CGAGCAAAAA TGGTGGTAGT CGGTCTGGGA GTCTAATGTT CTCAATTCCT
     541  GGAGTTGCAA TGTGTGCAAA ACTGCACCGG GGCGCTTGTG CCGGAACGCG GGGCTTGGAG
     601  CGCTTGTTCC GGCCGGAATG TAATACGCAC GGCAAAAGGA AACATCGTGC TTTCTCTGAT
     661  GCGTGTCAAA CGATATTGAT GGAAAGCTCT GGCACTGGAA ATCAGAGCAA TGGGAAATGG
     721  CGAACAGCAA ACATCCGTGT GCGCAGTAAT TTCGTGTTCG AAGGGCTTTA CTAATTGCAA
     781  ACAGTACAAA TTATTAACAC GCCGATGCCG CATCATTACA GGTTGATGTT GGATTTGGGA
     841  AATTAAATAG TGTGTGTGTG TGTGTATATA TTGTATTTAT GTGGTTTTTG TGTACTCCTA
     901  AACCTTGTCA GGTATCTTTT CCTGCTGCAA ATCATTTTAG GAGTTGGTAT CTACATCAAT
     961  TTGTATCATT AAATTCCTTT ACTTTTAAAA CTCAAGCATA TCCCTGCTGA AATAATAATA
    1021  AATACTCAAC TCTCTTGTAC ATTTAACTTA TCTTTCCAAT CCCGTCTTTC TCTTTGCAGA
    1081  ACGCGTCCTG GAAAATTCGT ACCACGGCTT GATAAAGGAA AATGAAACGT TCGTCGAGAT
    1141  AACGCCACTG ATTAAGGTGG ACGAGACGAA AATCTGCGGC TTTCACATAA TCAAAAAGAA
    1201  CAAGGAAATT CCCTTTCAGG TAAGCTTCCA TTCGTTACTT CCATTTTGCT CCGTTGAATA
    1261  ATGCATCGCA TAAGGAAGAA AACCAATTGT AGCTTGATTT GCTGGAAAAA TCAATAATCT
    1321  TCTGCTCGTA ACGCTACGAA ACGGCGTCAG ACGAGGCAAA CCTCCCCCCG TTTGTAAGAA
    1381  GACTCAAGCG CGATTTCTTA TCGATAACGA AGTGGAAACA CTTGGGCTAT CGAGCATTAT
    1441  TGGCAAAAGT GGCACGATTT GCAGAAAAAG TCGAGCGCGC CAGCGCGCAC GTGTCACCAT
    1501  CTTGTGCATT TTTCGTCGCC AATTCCTACG TGCCCCAATG GGGGGCGGAC CGGGTGAAGC
    1561  GCGACCGGGT TGCCATTTTC GCACATTGCG AAACGTCTCT TGTCCATGAA ATTCGACTAA
    1621  CTTCACTTTT AGCACGGTCA TATATCAACC CGGCGGTGGC GCGGGTTCGT CGTCTCGTGC
    1681  CATCTCCGCG CGCCGAAATA GCCAATGTCC GTCCACGGCA CGGTGACAGA AATTGTTTCG
    1741  CTAAAGTAAG GTGGAATCTG TCGGACATTT TTTCTTCCTC CCACTGTACA CCTCCGTTGT
    1801  CTGGGATGGG TTTTTTTTTA TAAACATAAC CACTTCCTTC TTGCGAGCGC TTCTAACAGA
    1861  GAAGGTGGTC CACAGACCAT CGTCATTTTC AGCGCATTTC TTATGCCGTC AGCATCGTAA
    1921  TCTGTTTTTA CTTTCCTCGT CGAGCGCTCT CTCTTTCTCT CTCCCTCTCT GTCTCTCTCT
    1981  CTTTTTCGAG GAGTTCGAGG ATAACCTCTT GGGGACACGA GATAAGAAAC TTCATCCTTT
```

2041 TCATCTGCTC GAGAGCAGCC ACCGGTACCA TGCTGAGTTT CGGGTCTTTG GCGTCGCTTC
2101 TTTGCCAACG TCCATCAAGG TCCATCAAGT ACGAAAAAAT CCTCCACCCC CGCTCCCCCA
2161 CTGGTTTCGG CACACCAGCT TACCGTGCTC TTCTTTAAAC GGTATCGCAT TTACTGACAA
2221 AATCCACTTG AAATTGCTGC TAAGCGCCCG GGAGGGATAA GGGCTGGCGA GCTTCCTTCC
2281 TTCTGTTCCA CCGGTTTAGG GTGGGACTAA GGAGGAGGGG GATCTGTTTT ACTTTCATCA
2341 CACTGTAGCC AAAATGTATG CAAACTTTCC CCCCTTAGCC TAAAGCACCG ATCCACCCGG
2401 CACGAACGAA CCAAATTTGT CCACCACTAA CGATGGCCCC GGAAATGGGT GCGGAACTGG
2461 GACGAAAAGG CGCTTTTTTT GCTGTATTTT TGGGAAGCGG TCGCTTTTTT TGTTAGGATC
2521 TTCAAAATTG GATTCCTGGT ACCGTACCTC GGGTGAGAGC ACCTCCCCCG GCGGTGACAT
2581 TGCGTCGGCC GGTAATTAGC GCAAATGATT ATTCCTTGAT GTCCACCATA ATGAGGAGCT
2641 CCCACCGGAA TTTGTGTGTC TGTCTTTTCT TTTTTTTACA TTCTGCGTGT TTTACAACGC
2701 ACAACATCTG TACGCCCTCA CCAAGAAGCG TAACTTGTGA AGGTCGCGCC ATTTTTTTTC
2761 GTCGTTGTGT CGGTGTGAAA GCGTGTCCGA TGCTCTTAGG CAAAGGTGTG CACAGACTTC
2821 CTTCGCGCGT TGTTTCCTGT ACCCAAGCAA GATATGGTAA GGGGAAAATA CTATACAACA
2881 AGCGACGACT GTGCTGACAG CTAAGCCTCC ACGGCCTTCC CTGATCATTA GCATAAGTGT
2941 TGCTAGGCAT CCACGTGGCA CTGGCAGGGG GGATGTATTG TGGGCCGGCG AAGAAAGTCT
3001 CTAGGGAGGT GTGAAGCTTT TCCCGGCAGC TTCCCCGGTG ACGAATGGGA ACATCATGGG
3061 TGGCTGCTAA GTTTGATGAC GCAAAGGGAG GCAGCCGAAC GTCTCAAGTT TGCAACTTCG
3121 GCTCTCTCAT TCTTCGCTCT TCCAGTGGGT AATTGTACCC CACACCCTCT GTGGTGTGGA
3181 GTTTTCTCGT TTAAGAATTT TCACATTTCC AATTTCTAGA AGGTACAAGT TTAAGTCTGG
3241 AGAGAGAAAT TTCGGTATAC AGGCGATTCT GTAACGTCAA CCATTAGCAT GGGAAACCTC
3301 TTGAGATGAA ACGGACCTCT GTATGTGTGC TTCCGAAGCA AATCAATAAC GGAAAGCTGT
3361 AAAGGGGAAG GGAAAAATAT TAAATTCATT CGTAAACAAG TGAAACTTTC TGCAGCGGAA
3421 AGCAACTTTG CGGGGCACCA CTGGGGGCTG CGCTGTGCCA TTTAAATGTC AGCCCAAGTG
3481 TGTTGTGTTG TGGGTGCTTT TATTAAAAGT TTAATTGGAA AACGAGGAAT GTTCAATTAA
3541 TTGACATTTT TTAGCTAACT CCCCCTTGTC GGTGCGTGTG TGATGTAAGG AATTTGATAC
3601 GCAACATTTT GGGGAAAAGG GCTGAAAGGC GAAGATGCCC CAAGATGACG CTACGGTTCT
3661 CAATCATTCA TGCCCTTTTC TCCGGTGAAG CTAACAGCAA AAAAAACAAG CGAACACGAT
3721 CCACTTGTCC GTTCCATTGT CCTTTTTTCT CTCTCTCATT GCCCCGAAAC CTTCCGTCAC
3781 GGGTCACGGC GAAGGATCGA AGTGTCCGAG CGTTTCTGAG GACTCATCAT GTCCGACATT
3841 TGCTTAGGCT GCATGAAAAA TTATCACCCA GCCCCCCGGG GTCGCAATTT GTATGTGTGT
3901 CCGAGCGCGC GCGCGTTCTG CGTGCATAAG AACCCCCCCC CCCTCCCCCC TTTCTCCCCT
3961 CGGACGATGC CGCAGTTGGC AGATAATTTA TTATCCATAA CACAAAACGC CGGTTTGATG
4021 GTTTGATGGT GAGCTGGTCC TCCGTGTGTC GTCAGTGGGT GCACGCAATG GGTGGCGGAG
4081 AGAATCGGTG CCAAACCGAT TATGCACACT CCGGCAATAC CGCACACAAA CCGTTGCTAT
4141 GCTATCAAGG GTAGTACGGT ATGCTGGATG TGCGCGATGC CTGGTGACGT CCTGTGGTCT
4201 GCGGTGTGCC AAGCGCGATA TTGTGGGCGA TCTCGCAGCA ACCCTGGGAG CTCCGTGTGG
4261 AGAATGGCGT GTTGGTGAAA CGTTATTGTG TGGTCTACCG ACAGGTTCGG CGTCAGGCTC
4321 TGTGAGAATC TCTCCTTTCA TCTCATCCCA TTTCTCTCTC TCTCTCTCTC TCTCTCTCTC
4381 TTCGCTCCCT TGAGCAGCCC ACAGGATCGT TAATTTAATG TCCGGAAAAG GAGGCTCAGG
4441 TTGATAGTAC AGGTGCTGCT GCACACGGTA GCGTCGGTAG CTCGCTGGTT GCTTGCTTTT
4501 TCCCCCCCGG CAATAGAAAA TGGCATTGCG ACCCGACGTT CCTGGCACTC GAGCCCGGGT
4561 ACACACAGTA GCAGATGTAT TATTGCATTT AAATGAACAA TTAACAAACC GCCCAAGCAC
4621 GGAGCGTAGC GCTAGCAGTG CCTGTGCTGT TATTCAGTGC ACTTTCTTCA AGAGCACTCC
4681 TTACCCGCAG CCGGTTGGCG GCGGCAAACG GTTTTGCCGC TTTGGCGACA CACGACAAAC
4741 GATTATTGGG TGGCTCGCTT TCCACTGTTC ACGGCTCGCC GTTGTACACT TTGGCGAAAC
4801 ACTTTAGCCT GCAAAACCCG TTTCCTATCT TGCCCATTAG TGGTTTTAAG GTTGCAATTC
4861 GCTCGAACCT CAATTGTTTT TGTGGTGTTT TTAAGAAGAA CCCAAGCACA CGAGTGCCAT
4921 TCTGCGTGGC GACGATAGCG ATGGGACGGA CACACACGTG TTATTCTAAA TTAGATTTTA
4981 ATTTATTTCG ATGGCATTTA CTTTCGCAAA GCCAAACAAC AGGGACACCC AGGGACAAAA
5041 TGGAGTGGAG CTAAATTAAT ATCCATCACC TTTTTCAGGC TGCCAAGTCC AGGACAGGGG
5101 TGGGTTTTAA AACCGGGGTC TCACACTAAA AGTGATGGGC TTATAGCATA TGGGGTTTTA
5161 ACCTCCGGTG GCTGTTCCGC TGACACGTGT TCGCGCAAAT GTTTGATGAA AGTGGCAAAT
5221 TATGCAAGCA ACAATTAATT AATTATACAT TCTGCTTGAC TTCTCGGACA ACGTGAAGGG
5281 GCCGCGAGTT GTGAGGAGAT CCTTGCGTGG AGAGAGAGAG CAAAATAAAA GCAAACTATT
5341 AATGCAACAT GTTCTGGAGC TGGAGTTCAT ATGCTGAATA TGTGTTGGAA TTGAATTTTT
5401 GATACTCATG CTCCAAATAG CAAACAAATC TCATGTTTGA GCCCAAAATT ATCGTTTAAT

5461 ACCTTTCCAA GTATTTTAAG CCACTCAATT TCATGTTTTC TAAAGTTTAG TTTCCCTCTT
5521 GCTGTATGAG GGCATATCGC AAAGGGCGTT TGAATTAGTG TAAACGCCCT TGTTCCCTAT
5581 CCTCCCTTCT ATTCTAGCAA TTTTAATTAC CTCCACACCA ATCCTTTGCT TTGCCCTTTG
5641 CTATAACCGC CGCTTTGAGT CTATTTCCGT TTAGTTGTTT CAAAATTAAA TGAGGGCGTC
5701 CCCCAACCCC ATTCGAGTCA AAACATTGGT GTGATTTGTT CCCTGCGCAA ATGCTCTTAC
5761 TACGCTATGG ATCTATTGCT AGCAATGCTC AATCTGCCGG AGAAAGCCAT CGAACCGCCT
5821 TTGATGACGT TTGACACAAG CAGCACATTG TGTTACAACG TGAAACACTC ATCTGCAGCA
5881 TCCAGCAGCA CAAGGAAGGT GAACAACACG GAGCAGCTCC TCCACCCCTT GTACCTTTTC
5941 CTAAGAAAAC ATTAAAGCGT AACACAACAC GGGCTTCAGA CCCTCCGAGC AGAAGTTTTT
6001 AATCTTCCCG CAGTCCCCGC TGTTGCTTGC TGTAGAAGAT AGAATATCAA TTTATTTTTA
6061 ATCGCCTTCG AATCACCGGT GCACCACCTG TCGTGACGTG AATCAAATAA CTAAAATGGA
6121 TTTGTTTCGA TGGCACCTTA GCGTTGGGAT GGGTGAAGGG AAGGAGGGGG GCTTTTGTTG
6181 TTGCGGGGGT CCCGATTTGG AAACATGCGT CGTTGGGGGG AAGAAAATCA AACCCTGCAG
6241 CGAGAAAACA ACTGCAACAT GGAGCATGGC AAGCGGCGTG CGGGGGGGGG GATAGAGAGG
6301 TACACAGTAC GGTGCCGTGA CCAGGATTGC GTTCGGATTA GCATTTTGAT GAAGGAACGT
6361 GAAATTCCCG TGAACGTTCT TGGGCGAGCG CGCGCGCGCG CGCCTTCGTG TGCGTCAGCG
6421 AGCGATACGT GTACGGTGAG TGGCCCCGGG ACGCCCATTT CTAAGCGCAT TCCATTCCTA
6481 CACACACACG CCGGCAGCGC ACGGTGTTTG GAGATGACAG TTGAACGTTG GTGTTACAGT
6541 TTTCGGGACG ATATTACGGG GTTGCTGGCT GCGCTTTTCT GCAGAGAGTA AGAGAGAGAA
6601 AGAGAGTGCG ATGGATGCAA TCGTGCCTTG AGGTTAAATT GGTGAAAGAT TGGAACGGGA
6661 TTTACTGTGA GTCTATTTTC TCGCACAGAA CGTTGAAGAA TGCAGGTTTC AAAGCTAGCA
6721 GTTAAGATTA GGATAGCAAT CTAGGGATAT TAATGATAGC AAATTATTTG ATTGTTATTC
6781 AACTTCATAT TTCCATATCG CTTTTCTTTC GTCTTGTTCA ATCGAGCCAA CATGCAGTTT
6841 GCGTAATTAT TGGTATACTT TATCCCATCT AGTAAAGCTA TTTAATTTAA TTGCATCTTT
6901 TCCAGCCCGG TGGTTTCGGC TTGAATTACA TCCCTTCGCC AAATGCTCAC AGTTGCTACA
6961 ATTCGAGAAG CAATTTTACT GAGCTTAATT AATTTATCAA ATTGCTCCCA GACCGTTACC
7021 ATTTCGGGGG GAGCGAGTTA GTTAACCTCA CACAATATTT AGCATCGCCA CGCTTTCACG
7081 GTCTGATCTA TACGCCCCAG GGCTTCGCAT TCCAATCCAG TTGGTGTCGA TTGCAAGTTT
7141 CAAAAGCGTA GTTTTGTCAA CTTTGCGCAT AAATAAATGT CGAGACAACA TATCCCGCAC
7201 ATGCTTTCGG GAGTTTTGTG CTGCCAGCAT AAAAAGAGTA GTGCATAATT CTCTGTTGGC
7261 AGTAGCTGTG AGCTATCGAA AATTCATTTT GTTTTTTTTT TATGTAACTC CGATCGCTTC
7321 ATTTGCTTTC GCAGATTCGG TTGCCGGTGC CGAATTTCGC AATTATGCAA AATTGTATCC
7381 ACGCCACGCG TGTGCACTCT CTGCCACTGC TCAACAACGA TATCAGGCGT TATTTAAACG
7441 ACTGTGCGCT TTATGCTCGC GCTGATGATT GGTTGGGTTA AGTTTTTTTG TTTGTTTTTT
7501 TTGTCCTTCA CTCTTCGTTT GTTTTTGTAA TTCTTTCGCG CGTACTCTTC GCCCGCAAAC
7561 GCTCACACAT TGGAGTGGAC TTTTGTTTGA GCGGTAAAAT AAGCGTGGCC CATCCCCCCT
7621 TACGATAGTG GTGGAATTGG CCATCGAGCA TCGACTGAAC ATCCAACCAA ACTGAGCTGT
7681 CCTTTTATGC TGCAGCCGTT CAGCGATTTC GATGGCCCGC GTGTACTAAC AAAAATAAAA
7741 CATTAATAAC TGGGAGACGA CCGTGGCAGC ACAACACACT GCGGCTCATC GAGGAACAAA
7801 AAAAAACGCA TATCTGCAAT TAACAGATGC AAGTCCGTTC GTGTTTTCCA CCCCACACTA
7861 TCCCTGATGG GGCGAATGGG TCAGTAAAAA AAAGCGATGA AATGAATTTA TGATTGCCCG
7921 TGGGTATTAC TTTTGTTTCG CTCATCTGTT TTCCTTTCAC ATCGTAATGC CACCGATAAA
7981 CACGTGGGAA AACAGGCGAT GCATCAATTT TACAATTACG CAAAGATGCA TGAGTGTTTG
8041 CACAGCAGGA AAGGGTGTAT GTATGTGTGT GTGGGTGCTT TTTCCACTTA CTATTTAAAT
8101 TTCAAACAAT CGAACGATGC CATTCAGCAG CGGCATCTAT CATTGTTTGT GTGCTGTTTT
8161 GTTTGGTGTT TCTGTCGAAT AATCAATATT TTTTGCAGAA CATAAATTTG TGATTCCACC
8221 CATGGGGATA GGGAGGGAGG GAGGGAAGGT GACATAAGTA TTTCCTCTGT TGCTGCGTTT
8281 GTTAGCTATT ATTTATTGTT GGGTTGATTA TTTTCGATCG GTACGGTAGA AATTAATGTA
8341 CGCGTGATTT AATGCCTGTT ATTTGCTGTA TTGTATTCTC TTTTCAGATT GAGCTTGTGA
8401 ACGAGCTTGG CATTCTGAAA GCGAAAAAGA CACTGAACTG CGAAAAGCGG AAAAACTACA
8461 AATTTGACAT TACCGCCGTT TTCTGTGATG GATCACATTC GAAGAGGTAA GAGACAACAA
8521 TTGTTCTATA ATTTATGGAA AAATAATGTT TTGATGTTTT ACGTAAACTA TCTCAATATA
8581 ATGACGTTTT TAATTTTCAT TTTTCTCGAC ATGAAATAAT TTAATTGCCG TTGTTGCACT
8641 TGTTGTTTAT TCATTTTTGA AGCTATTTTG AAGCTGCTTG GCTAGATTTG CTTCTAAATT
8701 CACTAGATAA TCTATTATGT TACAAGTACA TCATATTGCA TACCTTCTGG CGTTCAGCAG
8761 GCACAGTGAA ACGGAGTAAA ATTAATAATC AAATTTTGCC TATTTCGATT ACATTTTAAT
8821 CTATTTTTTA CATTTTAATA TTTGCTCGAC CAAATCGCGT TTGATTAATA TAATATATAT

```
 8881  AGGATAACAA GTCTTATTTC ATTACGAATC CAGCAACTCA ATTGATATGC CCCTATTAAT
 8941  CCCAAAAGCC CTACGAATTC CATAACTTTA GGCGCCCAAA CTTGAAACTT GTTTTGTTTT
 9001  TATACATTCA ACCAAAATTG TTCAATTATT TCCGCCTAAC TTAAACATCT ATTCGCCCCG
 9061  TAACTAACGT GCCTTCAAGC CGAGCAACAA ACGTCGGCCT TAAATTGCTC AGCAGTTGTG
 9121  AAAAATGTGT TCCATTAAAA TGTTGCTAGC GCCTACTTTT CATGTCAGCA ATCACGGCAA
 9181  TAAATGCGCC GCAGAACGTA CCACCCGCCC CCAAAAACAA GCGAGCGACA TACAAGCAAT
 9241  CGCGCATGTA GAGCTTGTTT TGTGCTGGTC CGTCCACCTT TTTGCTTCTT CTATTTACGT
 9301  TCGCTACCAT CGTTCCACTT CGCACATTGC CAAGGCTTCT TCACTTCAAT CACTCCGCGG
 9361  AGACACCGAA CCGAAAGTTT ATCTGACAGG CGAAATATCG AAGCACTTTC TCCCGTCGCA
 9421  ATGTTTCGCG CTCGCAGATT CCTCGTCCAT TGCGCTGAAC TTCCCAGGAA GCTTTGCTAG
 9481  AAATTACTTG ATGCGTACGT CTGTGCTTAT CCTTCCACCC AGCCGAACAC ACGAAATTAT
 9541  TCGGCACTTT GGTATCTTTG GAGAAATTTC TCTCCTTGTT TCGAGTTAAA TCCACGGCAA
 9601  CCCATGCTGT CCTTTTTTGT TTTGTGTTTG TTTTTTTGTT TTTTTTTTTT GGTTGGCAGG
 9661  ACTGATTACG ATTACGAATT TCTTATCCCA AAAACTTTTT TTTATCTCAC TCGCTTTCTC
 9721  TCCAATTCTT CCTTCCGTTT CAGTGCGAGC GTACATATCT CCGTGATTGA CATTAATGAA
 9781  TATTCGCCCA CATTCCTGCA GCCCTCATAC GTCACAGAGG TAAGTTTTGC AAGTAATGAA
 9841  CTTTCCGTCC GGCTTCTCAC TTAGCAGCCC GAGTGCGTTG GGCTCTGCCT TAGACTACTT
 9901  AGCAGCAGTG TTCGGCTGCA GTTATCAAAC ACATTCATCC CGACCTTTGG GTGGTGTGTA
 9961  CCTTTGTCAA CCGGTGCGTC AATATCGGTG GCTCTGACAC GGCCCTACTG CACCCGGGCA
10021  TCTCCAGCAT GTGCCAGTAA TTAGTGAATG CTTCTTTGCC GTCCCCGTAG CCCGCGTTAT
10081  GCGTCTTTGC CCTGCGAAAA CTCCAAAAGT CAGCAAACTT CCAATCCAAG CACATCGGTA
10141  ATTTATTGCT CGCCCAGCCC ATCCCCGGAG AAAATGGTTT CGGACCGTAT GAGATTGGTT
10201  TCGGTATGCG AAAGTGGAAT AAGAGAGAGA GACCGACCGA CCGCCCGAAA TAGAGCGCCA
10261  GAAGAAGTTT CTTAATATTC ATCATCCATC TTCTGGTTTG CGAATTAGTT CCACCTCAAC
10321  CACCAGGCGT CTCCATGGGG GCGTTCACAT TTCCATCACT TTCACCGTTC GCACTCCGTT
10381  GTAAGTTACC GATGGGGATG TCACAGTTAC TTTTGCACTT GAAGCAAATG AACATACACA
10441  CACACAAAAG AAACAAATAA AAAGTGGAAT GGAATGCCAC CCGAACCGGA CACGATCGTG
10501  AATAGATGGC CCACACGATT CATGAATGTC TTTTTCATTT TCTCTCTCTC TCGCTTTTCT
10561  CGACCCTCAC CGACGCCTCA TCAAACTCAA TTTTCGGCAC CTTTAGTCCG GTTCGTTTTG
10621  GTGTATTTGC TGCTCATCAC CCGGAAAGGT CTGCCCGTTG AGTATTTTGT TTCTCGTTTT
10681  TTTTTTGTGC CTCTCGATTC TACCGAGGCG TAACTATACA CCAAAGCAAT TCTGCAAGCC
10741  CCGCTGGCCG ATTTTCACAG ACTCGTAATT AATTCAATGT ACTGGTATTG ATTTTTGCTC
10801  CGTTACAAAA AAACCGGACC CTACATTGAG CGACGTCGCC CACAAAGCAA AGCAAAGGTC
10861  TTCTGACATC GCTTTTTTTG AAATAAAAAT GCTCCTCCGA AACACTTCAA TTATTATTGA
10921  ACACAGTGAC ACACACACAG ACACATTGAA CGCACGACTG CTTGCTTGTG TCGCATTGGG
10981  GGACTGAATT ACGAATTTGA TGAACGCCCC GAACCGTGAC AACCTGGACT TGGAGGGTTT
11041  CTGGGAAGAA AATAGCAACG ACAAAAGAAG CAGAAATTAT GACCGGTACG AATTCCAGTG
11101  CCTTTCTGGG TGTGCGAGAG AAAATCGATA CATTTTCCAA GAAAGTTGAT AACACACTGA
11161  AATGTCACAG CGAACAAAAA CACATTCTTA GCTTCAGATG ATGCTTCGGA AGTGGTTGAT
11221  CAAGATTGAT TTTTCGATCC CACTTTGTTT CGGCTTCACT TGCGCTTTCG CGTCGCATGA
11281  AACGGCTGCG ATAATGAAGC GCGATTTTCC ACAATTAAAA GCAACATTTA GCAAAAGAAA
11341  AACATAACCA TGGAGGGAGT TAACCTCCAC CCCCACTTGA TAAACGCCAC GAGTGGTGAA
11401  CCAAATTGAT AGCATTTAAT GTTTGATGAG CGTTAATGGT GAGGTTCCGT TTGCCACTTT
11461  CACGCTATTC TATCCATTTG CGCGCTGTAT GCGATTCCGA ACCCGTGGGA GGAAGGAAGA
11521  AAAAAACGTT CACTTTTAAC CACTTTGAAA GGGTCATTAA CGGGGCGGCA GCGGCAGGAG
11581  AAGATGGAGG TCAATATGAT GCAGCGGTAA AAAGCAATTA AAAAAGCACA AAAGCGTTTC
11641  CATCCATCCA TCCATCGGGG GCGAAACCGC GCATTCTGCT AGGGACTGGT GCGCGTAGGC
11701  GTTTACCAAA CCTGCCTCAC CTCTCTCCGT TTGCTGCAAC ACTTTGACGA AGGATTTCAA
11761  GTTGCTTAAT TCACGTCAGG ATGAGCTATG AAGTTTGACC AAAACAAAAA AGGCAGTTTG
11821  GACACGGCAA AAGTTTCCCC GTTTAGCGAG TCATTATCCA TAAACAGTAA GTGAACTGAA
11881  ACATCTGAAG TCGATGTCGA AGTCGCGCTG TCAGTTAAAT TGGATTGAAT GTCATTTAAT
11941  CCCTCCCAAA TGGCGGGAAA AGTATTGCAA ACTTCCTGAA AACTAAGTTA CAAAATAATA
12001  AAGGCCAGAT TTGATTGAAC CACCACAAAA TGGGCTTTCC TAACAAGAAT AATGTCCAAA
12061  CGAGCCACAG TAAATCACAA GGCATCAGCA TACTTGTATC GCGTTTAGCC CAACCGCAGG
12121  AGCTTAAATT ATCAACCAAG CACTTTTGTT ATCAATATCG CCCCTCTCTC TCTCTCTCTC
12181  TGCTATGTAC AAACCAAAAG GCTCATCCGC GTGCAAACTG CATCCTGCAA TATTTAACAC
12241  ATACATGCAC ACATACTGCG GATGGATTGC GTGTGAATAG CGCTTCGGAA CAAAACACAA
```

12301 AAGGCTCGGA GTTTGAAAAC AACGGCACAG CAATCTAGCG ATCGGCCCCG ATAATCGATA
12361 AGACAAATAC GTGCATCCAG GCATACGCAT ACAGACGACC GACCGCCTGT AGTTAAAAGA
12421 GATGCACCAG CAAATGGCGT TGGCAGGGAA ATGGGACGGA CCCTCTGCTC TTTCCCTCTT
12481 TCTGTGCCCC GCCCGTTCAT ACGCTGGCTT CCGTTTTGTG TGGCTAAGCT CCGGCTATTG
12541 TTTCCGGCTT CCGGTCCCGA TATCATTCCG GTTTTGAATG CGCTTGCCTA TGTGTGTGTG
12601 TGTGTGTGTG TGTTTGTGTT TGCAGGATGA GTGCATCCGC TTTTCGAATC GAAACTTTGC
12661 CTGCAAAACC GTAGTGCACC GTAGTGCTTA CAACAAAAGT AATATTAATT TCGAAACAAC
12721 ACCAACTCTG CTGCAGGTGA GTTAAGATAA AGTATGTCGC CCCGCGCCCT CATTTCACGC
12781 ATACGTGTGG TGGAAAGCCT TCCTTTGTGT GTGCTTTTGC GTTTTGCACC GGCGATAGCG
12841 ATGGGGCGTG TGTGTATTAA AATAACGTTC GCTTGCAACG TAGTGTGAAC GTACACGGGG
12901 CGCGCACTGA ACGTTCTGCA TCCCGTTTTT TTTGCAAAAT GAGGGAAAAG GTAGCGAGAA
12961 ATACAGCGCC ACCTAGGCAA GGCTGGCAGG GAAAACAGCA GGAGCTAGTG CAGCAACTAA
13021 CAACCAACGG CAGTAGCCCT GCTGCCCATG CATCATCGGT GGAAAGTTTC GCGCTTTTCC
13081 GTATTCACGC TGGCGTGCGG ACACTGCCCT TCCTTCCCGA CGGTGTGCAC ACACAAACTT
13141 TGCACGCGTT TGAGGTATTA ACGATTTCCT GCCGTGTGCC GGTTGGTGTG TACGGTAGTA
13201 GGGGTTTTTT TGTATCATTT ACGTGAGTAG TTTATGGAGC TTTTCCGATG TTTGGCCCCT
13261 TTTATGCTAT AGGTAGATGC GGCAAAAAGT TTTCGTAAAG GTAAAAAAAA AGAGACATGA
13321 TATCTTTAGG CTCAAGTTTT TAAAGTAACA ACGGAAGTTT AGCAAAAGCG GGTAACTCGA
13381 TCATTAGTAG ATGCATTGAA TAGGAAAAAC TGTTTTCTCA GAATAGGAAA GAAATCTCAG
13441 AACTAGCAAT GAGAATACTT ACTTTTTCAG TAATTTGAAT CAGAATAGAA GATTGAGATC
13501 AGAATTTTAA ACTCACTGTT TGAAGATTGA TAACTAGTAA TGCACTACAC AAAGCATCTT
13561 TACATTTCGC CATTACTTCT TTTCTATCCA ACTCACTCAT CTACTTTCTG TGCCGATTAG
13621 AGACTCACTC ATTGTGGATT CATATCACGC ACTCTTTTCC TGGCTGAATA GAGACCTCAT
13681 TCCTCGGGAT TCAACTCACT CTCTCACTGC ACGGATTAAG CAGTAAGCGG TCGCGTGGAG
13741 CCTCAAGGTC TCAAGGGGAG CCCGAGAAAA AGAATCTTCC TCAACTCACT AGTAAATTTG
13801 CTTCTCAATC CTTCTCTGAT TTAGAATTCT TCATACAACT TTTCGGTCTA CTCTTCAGAA
13861 AAGCCTTCCT TCGTGTGACC GCTTAAAGCC TCCACTTGTG TTAATCCGCC AGAGCTCAGT
13921 GTCGACCAGT AACGCGTTCT TTTGGGATAC TTTCTACTTG CTTATACAGA GTGAATAAAA
13981 TAGGAATTGA TTTATTCATA AAAATAGCTA AGCATTTTTT TAGTCTCTCT CAGCTGAGCT
14041 GTGTCCATAC ATTCACTTAC AGGATTAAAT TCCCTCATTG ATTCTCACTC TCAATGAGCT
14101 TTTTGTCTCA ATTCAATGCG ATGTTTACTG GTCGGTTCGG TAGTACAGTC GTCAAGTTGT
14161 ACGACCCAAC AGAACATGCC CGTCATGCGT TCAAATCTTG CACAAGCTTT ACCGTGCAAG
14221 AGTGGGTGCC TTATCTAGAA ATCAATAACA GATCATTGCT TAGTTGCTGC TATTATTACG
14281 CAACTATCGA AAATAACCAT TAATGATTTG AAGATAAATG TAGCTATTTT GAAGAGCCTC
14341 AATATTTCAA CCCAAAAGTT GTATTTTGAT ATTGAACGAA CGTTATGAAA ATTATTGTGA
14401 CATTTTTAGG CGATGAAATA AGCAGCAAAC GACAGCTTTG AAATGGCTGT GTGTACATGT
14461 ATTCTATTTC ATTTCCAAGT ATAAATTAAA ACGTAACAGT CTTACATTAA CATGCCCCTA
14521 GCAACTACAC TTAAGCTTAC TTCATTTACT CAAATCCCTA CAAAAACAAA CTCCGAGCAG
14581 TCGACGTGCT CGCACGGACC TTTTGTTTTA TTTTTAGCCC GAAAAAAGGC GCATGGTCCG
14641 TACTTTCCTG GTACGTTTTA CCATAACTTA ATTACGATGG ACCAATTTCG GTAACTTATT
14701 GGTGATTTTG TTTCGTTTCG CCGATCTGCG GAACCCTTCC TGGAAGTATC ATCCTAAGCC
14761 CGGGCTGCTG ATGCGTGTGT ACACTGTAGC TACCCTAGGC TCGTGATGGT GTGCTTCCTT
14821 TTTTCTATTC ACTTCTATCT TTATCTGAAA ACCTGAAGAA GAAGAATCAC CTGTCAAAGC
14881 CCAGCTCCAC CAAGCCTTCG ATTTTTCGAG GAGGGGTACA CCAGGAAAAA AAAAACAAAC
14941 ATTGGAGAAT GATTTTGATG GTGATGACCG TGACGCCGGA ACGAAAGGGA CACGATGCCG
15001 GGGACCGCAT CGGTACCTTT TTTGTTGTTG TTGCTTGTGC ACTGTAGTAT ATTGTTGTGA
15061 TTGTTTCCCG AAATGTGTGC TTTTCCTTCC TAAGCCAACA AATCCCGTGT GCGCGCCAAC
15121 CCGCAGTTTC AATTTGGTTC CCGTGTTCCT GAAATTCAAG GACAATCCGA TCATTTCCAC
15181 TTGGGTGGGT GTTTTTTTTT TGTGTGTGTC GTACCTAACC TTTCTGTCGC ATCGATTTTC
15241 GTTCGATGAG GGTTTCCGGG GGTCATAGAT AGACGGAGAG AATCCGTCCG TCCGACGAAA
15301 CAAATGAGGT ACGAAAAGCG GTCGGTAGCA CTTCTTCAAA CAGCGTTCAA GAGATGGTTG
15361 GGAGATTTTT TGGAAGAAAG TGTAAGAAGC AAAGGCTCAT CGTTAGATGG TTGAGCGTCT
15421 GAAGATATTC TCCCAGCTTC CCATCTCATG CAATCAGCCA ACAGGCAGGA CGTGCCCATT
15481 ATCTGCGCAA TATCCAGTGT CGTGTCTCGA ATGTCCCCGT CCGGAGAGAG AGCTGAACGG
15541 ACTGAATTTC GAAGGCGAAG CGAACCTTCG GAATGACTCC AAAGCGAAAA GATATCGGTA
15601 ATCCTTTTGA TGATGTTCTT AATCATCCAC TCAATTGCGC TGTATAAATT TCACTGTTCG
15661 CTTCACTTTG CTCCCGGGCT TCTTTGCCCT GTTTCCAGTA AACTTCGAAA CAAAACACGG

15721 TCCACCAAAT CCAATCCCAA ACTCTTTTGG GATCGTCCAC AACAGCATCG TCCCTTCTTT
15781 AAAGTCTGGT TTTTGGTCAT TCCTTTAGTA CGATTTACTT TACTACTCCA CCTGTTACAC
15841 ACCGCCACAA TCGTCATCAT CGTCTCGGAA CCGATAATCC TGATCATCCA AAGGTTCCGG
15901 ATCACCACCG CCGAAGGTCC TTCCTCGGTT TGGTGGATGA CGTTTCGACA GCAAAAAGAT
15961 GGCGAGCTTT TGCCCATTTT TCCGCAAATT CCACACATCC ACCAAAGATA TGGACCGCAA
16021 TAGAGGTGTG TGTGTGTCCT TTTTCTGGAA TAGGATGGCC CCGCTGCTGG CGCACGGCAG
16081 ATGAAATGCA ACGGTTTCAA CCGTAGCTTC TTGCTTCCCA ATGTTGGTTA TCGTTGGATT
16141 ATATATTCTT TACCTTTCGC TCTACAATCG CCTCGTCCTT TATTCGAATT CGGATGCGCT
16201 CTAGGTATAA ACGTCAGTGT GTTTTCCTAT TTCCTTTCTT TTTCACGTTG TCGTCACGAT
16261 AGCAAAAATA GTGGCTTCCT TTCGCAGAAG CTTTCGACGA AGGAAACGAA ATAGATGACG
16321 ACGCTGGGAA AGGTAAATGG TAAAATTCTT CCCCGAGCAG CATCCGGTAC GCGTGGCGAG
16381 GCGCACAAAT AAAACGCTTT CCTCCCATCT GTGTGTGTGT GTGTGTGTGT GTGTGGCTGT
16441 GGCTGGTTGT GTATTTCTTG GTTTGACTGT TCAAATGCAA GACAGTGGAA ACGATATTGC
16501 AATGCCTGAG CTGCCAAGAA AATGCCACAC AGACACGCCG AAGTATATGG GGAAATGGTG
16561 GTAGGCGCGC GTGGAGCGCA ACGGCGGAAT TGAAAACGAG AAAACTTCGA AGCTACTATT
16621 TCCTGCGTCA GGAACTCGTG ACTTTTTCCT CGCACCAACT CCCCCCCAAC TCTCATCCGT
16681 TTTTCCGTCC TTATCTGCCG TTTTTTTCCT CCTTCTACTT TGTTTGTTTC TCACTTTTCC
16741 GAGCTTCACT GACAGGTTTT GGAATGGGTT GGAACTGAGT GGAACGAGTG GGAATAAAAG
16801 AGAAAGAAAA AGTGGTGCCA AGTGGTGCCA AGTCTTCATG TTCTCCTCCG GTCCTATTCA
16861 GCGCCGCCGT CCCGGTGTTC GTTCATGTGG CCTTGGTGTT CCACTTGTGC AATTTTCCCG
16921 CTTCCCGACC AGCGCACAAC AATCATGCTC TCGCTCTGTC TCTTGGTGCC CGCTAAGAAA
16981 GCAAACGGCA AAGATCTTTT ACCGTGTCCT GCCCCGCTTC CTGCCATCCC GCTTGCCTTC
17041 TTCGAACACT TGTCCCTTCG TTCGAAAACC ACTAAACCCC AAAGGCTTGA CTTCACGTTT
17101 CGATTATGCT TCGGTGCGGT GTGCTGTACT CTTTTCCGCG CCTTTCGGGC AAATGCTTAC
17161 GGCTTACAGA GCAGAGCGAC CATCTTCGCG ATCCTTCCCC GGGTCGGTGC TTCTTCGCTT
17221 TGCAGGAAAT GATGGAATTA CATTTCCTTG CGCAAAAGCG GGGCAGGGTT TTGCCGTCGA
17281 GCACCAAATC GAGTTCGGAG AAAACGGTCG AGGACTACGC GCTCGCGCTC CCCCTTTGTG
17341 CGCTTCTTTG CAGTGGCCAT CCGCTCGCTA GAAGATAGCA GTTGTAGAGT TCACAGTTTG
17401 TCTTGTGGCC CTGGTTGGCT CCCTCGCAAC AAACGCAAGG ATTTCATCCC ACGTTTACGT
17461 GGAGAAGGCG GAAAGGATCC CACGCCGCAC GCTCTTGAGA CGTGCACAAT CGACTTTCGG
17521 TACTTGGGTA GGATCTCTCT GCACGATATC CTTCGAAGTG TGCGTGTGAC CGGTTCCGTC
17581 CGTCGTCGAG CAAGGGTCGT TCGTCGATCC AGATGGCGCT GACACGCTTC GAAAGAGAAG
17641 CGGAGGGCGC TTGGAATCCA CAGATATTTG TCATCGTTAC CGAGCGCCAG TCCGCCAGCA
17701 AGTCGATCGG AACGGCATTC CCTCAAACCA CACCGCATGG ACAGTTTTCT CCGAACGCAT
17761 CGAAACCCCC CCTCCGCCGC CGGTCACAGG TGTGTACGCG CGGGTGTGCG CGGAGATTCG
17821 TCAATAAAAG GGCCCAAACC ACAGGCACAA ACAGCATTGT GGGCGATGGT GGTGGTTTAG
17881 CCTCCCTCCG ATGGCGGCGA CGGCGGCGGC ACTGGGGCGA TTTTGCGAGG AAATATTGAT
17941 CCCCTTAAAC TACATGACAA TAAAATACCA TCGGCTACGC TCCATCAGAC ATTCGGAATT
18001 TGGCTCCCGG TTTCATGGAA AAGCGTTATT GGAGCGCTTC TCAATTTTCT CCTGCCACTT
18061 TGCTCTTCCT TCGCTTGCAC TCGGCCTGTA TTAAGCGCCA CTCGGAGCTT GCGACACATC
18121 CCCGGAGTGT ATCTTTAATC ACGTTTGACA AATAAATTAA TATTCTCCCT CAATTGGATA
18181 ACTTTTCGGG GCCGACGCGT TTGTGTCCCG CGCTATGACA GCCGAGGATT AGAAGGGCGC
18241 TGTTTCTGCG TGTGTCTGCC ACCTTTAGTA GCTCCTCGAC GCCAAACCGA CGCCTCAATA
18301 AATCCGTTTC TTTAGAACCG AACCATCGCA CGCACCATCT TGCAGTCTTG AAGCGCGCCA
18361 CGGTGTAGAA GTTTCGCACA TCGGCTTCAT TACACGTCCC GGGCGTGTGT CTTCGCAATC
18421 CTTTGGGATG TTGGGTATCG TGTCGTACCG TTGCCCGAGC GAAACGAGAG CTTCTGCCCC
18481 ATCCTCCTCC TTTCCCGGTT GACGATGTGA AGCAAACGAT AAGAGGCCGC TGATACACGG
18541 TTTTGTGTAA CGATGAGGCC GCAACCGCCA GCCAGCGCTG CGGGCTATTT CTTTTTGTGC
18601 CGTGCGTGCC TACCTTTAGC CGCTCTCGGG TCCTTGTTTG TCTAAGCGAG AACCAAGGAT
18661 GGCAGGAGGT ATTTATGAAC TCCGGTGCCT GCCACTCCTG CGAAACGGGG TGTAAGAAGT
18721 GTACTGGGTC GACCAGCATA TTTCTTAGCG GCTTCGTTTG TGTAGAGGTG TACGAACAGC
18781 GCCGTCCAAT TTTTGGACGA GCGAAGTAGA CGGAAGGTGT AGGAAAGGGT AAGCGCATAC
18841 TTAATTGAAT GTTAATCCAT GCAGGATCCT TTATCTCTCT CTCTCTCTCC CGGCACAATT
18901 TGGCCTCGCT TTTTGCAAAA GGCATTATAC TATAACAAAC ACATGAAGCA CTGGAAATGT
18961 TTGCTGATCT TGCCTCTCTC AGCACGGCGT TGTATGTGGC TGCTCATCAA AAGATAATGG
19021 ATATTGAATC ATGCCAAGCG AACGATAGAA AAATAGCATC GAACTCGGGG AGTGAGTGGG
19081 TGCAGAGAAC GAGACTGGAG ACGGCCTGTA TGGCCGCCTC GTGTATGTAG GAAGGCATGT

```
19141  GTCCAATGCA TGGGCAATAT GAAATGTTTA AGTTTACAAT GTTACTACAC AACCACGGTG
19201  GTGGTGGTAG CTTCGCCCGC TGCACGTGTT TTTCTGCTCG ACTCGAGGAT TGGTTTTTAT
19261  ATAGCGCGCC CGCCCCACTG ATGATAAGGA ACTTTATTGT TGATGTTTTT CTTACCCCCA
19321  TTGGGGAAAG ACAAGAGAAT TCACCGAGAG AGGAAGTGGT AATTTGATAC AGAAGACTGG
19381  AGAGGCTTTT CATTTTGTAT AATTGAGGCT ACAGGCTTGT AGAAAAAGCC GAAGGGAGAG
19441  CTACTTCATG GGGATTTATG TTCCATTGTA GATTATTGAT AAACAGATTA TTTGAGTTGG
19501  ATTGTTTTTT TATGAATTAT TTAATAAGGT AATACAATAA TGATGGTACA AATAATGTTC
19561  ATGTAATGGA AAGATAGTCA TATAAGCGTT TGGATAGAGT ATCGGAAAAA TGCCAGCGTC
19621  TAAAACTAGA ATGTCAGTTT GAATGAGAGA ATTGAACTTA CGACGTCCAG AGCGTAAGAC
19681  CGCTTGTCTT GCCACTGCAC CACGGCTCAG ATGCTCATGC CTATCGATTG GAGACCCATT
19741  TAAATCTCAC TTTCAGTAAC TGCTCTCACT GCTTCAGTTA CTTTAACATG TATAAATGAG
19801  AGCATGTAAT TACTTAGAAC TTACTTTAAA CCATTTTCTT TCGAATATTC ATCTTGTAAT
19861  GCTGAAATTT GTCACGAGCT CTCTTCCTTT CTCTCTTTCC ATCTCTTACC CTCCATTCCT
19921  CGCAGTTGTG AAATTGTTTG CCCAAAGAAA TTGAATTTTC CAACCCCACA GATCTCTTCG
19981  TGTGCGTGTC AAAAGAGGGG GCCAGTCAGG CATGCTTTGA TGTCGTCCTG ACCTGACCCG
20041  AAAGTTTGCT AGCAGGCTGG GGTGACTGCC TTATTTTTTT TTCTTCTTCT TAAGATGCCT
20101  CGTTGCACTA ACCAACCATA CCCTACCGCA GGGTGAAACT TTGCAGACGG GGAAGGCTCA
20161  CCGACCCGAC CCCTCCGACA CGGTGAAGCT CGTGGCACCT TTGAATCGTA TATCGCGTAT
20221  TCTACCCGTG GTCCGCGAGT TCGTACTTTG CAAAAGCTAG CCTGGCAGGC CTGGTGGTAC
20281  CAATTGGGCT AATCTACTTT GCATTTCACT CTGCATAATG TACGTTTATT TAGTCGGTTT
20341  TGTGTGAGTG GGTATTTTTT TGTTGGAGAC TGTTCGATTT GGAAGGTTTC CAATGTTTCC
20401  CAAAGAACTG AATGACACTT TAATATTTAG ACTTTTTTTT GTCTAAATAG AGCTGGTTCG
20461  ATAACATGTT TCGTACCACA CAGACACACA ACACAGGTTA CTTTGGCAAT AAATTGTAAA
20521  ACAACGACCT CGTCTTGGTG CCATTCGGTT TATTTCTTCC AATTGAAATT GCTTTTAATT
20581  CCGGCCGGCG CGCTCTACCA CCGGTGCTGC CCATTAACAC TTGATGCAAT AAATAATCCG
20641  CAACTTTTGG CGCATTTTCA TTTCGCAACT TTTCCATCGC GTGTGTGTGC CCTATAGGTG
20701  TATAGGTTGC CGTCTCAGAA CTTTGTGTTC TCTTTCCTCT CTCTATGCTT GCTACCTTCA
20761  ACATCTGATG ACTCACTTTT CACCTGTGAA GGCTGCCACC GTTCACACAA GGCGACCTCG
20821  TGGACCTCGT GCTTTCAGGT TGGGAATTTA TGAGCTTCCC ATCCAGACGG ACGGCTATAA
20881  TCGATGGCGG TGGTGCAAAG AATTCGATTC TTCGTTCATC AATAGGGTGT ATGCTTAGCG
20941  TTGTGCTTGG TAACGGAAGC GTTCGTATAG GTCGGTGAAA CCTCGGCAGC TTGTTAAATA
21001  GCACACGGGA AGTTTGTGAT ACGGCCGCGA CGAAGGTGAT GGTAACTAAT AGCCAAAACA
21061  ACCCCCAACC CCTGTGCGGG TAGTGTAGTA CATGGAGCCA CTTAACTGCT TGCTAGGTGT
21121  TATTTCCCAT AGAGGAAATT AGTAGGGAAC ACGTGAAAGA GGCACATGCT TGGGAGAAAC
21181  GGTTCAGAGC TTTATTCCCA CTTTCCGATA AAAGCTTCAA CCCTCTCAAA GGGCTGAGAA
21241  GGGACACGCT TCCAGCAATC GATCGAAAAG GTTCACCCAG GTCGTCTTGC GTAACTTCTA
21301  CTGCCCTGTT CCCCGCTAGG TTTGGAGCAT TCCTTCCCCC CTTTCTCCCA TCCTTCAATC
21361  AGCGCAATCG CTTTCGGGGA CTAGTTTTTT AGTTTGGAAC TGTTGTACAT ACGCAAACAT
21421  CAATCATGTC AACAAAACGG CAGTTTTGTG CTAGCAAAAG CAAGGTGTGG TACCATCAAA
21481  CAGCGGACGG ATGATGCCGC CCGTTCCCGT ACCGTATCGT CATGCCCCGC GGCGGTCATG
21541  CAGCGGTCGC GTTTCGTTTC CGTTCCGTTT GTTCACTGGT TTAGCTCATC GATCACATCA
21601  CGGTTCGCCG TACTGCGGGT CTTCGTTAGC TAGTTTTTCC TTCCCTTCAC ATTTGCCAAC
21661  CCAGCATTGA TCGTTAGAAA TCTTTCCGGG TCCATTCGGG TGCGAGGGAA AACGGGGGTC
21721  AACGTTTGCA TTTCAAATCG TCGACGGCGT TTGTGGTGGA CATGGTGCGC TTACGGTATA
21781  AGGGTCCGGT GTCCAAGGCG CCCTTGGAGC CAAAAATTTG TCACCCATCT ACCACTGTGT
21841  GACTGACAAC AACTCTGTGT GTATGTGTGC GTCGGTGTGC TTGTGGGTTT CTGCCTTTTC
21901  AGGTGTTTGC TTCTCGTTTT GATCGTTCGA TTAGTTCCGC CGCCCAAACG CCGGGGTGCA
21961  AATGGGCACA ACGAAAGTGA AATGTTTATG ACATGCTGAG CGTCAAATGG ACGAACTGTG
22021  GCGGAGATCG TTTCGATATG ATTTGTAGAA TAATAAATGT GAGGCTTTAT ATACTTTTGT
22081  TATGGAGTAG TACGCCAGCT ATGGCTTACA GGATTTTTCA GAACTTCTCA TAGTTGTGGG
22141  ACACTTTTTC GTCTCTTTCT TACAGGAAAT GAACTTTATA TGCTGAGAGA TGGACCCTTT
22201  TTGGACAAAA CTAATAGGAA ATTTGCATTG AATTTGACCT AAAAGGGTGT CATAGAGTCC
22261  AATATCCATC ATATAAAGCT TCTTTCAGAT AAGATACAGT GAAGAAAGTG TCCCACAACT
22321  TTGAGAACCC CTGAAAACCT CTTGAATATG TTTCATTATG TGGCTGATTC TGTAGCAAAG
22381  TCCTTTCAAT CTCGGAAGGT AGATGTCTTT TTCTCGTTTT TTTATGTTCA CTACCAATCA
22441  TCTATATATA GGTAATAGAA ACTTTTCTCT TTCTAAGATA ATAGAATCTA CATATAGGTA
22501  ATAGATATAG GTAATAGAAA CTTTTGTAAC TATTTGTATC TCGAAATGAT CCAACTTCTT
```

22561 GAAGAATTTG AAATTTTCTT TTAAAATTCA TAAAATTCGT TTCTGCTAGC AAAATTGAAG
22621 CACAACTCCA ATTTTTCATC GTGTTCGTTG AACTAAATTG ACCGCTAAAA CCGCATCCCA
22681 TATTTTTATA CCTTTTTTTA TTCACTAAAA ATAACCTTTG GTGCAACGAG AAAAACAGTT
22741 CCAAAACACA CCACCACCCA CTTACCACCC CAATCCTCTC TTTCTTTCAT CCTGATCGAA
22801 TTACGTGTTA CAGTTGTTCC AGCCCAGGTC GCTCTATTTC ATTCGTTTTC CCCTTACATT
22861 TTGCGCTTCT ATCTCTCCCT CGATCACGCG CACCTCAACG GAGAATAAAT TTTCGGTAAC
22921 GATTAGCATG GGAAAAACTA ATTTCCACGC TGGTTCACAT CCACTCCTCT GTCCGAAATA
22981 AGTGTGTGTT TTTTTCCTTC TTTACTGGAA GCTTTTCGTC ATCGTTCTCG TTTTGCTTTT
23041 TTTTATAGTC GCTTGTTTGG TACTGACTGG CGTCTGAAAG TTGTTATTTG CCCAGCCCAA
23101 TGCGGAGATT GAATCGTTCC CCTGGTGGTA GTGGTAGTGG TGGTAGGTGT GATGGTCAAA
23161 TCCTGACTAA AAACCTTTTC GGAAGCTTGT CGGTAAAAAT CGATATTACA GCAGCATACT
23221 GTGCTACCAC ACTCTTCCTC CGTCAATGAG GCCTGTCGTT ACGACGAGGC TGTGAAATGA
23281 AGTAACTCTA ACAAATGGCA GGCCCTAGCA GGCAGGAAAA CTACCTCCGT AGGCACAGTG
23341 CGAGGTAAAT AGAGAGAGAG AGAGCGAGAG AAGAGCTTTA TTACATCGGA ATTGGAAATG
23401 GTCTGCGTTT GAAGCGCTCT GGAGCGAAAC TTAATCTTAC CTTTTTAATC TGCCTGACCC
23461 CACCTTTACC CCACAATTCC CTCCTACACA CACAGCGCGT GACGCATTCC GATTGCCACG
23521 ATAGTAGTGG GCTTAGCCAT ACCGTAAATC GGGTTCGCCA TGGGGGGGTA ACGATCACAT
23581 TTCGGCCATA GTCATGCCTC CCTCCACAGC TCATTGATTG CTAAATGTCG TTGTATTTCA
23641 TGCCGTTGCC GAGGACTCTA TTTTTGGACA GGGGTCCAGC TCCTGTCCTG AGTTCAGAGT
23701 TTCCATCGGT GTTAAAGTCT AATCTCATTC CAGTTCTCCT CCTTCTGGCG GGGTCCCCGT
23761 GATGGAAAGC AAATCGAAGG ATATTTACCC GTTTTATTTA TCGCAGGATG AGGAATCAGA
23821 GGGAGAGGTT GGTGAAAATG TTCGACGCTT CGTTATCGGC AAACCCGACG AACCACAACT
23881 AGTGGTTATG TTTTACTGTG GCATATTGAA CGTCTGCGCT CTTTCTCTCT CCCAGCAGTC
23941 TCGGTTGTGT AAAGTTTCGG TTTCATTACA CGGTTTTACA TCGGCATATT ACTGTTTACG
24001 GGTCCCTTCA ATGGTGGTGG TTTTGTGGGG CATTTATTTA CGCTTCGTTG CGTGTACAAT
24061 TCATCTGCTG GGAGATGTTG CGTCATTCGT ACACGTGCTG GGCTGAACTC GTAATCCTTT
24121 CGGGAGGTTC AATTATCGCA AAGAGTTCTT AAAATAATCT AATCAGATTA ATGTAGGCTA
24181 AAGCTTATAA TTACAAGAAA TCTATTCCAA TCAGAGGAAA GCCAATGGGT CCAGTGGGTT
24241 CAGTTTAGTG ATTGAGGGAG CGCACCCACG ATTCCGATCC GATTATGGGA AAATGAGAGT
24301 CGGCGTTTTT TTGTCTTTCA CTTTGCGCGT GCACTTCGTA TCTTAGTTTC GAAGGCCAAT
24361 TCCGGACGTC TCCGACTCTG GACGACTCAA GACGACTCAG GACTAGTCCG CACGACTACG
24421 GAAGAGTTCT GACGACTCCG GACGACTCCG GACGACTCCG GACGACTCCA GATGACTCCA
24481 GATGACTCCA GATGACTCCA GATGACTTCG AACGACTCCG GACGACTCCG GGAGATTTCG
24541 GACGGTCCCA GACGATTCCG AACAACTCTG GATGACTTTG AACAACTCCG GATGACTTCA
24601 GGCAACTACG AATGACTCAG GACGCCTCCG GATAATACTA GGCGGATAAT ACTACCTACC
24661 GGAGTTCGGC GGAGTCGACT CCGGATTTTC GCCAACTTCA TCCATCACTA GTTCGATTGC
24721 TGCTTCAATT CTTAGTTCTA AGAAATTAAA CCATGAAACG ACGACTTGGC CATAATTGAT
24781 CCTCTTCAAA TCCTGAGATC TAATTACATC TTCTCTACTC CTCTTCCAGG TTGACGAGGG
24841 CCGCCTCTAT CAGGAGATAA TCCGCGTCGA AGCAACGGAC AAGGACTGTA CGCCCCTGTT
24901 CGGCGATGTG TGCAAGTACG AGATCCTCAC CAACGACCAG CCGTTCACGA TCGACAACGA
24961 GGGCTCGATA CGCAACACGG AACCGCTCTC GCACAAGATC TCGCACAACC ACATCCTCTC
25021 GGTGGTGGCG TACGACTGTG CGATGAAGCA GTCCGCCCCG GTCATGGTGA ACATTCGCGT
25081 GCGGCGCGTC TGCGAGGCGC ACGTGATGGG TGTGGCCGAG CGCATCGACT ACGTCGCTAA
25141 CTCCATGGAG AGTGTATCGC TGTTCCCGAA GATTCACCTC GAGCTGTGCG ATATGCAGTG
25201 CAAGGGCGAG GACATGGTCA TCCAGTCGTC GGTGTCGCTC AAAACCAAGC ACATCTCGTT
25261 CGGTTGCGAT CGGGACGCGA GCCAGTGTTT GCGCCGGCTG AGCGAAAGCG GCAGTGGCCG
25321 GGAGCGGGAC GGCGATGGGC TGGTCGATTT GCTGCAGAAG AACACGGACT GGACGAAGGA
25381 TCTGACGTAC GACGAGGGGG CGGAAGCGAT CTTCCACTTT GACGGTAGCT CGGGCGCGAT
25441 CGTGCCGAAG AATGTGGTAC AGCCGCAGGA CTTTGCCGCG CACCAGTTCA GCATCGTGAC
25501 GATGTTCCGG CACAACAGCA TCGTGAGCAC GGACAAGCAC ACGAAGGAGC ACATCGTGTG
25561 CAGCGCGGAC GACCACAAGA TGAACCGCCA CCATATGGCG CTGTTTGTGC GCAACTGCCG
25621 GCTGATACTG CTGCTGCGCA AGGACTTTAA CGACGGCGAT TTGAACATCT TCAGCCCGGC
25681 CGAGTGGCGC TGGAAGATAC CGCAGGTGTG CGACAACGAG TGGCACCACT ACACGATCAA
25741 CGTGGACCTG CCGAAGGTGG ACCTGTACAT CGATGGCATT CGGTTCGAGA GCAACGTGGA
25801 GGACCGGCAC AGCAATCCGG AGGTGATCGA TGACTGGCCG CTGCATGCGG CGCACGGCAT
25861 TAACACGACG CTAGCGATCG GCGCCTGCTA CCAGGGGTCG GAGAACCGGT TGAAGCACGG
25921 GTTCAGTGGG GATATTGCGG AGATTAAGCT GTCCACGCGC AAGGTGCTGA CGGCGAGCCA

```
   25981  GATCCGGTGC GGTACGGACT GTGCGGAGAA GTTGCTGCCA CCGGCCGACC ACTACCTGGA
   26041  GCCGGAGCAA CAGATCCAGT CGAACACGCA GATGAACAAG ATCATCATCG AGGGCAGCAA
   26101  CAAGACGAAC ATCGAGCAGC TGCTGCAGCA GATCCAATAC ATCAACTCGA AGGAAAGCCC
   26161  CACGATCGGG CGCCGCAACA TCCAGGTGAT GACGACGGTC AGCTGCCCGA ACAAGAAGGC
   26221  GATTCGACTG CCGACGATCG ACACGTACAT TATGGTCAGC TCGGCGGGCA GCAGCCTGAC
   26281  CGGGGGCAGC ATGCTTGCGT CGGCCACCTC CTCGTCCTCC TCGTCCGGGT TGTACAACAA
   26341  GATACTGGTG GATAAGCCGC TCGCCACCGA CCAGAAGCCG CAGATCGTGA TCAGCGGCAA
   26401  CTCGAACCAT CTCGTCTCGT ACCCGGACAT TAAGGAGGGC GTCAAGATAC TGGCCCAGAT
   26461  CAACATCAGC GTCTACACTG GGAAGGATAT TCGGCCCGAG CTGCAGAAGC TCGATTCGTG
   26521  CAATGTGAAC GTGTTCCCGA GCCTGAACCC GGACCACGAG GAGATTACGA TCGACGATAA
   26581  GGACGGTGCG GATGAGTCGC TGTTCAAGTT CGACATCAAA ACGCAGATCA GCAAGGAGGG
   26641  CGTCGAGATG ATCGGGTACG ACACGATCGA GAACTACCAG CACGTGCTGA AGTCGCTCGT
   26701  GTACATCAAC AAGAAGCCGG CCTACTATCT GAATCGGGTG TTCAAGCTGA CCTGCTCCCA
   26761  GGTGGACGAC CACTTCCGCA GCGCGGAGTA CACGCTCACC CTGACCGTGC TGCATCCGAA
   26821  GCAGCAGCCG ACCTCGTCGA TTGCTTCCGC CTCCGCACCA TCCTCCAATG TGTTGCTGGC
   26881  CATCTCGACC GCCCCATCGG AGCAGCATCC AGCCGTAGCG GTCGATTCGC ATGGTAAGAG
   26941  TCTACTGGAC GTTAATGTCT CCTGGGTACT TGGGTTACAT GTTTCTAACG TTTCTCACAA
   27001  CTTCCTTTTC TCTCCCACCG CCACAGACAA TGCAAACCAG TACGCGCACG TAATGCTCCA
   27061  CTCGCACGAC GTCCAGGAGT CGCAGTCGAA GGTGCGCAGC AGTGGCGGCG CCGCCCTGCT
   27121  CGGCCACGAG AAGCTCGCCT CGTCCCACTC GACCATGTTG ATCGTGGTCA TCTGCGGCTC
   27181  GTTCGTGCTG CTCATCTGCG GCGTCGGCAT TGCCCGTCTG CGCAACAGCA ACGCGGCACT
   27241  GGCGGGCGGC AGTAGCGGCC ATGGTGCAGC GGGCGGCGCG GCCGGCGGCA GTGCGTTCGC
   27301  GCGCAACCTC ACCGACAAGC ATCTGCCGTG CCCGAAGGTT GCGCCGGACC AGCAGCTCGA
   27361  CTGGGACGAT TCGGCGCTCA CCATCACCAT CAACCCGATG CAGCACGACG CCCTGTCGGA
   27421  CGAAAGCTCC GAGTCGGACA ACTCCGACTC CGAGGACGAG GAGGTGCTGA ACGGGCGCTA
   27481  CAAAAACGTC AGCCAGCTCG AGTGGGACAA CTCCACAATG TAAGTGCAGT GCGCCGGCGG
   27541  CGAGCAATCG GAAATCCCTT TCCTCGTGGG ATGATTGAAG AATCTACCGA ATGTGTCCTG
   27601  CTTTTTTGCG CGCTTGTGGG TGTGTGTGTG GGTTTGCACA AGGATGGACA GCTTTCCCAG
   27661  CAAATCCCCA TTGTTATTAC TATTCAAATG GGTTTGCCTT TTCCGTTCCG GGACGGACCC
   27721  CATGGATGGG GAGGGGTTTT GCACAACGTA ATCAACGCAA CCACTCCACT TTTCGTCTGG
   27781  TTTGCTGTCT GTCTTCGTTG CGGGACAACG TGACACACTC GCTTTCAGGC AATCGTTTGC
   27841  TTACGGAATC GTTGCGGAAC CGTCCTTTTC CTATGTGTGG TTGTGTCCGG GGTGTTTTTT
   27901  TTTTTGTTCT GTTCGTCCAT TCCATTCGGA ACGGTAACAT TTCAAAAGCA ACTCCACAGA
   27961  ACCCTTCCGC CTCAGAGCAC TCTCGCCGCA ATTGGATCCC TTTTGCGACT TTAAACCATT
   28021  CCGTGGGGCC AATTGGAAGT GACGCTTCCT CTTTTCTGTG AATCAACCAG GATTCACAGT
   28081  CGTTTGGACA TCGCTTTTTT CTCGAAGTAA TAATGAACTC CCAATCACGT TCAAGTGCGT
   28141  CCGCGTGAGC GTACACGTAC ACCCGCGGGT GTTGAGGTTT TGGGAAATAA TGGATAATGG
   28201  ATGGAAGTTT CGGACGTAAA CGACCGCACT ACAGAGTGGT AGTAGTAATA GTACCAACAA
   28261  GAAATGTCCG GGAAATGTGT GTGTCCTTTC GTTCGCTGAA TTGCATGTGA TGGATGAAAT
   28321  GGAATACCGC ATTGCCCTCG GCAAACAGAA ACCTGGTTGG TTTTGCAAAT TTTTCCAATG
   28381  AATTTCACAC CGGGCAATAT TTCACCGCGG CCTGTGGAGT GTTTCTTGTT ATATTGTCCT
   28441  GGGTTGCGCT CTGACAATGG TATAGTGTGG CAGATATTTC CCAATAATTG TACTGTGCTT
   28501  TCGCGCAGGA AACGATGTTA TTGTAGTAAA CTGGACTGTG TAATGCAGTG TGTTACATAC
   28561  GTTCAATGAG ATAATCATAA ATTTAATGCC TAATTCCAGC AATAGGAGCA GTCTAAAGGG
   28621  CTTTTGATTA TTGTGCAAAA TATCAACT

//

LOCUS          1     1555 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      1
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          join(500..554,838..1055)
                  /translation="MKYFLLVLICTTISAVFAYPYYQGAYNPYYQSSSDPYYQPFRL
                  WNSPTTYQIGTGVVPTYSAQSSYNPLYPYYQYNWDNFNTDFYNPYKN"
ORIGIN
```

```
       1  AAGTTTTCGT TTAATTTTCA ACTTCTTTTA TGCTGGTGTG AGATCGTTAA TTTCAACACT
      61  AATCTGATAT TTGCATATTA GAAAGTGTCA ATACTCCAGT TAATCGGAAT CGTGTCTGGA
     121  ATTAATCTGG ATTCTGAAAA TAACAGAAAA CCCTTTTGTT TTCTCGTTAG TCTTGAACTC
     181  AAAGTTTTTA ACATATATCG TATTACATAT TCAATGTACC AATATGTTTT CAAACAATGA
     241  CATCACATTA TCTTATTTTG GTGAAACTCG TCACTTCCTT TTTCTGTCCG AATCAATGAA
     301  TCAGGACTAA TATTATTTCA CGAAAGAGTC TTCAAGTGGT TGAAACCTAT GTGCATCATG
     361  CATTTTATCC CTTCTGTAAA CAATTGGCTC TGTAAAAATA ACAGGAAAAA AACGATTTTA
     421  GTGTTTGATC AAAGTGCCAT ATAAAGAAAA TGGACCGCTC ACAAAATAAC GAAATTTAGA
     481  CGTATTCCGA AACACTAACA TGAAGTATTT TCTATTAGTT CTCATTTGCA CCACAATTAG
     541  TGCTGTATTC GCATGTAAGT TTGATCCAAA ATATGTTACA AAAAGTAACA TTTTGTTTTA
     601  AACAGCTGAA TTATAAAAAC AATCCCAGAT TTAACAAGTT AGGATAAAAC GCTCACGGAA
     661  CGGTTAAATT TTTTGTCATA AAGACTTGGC CTTATTAGTC GTCTCGAATT AAGTTAATAA
     721  TCACTGTAAC TTAAGCTAGG TCCTCCTAAA GACATATTAC GCCAACTTTC AAGTATTCCA
     781  TTTTTGTGAC TGAGTGGAAG TTTTTCTGTA ATATGGAAAA AAATTAAACA TTTTCAGATC
     841  CATACTACCA AGGCGCGTAT AATCCATACT ACCAAAGTTC TTCTGATCCC TACTATCAAC
     901  CATTCCGTCT CTGGAATAGT CCAACAACAT ATCAGATTGG CACCGGAGTT GTCCCAACCT
     961  ATTCGGCTCA AAGTTCGTAC AATCCATTGT ACCCATATTA CCAATACAAC TGGGACAATT
    1021  TCAACACCGA CTTTTACAAT CCATATAAAA ATTGATTATC GAAATAAACT ATTGACAGCT
    1081  TTTAACTTTT AAAAAGATTA TTGCCACTAC TACAAGACTA CAATGTTCAT CTCGCTTCCG
    1141  AAATAGTTTA TGTTCTCAAA TATTTACTTC ACATCTCTGA ACCCATGTCA CAGAAGCAAG
    1201  CTCATTGAAA CAGAAACATT ATAAATTAGA AATTTTTTAA TTAATTGCGA AATTGAGGAA
    1261  GTCAATTTTT TATAACAACC CAATCAAAAC TTCATCATCA TTATTTTGTG ATCAAGTGGC
    1321  AAACTGTCAC TCTTTTCTTT TTTCTAGTAA ACACCGAGTG TAGATTGTTT GGATTTTCAA
    1381  ATCTTCTTCT CCACAGTAGT TAGTTGAAAA GTAAGAAAGT TAAATAGAAA GAGGCGGAGT
    1441  GTCCACGAGT CTATTTGCGT TGATTCATTT TGGCAACCGA TCATTGAGTA TATAAGTGAA
    1501  AAGGTCGTCG AAAAAGAACA GAAGTTTCAC GAATCATGCA GTTTTTAACG ATTTT

//

LOCUS          2     2217 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      2
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..580,1523..1717)
                     /translation="MTFFSILLLFCAMVVMMSKVSAIGRREENGMPFETRTLRPFPQ
                     MTDSEDDSTKTQVMVVLLVGFCVSLAITVIIRFLRKYSERQRLYSSFN"
ORIGIN
       1  ATTAAAAATA AAAATCATCA AAAAGAAAAC TCACCGTCAA AGCTACAAGT GTGCAGTGAT
      61  TCTCATTGCT AGGCATTTCC AATCATCTAA ATAGCATTGC CTGCCCATAA AAGAATAGAA
     121  GGAAGCACCA AAGTGTCATT TTTGTTCGTA AAAACTGGGT GTACATTCAA AAAGTCTTGT
     181  ATAGCTTAGC TAAGCTCTTC ATCTAAAAGC ATTGGCTTCA TCTTTTCACT CGTCATTCCT
     241  CTCAACAACT CCACAACAGT AAAGTCATGC CCTTCGTTCA TTAGTCAATT CGTTGAAGTC
     301  TATTAGTCTT TCGAAAGAAA CGAAATCATT CCATTTCACT AAATTTGACG TCACCAACTG
     361  TCTTCATACT CCTCAAAACT CCGCCCCGTG TGTTGTGCTG AACCGTTGCC ACGGGTTACT
     421  CTGGATACAG AAACATCTGC TCTTAACTAA TCTACCTTTT GCGTTTCTCT GTGCCAATTT
     481  TAGATTGAAA ACTGCTCACA TGACATTTTT CTCCATTTTG CTTTTATTCT GTGCAATGGT
     541  GGTTATGATG AGCAAAGTTT CCGCTATCGG ACGACGAGAG GTGAGAATTT GTATTTGAAA
     601  CATCTCATCG GGTTTTGGCC GAAACTTCTT GACCAAGTTT TTCCAAAAAT AACTTGGTTC
     661  AGAAATGCCA ATTGGCCAAC CTTTGTCTAG GTCTACTGTT ATTCAAATCT GAAACAAGTA
     721  TTGTCAATTT TTACCCAAGT TTTTACGGCC AGTAGAAACG TTTTGATGAA AATTGGATGA
     781  ACCTTAGCGC TGCCTAAACA ATGTCGTACC AACTTCGGAA CCACGTTAAC ACTTTTGGAC
     841  CGTACGAGAA GGCCAAATTC CAAAAACTTT GAATTTTTTA AAATTTCACA CTTACTTTAC
     901  CCCTATTTAT TGTAAAGATT GATATATTGA TGGGGGTGAG GGGTCCTATA TTTTTAATTT
     961  TAATTTTCAT ACTACAACAA GCTTTATTGA TAAAGCTCTC GTAAATAGTA GTATGTCATG
    1021  CCTTAGATCT ACAGTAATAA ATGAAATATC AAATTAATGT CGAGGTACTG TAGGAGTTTT
    1081  TCTATCCAAA AAAAAACCTA ACAGTTCAAT TCTTTAAAAT CGATTTTGCA ATAATTTGAG
```

```
    1141  AATTCGTGAA CTTAATTCAA CTTAACTTAT TCAACTGTAT AACTTTTAAC TTTCCACTTT
    1201  TAAACTTATT TCAGTGAGAA TGCACAGTTA ATTTTTTTAT AGGTGGGGCA TTTCAGCCTT
    1261  GACCGCGTGG ATTGTCCAAA CTTATAAAAA ACTTTAAAAA TAACAGCAAC TATTTATTTT
    1321  TGAGAAGGTG ATATAATCGA TAATAACTGA AATTAGTTAA TAAAAAGTTT CCTGGGGATT
    1381  TATTCGATGA ACCAAATGAG TATTGAATTT TTGCATTTCT TTTATACCTG CGTTTCTTTC
    1441  TTGCAAAATC AAGTTGTTCT GATTGATATA ATTTCCTTAA GGATAATTAT AATTATTTCA
    1501  AAATTATTTA TTAAAATTCC AGGAAAACGG CATGCCATTC GAAACTCGAA CTCTTCGCCC
    1561  ATTTCCCCAA ATGACAGACT CGGAAGATGA CAGCACTAAA ACTCAAGTAA TGGTTGTTCT
    1621  GCTGGTTGGA TTCTGTGTTA GTCTTGCCAT TACTGTCATC ATCCGATTCC TTCGAAAGTA
    1681  TTCTGAACGA CAACGTCTCT ATTCGAGCTT CAACTGATAC CATTATTCTG ACAAATACTC
    1741  TTTTGGCACA TCTCTCTTTC TCTCTTTAAT TCGCAGTCTC TCATACTTTA TATCCAATTA
    1801  CATTTTACTT TTATTATGTG CAATAAAGAA AATTTGATTA ACAATTTGGA ATATGTGGAA
    1861  ATAAGGTTAA GCTTTCATTT TGGGCCTGGA AAAACTGAAG ATTTGCTTAA AAAGCTGAAT
    1921  ACCTGGTACT GATAAAGCAC GGGCATGTTG AATTTTTCAA AATTTAAAGC GTAGGAAACT
    1981  TTTAGTTTTT TGCAAATTTC GGTATACGTA CAACTAGTAC TGTTCCAATT ATGTTTTGAG
    2041  TTAGTGAAAT TGTATTTCAG CTTCTTCGAA CACCGAATAG TTCAGACTGA AAGGCGGTAT
    2101  TTAGACAGGC AGGCACGTAT CCAGGCCGGT TTGTTAAGGC AGGTTTGAGG TAGGCACGTA
    2161  GGAACGTAGG CAGAATTTAG GTACCTATAG GTAAGAGATA AGCGTGACAT AGGCCTA

//

LOCUS          3     1429 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      3
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..764,916..929)
                     /translation="MESNLDVGSCLLDECSSAGSHRKLSALSAASSLHAPSESSMSR
                     KRSIRERMAEMADSRRNPKKPTPPSIFEAAQLIGKFSGQRIDLLSVTFV"
ORIGIN
       1  TGTTCCTCCT CAAATTTTCC CCCAAAAATA GGCGTTCCTC ACTTTTCCTA CGCGTTTTTT
      61  TCTCAAAAAA GGCCACCCAC CTCGCGTCTC AAACGACAAA ATAGTGAGAG ATGCGTCGCC
     121  GCCGAAAAAT TCAACAAATT AATTAATTGG TCCGGGAGAG ATGAGGAGGC TCAAGTCAGT
     181  CAACGGCGCG CGCCCCGAAA AACAGCTTCC TATACATCCG TTCTTGTCTT TGCGGCAATG
     241  CACCAATTGC TCTCGGGGTG CGAATTGAGC CATTAGCCGA TTCGGCTCCA CCCCTCAGCG
     301  CTAAGAAACG TGTGCTCGGG CGCGCGAGAG TGTGACACAG CGCAGGCGGA GCCTCAATAA
     361  CTCATCTTTG GACACCTTCT GTCCTACAAC TCCCGCTAAC TACCATCTCC GTAGTTGTGC
     421  AATTTCCAAT CCCAGTTATC TATATTTATC TATTATTCGC ATCTTCCCTC ATTTTATCTT
     481  GAATTAACTT TGTATAATTA TGGAGTCGAA TCTTGACGTG GGCAGTTGCC TACTGGACGA
     541  GTGCTCGAGC GCCGGATCGC ATCGCAAACT ATCAGCATTA TCAGCAGCGT CCAGTCTCCA
     601  CGCCCCATCA GAATCCTCGA TGAGCCGGAA GCGAAGCATT CGGGAGCGAA TGGCCGAAAT
     661  GGCAGATAGC AGGAGGAATC CGAAAAAGCC AACTCCACCT AGTATTTTCG AAGCTGCACA
     721  GTTGATCGGC AAATTTTCCG GGCAACGCAT TGATTTGTTA TCAGGTAAGT GATATTTCTG
     781  TATCTTGTTT TTCGCATTTC ACTCTGTTAT TTAAATTAAA CTAATCTAAA TAATATTATA
     841  AATGTAATGT GGTTTGATAC TATGATCAGA CCCTATAATC AAATTTCCAA ATTATATACG
     901  AATTCCGATC TGCAGTAACA TTTGTTTGAA CAGTCAAAAG AAAACAAAAT GTTTAGTTCA
     961  TATAATGTTG ATTAAGCCAA TTTTTGAAGA CATCCCTTCC CTTCGTAATT GCTTATTTTG
    1021  AGAACTTTTG GGAAAACAAA ATCTAACCAG TCATTACCCA TCCAAATTGG CAGCTCGTTA
    1081  ACCAAATCTT TTCTTAGCCA CTCCTTCCGT CTCTAAAATC CCATGAATCA TATTCTATGG
    1141  AATTTGGAAG TGGGTTCTGA GTAACTTTTG CTGCAAAGCT CCGTGTTCGG TGATGGGAAA
    1201  CACATAAGAG AAAGGGGAAG AAAAGTTTAA AAGAAAGCCC CTCTTTGAGA ATTCTGATTA
    1261  TAAATTCGGC ATATCCACTT ATTCTTCTTA TCCTTTCCGC CGCGAACATT TTCGAAAAAA
    1321  ATGCGAGCAG AACATAAAAA ACGAGTAAAG TTAGGCTGTC AAGAAGAAAG TCTGTTTTAT
    1381  TCATCAAATC CTCGCTCGAC ATTCACTTAC ACTCTTACTT TTAATGCAG

//
```

```
LOCUS          4     1335 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      4
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          complement(join(500..671,723..835))
                  /translation="MNVTLVFLLLFVCFISIHAQQQWSPQQLAAYRQQYYNWWYNQQ
                  RMQQQIQSVQAIQQAQTPSPFFGNSQSDSKGNYWHGDDNAKLLLIARSSWP"
ORIGIN
       1  GATCAACTTC AAATAAATCT GTAAAGCTTT GGATAAAATT TCTTGAACAA CGTTTCTGTT
      61  TCCTCTTTCT TGAACATACA ATATATGTTA CCCTGGAGAT CATAAGTTAA AGTGAGCCAC
     121  AAGGTTCAGA AAGGAGGGTG TAAGGTTTTG TTCACGAAGA GAAATTGACC ACACCGCTTG
     181  TTGACTTAAA TGTGGAACAC GGCAAGTGGA AAATTAATTT AAAACCATTT CTGTAATGCT
     241  AAAAGGGCTA ATTATAATTT GCCATTTGGT TTGGTAATTT CAGCACTATA GGAATTGTGT
     301  TTAATCATTT TTCCACGTAC GCATTCTTGA AACAAGCAAA CAACCTTTTC AACTATTTTA
     361  TTTGTGTAAG AAATTATTCT AATCATTTCA ATAAAATTTG TTGTCATTAC TTTTAAATTG
     421  TAAAACTAAT TTCATTGAAT ATATTTGTAA ACTAATTTCA TCGACTTAAA AGAATAAATT
     481  ACAACACGAA AAGCAACATT CATGGCCAAG AGGATCTTGC TATCAATAGA AGTTTGGCAT
     541  TATCATCACC ATGCCAATAA TTTCCTTTGC TATCGGATTG TGAATTCCCG AAAAATGGAC
     601  TTGGAGTCTG TGCTTGTTGA ATAGCTTGAA CTGACTGAAT TTGCTGTTGC ATACGTTGTT
     661  GATTGTACCA CCTGGAAACA ATTTCTTGAA ACAAAATGAA ACAATTAGCA GTCATATTGT
     721  ACCAATTATA ATATTGCTGC CGATAGGCTG CCAGTTGCTG TGGACTCCAC TGTTGCTGGG
     781  CATGTATAGA AATAAAGCAG ACAAATAATA AGAGGAAAAC TAAAGTCACA TTCATAGTTT
     841  TTTTTTGGAT GAAAAATGGT TACAAATGAT GTATTTTTAT ACTGATCAAA AAGAGGCGTA
     901  CCTGTAAACT ACAGGTGAAG TTAAAAAAGT TTCACCAAAC AACAATGCAA ACCAAATGTA
     961  ACGATTAGGG AATGAATACT TTATTCTAAT ACTCAATACT CAAATGGTCA GTTGTTTTGA
    1021  TTGTGCACCG TCATATACCG CATGGAAATT AAAATCTCGA AAATGAATTT GAAAAGTACT
    1081  GCATTTTGCA CAAATTAGCT GAAAAGAACC TTCACATTAA CTTTTAAAAA AACTTCTAAA
    1141  GATTCTTGAA TTTCAAAACT TTCAGTTAGG TGAATGTTGT GCACTACGTT CAAGTTGAAA
    1201  ATGTTTGTTT TTAGACTTTT ATTAAAAACG ATAAGTTTAA CGGAAATGTT TTTTTTCAA
    1261  AATTATCATT TGTTTCGGAA CATGTTTCGT TGAAAGCGAC TTGTTCCTAT TATACAATTT
    1321  TTTAGTATCC TACCA

//

LOCUS          5     2164 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      5
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          join(500..548,1215..1327,1533..1664)
                  /translation="MVVSLTIKRLLSSTVDGSAHLCVYDANHNGDDLTVTISCPLAS
                  GCCEKGCCKLPIQHIRHKEHEPDFVYHAYRPPQPKVSKNPPTVILTDYRNMTLE"
ORIGIN
       1  CCATGATTAT CCTTGTGGCA TTTTCGGAGA TTTCCCAAAA ATTTCATAGC CGGCAATTTA
      61  GGAAATTTCG GATTGTTGCG CACGTTAGAA GTGTTAGAAA TTGTAACTGT ATAATTCCTG
     121  AATAGATAAT TGATGCGCTT GTTCTCGAAA AGCATGCAAA AGTAAATAAA TTATCAGTAA
     181  GGTCAGTTTT TTTGGTTTTC AAAGTATCAT GTCGCATTAA AAAAAACTTT TAAATAGACA
     241  AAGTCCGTTT TTTTTAAACA ATTTCTGTTT CATAGATTTA CCAAAATACG ACGAAAAAAG
     301  TCAACAGTTT TGTTTCCAAA TGTACATAAA TTCATTCGTG CGATCAAATG ACACATGTGC
     361  GAAACGCAAT GTGAACAGAT CGTGAAAACG GTCGACAAAC AGGAGGAGGA TTCGGATATC
     421  TTTTAGGTCA TTGTTCACAC CTCTGCTCAA TCCCCCTCGA AATCTACCAA TCTACTTCAT
     481  TTTCTTCGTG TTCTGACATA TGGTCGTTTC TTTGACCATC AAGAGATTGC TGTCTTCGAC
     541  TGTGGATGGT TAGTTTTTTG AATGGATGAA AACTGGAAAA AATTGTAAAA TCAGCAAAAA
     601  CGACTGAAAT TAATTGTTTT TTGAATTGCC GCGCAAATGA CTTTACGCGG GAGCTCTGCT
     661  GAAACTTCGT TATAATTTGC GAGCAATTCA CTTATAAATT TTCAAGATTT TTCCAACACG
     721  GTATACGCAA TTTTCCGATT TGTTATACCA CTGTAAGTAC ATAAATATGT ACTATAGTTT
```

```
     781  TCAGTAATAT CCCAACACCC AGAGCTCCTT CTTTAAACAT TTCAAACCAT CATTCTAAAA
     841  CTGTTTCTAA TTTGGCACAA CTTCTACAAA ACCTAATTCC AATAATATCG CATTTTCCCG
     901  TCTTTTTTTC AAGTTTTCAC TTGTCAAATT TTTAAATATA TATTTCCGAA AATTGATCGC
     961  TTCAATGGTT ACGGCCAGCA AGTTTAGAGA TTAAAGTGTA ATGCAATGGT AACTATATCA
    1021  TTCTAGTTTG TTATTTTCTA AATTTTCATT CTAAACTTTC CCAATGACAA AAATCATGTT
    1081  TTTTGATTTG TCTTGTGAGA AAAGTATTTT CAAATTATGC TGCATGTTCC TGGGAAAATT
    1141  AAAACTTTAA ATTCCCACGA AAAGCATTTA AAAAGTATCA AAAAACAATT TTTAAAAAGT
    1201  TTCCCACCTT CCAGGCTCAG CACATCTTTG TGTCTACGAC GCAAATCACA ACGGAGACGA
    1261  TCTAACAGTC ACAATCTCAT GTCCTCTGGC TTCTGGTTGT TGTGAGAAAG GTTGCTGTAA
    1321  ACTTCCAGTC AGCTTTATCA CAAATAGTCA CATGAATACT TCATCATTGT CCGAGGAACC
    1381  AATATCAAAC ACCAGCCAAT TCAATGAGAC TATCTTGTTA ATCACATTGA CTCTTATGAT
    1441  ATTTTTCTTT GGATGCATTT CCATTTTGTT CATTTTCAGA AAGGCAAGAT ATTTTTTATA
    1501  GACTCAACGT TCAACATTAT GTCTTTTTTC AGATTCAGCA CATCCGTCAT AAAGAGCACG
    1561  AGCCAGATTT TGTGTACCAC GCGTACCGTC CACCACAACC AAAAGTAAGT AAGAATCCGC
    1621  CGACAGTAAT TCTAACCGAC TACCGAAATA TGACCCTTGA ATGACATGTT CCCTTCACAC
    1681  ACACACACAC ACACATTTTG TTGCCTCACT TTCCAATTCA GATTCTAATA AATTTTTTGT
    1741  TTAAATTAAA AAAAGAAATC AACAAGGGTA GTATTCCAGA TGATGATTCC CCATAATCTG
    1801  TATGGATCGC CTTGCTGAAT TCGAAAAGAT ATTGAAAGCT ATTGTAATAA AATAAATAAA
    1861  TAATTATTGA ATGTTCTTCT AAAGTTTAAT GAACTATTTT TGGTTGATCA ATACACACAC
    1921  AGACATATTA AAATAAATAA GTTTATTCAT CGATCAGAAT TCAGGGTGAT AACTGATATC
    1981  AGCGTGTATT GATAAGAAAG TCATGTGAAA TCAAAATTGC TTCCAAAAGT TTTTAGTTGT
    2041  CAACGTGGTA TTGCTCGTAG AACATGAGCC ACTCGTCGTT TTCTTTCTTC CAAATGTGCA
    2101  TCACTTCTCC ACGAAGGATC ACCTCTTTTC CTTCAACTTT CACCAATTCC ATTTTTGCCT
    2161  CGAC

//

LOCUS          6     1900 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      6
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          complement(join(500..724,1326..1400))
                  /translation="MGFFSHAAMLGVGYWIGKSYCFRYKIHQLPTDPDASEISFRIE
                  RKEGAGPFFGHPWWNHHHQCWHQYHHRTENTVNSQNAPSQVIPKKNVELLEVETNQ"
ORIGIN
       1  TAATTAATTT GAAGCATTGA AATATTTAAA TAGTCGTGTG AATGAATAAA TTATTATGGT
      61  TTAAGATACT TGTTGATGTA GATAAGTTTT TATAAATTTC TGATTGTGAG GTGTCCGAAG
     121  GCAGAAAATT TGGGTCCTGA AACGAACGAG GCTTTTCAGC AACAGTACCC CTTCATAATT
     181  TGGGTCTAGT TAGCAAAGTT TTTAGTAGAT TTCAGGCAGT CATTTCTCAC TATTTTTCTA
     241  TTGTTTTTAT TTATTTTCAT GTCAAAATTT ACATTTTACG CGTTGTTAAG AATTTTCCAA
     301  TAAGAAGTTT TCATATGTAT TGAAAATTAT GAAAAATCAA GAGTACGGTG GGAAACGGTA
     361  TCTGAAAAAT TTCGCTTAAT TCTAAAATTG TTCAGAAACA ATTTGTTCTC TGATACGTGA
     421  GAGAAACAGA ATATTCAAAC CAAATTTATT TCAATTTTCG GGAAAAAATA TTATTTACAA
     481  AATCTTAATC TTCTTCGTTT TATTGATTGG TTTCCACTTC CAAAAGTTCC ACATTCTTCT
     541  TCGGAATCAC TTGAGATGGC GCATTCTGAC TATTTACAGT ATTCTCAGTC CTATGATGAT
     601  ATTGATGCCA ACACTGATGA TGATGATTCC ACCATGGATG TCCGAAAAAT GGACCAGCAC
     661  CTTCTTTTCG TTCAATGCGG AATGAAATTT CAGATGCATC CGGGTCGGTG GGAAGTTGGT
     721  GAATCTGGAA ATATTTGGAT GTTTTATTAT CTAAATGTGA TATTATGTGA CTGTTTCCTT
     781  GCTGGGTCTC AACACGACAG AGATTTCTGC AATTTATAAA AGATGTGCGC TTTTAAGGAG
     841  TACTGTAATA TTAAACTTTC GTTGCTGCGG AATTTTCATC GATTTATCAT ATTTTTTTCT
     901  CACGATTTGG TGAAAAATAT GTTTTTTTTT TAATTTTTAA TTGTAGTGTT TCAAAAAATA
     961  AATATTTTTG AAAAATGCTG GCAACAAAAA AAGTTTGAAG CTACAGTACT CTTTAAAGGC
    1021  GCAGGCCTGT TTGTATTTAA CAAAAGTTGT CGTGTCGTGG CCAGGTACTG TATTTTAGCG
    1081  CAAAATTCAT AAATTTTCGC GACTGTGCAA TAAATAGTGT CTTTCCGAAA CTCAGTTTGG
    1141  CCCAAATTTC GGCCTTTAAA GGTACAGTAG TTTATCTTTT CATTAATTTT TTTGTGGAAA
    1201  AAATACAGCA AAATAAGAAA ATCAAAGAAA ATTCCAAAAA AGAAAAGTAC TGTGGTTTTT
```

```
    1261  AAAGGCACAC TCTCCTTTTC TCAACTCCCT AGAAATTTTG GGTCGTGCCG AGACCTAAAA
    1321  CTGACCTTAT ACCGGAAACA ATACGATTTT CCAATCCAAT AGCCAACACC AAGCATTGCA
    1381  GCATGAGAGA AGAAGCCCAT TGTTTTTCAA AATTTGAATA GTATCAAGAA TAAATATTAA
    1441  TAATTCTTAA TTTTCCATTC CGATTCTCTC TGTTTTTATA GTGTGTACAA AAGGGGTGGG
    1501  GCGTCCTGGC CGCGATTGCG AAATATTTTG CATCACGAAA AGTGTGTGAG AGTGGAGTTT
    1561  GTGTTGCTGG GATCAAAATG TGACGTTTAT TGAAACGAAT AAGTTACAGG ATTTTAATAA
    1621  ACAAAAGTCA AAAGTAGAAA ATTGACAATT AATAAAATAT TTTTTTTTAA TGAAAATCGG
    1681  GAAAATTCAG GAAACGATGA TTTTAGTCGT TGTACTGGCG GCGGCGATCT GAAAATTGAT
    1741  TTTTATTGGT TTAATATTGG AGTTTCATGA CTAAAAATAT CAGTTTTTGA AAAATTTTTG
    1801  ATTTTTCGCC TACTACTTTG AGTCCTGGCA CGAGAGAAAG TTTTCAGCTA CCGTACTCGT
    1861  AAAACCACGG TACCTGGTCT CGACACAAAA ATATTATGTT

//

LOCUS          7     1305 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      7
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          500..805
                  /translation="MLSAVASTSSLIQFTTTNSSSSPLVSRRKISEISNPTETLYRK
                  WSHLGTPPRSSTMIAGSQHHGLMQRGGPLQSTLSMSAKDRARKFSEKIASWVSMSIC
                  E"
ORIGIN
       1  TTTAAAAATC TAAAAATCTC GCCGTCCATC GACTTTATAA TACCTAAATC GAGTTGAAAA
      61  CGCAAGATAA TTACACTGTA TACTCAAAAT TAGACGGTAA TTTCAAAAAA AAAATTGTTT
     121  CCAGCCGCTT GACGAGTCGG TAAAATTTCC AGTTTTAGCT AATTTTAGGT CATTTTTTGA
     181  GCCGTCATAA CTTTTTTTTG AGAAGTTTTC AATAAGTTTC GTTATGAAAT TCGGCGTTTT
     241  CAGACAATTT TGAGTCTAAT CAAGCAATAA AAAGAAATTC GACTACACCA CCTTTAATCC
     301  GAAAGTGATG TTTCTCGATT AGTTTTTCCA TGGCTTCCCA CCTCGATGAC GTCAGTCGAA
     361  TATAGTAGTT GTTTTGTGTG TTATCGCATA ACATTGCATT TTTTCTTGCT GAACATTTAT
     421  TTTTTCGGCA ACTATTCCTC TTAACTTTCT CGTTTCCCTC TTCAACAATT CCACTTCCAG
     481  ACGACTTCAT TTAAAAAATA TGCTATCGGC GGTCGCATCA ACCTCGTCTC TTATCCAATT
     541  TACCACCACA AACTCTTCCT CTTCGCCGCT AGTCTCGCGA CGAAAAATCA GTGAAATATC
     601  AAATCCGACG GAGACGCTCT ATCGAAAATG GTCGCATCTT GGAACACCGC CACGGAGCTC
     661  CACAATGATT GCCGGGAGCC AACATCATGG ATTGATGCAA CGTGGTGGAC CACTTCAATC
     721  GACGTTATCA ATGTCGGCGA AGGATCGGGC GCGAAAGTTC TCCGAGAAAA TCGCCTCGTG
     781  GGTTAGCATG AGCATTTGTG AGTGATTATT ATCTATCACA ATATGCATAT ATGTAGATGC
     841  CGAAAATATA CGGTAAAACG TAATTATTAT TTGAAAATGG TATAATACAA CATATCAGTT
     901  TTAACGCCCC ATTTATCCCA GAACTCACGT GGTGTCAGGC TGTCCCAATA CAGTCTGATC
     961  TACAAAAAAT GCGGGATTTT TTTCCCAAAA AAATTGTGAC GTCAGCACAC TCTTAACCAT
    1021  GCGAAATCAG TTGAGATGTC TGCGTCTCAC CTCCCGCGTT TTTCGAAGAT CAAAGCGGAA
    1081  TGTGGGACTT TCTGACTCCA CGTGAGAACT GGAATGTTAA GTCGACCTAC AAATAATGTA
    1141  CCATTGGCAA ATGTTATTAA TTGGGTATCT AGTGATATCT AGTATTGCAT ATACTACATT
    1201  ATGCACCCAA ATACACTATT GAAAATACCT ACAAAAAATG TATCATTGGA AAACGGCATC
    1261  GAATTGGTAT ATACCGTATA TTCTCGATTT TTATTCCGAA TCTAT

//

LOCUS          8     1887 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      8
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          complement(join(500..750,1330..1387))
                  /translation="MVCPSSVALSVLFTILVANSLRCEATSVVTCMLCTVVSESLER
                  EMTPTATVNSMFKKCDKMGLMEPVCVQFVSENVKEMFQRVRQGIPSTSVCQALRFCD
```

```
                LQ"
ORIGIN
       1  AATGGAAAGA GTGGAGAAGC AAAGATCAAC AAGAAGGATC CATTGGAAGA TTTGGAGCTG
      61  GAAGATTTGG AAAAACTATC AAAATCTGAT GACACAAAAT CTGGAACTTC TGAGGCAGTC
     121  ACTTCTGATG ATGTTGATGT TGAACCAACT GAGATTAAAA AATCAAAGTT AGGAGATGAT
     181  GAGGACCTGA TTGACGATCG AAAAAAAAAA GCAAAAAGAA ACTTCCTCGT CAAGTGGGCT
     241  TAATTAACTA AGAATTTCAC TATTTTTGTT AAAATTATTA TTTATTTTTT GTTATTGAGG
     301  ATCTCAGGGA TAATTGACAA TTTGCTGCAA TTCGAATTCT CTGCAATATT GTAAATAAAT
     361  ATTTTTTCAC TTTTAAAAAC GGCCGACAAC AATGAAAACG TATGTTTCCA CTTCATATAA
     421  CCACAGAAAC GATATGCACG ACATCAAATA AACATTTCAA ATGACTTTAT TACAAAATGC
     481  ATTTCCTCCA CTCAAAGGAT TACTGTAGGT CACAGAATCG CAACGCCTGA CAGACACTTG
     541  TTGATGGAAT ACCTTGTCGA ACACGTTGAA ACATTTCCTT CACATTCTCG GACACGAATT
     601  GCACACAAAC AGGCTCCATG AGTCCCATTT TATCGCACTT TTTGAACATT GAATTAACTG
     661  TTGCAGTAGG TGTCATCTCT CTTTCAAGAG ACTCACTGAC AACTGTGCAC AACATGCATG
     721  TCACCACGGA AGTTGCTTCA CAACGAAGTG CTGAACAGAG TATTGTTATT AATTGGTACT
     781  GATCAAGCTT AAGCCTTTTC TAAAACTGAC TGGGAAACAT CTAAAACCTA GTGAATCATA
     841  AATTATCATC ACCTTCCGTT CCAGGTGCCG ACCTACTTTC GGTTTCAATC GTTATCTCCG
     901  ATTATCTTTT GCCATCCGAA AGATAAATGC CAGTTGCATA TACACAAAGA AAAGTGAAAG
     961  AACGTTCATA AAGACGATAG CTTCTCATAT TTCTGGTAGG CGAGGAGAGG AAGTTTTTGA
    1021  AATTGACTCT TTTGGCTTTT TTGTAGTCAA ACAGAGCGTG CTGCAGCTTG TGACTCATCA
    1081  CAATACTTGA GCCATTTGGT TTGACACCAC ACTTCACAAA ACCAAAGTCA CAAATCATAG
    1141  AAGTTGAGGG AAATCTTTCT ATTCGATGGC TTTCAGATCT AGTCTTAAAT AGCGACAATA
    1201  TTTGTTCAAA AAGAAACAGA ATCGTTCGAA ATTCTGATAT TTATCTTAAA TCAAAGCGTT
    1261  ATTTGGCTTT TTTTTTTAAA GATCTCTATT AACAGAAACA CCACGATGAC GCGTGAGTTT
    1321  ATAACTTACA ATTGGCAACA AGAATAGTGA ATAAAACTGA CAAGGCTACA CTTGACGGGC
    1381  AGACCATCTC GGAAGACGAC GAAACGGACA GAATGATCTA GAAGAGTCTC GTCTGCGGGA
    1441  TTTCGACTCA GCGTCGTCAT CCCTTCCGGA ACCTCCATAT CAAATAGCAC CGTTTCTCGC
    1501  TTCTCCGCCT CCCCAGGCAC TATTATGAGC TGTTGTGTGT GTGCAAGCTC ACATCATACA
    1561  AGAAATCTCG AATTCCCACT AATAATAGAC AATGAGACTG ATGTTTTGAT TGAGTTGAGA
    1621  TCGTTTGATT AGTCAGAATA GACGGAAATT GGATGGACCA ACAGAAAAAG AGAGGAACGC
    1681  GAATCGAAAA ATATAACTGT GGAAATCGGC AAAAAAAAAG AATGATAACA AAAGGGAAAA
    1741  GCGCGTGGCA TATTCTTCCA ACAAAATATG TGTTTTTTTG GCGACCGACT GTGCAACTCT
    1801  CTCATCATTT ATATTCTACA CAAAAATAAT TCGGAATATC CAAAAACATG CATAAAGTCG
    1861  CGGAAATGTT ACGAATGTCA ATCCGAA

//

LOCUS          9     2327 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      9
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          complement(join(500..638,794..876,1711..1827))
                  /translation="MQYICLVSALFVLVLSRGIPTIPQPRIDVWSGPNCHMCEVILN
                  NVRYEYRNNFTNVTPDQLRQSLYQQCDLNLNGFTDQECHKIADEDSADILQQLQEGT
                  SSYRICQRELLC"
ORIGIN
       1  ACATAATTTA ATTAAAACTA AAAAGTAAGT TTGAAAATCG AGAACAAAAA AATAATCAAA
      61  ACGTCATTTG AAATCGTTTG ACAAGAGGAA ACAAATGATT TTTTGTCTTT TTTGTGTAAC
     121  ATTTCACAGA GAAAGATGTC TTGGCAAACA TCGGCGGGAG ATTCAAAGAA ACACATGGCA
     181  GGGAAGAAAA ATACTTTTTG GAACTTTTTT TCTTGAAAAA ATCTGAAAAT TTTCTTCATC
     241  GAGTTGGGTG AAATAGTATT CAGAAATAAT ATAACTTTAA TAAAATGTTC TTCAAAAAGC
     301  TTCTAGATTG TTTGCCTACC TCTAAACAGT CTGAAAATCA TCAAAAGTAC TTAGAAACTT
     361  TTTTCTGAAA ATAAAATGTT TCAACAACCC TGTCAATCAG TAAACAAGCA GTGTTTCAAG
     421  AAGAACAACA CCATGAAAAT TATATTTGAA TATTTAAATT CAATGAGTAT AAAGACATTA
     481  GATAAACAAG ATAAAAATTT CAGCATAAAA GTTCTCTTTG ACAAATTCGA TATGAAGATG
     541  TTCCTTCTTG AAGTTGTTGA AGTATATCAG CAGAATCTTC ATCCGCAATT TTGTGACATT
```

```
     601  CTTGATCAGT GAAGCCGTTC AAATTTAAAT CACATTGTCT GAAAATTGTT TGTGTTGGGA
     661  AGGCAGTTGT GGTCGGAAGG TAATGAGCCG GAACTTTGAT AAAATATCGG GTACATATGA
     721  CTGTTCAAGA AATGATAAAA TAATTAAAGT GACGAAGAAA TATGAGAATA GACCAAAATG
     781  TGTTAGTACT TACTGGTAAA GAGATTGCCG AAGTTGGTCA GGAGTTACAT TTGTGAAATT
     841  GTTACGATAC TCATATCGGA CATTATTAAG GATTACCTGA AATCGGAAGA AAATTTAGAT
     901  TCCTATTAGG TTTTTTGTTG GAATTATTAT AATAATCGCA TTCAATTTCT AAAAAAACAG
     961  GTGTTGAAAT GTTGAAATGT TGAAAATTTG CAACAAAAAA TATTCCAAAA GTTGTTTCTA
    1021  GAAACTTGCA AATTATTGAA AAACTTACAT CAGCTTCTGT AACTGACGGA AGAACTCAAT
    1081  TTTGAAGCAA AAGTTGGGTT CATTTTTTGA GATTTTTCCT AAACTTTTGA ACTTGTTGCT
    1141  ATTAGGACGT ACAGAACCCC GTCTTTGTTT ATTTCTAAAA ATCTGATTTT TTCGGCATAT
    1201  ATTTCAAAAG CGAGCATTAT TGGAAATGTC TCCAAACGTT TTTCTAAGTC TATTTTATGT
    1261  TTAGGATGTT TTGTGAACCT GCGAACAACT TGAGGTGTTT TCAAGAACTT TTATACAAAA
    1321  AAAAGAAAAA TTCCACAAGT CTTCCAGATT GTTTGGGTAC CTCAAGACAT TTCAAAAAAT
    1381  TATAAGTCTG CATTGTCGGC ATTTGTAGAT TTTGCAAAAA TCCACCTTCT CGAACAACTT
    1441  TACTGGCCGC GAAAATAGCA AAAAAGTTGT TTTAAAAAAT AAATACAGAT CAAGCTGTTT
    1501  TCTAATTTAA AATGCCCGTC TGTAGTTAAT TTTTTTTTGT AAAAACCTAT ATACGTTATT
    1561  TTTTTCAAAT GGCATAAATG TTATTTGGAA GTACAAAAGT GAGCTTAACA CTTTCTTCAA
    1621  AAAAAGCTGA ACAAATTATT TTTTAACTTA TCAACTTCCC ATATTTCCTT TTAAAAAAAC
    1681  AACCGAGACC GAAATATAAA ATTTGCTCAC CTCACACATA TGACAATTAG GTCCTGACCA
    1741  TACATCAATT CTTGGTTGTG GAATTGTAGG AATACCTCGA CTCAAAACAA GAACAAAAAG
    1801  AGCCGAAACA AGACAAATGT ACTGCATGAC ACTGCCCCAA AAACTGATTC GTTTTCTTCA
    1861  TTTTGTTCCT TCTTCAGTAG CTGCTGCCCC TTGTCAATTT GTTGACGTCA TCATTTTATC
    1921  AGTGGTCGTG AGCTGTAAAT TTACTAGACA TTTTCGTGCA CAGCCACATG CTTTTTCAGG
    1981  TTTTTCGAGT GGAACGATCT AAATGGTTTT TGATATGTAT TTTCATGTAC CTGGTTATTG
    2041  CTCTTTGTTT TCAAAAATAA ACTCCTTATT CTAAAATTTT AAATGTTTAT CTTGTTTATT
    2101  CAAAAGAACC CTGATAGAGC CTACAATGCA TTGACTGAAT ATATTTTAAT CATTATGGGT
    2161  AATTGGAGTT TGCAGCCGAA TGTTTTGCAC AAAAATCAGT TTGTTTCGGC GGAGCCATGT
    2221  AAGCCCCTCT ACTGCTCTTG AACATTTGTA AAATATTCAG AAATTTCAGA AAAGCGATTT
    2281  AAAACGAGGT GGAGTCATCC ATAACATCCT TAGTTCTGAA TGTATCA

//

LOCUS          10     1867 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      10
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          join(500..551,1081..1367)
                  /translation="MFKIIVLLATILIVTVYGGCPGISSIYLPRRVNCPLCDFAPFS
                  QAVTGSGNQYSIKCPAGSDVWYYNEQGREFVTPMSKVGSKGSIKCDGSKWKISVPGQ
                  SQTLSAFACGSS"
ORIGIN
       1  ACTAGTAAAA TCGGCAAAAA ACCATAAAAA TTCCCGGTTG TCTCGCCAAT ACCTGACGAG
      61  ACCCAACGTT TGGGGGCGGA TCCTTTTCAG TTGTCCGTGT CAGTCTGGAA ACCTATATTG
     121  GCTACTATCT CGAAAACCAT CACAAAATCG ATTTTGCGAT GCATATAAGA AATGACCGCA
     181  ATGAATATAC CTATCTTTAT TTGTGATGAA TTTTCGATTT TGTGCTTCCT GCCCAAGTTA
     241  TGCACGTTTG TTCGGTGGAG TGCGTTTGCA CCCATCTAGC AACTGAACCA CCGTGAAAAG
     301  CGCATTTTTC TACGTCTCAA TAATATTTCC CTTTGTTCAG TCTTTTTTTC CCGTTTCAAT
     361  TCAAAAACTA ATCAATCATG ATGAAGGAAG TTTTGTGTGT GGGGCGGGGT GAGCGTTCGT
     421  CGCCGAAAAA AAAACAGCAG TAACGATATT TAAGGCTGAG AATTAGGGTT TGAGTACTGA
     481  TAGATAGGTC AGCGGCGAAA TGTTCAAGAT CATTGTTCTA CTGGCTACGA TTCTGATCGT
     541  CACTGTCTAT GGTGGGTTTT TTGAGATTAT TGATTTTTCA ACAAAAAAAA ATTCGAATAA
     601  TGTTTAATCA CCCAATTCAT TAAAAAAAAG AAGTGTAATG GGATTAATGA AAAAATAGGA
     661  TTTTTTATAA ATAGTTTTTA AAATTTGACG AATTTTTCTG CTTTTCGCCA GATTATTTGA
     721  AAAGTGAATT TTCGTAAGGG GAGTCAGAAT GTCCTATTTG GGTTTTGATC TTCAAAAAAT
     781  GCGGGAGAAG AGACGCAGAC TTCTCATCTT ATTTCGCCTG GTTAAGAGCC TACTGACGTC
     841  ACAATTTTTC CGGAAAAATA TTCCCGCATT TTTTGTTTAT CAAACCGCAA TGAGACAGCC
```

```
     901  TGACACCACG TGTTTTCAAA AATATCAATT TTGTACAGGG AATTTGTTTA TTTCCCCAAA
     961  AATTAAATTT TTTTTCGAAT TTGCGCGTCA AACATGGTGC ACTGTCGTAG TTCACTGACC
    1021  CAATGCATCA AATTTTACGC GCAAATGCGG CCCTAAATGA ATAATTTGAA TAATTTCCAG
    1081  GTGGCTGCCC CGGCATTTCC AGCATCTACC TCCCCCGCCG AGTCAATTGT CCCCTCTGCG
    1141  ATTTCGCTCC ATTTTCTCAA GCTGTAACCG GCTCCGGCAA TCAATATTCG ATAAAATGCC
    1201  CCGCTGGCTC TGATGTTTGG TACTACAATG AACAGGGACG TGAATTCGTG ACCCCAATGA
    1261  GTAAAGTTGG ATCAAAAGGA AGCATCAAGT GTGATGGATC GAAATGGAAG ATCAGTGTGC
    1321  CCGGACAAAG TCAGACACTT TCTGCGTTTG CGTGTGGATC CTCGTAATAA GATATAAATA
    1381  ATCCAAAAAT TGTTAGGAAA CTTATCAAAT GAAACTAATA AATACTAGTA CAAATGAATT
    1441  ATTATTCACA GAAAAAAATA AATAAAAATA GACAAATCAC AGAAGCTCAT TGGAAATTAT
    1501  TTGGTACATT TAAATATTTG TGGAAGCAGA AATTGGAGGA AAAAAATAAA TAATTAAAAA
    1561  AAACCGTAAA TCAGGCAATA GAAAATATTA AGAAGCTGCT GGCCAATAAT TGTTGACTAG
    1621  AAAATTAGAA AATACATGAG AAAAAAAAAA TCAATTAATT GTGTTTCGGT CTAAACGAGG
    1681  CGTCGGAAAT ATCTGTAATA ATATCGGGAA TATCGATTGA AACGGAAACT TCTTCGCTCG
    1741  AACGACGGCT GTCTCGTCGT TTTGGTGTCG GTGCCATCGA GTCCGTTCGA CTCTGAAAAT
    1801  TTTGGTTTTT TGTTGAAATT TGGAGGCAGA TACTGTTTTA CTACTTGATC TTTTACTACG
    1861  TATCTCT

//

LOCUS          11     2491 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      11
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          join(500..544,1126..1182,1243..1371,1884..1991)
                  /translation="MSIDRARAGTIERRQRLHDLVDLREELRDLQANLEHEELRFAD
                  AVKERSSIMNKNDQEDSEDEDHFLESLRLTEENIHSHRQKILEIQGKIIDKKESILS
                  LEILVKLHENDK"
ORIGIN
       1  TTGCCAAAAG TTGCAAATTT AACATTTGTT TCAAGTAAAA TTCCTGTGAC TTTTGAAATT
      61  ATTCAGACCG CGACGCGACC GCAACGCAAC TCCTTTTTTT CAGATTTGTA AGTCGGCTCT
     121  CGGTGTTTTT TTTGCAATTT TAAAATTTTA AACAAAACTA TTTTAAAGTA ATTTTTTCAA
     181  AAATTTTTGT ACTAAGAAAA CCACAAAGAT ATATCGAACC TGAAAATTCA AAGAATTTTT
     241  TGAACTCTGC CACTTCCCCA TGGTTCTTTG ACATAAAACC CTAACAGAAA AAGGCTTATC
     301  CAATATTCAC ATTCACTTTT ATTCGATTCC TCTTCTTTTC TTCCCCCTTC TTTTTGCTCC
     361  TTTTCGCCTT CCACCTTTCT GCTCCAAGTG TGCTCTTTTT TGTGCATTTT CTTGTGCGAA
     421  TCAACAACAG CTTAGCTTTT ATGACTAGTC CCACATTGAA AAACAAAACG AAAAACGATT
     481  TCCAGATTCA CACGCAGAAA TGTCTATTGA TCGTGCACGT GCTGGAACTA TCGAGAGGAG
     541  ACAGGTAAAT TTTTGATTTA TTCGATAAAT TGAAATGTTA TTTGAAAAAT ACAATGTCCG
     601  TTAACTTTTG GGTAAGCTGG AAAATTTCGA TGAATTTGGA GCATTTTGAA CTAAATTACA
     661  GCCGTTCAAA TTGTGAGGGG CGCGGCGCGG TCGCGTCGCG GTCAGCTCAA AAATTTTGAG
     721  CGACTGTAGG TTTGCCCCCA AAATTCGTAT TGCGTTGAAA ATTTTTGGGG ATTCTTAAAA
     781  TACGATTTTC CATCAAAAAC TGAGTGGCTT GGAATTATTG AAAATCAAAG TTACCTTCGG
     841  AATTTTCGGA ATTTTCGGAT TTTTAAAAAA AATTTATTAA CAAGTTATAT GAATTTTAAA
     901  TTTTTGAATA AAATCATGCT TTGGGTCAGC TGAAAAAATT TCAAAGAATA ATAAGCATTT
     961  TAAGCAGAAC TACAAGCACC CAAAGTTTGT AGGACGCGGC GCGGTCGCGT CGCGGTCAGC
    1021  ACTCAAACTT TGTGCAGATA TAACATTACT CAATGTTGTC GTAGATGTCT GAAAATTTTG
    1081  GAGCTAGTGA GCAACTGGTT AAACTACAAA ACTTTTAATT TTCAGCGTCT CCATGACCTT
    1141  GTCGACTTGC GGGAGGAGCT GAGAGATCTT CAAGCTAACT TGGTGGGTGC TTCTAGTTTT
    1201  CAGTGCTTTT TATAGTATTT CAGTATTTAT TTTATAATTT AGGAACACGA AGAACTACGG
    1261  TTTGCGGATG CCGTAAAAGA AAGATCCTCC ATTATGAACA AAAACGACCA GGAAGATTCA
    1321  GAGGACGAGG ATCATTTTTT GGAAAGCTTG CGGCTAACTG AAGAAAATAT TGTAAGTTTG
    1381  GAAGATGCAC AAAAAAAATG CAGAGTTCTA GGATAATTTT GAACAAAGCT ATAAGCTTCC
    1441  AAAGTTTTGA CCGTGCGACG CGGTCGCGTC GCGGCTAATC TCTCATTTTC AATTACATAT
    1501  ATCTCAGCTC TCAGTTGTTC AATTGACTTC TAAATTGATA GGCGGTCTAA TCAAGTTGTG
    1561  CTCTAGTGGA TTATCAAAAA AAAAGCAAAA ATTCGAAGAT TAAAAAATTT TCTGTTTAAA
```

```
    1621  ATTTAAAAAA AAAGTTGACT CTGCTTATTA AAACCTAGCC GTTTAGAAAG CATTTCTAAT
    1681  TTCCGGTAAA ATCATCAAAT TTTAAAATTC TAAAATTTTT TGAGCAAAGT TAAAGGCTCC
    1741  TAATGTTCAA AGCGTGCGTC GCGGTCGCGT CGCGGCGGCA GTCAAATCTT GGCTCCTATT
    1801  GAAGCTAGCA CGCTAAATAT TTTTTTGGCG ACTGGGAGCT TAATTTCCTT TTAAATATCG
    1861  ATATTAATAT CTAAATATTA CAGCACTCCC ATCGTCAGAA GATTCTGGAA ATCCAAGGGA
    1921  AGATAATCGA CAAAAAGGAG AGCATTTTAA GCCTTGAGAT TCTAGTGAAG CTTCACGAAA
    1981  ATGATAAATA ATTATCGGGT TTTTTTTCGT TCTTGATTTT TCGCACCTAA AATTAAATTG
    2041  TTGGTTTTTT GTCCTCGCTA GAAATAAAAC AAAAAATCCA AAACACCAAT TTACCCGGTT
    2101  CCAAAGGGAA AATACTAGTG CAATCTCTAG AGAAAACTGG TATATTTACA GAAAAAATTC
    2161  AGTTCTTCTT TTTCGTTAGT TTTTGCTCTT AAAATGTAAC GGTTTTGTCT AGAAGCTTTA
    2221  AAACTTTAAT ATCTTAAAAC CAATAACTTT TGTCAAAATT TGCTAACTAT CAAATTAATC
    2281  TACGCAACAT TTCCTATATT TTGTTAGTTG ACAACTTTTT TCTAGCTATT TTCCCGGAGG
    2341  AGTTACGGGG ATTCAACGGT GGCAAGGTAA GAGAGTGTGG GAGAGAGAGA GAGAGAGAGA
    2401  CTCAGACAGC GTGCGCACTC TCCCCACCTG CGTCTCTCTC CTCTCGCCGT GATACTTTTA
    2461  AACGCATATA ACTTAGTGTA GGAGAACGCT G

//

LOCUS          12     4431 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      12
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..622,3660..3720,3771..3931)
                     /translation="MKLLILQLTLFQLNSIFAQDNAADLEFSPYVFKHVDIPFKNNN
                     YAKDWFGSDVALVQHPLLAPAKSVKPGNPVSLPTIAPLIVKWSDLNNVKVNEQIGQN
                     FEKRIDSHVWWPNK"
ORIGIN
       1  AAAAATTATT TTCAATAACT ATTAAAAATT TGAAACATTA ATCTATGTCT TTAAATGTTA
      61  GCCAACAATT TTTTGAAAAA CGGCTAAAAA TTAAAAAAAA ACTGTTTCAG ATTATTAGTA
     121  AGTTCATTAT GTTAAGAAAA CTTTTCATTA TGTTAAGAAA ACCCTCGACT TTTCTATTTA
     181  CCTTCTCAAA TCTTCATTTT TGAAAATTCA TTTTGTGAAA AGTTTTTTAA ATAGCACTTT
     241  TCAGTTTTCT TCAACGTATT TTAACTAACT CGCTGTGCAC TACTAATAGG GGATGACCAA
     301  ATCTTCAAAA TAATTCCAAT TTCCCATAGC ATGTTGCCCA ATGATAAATG TATACCTAAA
     361  CAGGTTTCCT GTTTCACATT TATCTTATTA CACTGTTCAC TTCTGTTTCT GCCCCTGTGA
     421  TCAACATCAC ACAAAAAAGT TATTTCTCAG AAAAATATTA TTTTTCTGTT TCATAATCTT
     481  CATATATTTG TGTCTAAATA TGAAATTACT GATTCTCCAG TTGACTCTAT TTCAACTCAA
     541  TTCAATTTTT GCACAAGACA ATGCGGCGGA CTTGGAATTC TCCCCGTATG TGTTCAAACA
     601  CGTGGATATT CCATTTAAAA ATGTTAGTTT GCATCGGTTC CGTTACCGAT TTAAAATAGA
     661  AAAGTCGAGG GCCCCGGAAA ATGTCGGGCA TTGTGAAATA TTATTTTTTA GTCTTCAATG
     721  GCCAATTGAG ACCAATTGAG TAATTTTTCT TAACATAATG AACTTACTAA TAATCTGAAA
     781  CAGTTTTTTT TTTTAATTTT TAGCCGTTTT TCAAAAAATT GTTGGCTATC ATTTAAAGAC
     841  ATAGATTAAT GTTTCAAATT TTTAATAGTT ATTGAAAATA ATTTTTTCCA AATATTTTGA
     901  AGTGCGAATA TGTGTAGTAC AACACTGAAT GCACCAAAAA ATTGCAGTAT TTCAAATTTT
     961  GAATTTTGAA AAAATTTTGA AATTATAACA TTCTCAATAC ATTTTATCTG TTTCAGTAAG
    1021  TTTTTAACCA CGCACTAATT TTTTTCGATA TAGTTTTCTG TTCATGGAGA CAATTTTGAA
    1081  AATCAAAAAA ATCGTTGACA AAAAGTTATT GTAAGCTTCG ATGGAAAAGT ATTTTCTATT
    1141  TTTTTTTCTA AAAATAAATT TTTCAGTTTG AAAAAATTAG TATGAATTTT ATGGACATAT
    1201  TGATTGAATT CAGACAAAAT TAAAATTAAT AACTAATTTT CATGAAACAG TAATTTTTTT
    1261  TAATTGAAAA ATGAATTTTT AACAATACGA AGCTGTTTGA TTTAGTTTAT TTTTGAAATA
    1321  GTTTGATGTC CTGATAGTTA ATTTTATTTC CGGAGGGAAA TCTATTTGCA TTCACAAAAT
    1381  TTGACTATTT CAAACCATTT TTAAATTGAC AGCAAAATTT CTAAAGATAG TGATTGTTTG
    1441  GCCACAACAT TTCATATAAA AAAATTAAAT TTTAAATTAA TGACAGGAAA ATATATATGT
    1501  TTCCAGAAAT TTGAAAATTG ATTTTTTAAA GAAAAAAATT AATTTGAAAT AAAATTTGTA
    1561  TTTTCTTATT TTAACCACAA GAAATAAAAA AACCGCTTGT TCAGACATAT TTCATGTACA
    1621  TGTGAAGCTA CAAATTGCTA ATTTTTTTTC GGTTTCGTCC GTTCTATAAT AATAAATAAA
    1681  AAATAGAAAT CCTTCTTCAA CTAATGGAAA TAGTGAAACT GAATTTTTTT TTTGCAAAAA
```

```
    1741  AAAGTTGATC TCACCACACA TAACTTTTTA CGGAGATTTC AAAATGCAAG TTCCACACAA
    1801  ATTTAAACAC TGAAAATTCC TGCAACAACA ATAAATATTA TAATGACTAC AGTACTCTTT
    1861  AAAGGCACAC GTATTTTTGG ACCTGAAGAA ACGCGAATCG TAGTGTTTTC TTTTTCAAAC
    1921  CTTGAGAATT AATCACAGTT TTGAAACAGT AATTTTTTGA TTTAAAAAAT CCGAAAAAAT
    1981  TTCCTGTTTC AAACAGCTGC CAAAATCAAA AAACTTTCAA ATAATTATTA TCATCTAGTC
    2041  ATTGATTCAT CGATTTTGTT AATAAGCTTT CCATTGTTTT CTTTTCAAAT TTTTGGCGTT
    2101  CGAAAGTGAA AAATCTTTTT TCTTAAAAAA TGTCTTTGTG ATTTTTATTT TATCATTTTC
    2161  GGGTGAGAAA TAATGTTTTT TTAAGCAATG AAACATCTTC CAAAACTTAT TAAAAATGTT
    2221  TGAATTCACT GGTTTACCGC CAAATACATA CACATCGAGA CCCGACTGTT GATAACGGGG
    2281  GCGTTTTTAA TTGGCCGTGG AGCACATTTG CAATTTTTAA ATTGCAATAA TAGCAAGTTT
    2341  TTGGGGAACC TAACTGTTTT TCTTTTGAAA TTGGGTAAAT TATAAAGAAA TGCATTTCGA
    2401  AATGTTAAAA CTATATGTAG TTTCTTTTCG ATTTTTGAAA AAAAAACATC CTGTTTCAAA
    2461  AATTTCAGAA ATTAGTACTC AAAAAAATTT CTATGAGGAT TAATTGTATT TTAGGGCAAA
    2521  AATCAGAAAC AATATATTCA ACAAGTTAAA AAGTCCATAA TACTGTTTTC TTTTACAAAC
    2581  TTTGAAAATT AATCACAATG TTGAAACAGT GGATTTTTTT TTCAATGAAA AAAATCCAAA
    2641  AAAAATTTCC TGTTTCAAAA AGTTGTCAAA ATCAAAAAAC TTTCATCTCT CTATTATCTC
    2701  ACATCTATTA TCATCAATTC TTGATTTGGT TAATACAGTG CATATTTTCC AACTTCTACG
    2761  AAAGGCGCGC GCATTTTCGC GCAATGGTCT CGCCACGCGC ATCCCATTGA TTTGACGGGA
    2821  CCATTTTGTA TAAATGCGCG CGCCTTTAAA GTCGTAGAAG TGGAAAAAAA TGCACTGTAA
    2881  GTTTTCCATT GTATTTTTTT CAAATTTTGG ACGTTCGAAA GTGAATAATG TTTGTTTAAA
    2941  TAATTTGTTT ATTTTATCGA TTTCGAGTGA GAAATAATGT TTATTTTAGC AATGAAACAT
    3001  CTTACAAAAT TAATTGAAAA TGTTTGAATT TACTGTAGTA CCTACAAATA CCCATCGAGA
    3061  CCCAACTGTT GGGGGCGGGG GCGTTCTCAA ATGGCTGTGG AGCACATTTG CAATTTTTTA
    3121  ATTGCAACAA TAGCATGTTT TTAGGGAACC TAACTGTTTT CCTTTTGAAA TTTGGTTAAT
    3181  TATAAAGAAA TGCATTTCGA AATGTTAAAA CAACATGTAG TTTCTTTTCG ATTTTTGAAA
    3241  AAAAAATTCC TGTTTCAAAA ATTTCAGAAA TTAGTACTCA ATATAAAAAA GACAACAAAT
    3301  CGGCAAAACA AATTGAAAAG TGTTTAATTT ACTGTTCTGC CGATAGATAC CTAACGAGAC
    3361  CCAACTATTG GGAGGCGGAA CGCTTTCAAT TGACCGTGGA GCGCATTTGC ACGTTGAAAA
    3421  ATAAAAGAAA AATTCCTCGA ACATGGGAAA TTAATGCCTA CAGTACTTAT TAAAGGCGCA
    3481  CACTTTTCCG GAATAAATAT CAAAGGAACT GTTTTCCTTT TAAAATAGGG GAAATTAATA
    3541  AGATAGTTGA AACAGTAATT TTTTTCTTTT TTTGAAAAAA AAATCCTGTT TCAAACAGAT
    3601  CTCAAATTAA TATTAAAATT TCTATTACGA TTTTCCTATC AAATTATGTG CCTATTCAGA
    3661  ACAACTACGC AAAAGACTGG TTCGGCTCCG ACGTGGCTCT CGTACAACAT CCATTACTAG
    3721  GTGAGAAGCT TCATCTCCAG CTTAAAATCC AAATTTCAAA AAATTTCCAG CTCCTGCAAA
    3781  ATCAGTAAAG CCAGGAAATC CGGTGTCACT GCCAACAATT GCTCCATTAA TTGTGAAATG
    3841  GTCAGATTTG AACAATGTCA AAGTTAATGA ACAAATCGGA CAGAATTTCG AAAAGAGAAT
    3901  TGATTCACAT GTTTGGTGGC CGAATAAGTA GTTGAAATTT CCAATAAAAA TCTTTTGTTT
    3961  TAAAATGAAT AGGCAATCAA AGGCAAATAC AAAGAAAAGC ACCTGATCTG TGGGGGAAAT
    4021  GTATAAGAAC AGAGCGAGAA ATTGTACAAG AATGGGTGAG GAGAAGTGTA TTATGTTATA
    4081  AGTACACCGG AAACCTGTAG GGAAAAAAAC GGAAAAAAAA GGAAAAAAAT AAAATAATTT
    4141  AGAAAATAAG GCATCATCAA ACACGATCAT AACACACACA AAACGTCGCG AGATAATAAT
    4201  TTTTGTCGAT AATACAATGC GGAGTGTCGC CAAACTTGTT CACGTGGCTC ATTGATCCAA
    4261  AATCCAGCTT CACCTCATCT CGTACAATAT CGTAGAACAG GGTAACTTCG TAGATCGCTT
    4321  CTCCGGGAGT TGTCTGCAGC TTCAGTTGAT AATGAGCGAA TGCGGCTTTC GTGGAACCTG
    4381  AGAGGTCAGG CACAAAACCA TCGGCATCTT TCACGTTTTT GTAGGCGAGT A

//

LOCUS          13     1489 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      13
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..576,707..989)
                     /translation="MSKHLYKQYVRIAGKWPKDAFRAPERDFSNFLVKEIERQFQPG
                     VVASQAICEKRLQALEQLLNNEVLKKHPNSYSSGVFGMRLADLQAASSEESRKQMGL
                     QPKVSIFKRLYRSVVPEKK"
```

```
ORIGIN
       1  GAATTTTCAT CGAATCCGGA GAATTTCGCT ACATTTTGAA CGATTTAATG ATTGCCGACT
      61  ACATTTTCTG GATTCAAACT GTGCCCGACG AGATTCTGAC GCGAATTCAG ACGGATTTGA
     121  CGGAAATTTC GGGAAAAATC GATAAATACG ATATTGGATG GGATTTGGAG GTGTTGGAAG
     181  CCGAGGCGAA ACTGGCGAAT ACTCAACTTG ACAGCGATGA TGAGCCCGAT AATTGATATG
     241  ATTTTTAGAA AAATCGATAA TTTATCGATT TTTTTTTTCA ATTTTTTTGT TATTTTCTAT
     301  TATTACCCAC ATTAATTGAT TTTTCCGCTA AAAAATCGAT AATTTCACAA TTTTTTCTGT
     361  TGTTGCACAC TTTGTTGAAT TCTAAAATAA TTACTCAATT TTTTCTCCAA ATATCGATTT
     421  TTCGGGTTCG AAGCTCCTCA CTCTACCTTT TTTCACCGAT TCATTGATTT TTCAGCTAGA
     481  AAATCCGATA TTCTTACAGA TGTCAAAGCA TCTCTACAAG CAGTACGTCA GAATCGCCGG
     541  AAAATGGCCG AAAGACGCAT TTCGGGCGCC GGAAAGGTAG GGTTATACGC GACGCGACCG
     601  CAACGCGCGC TTCACTAAAA AATAATCAAT AAAGTATCGA TTTTTCCGCT AAATCTTCGC
     661  TAAAATCTAC AAATTATCGA TTATTTCCAA AATTGCCCTT TTTCAGAGAC TTCTCCAATT
     721  TCCTCGTCAA AGAAATCGAA CGGCAATTCC AGCCGGGAGT CGTCGCCTCT CAAGCGATTT
     781  GCGAGAAACG TCTTCAGGCG CTCGAGCAGC TGCTCAACAA CGAAGTGCTC AAAAAACACC
     841  CGAATTCGTA CTCCTCCGGC GTATTTGGAA TGCGATTAGC CGATCTACAG GCGGCCTCGA
     901  GTGAAGAAAG CCGAAAACAG ATGGGTCTAC AGCCGAAAGT ATCGATTTTC AAGCGATTAT
     961  ATCGATCCGT CGTGCCCGAG AAGAAATAAT CGATTTTTAT TGGATTTTTT TAGCTTTTTG
    1021  AACGATAAAA TTGCTTTATT TTCTGATTTC CCGTATTTTT CGATCGAAAA CTTAAAAAAA
    1081  AAATGTAAAA AACAGTCCGC TAAACAGGGC GCGCAAGCGG AATATCGGAG TGTCGTTTGT
    1141  GTTTTGAATT TTAAAAATTT AAATTTTTGT TTAGAATTAA TTTTTTTTGA TTGAAATTTC
    1201  GGTTTTTCCG CGGTTTCATC AGTTTTGTAG TGAAAAAACA TTAAAAATAG CGGTTATTCA
    1261  TTAAAAAAAA TCTAAAAATG ATCGGCTGTT TCGATGATTT TTACGAATTA CGCGCGTGAA
    1321  ATTTTAATTT TGTAGCGAAA TTTGAAACTT TTCCAACCGA ATCGCTTATT TTTAAGCGTT
    1381  TTTTTTGTTA ACTGTTCGCT ATAAAAAATA AAAAATTCGA CAGAATATCG GAACATCGTT
    1441  TTTTTCCGAA ATTTTTATTT TTAAAGCCAA ATAGTTTATG AAAAATCAA

//

LOCUS          14     1782 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      14
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          complement(join(500..624,982..1123,1169..1282))
                  /translation="MRSSFLLVLLGIAFCQKASWNKATANERLDDAITNEFNVTDMD
                  GNGLVTIEEIREYYENHLNKLSNEGIDFIDRNFHNEKSVDIDTYRKLVKEWLTEKKS
                  VEKILVGLFGNRREQLQESSSSTVSS"
ORIGIN
       1  ACTTCCCACG TAACTACAAT AACAATGCAT TCTGCACGTA CCAATTGTCT ACACTAGGAT
      61  CATATAACAT ATTGCTGAGA TTTGCACCAT TTTCCCTTGG TACTTCTGAT TATGTGATAG
     121  TTTACGATGG AGAATCGGCA AGTAGTCCGG TTATTGGAAA TTATAGTGGA ACCTTGGTGC
     181  CGTTTAGCTT AATTTCAAGT TCGAACAATT TGCTAGTGGT ATTTAATGGA AATACCGGTG
     241  GCCAGGGATT CAGTGCTAGA TATACATCAT ATGCTCCTTT AATCTAGTTA TTTACACTTG
     301  TTGGCATAGA TTTCAAGTTT CTTTTTTGTT TTGTTCTTAT TATGCATCGC GTAAATACAA
     361  ACTTGCGAAT TTTTATCAAA ACAAATCATA TTAATCATGT ATTTTTTCTG TATAAAAATA
     421  CTTGCAAATA TTTGTTGCAT TAAATTTCAA TGGATAAACA GTTTAAAACA CGTTTGAATC
     481  AAGAATTGAT AAAAAAAATT CAAGAAGACA CCGTCGACGA TGAGCTCTCT TGTAATTGCT
     541  CTCTTCTGTT TCCAAACAGG CCGACAAGAA TTTTCTCAAC GCTTTTCTTC TCGGTCAACC
     601  ATTCCTTCAC CAGTTTTCGG TACGCTGAAA AATATTTCAA AACTGTTACA ATTTATATAT
     661  TCTGTTTTTC CTACCCTACC CCTACTCGTA TTGCACCAAT TGAAATGTTA AACCTAAAAA
     721  ACAGTTTTGT AGTCACTGAT AGAATTGCGT CTATAATTTT CTGGCAAGGC TCAGTGATGA
     781  AATATCAATA AATTTTTCGA AAATAATAAT TTTTTAAAAT AATAAATTTT TTGAAGAATT
     841  ATATTTTTGT GCATCGCAAT TTAAATAAAA ATTCTAAAAA TTGATAAACT TCAATGTGGG
     901  GATGCATTTG AAAATTGTAC AAACTAAAAA TTTTAATTCG AAATGCTTGC CAAAATATCC
     961  ACATTTTAGA GTAAAATTTA CTGTCAATAT CAACTGATTT TTCATTGTGG AAGTTTCGAT
    1021  CAATGAAGTC AATTCCTTCA TTGCTCAATT TGTTCAAATG ATTCTCATAG TATTCCCGAA
```

```
    1081  TTTCTTCGAT TGTTACAAGC CCATTACCAT CCATATCGGT TACCTATATT CAAAAAAGTA
    1141  CTTGATTTAT TCTCAAAAGA AAACTTACAT TGAACTCATT TGTGATGGCA TCATCGAGCC
    1201  TCTCATTTGC AGTTGCTTTA TTCCAAGATG CCTTTTGGCA AAATGCAATT CCAAGTAAAA
    1261  CCAATAGGAA TGATGAACGC ATGATATTTA TAGAAAGGGA ATACAAAAAA GCGGACTAAG
    1321  ACTTACAAGG TATCAAAAAT CTGAAAACAA AAAATAAAAA AATGTTTTAA TTTTTTGCTC
    1381  ATTTTTTGTC TGCGTTTCTG CTCATTCGAT CTCTGAAAAA GTGGCCTCAT TTTTCTTTTT
    1441  ATAAAATGTT GTCACAAATG TTTTCCTGAA GATGTTTTAC TGAAACTTCT TTAATTTTTC
    1501  TACTTCTCAC CCGCGTTATC CCGCCCGTGA CCGATGTTTT TTCCTACTCA CCAAGGAATT
    1561  GTTTCAATAC ATTCATTTTT CTACTCTCCC GATAAACATG TATCTCTATC TATTACTCGA
    1621  ATGGGTGGAG TTTGACCACA CACGAACTAA ACTCTTCGAT TTCTTCTCGT TGCTGGTCAG
    1681  GGCTCATTGA CAATCGAATC AAACTTTGGA ATGTCTTCCA AGTCGTTTCT GATTCTAACA
    1741  CTCTTTATTG CTATTTCTAA ATTGGAAGCA CTCATTTGCT AC

//

LOCUS          15     2191 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      15
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          join(500..616,1071..1212,1558..1691)
                  /translation="MRQSILLILFGVAFCQKKNQWNSETADQRLDDVITNEFKARDM
                  DHNGLVTTKEIREYYESHPNKMSKEGIEFIDRNFHSGGAVDVDGYRKLIKNWLTEKG
                  SVEKILVGVFGNKTKPASTNNVGRATTVST"
ORIGIN
       1  ATCCCCAACG TTATAGTCTC GAAATTTGGA AATTCGAAGC CAACTTTAGT AGTGATTACT
      61  GATAGTAATG ATAATGATAG CGAGATATTT CTGTGATTTG TGTATGTATT CGTTATATTT
     121  TGTTTTTACA ATTAATCTAT TCTCCTTATT TGATTCACTA TTTCTGAACA CAAATATTTA
     181  TGTTTTCTTT TCTTTTCATT TTTTCGTTGC CACCTCCCTT CCAATCACCT CATCTGATCT
     241  CCATTGTGAT TACTTTTTAC TGATTACTGA GTAATTGTTC ATTTTGACCA TCTCTCGTTG
     301  CCACCGGGTT ACCCTATTAC TTGCCATATC ACGTGTCAGG CTACTTCAGT CTTTTTGTAT
     361  CATTTTTTCA GTTTTTCACG CAATAAATGA ATTCAATTTC AGTGTGTAAA GTGTAGTAGG
     421  AAACAAGAAT ACGGAGAGAG AGAAATAGAA GACAAGTCAC TTGCCATTTT TCAATTATTG
     481  AGAATCGATT CTTTAAATCA TGAGGCAATC TATTTTACTT ATTCTATTTG GAGTCGCATT
     541  TTGCCAGAAA AAGAATCAAT GGAATAGTGA AACTGCAGAT CAGCGACTAG ATGACGTCAT
     601  AACGAACGAG TTTAAGGTAT ATTTATAATA ATAATAGTTA ATAGTTAACA GTGTCTCCAG
     661  AATGTACCAT GACCTCCAGG AGTTTTGTGA ATTTTGCGCC AATCCGGTCT CGGCACGATA
     721  AATGTTTGTT AAATGTGAAA AAAGTGTGTG CCTTTAAAAC TACTGTAATT TCCAAAATTT
     781  GGTGCAGCGG AATTTTTATC AATTTTCATA GTTTTTTCTA ACGATTTGTT AAAGTTTATG
     841  AGTTTTTATT TTTTTGTTCT CATCGAAAAA GTCTGCAAAA ATTGAGCAGC GATACTATTT
     901  TGAAATTACA TTTAACAAAA TTAACAAAAA AACTTCGCGT CGAGACCAGG TACCGTACTT
     961  TTAGCGCAAA AATGCCAAAC TTTCGCGTCT GTGTAATACA CCAAACGGCT TGTACATTTG
    1021  AAATCACTCT TTGGCTTTTC TAAAATTACT GTATAGCATT TCAAATTCAG GCCAGAGATA
    1081  TGGATCACAA TGGCTTGGTG ACGACAAAAG AAATACGAGA ATATTATGAA AGTCATCCGA
    1141  ATAAAATGAG CAAAGAAGGC ATTGAATTTA TCGATAGAAA CTTTCACAGT GGAGGGGCTG
    1201  TTGATGTGGA TGGTGAGATT TTTAACTTTT TTCGATCAGT TTTGTGTCTT AGCACGAGCG
    1261  AAAGTTTATG AATGGCATGT GTCTTGAAAC AGATACGGTA TTTTAATTTG GATTTCATTG
    1321  AAAAAACAAA TTCAACGTTT ATAGCGAAAT CCTACAGGAC ACGAAACGAA TTTAGTGAAA
    1381  ATTCATAAGA ACTAAATTTA ACGTACCGTA TCCCTTTTAA GACTGCTTCA ACACCCATAA
    1441  TGTTGAAAAA CCATAGTGAT AAATTATATA TTTCTAGAAA ATGGGCAATC TTAAATTTAA
    1501  AATTTAAAAA AAACTCAATT GTACAACTAA ACATCAAAAT TTTTTTTCAA TTTCCAGGAT
    1561  ACCGCAAGCT GATAAAGAAC TGGTTAACTG AAAAAGGAAG TGTCGAGAAA ATTCTCGTGG
    1621  GTGTTTTCGG TAATAAAACA AAACCTGCTT CAACAAATAA CGTTGGACGT GCGACAACAG
    1681  TGTCTACATG AATGCACTAA ATTTATTGAC TTTCTTTGAT TATGAAACAA GACCAGACTG
    1741  AACATGTTGT AGACCACAGA ATAAATGTAA TAAATACGAA TTATCTCCAA AAAATTATTA
    1801  GGGGTTATAC CAGTCAATGA AGAGAAGTGC GAACGAAAGT ACAACAATTG AAAATAGAAG
    1861  GAATGCTCGC TGTTTTCCAG AGTTTGTGAT AGCTTTTCGA ATCCATTCAT TTCCGTCGGT
```

```
    1921  AATGTTTTCG GTTGTATGTA AAGCATGATC GTGTATTAAT TCAACATCAC GGTCTTGTTC
    1981  CAAAACCTGA AATAAATTGG TTTTGAAAAT TTTGAAAATA AATTTGCGAA TACCTTTTCT
    2041  GAAAACACCG TGTGTAAAGA TGCGATCTCC GTTAAATGCT GCTGAAGCTT GAGAATCTCG
    2101  TCATCCTGAT CCTCCTTTAC AAACTCTTTC TTTCGAAGCT CAGCTTGTTG TTGGAATTCC
    2161  ACGGCAAACT GCTCTCCTTG CTCATCTCGC C

//

LOCUS          16     2276 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      16
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..518,750..841,1015..1104,1524..1698,1754..1776)
                     /translation="MGNILCCKKNENEESMLTTAEKRPPKKKVRHSRKKQLVIYDPE
                     GISVHDLALLGQKNVADGLVKPDLKPKKMGRVTLMPDIQPHLTLKAGTTTISSQKPS
                     PASKSTDDFEISEVEDYVDDLYLQLMDPDRDL"
ORIGIN
       1  ACGCTGAATT CAGAAAAACT AGGTTTACCT GGAAAATACT CTTTTTTTAC CACATTTTCC
      61  AATTCCGATT TTTACTACGT TCCATATCTT TTACTACGGT AGCTTGTTTT ACTACAATGG
     121  CTGGTTGGAT TTATTACTAC GTTTTGTATT TTTTTACTAC AAATAACGAT TTTTTTACTA
     181  CCTATTGGAA AAAAGAGTAT CAAAAGTTGG AATTTGAAAG TTACCTTGTA GCAGATTTCA
     241  GGTATAACAA CAACAAAAAT TTTACAAAAA TTTTATGTGT TGGGGACGGA CGTAAGAGGT
     301  GCCACCCACA AAATGGTTAA ACTTTTTAGC TGGAAAAAGT TGGTAAATTT TCAAAAAAGT
     361  TAAAATAAAA ATATATATAA AATATTTCTC ACGTTATTTT TTGTTCGAAA TTCGAAAATT
     421  TCAATCGACC AAATGTATTA ATTTATTTCT CAATTTTCAA AAGTTAAAAC AAAGTGTCAT
     481  TCTTTCAATC GGTAACATAA TGGGAAACAT CCTCTGTTGT AAGTTTTGAG CGCCTACCTC
     541  ATACCTGCCT ACCTAGGCAG GTATTTAGGC TAAGGCTGCT TCAAATGCAG GTAGGCAGGC
     601  GTAGGTTTCC TTGACTGTAG TTAGGCGAAC GTTCGACGCC TATAAGGCAG CTTGGCAGGC
     661  GTAGGTAGGC TTGAAGGCGT ATAGGCAGGC ATTTGTTGCC TAAGGCAGCT TGCCAGATTT
     721  TCGTATCTAT TTGAAACCCT AATTTGCAGG CAAAAAGAAT GAAAATGAGG AAAGTATGCT
     781  GACAACGGCG GAGAAGCGTC CGCCGAAGAA AAAGGTCCGG CACTCGAGAA AAAAGCAGTT
     841  GGTACTTGAT AAAGATCGAA AAATCATTCA GAAGGAGAAG GAGTTGTATC AGGCAAGTGC
     901  GCTCCATTGA TAAGCAACGC GTGTACTCCA CTTGGAAAAA TTGGAAAATT GCGAAATATC
     961  GATTTTTCAA AACCATTTTC CACGGGGAGT ACACAGTTAG ACAACAGGTT TTAGGTGATC
    1021  TACGATCCGG AAGGAATTAG TGTTCACGAT TTGGCTCTCC TGGGACAGAA GAACGTGGCG
    1081  GATGGGCTGG TGAAGCCGGA TTTGGTGAGG AGTTGGGATT CTAGGCCACG AGTGACGTTA
    1141  TAACTTAGCT TGTCTTAGCT GTGGGAGCAC GGATTTCTGG CTTCTCTCAT GAATTGAAAT
    1201  GGAAGAGTTT TTGCCGAACT AGGCCATTTT GGCTCGGCCA TATCTTGGGT AGATTTACTG
    1261  CGCGTTGCGT GTCGCGTCGC GGCTCGATTT TAGTTGTAAA ACTAAATGTA TTTGTCCGTG
    1321  TGGAGTACAC AACTTTCCCA CGCGTTGTCC GTCAGGCAAT TGTCAATGGA GCGCGAAAAA
    1381  TTCAATGAGG AAGGCCAGAA CCCAGTGTAG GAGTTTCTAG ACTAAGGGTG ACGTCATCAA
    1441  TGTCTCCAGA CCACTGATGA CGTCATAACT TGGCGAGTGG GAGTTTGGCG ATGTCACTAG
    1501  AGTAATTCCG GCCTTAATTT CAGAAGCCGA AAAAAATGGG CCGAGTAACA CTGATGCCCG
    1561  ACATCCAGCC GCACCTCACA CTGAAAGCCG GCACGACGAC GATCTCATCA CAAAAGCCAA
    1621  GCCCCGCCTC CAAATCGACT GATGATTTTG AAATTTCCGA AGTTGAGGAT TATGTCGATG
    1681  ATTTGTACCT GCAGCTGAGT AAGTTTTGCA AAACAAAAAA TTCAATATTT ATCGATTTTT
    1741  CTGATTTCTT CAGTGGATCC CGATCGAGAC CTATAGACCC GCCTAACTTT TTGAATGAAG
    1801  TTTCGAATAA AGTTTATTGA TAAAATAATT GTGTACTTTG TGTGGACAAG CAATGACCTT
    1861  AAAAAAATTG CTATTGTTGG GGACTAGTTT TTGAACTGGT GGCCTAGAAA ACTCAGAAGT
    1921  TTGGTCACAT TTATGAAGAG GTTTCGGTGG AGGGCCACTA TTGGACAAAA AACGCTAGGG
    1981  ACTAGTTTTC CACCAGGCCG CCTAGGAATT CCAAAAGTTA AGCCACTCTT AAGACTAACT
    2041  TCGGGGTACT GTAGAAGTAC TGTAGTTTTA GAAAAATTGA CTTTTCGTCT TTTGAATGGA
    2101  TATTGGTTTG GGGTTATTGG AGGGATATGG TCGGGGTACT GTAGTAGTAC TGTAGGGGTA
    2161  CTGTAGGATT ACTGTAGTTT CAGAAAAATG TACTTTTTGT TTTTTTGAAG GGATATTGGT
    2221  ATTGGGTTAG TGGAGGGAGA TGGTCGGGGT ACTGTAGTAG TACTGTAGTT TCAGAA
```

```
//

LOCUS          17     5257 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      17
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             complement(join(500..669,2113..2320,4728..4757))
                     /translation="MYPNLKRKRPPMIPTMSAPPEYCETVDETIASHSKDEMPPIYV
                     MDVSNLTWHRAFPPTYDDEERQIDEEPMRSIEAEIDREQTRMYRCGPCVLSETKSFL
                     LCFFTVFLGLLILLPTLFFVSVLSKDQDPNISPSI"
ORIGIN
       1  TTGTTCTTTT ATCTTTCAAA TAAATCTTTT AAAAATTTAG GACATTGAAA CAGAAAAGAA
      61  GATAGAAAAA GCGATGTTTC ATTTTTTTCA AAGGAAAAAA TTACAATTAA ATTTTAAATT
     121  TGGTTGTGAG ATCAAATAAA AATACCGTAT TCTTTTCAAT TGGTACCATA AAAAGGGGTG
     181  TTACAAAAAC ACGCCAGGGC CTTACATTGT AGGCAGGCGC AGTTTCAGGG CGCAGTGCCT
     241  GCAGGAAACT TTACGAGCTG CTGGTCTCTA AACAGGAGCT TTGACATCTA TGTACATGAA
     301  TCCCCACGCT ATTGATTAAT TGATTTTGGT CCTTGCCAAA AATAATAGTT TAAGCAAAAA
     361  TGCTGAAAAT CGGAAGTTTA ATTTTTAATG CATAGAAAAA CTTCTCTCAC TTGAATCAAT
     421  TTTAGGCAAT CACACTTTAC TGCTGCTAAT CAAAAAGAAT ACGGTTTATA CAAGAATTGT
     481  TGCTAGAATA AGTTTGTTAT CAAATACTGG GCGATATGTT TGGATCCTGA TCTTTCGAGA
     541  GAACGCTGAC GAAGAACAGT GTTGGCAAGA GGATTAGAAG ACCAAGAAAC ACGGTGAAGA
     601  AGCAGAGAAG GAACGACTTT GTCTCTGAGA GAACGCACGG ACCACAGCGG TACATTCGCG
     661  TTTGTTCTCC TGAAAAATCA TTGTGAATGA GAATACTGGT ACCTGGGTGT ATACAACTTT
     721  AAATGTCCAT ATTGTGAATG CCCAACCAAT TTTGATAATT TTTCACCGTT TTAAAAAATT
     781  ATATTTTCAA AATGAAAATT TATATTGAAA CTTTTCGATC TAATAGTAAG ATGTTAGGTA
     841  AAAACCAAGA ACAATAAATA CCGTATATCT TCTATTAATA TGGCCTCCCT ATTAGACTTG
     901  CACTCCCTAT TAGTCTTGCA CTCCATAATA GCCTTGCGTT TACTCTTTTT TATTGCAGGC
     961  CCGTCTTGTT AACTTCAGGA TTTCAACAAA AAAATTTGCC TATTTTGCAA CAGTTTTTCA
    1021  CAACATTTCT ACGAGATTGA AGTTTTTGAA CTTATATGTT ACAATTAAAA TCAAATTTTC
    1081  TGAATTTTTA CATAAGATTC GAGCTTTTTC AGAAAAAAAC TAGTAAGCTC AAAGTCAGAA
    1141  AAATGTTCTG AAAATTAGTA TTGCACTCCC TAATAGTCTT GCACTCCCTA TTAGTATTGC
    1201  AAGCGGGAAG GTCATCGAAA AATAGTATTG CAGCTATATT AATAGAAGAA ATACGGTAGT
    1261  TAGCTTCCTT GAAACAATCT TCAACACTGT TACACTTAAT CTATTATTTA CAAATCATTT
    1321  CCCGCTAAGC TCTAGTCTTA TCTTTTTTTT TCTAAATTGC CTATTTAAAA AATGTTTTTG
    1381  AGTGTACTCA GACGAATTTC AGCGATGCTT TTTTTCTTTT CAGTTCGGGC GGTGTTGAAG
    1441  TATTATAATA GTCAAACTAT AGACTAGAGT TTCGTACATT CGTCAATGGA AAAGAGTTAC
    1501  AAAACGTGAT CTTCGAAAAT GTTGGGAGAT TTGAATATTA TTATGTCCGA ACGCGTAATG
    1561  TGTGAACATT ATACTTTTTT AAAAATTGAA TTTTAAGTTT TCTCATGTGA CTTTTTTTTT
    1621  AATTTTAAAA CTGATATGTC GCGGAACAAA TGCAATAGAA AAAACTGTTT CGTTTACAAA
    1681  TAGGCACAAA AATATCTGCC TGTCTGCCTA CTTTCATCAG TACATTTTTT TCGGTACATA
    1741  TTTGCAAAAC CTCCATTTTC TTACTGTAAT GATAACGTTT TTATTTGGAA ATTTCAATAG
    1801  AACTATTGAA AAATAATTCA ACTTTTTGAG AATTAATTTA TTTTATCAAA AGTTTTGAGG
    1861  AGTTGCTCTA AATCAAAATA AAAGTTTGAG GCAATGATGA GAAAATGGAT GTTTTGTAAG
    1921  TGTTCCCCGT TTTTAGTATT GCTACATATT TACTGCAAAA TTTTGTGGAT AGACTGTTTG
    1981  TAGGCACGTA GGCACAATTC AGGCACGTAG GTAAAATACG TGCCTACCTA CATAGCATGG
    2041  AAGTCTCAGA TCAAACTTGT TTAGGAAAGT TCAATAGAAA ATGGTCGAAA AATAAACTAA
    2101  AAAGAAACTC ACGATCAATT TCAGCTTCTA TACTTCTCAT TGGCTCCTCA TCAATTTGCC
    2161  GTTCTTCATC ATCATAAGTC GGCGGAAATG CTCTGTGCCA CGTCAAATTG CTCACATCCA
    2221  TCACATAAAT CGGAGGCATT TCATCTTTTG AGTGTGACGC AATTGTCTCA TCAACAGTCT
    2281  CACAGTACTC TGGAGGAGCG CTCATTGTCG GAATCATTGG CTGCAATCAA GGGGGAAAAG
    2341  TGAAAAAGAT GATGATGATA TGAGTTTGTA AATAAATTAT GCATGTACAT CAGGGTTAAT
    2401  ATCGGTGCAG TGATATTTAT AGGTTTTTTC GTATGATTGC AACTGATAAG AGAAAAGTTG
    2461  GGCGGAGGGT CAAATGATTG AGAATCGGGT GGAAAAACGT ATGGAATTCG AATGGGAGTC
    2521  GATCAGAAAT CTACAATTTT AAATTGAACC TTTCAAAATC TGGGGTGTCT AGACCATTGG
    2581  TGACGTCATT TTTCCGTCAA GAGCACAATA CTTATGAAAA TACAAAATTG ATCTAAATAA
    2641  TAGTTCAATT TAAAATTCTT TAGACCTTGT AACAATTTCA GTCATAATAT CGGCAAACTG
```

```
    2701  CCAAAGTTTT GAAAAATTTA GAGCATTCTT ATGCGTTTTA AACCCTACAA TAATCTTTTT
    2761  TTTTGGTTTA TGAAAATTTT TGTTACAATT TTTGGTATTT TGAGCATCAT AAGTAATTCT
    2821  AAACATGTTC CACCATTAGC AAACAGTTTT GAAAATTGTT TAAATTGTTT AAAAATAAAA
    2881  AATAAACTGG GAAAAACTCT CAATTTAGTG TCCACATCAC TCCTAAATAC CTCACTATTA
    2941  ATAGAAATTT CACATCAAAA TTTAGATAAG AACGAGAGCT GTAAGAGGGT CAAATAGTTA
    3001  AAAATAGAGA AGCATGTCAA CGAAATTAGC TCAAAAATAA ACTATTTACA ATTTCAAGAA
    3061  TTATTTTTGT AGAAAACTGG TAAGTGAAAA ATTTTTGCAA AAATATTACA AAAAATAAAT
    3121  ATAAGAAGAC ATAAAATTGG AAGGAAAACC TAAAAGTTTG ATTGATAATT CAGTAGCCAG
    3181  CTTCCTTCAT CATAACACGT ATGAATCTTC CATATTTTTA TGTTTTTCTT CGGCTACTCA
    3241  TTTCCTCATT TCTCTTTCAA AAAAACGAAC ATCTGTCTGG TTGTAATGAC TTGTAATGAG
    3301  TTTAGAACCA CATATCTTTC CTCTAAAAGT TTTTTTCCGG CAGAAAAATG TAAAAACATG
    3361  TCTGAGAGTT GGTTGGAGTA GGGAAATAGT CTGGCTGAAA ATTATGATTC GAAAAGGACA
    3421  TGGGAAGTTT TTTTTGAGAC GGTAGACAAC GCAATGTCAT TTACTCAAAA GCTTAAAAAT
    3481  TTAAACTTTA AAGGTTCTAC TTAAATTTCA AAACAACTTT CACCTGCCTC GCGACTGATG
    3541  CAAAAGTATG GGTGAAAATC GACAATTTTG GCTGAATCTA GGCAATTCTT CAGATAGTAA
    3601  TTGTTCGGGA ATTTCGTTTA ATAAAAACGG AAAGAAATAA CGTAGTTTCC TATTTTTTGT
    3661  CTAATTTCTA AATCAAGTAA AACACAAGTA AAATTTTCAG ACATTTTCTT CCGCAAGCAA
    3721  AACAAAAATT AAAATAATAT AAAATAATCG GTTTTGATTC CAATTTAATA ATGATTCAAC
    3781  CTTTTTTGTT ACAAAATTGA ACAAAATTTT TTTTTAAATT AGAGTTGAGT GAAAAAAGTT
    3841  ACCTTCGGAT ATCTCAGGGT GAGTTTAATT AATTTTTTTT GTTCGGGAAA AACCCCGCAG
    3901  GTGTAACTCC GACTTAAAAA TTACAGCTGT TTCTAAAATT GATATCTGAA TTTTTATAAC
    3961  AAACATCCTG TAGAATTGCG CTACAAAATC TGAAAACAAA GACTGAAATA CATAGTCAAG
    4021  TGTCCTTATT TCTTCCCCAT GCCATCCTCT ACTCTACTCC ATTTACATTT CTCGCCTCCA
    4081  ATTTCCGGAC CTCGTGCCGT TTTCTAATAG AAGAAGAATA CAGAGGTTAA CCAAAAATTT
    4141  AAAAAAGGCG AAAAATATGC ATAAATTTTT ATTACAAATT TACGACGTGA ATCACACGGG
    4201  GAGCATGAAA AAACTGACAA AAAAGTACGA AAAAGCAAAA AAAAAAAATG AGAAGAGCAT
    4261  CCGGGAGAGT AAGAAAAAGG TCCTAATTGC CAACTCGATG AACTTTTGAC GGACGAGAAG
    4321  AGGACAGATG AAGTGACTTT TTTTGCATGA TGGAATGACT TGGCAGGGAG GAAAAATAAA
    4381  ATAAATGTGT TGGTTTTGTC GGCTTGACCT TTTAATGGGT GTTAGTTATT GAACTCGTAA
    4441  CTTTTTGAAA TGGAGCAATT TTCAAATTAA ACTTTAAATA TGGAAATACC CTTCAATAAG
    4501  GCAAATTTTG CCTATTATGC GCCTATCAAC TCCTAAAAAA ACTATGACCA TGTTTTTTTT
    4561  TAAATTTTAA CTTTGGAAAG ATAACTTTCA TGTTCCGGAG AGCTGAACTA AACTAAAGAA
    4621  TGACTAAACT TGAATAAACT TAAAACATTG GAAATTCCAC CAGAATTCGA AAGTTGACAA
    4681  TTTAATTTAA TTTTAGCCAC CTCTTTTCAC TTTTTTTTGG CACAAACTGG TCGTTTTCTT
    4741  TTCAAATTTG GGTACATTTT TAGCTGAAGT AGTAAAGCTA TAACAGATTG CGTTACTTCA
    4801  ACTTCAAAGG TTTTAAAACG TTCTTTTTAT ATATTTTTTG CTTTGATTTT CAAAAAAAGG
    4861  TTTTACCCAC CATAAATCTT TTTCTGAAAA AATATATTCA AAACTGTTTC AATTTAAAAA
    4921  TGCGTAACAA TATACGTATT ATGTAAATGT TTTAGAAACT GATTGCTATC CAAAGTTTTT
    4981  CTAAAAGATG GGATCTGACA AATGAGTGTT TCTCATTTTC TTCTACTTGA AAAACACTTC
    5041  TTACTCGCTA GGAAATTGGA TTAGGAGTGT TCAAATTAGG TTGTAGATTC TTCTAACTTG
    5101  AGAGGTTAGC TGAACTTTTT TTGGCGATCA TTTATTCAGA ACTAGGGGAC ACCAAACAGA
    5161  GCATAAATTG AGCCCAAATT TGGGAAGTTT CATTGATGTC GTTAAGTACC TTCAATGTTT
    5221  AAATGTGAGC TGATTTAATT AAAGTAAGCC TTGAAAT

//

LOCUS          18     1693 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      18
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS          complement(join(500..718,765..920,1155..1193))
                  /translation="MLGVVGFFIYAKENSRLFDSVNPRYAQQLLEIEMFTFCSLDIV
                  DEKSTKASEMFLGQLYNDQKWRSFGYITNTGVRMILVLDATSAASLKDQEIRLIFKR
                  FHGHYCNTISNPFYEIGTPMQSKWLDEGINDLYSISE"
ORIGIN
       1  TTTTAAACAT TCAAAATCGC TGGATTATCT TTAATATAAC TTGTTTTGGA TTTATTTTTA
```

```
       61  TTAAAATTTA CCAGAAATTT CAGAATTTCA AAAAAAAAAC AATTCCTGAC AATACAAAAT
      121  TGTAAAACGA AATTTAATTT CGAAAAAAAT CAGAACATTT TCTGAAACGT TTTTTAAAAA
      181  TTCGAATTAA AAACAAGAAT ATAAACTTCT CACAAAGTAG TTATACTTTA ACTTTTTAAC
      241  TTTTGAAACT GGAGAATTAA CAGACGTCAT TTTTTAAAAA AACTCACCAA ATTCGGAGCA
      301  ATAAAATCCG AAACATTTGA AAGTCGAACG ACACCACTGA ATCCAGAATC TTCCATAAAA
      361  CTTCTGGAAA TATTTTTAAT ATTATCTGAA TAAGACAATT AAAGATACAG TGACGTCAAA
      421  GACGATTGAA AATCATCCAT ATTCGTAGAT TCACAAGAAA AAAACAAAAA TTATTTTATT
      481  TACAAAATCC GTTTCATTAT TATTCACTAA TCGAATATAG ATCATTGATT CCTTCGTCAA
      541  GCCATTTTGA TTGCATTGGG GTTCCGATTT CATAAAATGG ATTTGAGATA GTATTACAAT
      601  AATGTCCATG GAATCGTTTG AATATCAATC GAATCTCCTG ATCTTTTAAT GAAGCTGCTG
      661  AAGTTGCGTC AAGAACCAGA ATCATACGGA CTCCTGTGTT GGTAATGTAG CCGAATGACT
      721  GAAAAATGAA GATATAAAGT AGTTTTCAGT CAAATATTCT TTACCCTCCA CTTTTGATCA
      781  TTATAAAGTT GTCCCAAAAA CATTTCCGAT GCTTTTGTTG ATTTCTCATC TACAATATCC
      841  AATGAGCAGA ATGTGAACAT CTCGATTTCA AGCAATTGCT GTGCGTAACG TGGATTCACC
      901  GAATCGAATA GACGGGAATT CTGAATTTTT CTTTTACGAG GAGGCAGGAC AGGCCGAAAT
      961  GATCGTCAAA ACGCGCTCCA CGGTCAATTG AAGACGCTCC GCCCCCAAAC GATGGGTCTC
     1021  GTTAGGTATA TGGTGGTAGA ACCGTAAAAT CAAAAAAACA GGGGCGGAGC ATTTCCAATT
     1081  GGTCGTGGGG CGTATATTCA CATCGACCCG TCTTCTTGCT GCAACCGGTT TCGAATTTAA
     1141  AAAAAAACGC ATACCTCTTT AGCATAAATG AAGAATCCGA CGACGCCAAG CATACTGAAT
     1201  TTAAAAATTT ATTTTTTGTA ATACTATTTT TTCTCTATTG AGAATATTAA TCATATTCAA
     1261  GTATTTTTCA GCGATGTTTG ATGTACCACA CACGTATTCT ATGACGGACA GCGTTATAAA
     1321  ATAACTCTCA ATTCATTGGT AACCATAAAG AAAACAAAAA AAGCACTCGA AAAGTTGAAA
     1381  CCAGAATAAA TAAACGAAAA GAACAATGAT TATACTTTGC TGGAAATGTA AAAAGGCAAT
     1441  ACAAAAATTT AGATAAACAT TCAATTATTC AGTTGCGCTC GCTTCGAATT TTGGATATTT
     1501  GGCCTGAGCA TCGGACCATT GTTGACGGCC AGCGACAAGA TCAGAAATGA GTCTTCCGAG
     1561  TTCTGAGACT TCTAATCGAA TCTGTGACAT TGTTTCAGCG TGGCGGATGG TGACAGTGAA
     1621  TGGAGTGGTC TTCAATGAAT CAAAGTCAAC AGTGATTCCG AATGGAATAC CAATCTCATC
     1681  AGTACGTGCG TAA

//

LOCUS          19     1461 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      19
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          join(500..653,696..961)
                  /translation="MKTLLLLFLFTASVAAIPFRSTTVVARVILRCKHPKDPSAKVF
                  LMEKDFGDHEIDEDDTMDFAEFKFGQIPEKPLQLKGEEFEFSGLEPYIYISHACTEN
                  RVEAWDSFNVTLETSMYLSATFDVIIDLDTKKSTVIKNI"
ORIGIN
       1  AAAAGATTCA AGATTTCAGC TAAAATAAGC TCTTTTTTCG AAACTTTGAA CCGCGCTAAA
      61  AAAAAATAAA AAAATAGGAC ATATTTTTGA GCAGTTTCAT AACGAAGTTT GTTCATTTGA
     121  ACAAATTCAG GCCCTGTATG TTTAAAACAG TCAGAACCGT TACTCCTCCT TTGATAGGAG
     181  AGTGAAAACA CAAGTGCCTG AAAGTTCTGA ACATTTGGGT AGAAAACTAC TAGAAACCAA
     241  GTAGCCAGAC CTAATTTAAG TTTTGAATAA TCACAGTAAA ATTAGTTTTC AAAAAGACGA
     301  AACTTTTAGA AAATCAAAGA TCAAAATTGC ATTATTGAAT TTTGTTTGTA TATTCGTAAA
     361  TCCATGGTTA TTCCTTTTTG ATATATTGCT CATATCAAAT ATATGTTTGC TAAATAAAAT
     421  AGAAGGCGGG CCGGGCCAAA AACAAAATAT TCAAATGCAT ATAAAACACC ACGTTTGAGT
     481  TTTTTCTCAC AAAGTTACCA TGAAAACTTT GTTATTGCTA TTTCTTTTCA CAGCCTCTGT
     541  GGCAGCAATT CCTTTCAGAT CAACAACTGT TGTTGCGAGG GTGATACTTC GTTGTAAGCA
     601  TCCAAAAGAT CCATCTGCAA AAGTGTTCCT GATGGAAAAG GATTTTGGAG ATCGTTAGTA
     661  AATTGATCAA TACACATGTA TTAAAATAGT TTCAGATGAG ATTGATGAAG ACGACACAAT
     721  GGATTTTGCA GAATTTAAAT TTGGGCAAAT CCCAGAAAAA CCACTTCAAC TCAAAGGTGA
     781  AGAATTCGAA TTTTCCGGTC TTGAGCCATA CATTTATATT AGTCATGCAT GCACCGAAAA
     841  TCGCGTGGAA GCGTGGGATT CGTTCAACGT AACTCTGGAG ACATCCATGT ACTTGTCGGC
     901  TACGTTCGAT GTAATCATTG ATTTAGACAC GAAAAAGTCT ACTGTCATTA AGAATATTTG
```

```
     961  ATGTTAATTA ATTTTCATTT GGAATAAAAC TAACAGTTAA CGGATTATGA CAAACGATGG
    1021  TATGAATTAA AACAATGAGA TGTTCAACAT TTTCATGAGG ATGAATTGAA GGTAGTTCAA
    1081  AGTCAATTAA ACAATAAAAT GAGAAGAAGA AAACAGGAGT TCAGATGAAA AAGAGTAAAT
    1141  CAAAAATGAT TGGTTTTGAA AATCTAAAAA AAATAAAGAA GGAGTAACGG TCTCTTTCTT
    1201  TTCCTTCAAG GTATTCATAA CGTTCGTCAA TTTCTCTTTA CTTTTGATGG ATTGCGACGA
    1261  GTTTTTCTTT TTTTCGGTCC ATCTAAAAAA TGGCATTTAA AATAATTAAC AGTCTTGAAA
    1321  AAAGAAATGC ACAAACCTTC TTCTTCATCG TCGATAAATG GAATTTTGAG CATACTGATC
    1381  GGCGGATTGG GTAGACCATT TTGGAAAGGC GAATCTCTTG GCAACCCGAC ACATTGAATT
    1441  GTCTCGGCGA ACGCTTCAAT G

//

LOCUS          20     1849 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      20
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          complement(join(500..829,1257..1349))
                  /translation="MDISLPDNERSIRTAFHKKKEELINEKRMFMDCPKCLMIFFDM
                  IMYRMHMAMHIPLPSTDKMSLESSDTWKCSLCKREFSDRREFYIHTVKYSHSLIPRG
                  FFVPPALPYHFGLPDLASLCQPSDLLKELAELVKANATDK"
ORIGIN
       1  GAGACAAGAA GTTGCTCCTA AAAAATAACG AATTAGATTT TAAAACTTGG AAATTATTTC
      61  ACCTCGGTTG GAGTTTCAAA GCTTGATAGC GCCGTCGTGT AATCGAAAAG AAGAGTCTCG
     121  TTTATTTCAC GGGCTTGTGG GTCTTTCACA ACCATTGGGT AAACGTCAAT TTTTACCTTT
     181  TTGCGTTCAG GGGATGCTCC TGGAGACTTG ACCATTTTTA ACGCTGAAAT TAATTTATTA
     241  AAAACATTCA ATGACTTGTA AATAAAGCGA GGAGCACGAA AAGGAAATCG TAAAAACAAA
     301  AGAAACAAAA AGAGTACCTA GTCGATTCTC TGCTGCCTCC TCGCTCTCTT CGCGTATCAA
     361  TGGGGCGAAG TTGCCTTGCA TTTTAGCGGG TATTCAAATC ATATTTTCCA TTTTTGTAAT
     421  TTTTGGAATT AAAAACGTTT ATCACATATA TCCGATGAAA AAATTATAGT TAGAAAATAT
     481  TACAATTCGC AAAAACACAT TATTTATCTG TGGCATTAGC TTTCACCAAC TCAGCGAGTT
     541  CTTTCAATAA ATCACTTGGT TGACAGAGTG ACGCCAAATC GGGAAGACCA AAATGGTAAG
     601  GAAGGGCTGG AGGGACGAAA AATCCACGTG GAATCAAAGA ATGTGAGTAT TTAACTGTGT
     661  GAATATAAAA CTCTCGTCGA TCCGAGAACT CTCTCTTGCA CAACGAGCAT TTCCACGTGT
     721  CAGATGACTC AAGGGACATT TTGTCAGTAG ATGGTAGAGG TATGTGCATT GCCATATGCA
     781  TACGGTACAT GATCATATCA AAGAAGATCA TGAGACATTT CGGACAGTCC TAAATCATAA
     841  TTATGATTGT TTAAATTCAA AATCTTCATC TTGGTAGTCA GTTGCAACTT TTTGGCAACA
     901  AAAGTAATAC TTTTCCAAAG GTGGAGTTCC ACAGTCAGTT TTTGAAACTC TAATTTTGGA
     961  TGTAGACCAA AACAATTTTT TAAAATTGGT TTTAGACCAA TAATAACTTT AAACTTGCAA
    1021  GAATGCTCTA AATTACTCTA ATTTACCAAT TTACCAATTT ACCAATTTAC CAAAATTTTG
    1081  ACTGCAAGAA GTTTGAAAAA TCTAAACATC TATAAACTGA ATACCTACGA TATTTGACCA
    1141  TGCTAGATTA TTATGAGCGC ATTTATACAA GAAACATGAC TTGTTGGCAC TTGAACCTTA
    1201  GAACACCGTA GTCTCCGGGT AGGATTCGAT ATGATCTCTA CGAAAACCTG ACTCACCATG
    1261  AACATTCTTT TTTCATTAAT CAATTCTTCC TTTTTCTTGT GAAATGCTGT TCTTATACTT
    1321  CTCTCATTGT CCGGTAATGA GATATCCATT CTGTAAGTTG ATTTTAATTT TGAGCCAAGT
    1381  AAGAAATATA ACGTTAATAG TTTAAGAAAC AAATAACACC TGCACAAAAT CCAACTAACT
    1441  GGCCTGAAGT GATTGACATA TAAACACTGA AGTTAAAAGT AATTAAATGT TCACTAACTT
    1501  AGAATTACAA CTAATCATTC AACAGCTGCA CCTAAACTGA ATTTGGAAAG ACATGCTATG
    1561  AAATGCTGCT TAAATTTCAA TTATTCAGCA AACCAGTCTC TAAATTTAAG AGTGAAACCA
    1621  GTTCTCGTAA ACTCATCAGA CCATAGCATA AAATTTCTGA TTCAGTTTTC CGGTGAATTC
    1681  AGTTTTGAAA AATAATTTGG GAGGTAAAGA GCATTCATAG TTTAAGATGT GAGGAACTTG
    1741  GTCTGACAAT TTATTGATTA ATTTTCCGAA AAAAAAACGA ATACTGTAGC GCGACTGACA
    1801  CCCTCCCTTC TCCATCGCCT TCTCGCCTAG GAAAATGCCC CCGAAAAAG

//

LOCUS          21     1674 bp     DNA    linear PLN 25-FEB-2004
```

```
ACCESSION      21
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             complement(join(500..581,831..1174))
                     /translation="MSALIQVFLLLAVFAGISEMCGSVQHTETTTTTVAPATGGARK
                     RRSVEGSGLVHIEFQTTLTSAGDAEFSKIDEYVREEFGELLNKAHQQVATGVEGGSV
                     LKYMAINADCGAAQLFAQAVKDLTSFVTSATVRCDTQTSTI"
ORIGIN
       1  GATAGAACAC TGAAAATGAG AATTTTTCGT CTAAATTTCA AGTTTCGGCG TTTTTATGAG
      61  TGAAACTGAT TTTTTTTTGA AAAATGGCCT GTAACTTGGC TCAACATGAA AATATCAACA
     121  CACTTGAAAT TGAAAAAAAA AATGTGTCAA AATATTTTTA GTTACTACTC AACGAAATGT
     181  TTCTTCTGAA ATTCAATGAA TCAAAATTAA GAACAAATGA ATCAAAATTA AGAACAAAAA
     241  TTGAGTATTT GTTGCAACTT GAGAGAATCT GATGTCTCAG CTCTTTTTTT TTTTGAATTG
     301  GCAGTAGATC ATTCCGTGTT TGTGTTGGAT TGTGTTAGTT TTAAAATGCA AATGTGATTA
     361  TGAGATTTGT CTATTGTCTT CAAACATAAA CTTGTCAGTT TGCGAACAAT TGTTGATCGG
     421  AAATGTTTAT TGATTAATTT AAAACAAGAC TTTGAATCAC CATAACCCCA AATAATAAAA
     481  AATGACAGTT TGAATAACTT TAGATTGTAG AAGTTTGGGT ATCGCATCGA ACAGTTGCAC
     541  TTGTGACGAA AGAAGTCAGA TCCTTGACGG CTTGTGCAAA ACTGAAATAT ATCAATTAGT
     601  GTGGTTAAAA TGAGAAAGGT AAAATTTCCG AATTAATTAA TTTATTTATT TATTTATTTA
     661  CGCTCGCGAG ACGTGGAAAA AGTTGAAGAA AAAAGAAAAA TGTACACATA CAATAGATCA
     721  ATACAATACA CTGAATAAGA GGGATAGAAA CAAGAAGAAT AAATAGAATA AATACTATTT
     781  CATGCATTTT CAAGTCAGAA CTAATCTGTT TCATTAACTT TTTCACTTAC AGTTGAGCAG
     841  CTCCGCAATC AGCATTGATG GCCATATACT TCAGAACACT TCCACCTTCA ACTCCAGTAG
     901  CAACTTGCTG ATGAGCCTTG TTCAAAAGCT CTCCGAACTC CTCTCGGACA TATTCATCGA
     961  TTTTTGAGAA TTCAGCGTCT CCAGCAGACG TCAATGTTGT TTGGAATTCG ATGTGAACAA
    1021  GACCGGATCC TTCGACTGAT CTACGTTTTC TTGCACCACC TGTAGCTGGA GCAACGGTTG
    1081  TGGTTGTTGT TTCAGTATGT TGCACGCTTC CACACATCTC AGAGATTCCA GCAAAGACAG
    1141  CGAGAAGCAG GAAGACTTGA ATTAAAGCAG ACATTTTTTT AGAGACGGAT GATATTCTGC
    1201  ACAAATACTC TGTCTTTTAT AGGATTCATG TATAGACACA TTATTTTCAC TTTTATAATT
    1261  TTTATGGCAT ACCGTGAATA GAAAATTTTT TAATGGAATG TTGAGAAACT GACCAGTACA
    1321  ACGACAACTA CTGTTCAAAT AAGGAAAAAT AATAATTTTC AGTCTGTGTA CATGTTTGTA
    1381  GGGAAACGGA ACAAAGTCCA GTGTTGATCA AGTTCAAAAA GCTCGACAAC TTTTTCCTTT
    1441  TTCTTAAAAA AAAATTTTCA GCGTATTTTT AAGTGAGAAA AAGAAGGGCG TAGGGCGATC
    1501  AAGAGACTAG AGGAGAGCAA AACAGTACAA TTTCGAGATA ATCCAAGTTT TGTGCCTGAC
    1561  AACCGCACGA TTTACACACT TTTCGAAATT TTTAGCAATC TCGAATATAG CCTGTGCTTT
    1621  TTAAAAATGA ACATCTTAAT TTTGTCTACT TCTGAATTTG TTTTCCGTCC ACAT

//

LOCUS          22     1567 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      22
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..538,678..1067)
                     /translation="MSYMLPHLENGWQVDQAILAEEDRVVVVRFGHDWDPTCMQMDE
                     TLFKIAPKVKNFAVVYLVDITKVPDFNKMYELYDPCTAMFFFRNKHIMVDLGTGNNN
                     KINWAVTDGQELIDIIETVYRGARKGRGLVISPKDYSTKYKY"
ORIGIN
       1  GGGTGGTTGG CGGCGCGGAA ATCTGGAAAA GACACAATTT TTCTATGTTT TTTTTTCGGT
      61  TTTTGAGGAA GGAAATTCTG AGAGAATTTT AAAGCGATAA ATAATTGCGA AACGTCCATG
     121  GGGAGAGATT TGCATCATTG TTGCTCTTTT TCAAGTTGTT TTGTCAATTT GCACCTAACC
     181  GAGGTTTTTA GCTTAAAATC TTCGAATTTT ATCTATTTCA TCGCCATTTC TATCAAATTT
     241  CATTTGAAAT TCCCTCCAAA ACGATGCGAA ACGGGAGAAA TTGAAAGAAA TGAGAAGAAA
     301  AAAGAAAAAA ACAAAAATAA ATGTGTAACG ATGCTCCGAA ATGCCGCATC GCGTGTTGTT
     361  TAGCCAGCGG GAATTTTGTT TCGAAAATCA ATTTTCCACT GGTTTTTCTG TTACTTTCTC
```

```
     421  ATGTTTTCAG CCATTTTTCC ATTAAATTTA TCAATTTTTC AGATTTGTTG GCTCAAAAAC
     481  TCACTAATTC AACGCAACCA TGTCGTATAT GCTCCCTCAT TTGGAGAATG GTTGGCAGGT
     541  AAGGGAAATT GACAGTTTGA TTTTAAAACG GCAAAAAAAC CTGCAAATTT AATGTTTGCT
     601  GTTCAGAAAA ACAGAATTCT GTGATGAAAA TTTGTTTAAA ATGCTCAAAA CTGACAAAAA
     661  ATAACGATTT ATTGCAGGTG GACCAGGCGA TTCTCGCCGA AGAAGATCGC GTAGTCGTCG
     721  TGCGATTCGG ACACGATTGG GACCCAACTT GCATGCAAAT GGACGAAACG CTCTTCAAAA
     781  TCGCGCCAAA AGTCAAGAAT TTCGCCGTCG TCTACCTCGT CGACATCACG AAGGTGCCGG
     841  ATTTCAACAA AATGTACGAG CTGTACGATC CGTGCACCGC CATGTTCTTC TTCCGTAACA
     901  AGCATATTAT GGTGGATCTG GGAACCGGTA ACAACAACAA AATCAACTGG GCAGTGACCG
     961  ACGGGCAGGA GCTGATCGAC ATCATCGAGA CGGTCTATCG TGGAGCTCGT AAAGGACGTG
    1021  GTCTTGTCAT CTCGCCGAAA GACTACTCGA CCAAGTACAA GTACTAATTA TTTTAATTAT
    1081  TTTCTCTGTT CTCTGTGTTC TGTAAATTCT TAAACTGTAT CTAAAAATGT TTCAAAATAA
    1141  TTTACATTTT ATTCACTCGA AAACTGAGCG ATTCTATACA AAAACAAAAA TATATAGGCG
    1201  GCTTGAGAAA ATTGTGATTT TGATAATAGA AAAAACAAAA CGCGAAATGA GAATCGATTA
    1261  GTCCTTCTTC TTGTCACCCT TGGATTCGGC AGAATCTCTG AGAAGATCGG AGACGCGGTC
    1321  GAACTCGCGC TGGAGCCCTT CGGATTGCTT CTTGAGAGTG TCACGGTCAG TGTTGGCGGA
    1381  CTTGAGCTCT CCCTTGAGCT TCTCGATTTG CTTTGCCAAT CCGGCGGCTT CTTCGTTGGT
    1441  GTCAGCCTGA AAATTTGGAA ATTTACACGG TTTTTAAGGA ATTTGAGCCT ATTTTCTCAC
    1501  CGAGTTCATC AAAGTCCTGG CGGTCTTGGT AGCAGATTCG GCTTGACGCA TTGCGGCCTC
    1561  GCTGGCG

//

LOCUS          23     1484 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      23
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             complement(join(500..848,902..984))
                     /translation="MSQDPTISNFLNGKPIIAPVGPSSLLTRLQSFLPQMEEANRNM
                     PENASADAFHIENVEDDSSDDSDSNSESISEEVPVTTDEKSTASSTQRIEIDLDVFK
                     EMNSAVDDRDVGIQNVEALPEAFQSKGEDSKPEITKKPLIEEL"
ORIGIN
       1  GACGTCATCA GATACCGATC CGTAACTTCC GGTCCTTTTT TTCCAATCAA GCTGACACGT
      61  TGCTGACGAG TCAACTTATT CAGCTGCATG GCACGGTTAT CCCAAGTCGT GGTGACTTTG
     121  ATGAACTCAT CTCCTGACTT CTTCTCGATT CGTCCTTGAG TTGGAATCGG AATTTGTGCC
     181  AGATCGCCAT AAAGATCCTT GAGCTTTTCC AACGCTCTCA CATCATACTC AGCTGCACTT
     241  GTCACCACAG CAGACGACAT TTTACGAAGG AGAAGACGCT GCATTTCGGA AAAGTTTTCG
     301  GAGAATTTTA GAAAAGGCAG AGATAAATAA AAATGGGAAG GTAATTCAAA AATGGATAGA
     361  AACCAATGAG ATAAATGAAT AAAAATCTAA AATCGAAATT TTATTGATTC ATTTCGCAAT
     421  GTGAGAAAAT GTTTCGAGGG GTAGGTCAAA AAGTAACAAA TTCAACAAAA AATTCACAAC
     481  AATAGAAAAA ACATAAATTT CATAATTCTT CAATCAACGG TTTCTTTGTA ATTTCCGGTT
     541  TCGAATCCTC TCCTTTTGAC TGAAATGCTT CTGGAAGAGC TTCGACATTC TGAATTCCGA
     601  CATCTCTGTC ATCAACTGCG CTGTTCATCT CCTTGAACAC GTCTAAATCA ATTTCAATCC
     661  TCTGTGTACT ACTCGCTGTG CTCTTTTCGT CAGTTGTGAC AGGTACCTCT TCCGAAATCG
     721  ATTCTGAATT CGAATCGCTG TCGTCCGAGC TGTCATCCTC GACATTTTCA ATATGAAAAG
     781  CGTCAGCTGA CGCGTTTTCC GGCATATTTC GATTGGCTTC TTCCATTTGA GGCAGGAACG
     841  ATTGTAATCT TAAATTTCGA ATTAATTTTT CTTAAAAATT AAATAATTGC TTCAAAGTAA
     901  CCTGGTTAAT AAAGAAGATG GCCCGACAGG TGCGATGATG GGTTTTCCAT TTAAAAAATT
     961  GGAGATCGTG GGATCCTGTG ACATCTGAAA ATGGTTTTTG GAGAAGCACA ACTGAAAAAA
    1021  GATGCAGTGG GCGGAATTTC GGCTTCTTCT TGATAGAGTG AATAATTTTC AAAATAAAAA
    1081  TCGATATACA ACGAATAGAG AAACAGTTAA AGTGCGGTTC TGTACAAAAT CTTTAATATG
    1141  CTGCTCGAAT CTTTTTGCAT CTTATAAGTT AAATACAACG AAAGCGATTA GGCGAAATAA
    1201  AACATCAGCA GTACCGGAAA AAAATAATGA CAAACATATC ACACTAATAG CCGGGAATTA
    1261  TATTTTTAAA TTATACAAAA ATTAGTCGGA AGCTTCAAGC TCCTGGATAA TTTCGATATG
    1321  CTCATCGGCA ACGATTTGTC CATAATCAGA CAGGTTTTCA GCGAAGACGA CATGAAGAAG
    1381  AGCAGCTGGA GCAAGACTGA GTTGAGCCAA CGATGCAGAT TCATCCTCCA CGAGTTGACT
```

```
    1441  GGCAGCGTCA CGAAGAGTGA ATTCGGATGT CGAAAGAGCA TCGG

//

LOCUS          24     1872 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      24
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..643,1079..1372)
                     /translation="MAEKKESSASSKKSESSGNSSSGREDLEGSGNTGVVIKPRKSK
                     KKQKKVPKVSPASRKEDIPGCATVEKVSIRSGSEISQSEWLDNLVPGTLAKPQSRWE
                     SITGFIDNYKYPIISSVVLAGVLTYFIVKKGKAGAKCPFTGKKAN"
ORIGIN
       1  CGATGAGGAT TCATTTGTAG AGTCAGACGA GCTTATGGTA GTTTCTGGAG TAGTTGGAGA
      61  AGCTCCTGAT GAATTGAATC CAATGAATCT TTTTGCAGGA TCTTTGGAGT TATTCGGAGA
     121  AGTTGGTGCA TGTTTCTGAA AGTTTTATAT TTTTTGAATA ATCTCTCAAA AGTTCTTGAT
     181  GGCTTTCAAA TTTCTAGCGT AAGGCGTCTA ATATGCACTG AAAATATGCA ACTGAAACTA
     241  ACCTTCAAAT CCACAACGCA CACGGTCACA AAGAAGATTG ATAACAAAAA ACAAGAAAAC
     301  TGCTTTGAAA ACATGGCGGG AAATTTAAAT TTAAAAATTC TTCCGCATAA AATTCGTAAT
     361  AATATTTTTT AAATTATTTA TTTATAAATT GTTCTAATTA TACAAAAAAA ACCACGTCGC
     421  CTTTTCGTGG TTTTTTAAAA TCAAAAGTTG AACAATCAGT TTCAAAATAA TTCATTTCGG
     481  TACAATTTCA TCGTCAAGAA TGGCTGAAAA AAAGGAATCA TCAGCTTCAT CGAAGAAAAG
     541  TGAGAGCAGT GGAAATTCGA GTAGTGGAAG AGAAGATTTG GAAGGATCAG GCAATACTGG
     601  AGTTGTCATC AAACCGAGAA AGTCGAAGAA GAAGCAGAAA AAGGTATAGC ACTTGGTTGA
     661  GTAGACACTA AATTGTTACT CCCATAGTGG GCAAACCGGG TAAAACCGAC CTTCCGTTTT
     721  TGACCGGATT CTGGGGTTTT CGTATTCGGA GCCGTTTTTG GGTAGGCCAT CCGGTTTCTC
     781  AAACTATTTC GCATATGATG CACAGGAGAA CAGGAACTCG TAAAGCATTT AGGCACTGGG
     841  GTAGCCATGT AGGCTGTCAT GCTTCAATTA GACACATATG TACACAGTAA AGACATATGT
     901  ACACAGTAAG AGCATCTGGA TTTTCCATTT TCAATTTCTC TCCTTGACTT TGCTCTTTTT
     961  ACAACCTTAA GAATTATCAT CACGATTAAT TAATATCTAA GATCTTGGAA TTGGAAAATG
    1021  AGATTTTTCA TCTTAAAGAC CAAAACTAAT TTTTAAAAAT ACATACATCA TCTTTCAGGT
    1081  TCCGAAAGTA TCACCTGCAT CACGAAAAGA AGACATTCCT GGATGTGCAA CAGTCGAGAA
    1141  AGTATCCATC AGATCAGGAA GCGAGATTTC TCAATCGGAA TGGCTTGACA ATCTGGTTCC
    1201  AGGAACTTTA GCCAAGCCAC AATCTCGATG GGAATCAATT ACTGGATTTA TTGATAACTA
    1261  CAAATATCCG ATTATTTCCA GTGTTGTACT TGCTGGTGTT TTAACTTATT TTATTGTTAA
    1321  AAAGGGAAAA GCTGGAGCAA AATGCCCATT CACTGGAAAG AAAGCTAATT GAAGAACTAG
    1381  ACGATTTTGA AATGCTATGA AGAGAATAAA ATGTTTTATT TGATACTGAA ACCAATTTTT
    1441  GAGTTTTTTA ACATTACTAA CTAAATGAGA AATGTGAAGG AACATCAAAA GTTCGAAAAT
    1501  TATTTTGAGA TAGATATTTC AAGAACAGTT ATAATTGAGG AATTAGAAAA ATGAGTAAAT
    1561  TCATTAAAAC CATACAATGA AGTCCAAATA TCTGTTTTCT CTTATTCCAA TAGATGCAAT
    1621  TTGGCAATCT CCCTCTATGT CCATATAACT GGTACAGTCA ATGTTTAAAT CTTCTTTTTG
    1681  ATAATTGGAT GGTCTTCTTC TGGGATTCCA TGGTTTTGCA TTCATTCCTT CTAGAACCAA
    1741  ATCAAAGTTC CAGTCTTCAG AAGCCAATAT TTCCAGATGT TCCAGACGAG TTTTATTACC
    1801  CTTCCAGTTG CTAATTAATG ATCTTACAGC GTTAATAGCT GAATTCTCTT TCCAATATAA
    1861  TTTTATTGTT TC

//

LOCUS          25     2172 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      25
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             complement(join(500..856,1571..1672))
                     /translation="MTELISMNTLFSCIFTHFIIGATLIAFCTKQKAKSTIKEEPSQ
                     AENRPQGGNETFHFTPPHIKTAVPTRSLLRQKHQQCPVHYEFSEDATTTTFVDSAH
```

```
                SSDIQFDSTQSASDSSISFLKRCTCFCAQKRVRFQCPPTTTTTTTTTSFGFI"
ORIGIN
       1  GCTGTTAAAT CTACAAAAAG TATCAAAATG TATATTTCCG GCAGTTCTGA TTTCCGGCAA
      61  TTCCGGCAAT TGCCGGTTTT GGGAATTTTC GGCAATTTCG GCAATTTTGC GAATTGCCAG
     121  TTTTCAAAAT TTCCAGCAAT TCCGGAAATT GCCGGTTTTA GGAATTTTCG GCAATCCCGA
     181  CAATTGCTGG TTTTATGAAT TTTCGGCAAT TTCGGAAATT GCCGATTTCA GTAAAACTCA
     241  TGTTCAGGCT CTTCAATCTA CAAAAAAGCC TCTCAAAATT TATATTTCCG GCAATCTGCC
     301  AGTTTGCCGA TTTGCCGATT CGCCGGAAAA AATCGTTTGC CTACCCCTGA TTTATATTTC
     361  CCGTTTTCAA ATCTTCCGGC TGCTTACTAC ATTTTAGTGC AAACAAAAAC CTCTTTCCTC
     421  ATCTTAATAT TTCTACATTT ACAATTTAAA ATCCAAATTA ATTTATTCAA AAACTACAAA
     481  AAAAGAAGAA ATCGTTCGTT CAAATAAATC CAAAACTAGT AGTCGTCGTA GTCGTCGTCG
     541  TCGTCGGTGG GCACTGAAAT CTAACACGTT TCTGAGCACA AAAACACGTG CAACGCTTCA
     601  AAAACGAGAT GCTAGAGTCT GAAGCCGATT GTGTTGAGTC GAATTGAATG TCGGATGAGT
     661  GTGCAGAATC GACAAATGTC GTCGTCGTCG TCGCGTCTTC ACTAAACTCG TAATGAACGG
     721  GACACTGCTG ATGTTTTTGA CGAAGAAGTG AACGTGTTGG GACGGCAGTT TTAATGTGTG
     781  GAGGTGTGAA ATGAAAAGTC TCATTTCCTC CTTGTGGACG ATTTTCGGCT TGAGACGGTT
     841  CTTCTTTGAT TGTTGACTGA AACAATTAGA ATTGGAAATT TGAAGTAGAA TTGCACAAAA
     901  ATATGCATTT TTAACCGAAA AAAACTCGAA AAAATTGAGT TTTCAGCGGT TTTTTGCACA
     961  ATTTTTGGTT GAAACCAGGG GTAGAAAATT TCCGGCAAAC CGGCAAATTG CCGAAACTCA
    1021  ACATTTAGAG CAAATCGACA AACCGGCAAA TTGCCGGAAT TGAACGTTTC CGATAAACCG
    1081  GCATATTGCC GGAATTGAAA ATTTCCGACA AACCGGCATA TTGCCAAAAA TGGAAAATTT
    1141  CCGGCATACC GGCAAATTGC CGGAATTGAA AATTTCCGAC AAACCGGCAA ATTGCCGGAA
    1201  TTGAAAATTT CCGACAAACC GGCATATTGC CAAAAATGGA AAATTTCCGG CATACCGGCA
    1261  AATTGCCGGA ATTGAAAATT TCCTGCAAAC CGGCAAATCG CCGGAATTGA AAATTTCCGA
    1321  AAAACCGGCA ATTTGCCAAA AATGGGCAAT TTCCGGCAAA TTGCCGGAAT TGAACACTTC
    1381  CGACAAACCG GCAAATTGCC AAAAATGGAA AATTTCCGGC AAATTGCCGT AATTGAAATT
    1441  TTCCGGCAAA CCGGCAAAAT ACCGAAATTG AAAGTTTCCG ACAAACCAGA AAATTCCGGC
    1501  AATTGCCGGT TTTCAAAATC TGCCTGAAAC TCAGTTTTTC TGGCTAAATT TCCCATAAAA
    1561  ACTCCCCAAC CTTTGCCTTT TGTTTCGTGC AGAACGCAAT TAGTGTTGCA CCGATGATAA
    1621  AATGCGTGAA AATGCATGAG AACAGTGTAT TCATTGAAAT GAGCTCCGTC ATATTTGAAT
    1681  CTGAAAAGTG TCTTTCAAAA TTGTTCTTCG ATTTTTGCTT TCCAATAGCC AAAATAATCT
    1741  GATCTCTCTG TTTGAATTCC CTCTAAAATG AGAACTACAC AAAACCAATC AGCAATTCGC
    1801  TCCGCCCACT TCCAACCAAT CAGGCGGCGT GGGCGGAGCT GGAAGGCGCT GATTGGTTTT
    1861  GTAGTTCTCA TTTTTAGAGG GAAATAGGCA GTTTAAAGCA AAATCATTTG AAATAAAGAA
    1921  ACTCACAATT GTCGATCATC GTTAAAGTTT GAAATGAATG GATTTCTATT TTTTGGGCTT
    1981  TTGAATTTAA AATAAATTTA AAAAAAAACG AATTTTAATA TTTACACGTG ATAATTTTGG
    2041  ATTTTAAGAT TTGTTCTTTT ATAATTAAAA TTAAAAAAAA AAAACAGCGT AAAAATCAGT
    2101  TAAAAATTTA ATTTAATTAT TTGAATCGAA AATTACCAAT TTTCGCTTCT CATTACAAAC
    2161  TTCGGGAGTT TT

//

LOCUS          26     4089 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      26
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..565,2579..2785,3101..3172,3470..3589)
                     /translation="MIDWEVKKKKKRWGSREIRSNSNNMRLQLLIFVLFANLVISRK
                     HYKGRKYDLTLFVRNHNVSDATFFRLVGVDNTADFNVSAGNTFEHTFEVRRKGYWTL
                     FVEFPGDRYAKSPRKIISRDSYRKWVMEVYYYPGAVGFSEWHKLSRRGEMHIHG"
ORIGIN
       1  ACAGAATAAA TTCGTTTATC GAGAAGTTTC ATACCTTTTT TGGCGTAAAT TTTAATTACA
      61  CTTCCTTTGA TTTGCTGAGT CATATTTCTC ATAACTCAGA AAAAATTGGT GAGTCATTTG
     121  TAATTGTTTC AATTCAACAA TAAACTACTT TTAAGGACGA TCGAAAATTT CACAATATTC
     181  ACCAAAAAAC TGTCATTTCT GGGTCAATAT GGGCTGGGTC TATGATAACT TTTTCGTGTT
     241  CTAGAATATT TCTCACCACT GTTAGGCACG ACAAATAATG TATATAGAAT CAGAAAAATG
```

```
 301  GATTTTTCAA TTTCATTTTT TTGATTGCTT AATTTTAGGT TCCATTTTTG ATCTAATATT
 361  ATCTTAACTT TCTGAAAGTT TAATGATATG CTGGTCCCTT TCTCAAGTTT TACTTAATTT
 421  TAGAAGACTT TCCAGGCATT ATTTAGACTT CTTTGTCTCC TTTTTTAGAC TTTCACTGCT
 481  TTTTCCAAAA AAAAAAGGAA TGATAGACTG GGAAGTGAAA AAGAAGAAAA AGAGGTGGGG
 541  TTCCCGTGAA ATTAGAAGTA ATTCGGTTAG TAAAACCACA AACTTTTCCC CACTTTTGCC
 601  ATCCCATTCA ATTGTTTTTG TTTTCATTTT TGTTGTCTTG TTTCTTTTTT TTTTTTTTT
 661  CTTACTCCAC CTCTTTCACC AACAACTTTC GGGTCTTCTG ATGTCTGAGT TTCGGCTTCT
 721  TTGGCTTCCT CGACGAAGCA AGACAAACCA TAAAATTATT CAAATTTCTG TCATTTTACA
 781  CACCCTTTTG CCAATTTATA CAACTGTCAA ATGGTTTGCG CATGAGGAAA CGTGAACGGT
 841  GAAACCTCCC ACCAGCTGCC ACGTCATTTG CAAGAGACTG ATGGATGTTT TTGTTGCAAG
 901  TTTCATGGGA ATTGTATTTT ATTCAATGAC ATTGCCAATT ATGATTTCCT TCTTGATTTG
 961  TCACTTTTTG AGATTTTCCA CTTTTACTTT TATAGAATTT CTATGAACAA AATTGTACCA
1021  ATTGCTAAAT TTGGTATTTT TGTACTCCGA AGTACGGTAC CCGGTCTCGA AACGACTGAA
1081  ACTATAGAGT TAGCTGGGCC GCGAAAAGCT TGTCACAACT GTAGCTCCAA AGTTTGGGTC
1141  GAGCAGAGGA ATGTTTTTTT TAATATCAAA TACAAACTAC CCTAAAGCTC TTCTCTAGAA
1201  ATTTCTATTT TCCTTTCCGT TGCCAACAAT AAACGTTTAA ATTTAAACTG TACTATAAAC
1261  ATTTTGAAAT ACCCTATAAG TTTTCTGATT CCTGATCCTT CTATCTATAT TTCCATGACT
1321  CATACTCATT TTCCCGTGAA TCAAAAACCC CCGAAAAACA TACGAATTAG TTCAACTCAT
1381  TAGGTTCTTC GTTTCCAATA ATTCCACGTA TCTCCAATCT AAGTTTCAAT CTCTTTTTAC
1441  CAATCTCACC CCAATAATTT CACCAAATTT TTGCCTCCTC CTCCATTTTC TTCTCCCCAG
1501  TCCTCGAATA GTCACTTCAA ATCAAACTCA GTCGCAGTTT CAATTCTCTT CAAAAATGAT
1561  TTCTTTGTGT CTCTTCGTCT TCAATGCTAT TTCATCAAAT AATTCCAAAT TGGTTTAGTT
1621  CAGCTATGAG ATACGCATCA GTAGTATTTG AATTGACTCG AGATTAATTT GTCACCTCCT
1681  CCGTTTAATT TGTAATCGTG ACCGCCCTTT CACTGGGTGG TCTTGAATGT TCTTTTTTCC
1741  AATGTTTTCT GAGTTTTCTG GCATTCTTGA CCTAATTAAC CAATTTTCAG ATTAATTTCT
1801  CTCTGGGAGA GAGAGAATCG AACGACTGGC ACATAGGTAC TAGAAAAACA ACAAAAACCG
1861  ACAAAAAAGT AGGAAAGGAG GAATGATCCG ACAGCTAAAA ATTGGTAAGT CGTTGCCACG
1921  ACAACCTGGA CAAAAATTTC AATTTCCTGA ATTTTACTGA AATCTCTCAA AATAAAATGT
1981  GCAATTTTTA GTTTCTCTAT CCTGTTTCCG GTTTCTTTTT GTTTGTTGGA ATATTGGAAA
2041  TCATACTTAT TTTTTGTTGC TTTATTTTTC ATGTTTCTAC TCTTGTTACA TCCATTTTCG
2101  TTGAGAGCTT TTCCCGTTTC TTTTCGATGA TAAGTTGACA GCTACATAGA CAAGAAAACA
2161  ACTAAGAAAC GAATAGAAAA CAAAGACGTT CATTTCTTTG ATTCCTTTTG GCTTTGGAGC
2221  TTCGACAACT GTTGGGCTCC AGATTTAGAG TCGTATCAAG AATCTGATTT TCGGACGACA
2281  ATAGTTGAAT TTTACATGTC CGAGAATTGT GTTCAATTTA ATGTTACTAA AATTTTTGAT
2341  TTCTGTTCAA AAACTTGTAA CTTTTGAAAT TAAAATACTT TCAAACTTCT ACTAGACTTG
2401  ACAGTTGGGT AGATTTTATG CCTACAGCCC AGAGGGATCA CCACCAGGCA GTGTACCTGA
2461  CTCCCAGATC CGCAGTGCAT AGCACTTGAA GAACGGATCG TTCTTTAATC CTTTTTTTTA
2521  AATTGAAAAC CAAAGTTAAT ATTCTCAATT TCTAAATTCC CACAAAAAAA TTTTTCAGAA
2581  TAACATGAGG TTACAATTGC TAATCTTCGT GTTGTTTGCC AATCTTGTAA TTTCACGTAA
2641  ACACTACAAG GGACGAAAAT ACGATTTAAC GCTGTTCGTT CGAAATCACA ATGTCTCCGA
2701  TGCGACATTC TTCCGGCTCG TTGGCGTTGA TAACACTGCA GACTTTAATG TATCAGCTGG
2761  AAATACTTTT GAGCACACTT TTGAGGTAAG TCACGAGAAT TGGATAGGAA ATGGGATTTT
2821  AGAATTTCAT TTCATACAAT AAAAGGTTAT ATGTAGAGTT CTTTCTTTTT TGCATTCTTT
2881  AAATTTTAAT GACTATATAA TAATATGAAT AAGCTTTTTG ATTAAAAAAG TGAATCTTTC
2941  AATCTAGTAC ATGACATTCT GCAATAGTAT ACAGGACTAC GATAGGGAAA AATCACTGAA
3001  AAATCAGGTA CTTTTCTGGA AATTTTGTGC TCTTCTATGC TCATGTTCAC ATATAATATT
3061  CTGAACCAAA TATTGTTTTA ATTCAATTTA TGTATTGCAG GTGAGAAGGA AAGGCTATTG
3121  GACGTTATTC GTTGAGTTCC CCGGAGATCG GTACGCAAAG TCACCAAGAA AAGTAAGATT
3181  TAAAAAACAT AAAATCCAAG TAAAACCTCT TTACAATTTT TGGGTTACTG TACCTCCACT
3241  GTACCTCGCG AGATTTATAG CCTAGCAAAA ATTTGACGTT CCAGTTTAAA GGCACACGAT
3301  GTTCTTCTGA TTATGATCTC ATATGATCTC ATTTATAACG GTCGCTTGAC AAATTTCACA
3361  AATTTGATTA TTTGGGATAG AAAATGAAGC TTTATCTTGA AAAATTTAGG TTGTTGGCAT
3421  TTCGAATTTG ATAATAAATT TATTGAAACA AAATCAATCA AATTTTCAGA TAATATCTCG
3481  TGATTCGTAT CGCAAATGGG TGATGGAAGT GTACTACTAT CCGGGTGCCG TTGGTTTTTC
3541  TGAATGGCAC AAATTATCTC GACGAGGTGA AATGCACATT CATGGATAAT TTACTGCTCC
3601  GTTCATCACA CTTTTGTCCA TTTTCCCATC TCTTCCGTGG ATCTCAATGT CCTCCATCCG
3661  GTGGACCACC TTTTCGATCA AATAATTGTT TTATTTCTCT AATTCTTCCC CACAGTTCCT
```

```
    3721  TTCATATAGA ATTTTTATAT TTTCTATTAT ATATGAACAT TTGTAAACGG TGATCTTAAG
    3781  TAATTTTCAA AGAATTTGAA ATTTGAAAAA TTTAGCTCTA TTTCAAAATC GCTGTAGCAG
    3841  TTTTATTCAA ACTTACTAAT TGTGTATTTT TGTATGTACA AATAGTTTTT TTTTCGTGTA
    3901  GAGCTTTATC CTGTTTCAGT TCTTCCGTTC AACTACTGCC ACTGTTTTCC TGTCAAAATT
    3961  GTCGATGACT CTCTCACCCG ATAAAACTTT CTCAAGTTGA ATTGATTTCA GGTTGACTGG
    4021  ATTTTATATA GTTCACCGGT GTATCAAAAT TAAAACATGC AGTAAAAAGG TACAAAATAT
    4081  ACAAATATA

//

LOCUS          27     4244 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      27
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..533,582..739,3415..3591,3640..3744)
                     /translation="MFAVFLAVSVFVLSTAIPTQLKDTNPVTIPACPAGSRPLLRPD
                     GQPRKCLPHQNSLCINALPDKPNADTVCCYHNQVDYYCCLDTTEQQCPDYHQVTVVI
                     HNSFPQNPFALKSFFFKDGIEDDIIDATFGRGDEGLQNENGLLVRHQQPQGNEILKQ
                     "
ORIGIN
       1  TTTTAAATTA TAAATTATAA ATTTTATATT TAAAAATGTT ATTAAACAAA AAATGTAAGG
      61  ATATTGGCGG AACAAGGCAT TTGGAAAAAA AATTGAAAAA AAAATTGAAT ACTTACTTCT
     121  GGATTTTAAA TTCAAAATTT TTTTAAACCG GATTTACAAA AGTGGAGGTT TTTGACATTT
     181  TTTTTCAAAC CTTGAGATTT TTTAAAAATT TTGGAAATTT TTTAAAGAGA AATTTCAAGA
     241  AACAATTCAA AATTTTTGGA AAATTTTTAG GGTTCATCCA AAATTTTCTA TATATTCTCA
     301  TTTTCCCGCA CGCTTTTCCC GCTAGCCTTC TTCTTCAAAG ATTTGATAAG ACGATTGATG
     361  AGATCTCAGA TCTCCGCTGC GTAAAATGTG AAAGAAAGTG AAAAATAGGT CAAATTTGAA
     421  GAATCGTCGT CGGAAAAGTG AAAATTGACG GAAAATCATC CGAAAAGGCT TGTCAGATTT
     481  TTCCCCGTGG AAAATGAGAA TGTTTGCAGT TTTTCTAGCG GTTTCTGTTT TCGGTAATGT
     541  TATTTTAGTA ATACTTATTA ATTTAATAGT CTGAATTTCA GTCCTTTCCA CAGCAATCCC
     601  AACCCAACTA AAAGACACAA ATCCAGTAAC AATTCCTGCA TGTCCTGCAG GATCCCGACC
     661  TCTCTTAAGA CCCGATGGTC AGCCAAGAAA ATGCCTACCA CATCAGAACT CATTGTGTAT
     721  TAATGCGTTG CCGGATAAGG TAAGAAAATC ATTTTTGCTA ATAAACTTTC GATTATGAAT
     781  TCATATTGAA TTCATATTAG GTCTCCAAAT AAGTTCCGGG TCAAAAATCA TAACTTTGTT
     841  CGCTGTGTAT CGATTTTTAT GAAACTGTGG GATTTTACGT TATCAACCAT GATCTTTCAT
     901  TTTACAATAG TCACAAAATT TTTTGGCCGT CCGAAGTGCC CTAACTCGGA GCCAATTTTT
     961  TCAGACATTT TTTAGATCTC GCTTCTTATC AGGTTTCAAT TGAGGTTTGT GTGCGGATTT
    1021  CGCTAAGTTT AGTACACAAT GTAAGAAAAC AAAAAAGTTT GGAAAAAAAT CCGTCAAAAA
    1081  AAAATTTTTT TCGTCGGTCG TCAAAAAATC TTCAAAAAAA AATTTTGTCG AACAGAAAAC
    1141  ATAGCTTTTT CGAAATATTT TTGAGTTTTT GGAGTATTTC TCGAGATCCA AATGTTAAAT
    1201  TGAAATGTTT TGTACGTGTA AAAATAGTTT GCACATGGTT ACATCATTTT TGTATGAAAA
    1261  ATCAAGAATT TTTGACAAAA AATTTTTTGA AGATTTTTTG ACGACAAACC AAAAAAAAAA
    1321  TGTTTTGGAC GAATTTTTTT CCAAATTTTT GTTGTTTTCT AATAATGTAT TTTAAACATA
    1381  GCAAAATCCG CACACAAGCC TTAACTCAAA GCTGAAAAGA ACCGAGATCA GAAAAATGCC
    1441  TGAAAAAATT GGCTCCGAGT TAGGACACTT GGGGTGGTCA AAAAATTTTG TGACTATTGT
    1501  CAAATGAAAG ATCATGGTTG ATAACATAAA TTCCCAAAGT TTCATAAAAA TCGATACACA
    1561  GCGAACAAAG TTATGATTTT TGACCCGGAA CTTATTTGGA GACCTAATAA TAAACCTTCA
    1621  ATTATGAATT CCTATTGAAT TTATATTAAA CTCATGCTTT ATTTAGACAG TACAAATTAC
    1681  CCATTGTTAG TATGGCGTAT TTCACCAGAT CACTTTTTTT GGAAGTTAGT CTTGCAGTCT
    1741  TCAATAATAT TGCATTTACT GTGCTTGTTG AATTAGAAAA TAATTTTATT AGTTAGTCAT
    1801  TATTAATTGT TAGTTATAAT AGAAATTCGC AATGTTGTCT GTAAAAAATC CTTTGAGCTA
    1861  TTCATTTTAA CAGGCTAATA AAAAAAAAGT ATACTTTAAT TATATTACTA CTTCAACTTT
    1921  CTTTTGGCCC CAACATCTCC TTCTAAATAT CGTTGAAACA AACTATTTAC ATAATAATGA
    1981  CATACATTTT TATTTTATTA CTTGATTACT TAGGCTATAA AAGTATCATT AAAGGTGGGG
    2041  TACCGAAATT TGAGACTTTG CTTTTTTAGA CCCAAAATGG TACAAAACTA CCAAAATTCG
    2101  TAGTGAGACG TTCTGAAAAT GTTCCCAAAA AAAGTTGTGG CCCCTCAAAG TTTTGGAAAA
```

```
    2161  TGGCCAATTT TTAGCTAAAA TCTCATCAAT TTTGGCAACT TCTCGGTGTC GCAGTGATGC
    2221  TAAAAATAGA CCATTTTTTC CAAAACTTTG AGCGGCCATA ACTTTTTTTT TTGAAAAATT
    2281  TCCAGAACGT CACATTACAA AATTTGGTAG TTTTGGACCA TTTTTGGCCT AAAAAAGCAA
    2341  ACTCTCAAAT TTCTTTAATA TTTTAAATGA CTCCAAACTT GCTCCCTGAT TCCGAATATC
    2401  TAAGTGAAAA AATTTAAAAA ACTTTCCCTG ATTTTATATT TGAGCTTGAA ATAGTGAATT
    2461  TCATTTGTGC CCCGGTTTCT TTTTTCAAAT ACGCGCCGAA ATAAATTTTC CTTGGAGCGC
    2521  GTTTGCTTCA TTTAATTTTC TCCATATATT TCCACTTCTT TCAGCTATTT TCATTCATTT
    2581  TTACACGGGG TTCTGGCCTT TCTCATTGAA TTTTCGCAAA TACATCAATT TACAACTAAA
    2641  ACGAGCCGCG ACGCGACACG CAACGCGACA CGCAACGCGC CGTAAATCTA CACAAAATCT
    2701  CTCCGGCCCA AAATGACCTA GTTCGGCAAA CTCTGCCATT TCAATTTATG AGGGAAGCCA
    2761  GAACCCCGTG TAAAAATGAA TGAAAATAGC TGAAAAACTG AAAGAAGTGG AAATATATGG
    2821  AGAAAATTAA ATGAGGCAAA CGCGCTCCAA GGAGAATTTA TTTCGGCGCG TATTTGAAAA
    2881  AAGAAACCGG GGCACAAATG AAATTCGCGA TTTGAAGCTC AAATATAAAA TCAGGGAAAT
    2941  TTTTTTGAAT TTTTTCACAT CGATATTCGG AATCAGGGGT AAATTTGGAG TCATATGAAA
    3001  ATATTTCCCA AATTTCAGTA CCCAACCTTT AATATACAAT AGCTTTGTTA ATGATGGTTA
    3061  TTAAAATGTG AAAGGAAATG TGAATTGACA AAACTTTTTG CTCATTCCTT GTGATATTGG
    3121  TATTTCACAT ATCCTTTTCT ATAAACTGTG TGTATAACAA CACATTTTCA AATAAACACA
    3181  CGTGGTAGGC TAGATTTTGA TAGAAAAAAT TGTTAAAAAT TAGACCCATA GGCATAGTTC
    3241  ATAGGTTTAG GCCCAATGTC TTAAGTCTTT AAGATTAGGT ACAATAAGAT GTGGATATTC
    3301  GGAAAAAAAC TGTCTGTAGA CTATAACTTT TCTACAACTA TAGTTACACT GAAAAAATAT
    3361  AGGGTTTTCA AACGGAATTT TAATTTATTT TTGTTCATCC TATAAAATTT CCAGCCAAAC
    3421  GCTGACACCG TCTGCTGCTA CCACAACCAG GTCGACTATT ACTGCTGCCT GGACACTACC
    3481  GAACAACAAT GCCCAGACTA CCATCAGGTT ACTGTAGTCA TTCATAACAG TTTCCCTCAG
    3541  AACCCGTTTG CGTTGAAGAG CTTCTTCTTC AAGGATGGTA TTGAGGATGA TGTGAGTTTG
    3601  TTTGACTAAG AACTTTAGTC CTAACAACCA AAATTGCAGA TAATTGATGC AACATTTGGC
    3661  CGTGGAGATG AGGGACTTCA GAATGAAAAC GGGCTTCTTG TACGTCATCA GCAACCTCAG
    3721  GGAAATGAGA TTTTGAAGCA ATAAGAATTA ATCACGACAT GTGTAATTTA CATTGGTGTT
    3781  GACTATTTTG AACATATTAA TAATTATTAT TTTAATAAGA ATGTTAAAAA ATAAAGTAAT
    3841  AACGAAATGC TCAAAAATGC TCAATTTGCC TTAGAGCACC CTTTGTTTCC ATATCCATCC
    3901  AAATCGGTAT TATACTAATG GGTATGTTCA TTCCAGCTCA AGTCAGCATG CAGGTTGGAG
    3961  CAATAGAAAT GATCGGCTTC CTGGCACACG TTTCAAACGG TTTAACGCCT CTGAGAAGTT
    4021  TTTGTGTCGA ATTTCAATTT CAACAGGTAC TTCAAAAGGT ACAACAGGTA CTAGCCTTGC
    4081  GAAAAAAGGG TGGCACAAAC ATTTTGGAAT TTTGGTAGAA TTTTCATAGA TTTGGATTCT
    4141  AAATTTTTTA AAAATATTAT TGGTACGTTT TTGGACCAAG ATGTCAATTT TCAGACGCGC
    4201  TAGAGGTTTG GATTAATATA CGCACTGAAA GCCACAAAAA ATCA

//

LOCUS          28     2446 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      28
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          join(500..589,747..941,1749..1946)
                  /translation="MNGNFGQYINNCDNLDEQISVTDFIKTLKKVESSDTLNGRESV
                  SSSSSTDKLSKLPEDGRRVDKKNSANDLLQTMRKRFVHRATEPIFHALPEAQERRLV
                  SGRDCGTSSSSISLGRKPKVFEGGCGANDDLLKMPSFESTLDRKTDSDPTLSKTHKY
                  HVS"
ORIGIN
       1  TCAACAGTTA AAATAACGGA CGAATTTCAG AAATCTGGAA AACGTTATCA TCGTAAACTT
      61  ACAAGTCTAA TAATACAATA GATTTTTTAT CGTAACTTTT ATTTTATATT CTTTGTGACC
     121  TGGCATCAAT AAATCCAATA AACCTCTACT CTCTCAGAAA TTTATTACCA ATCATTTGTA
     181  TCTACAAATC AACATCATTT TATTCACACG CTCTTTTCGT GGGATTTGCG TCAATCTGAG
     241  TTGTTATTTG AAGAAAGTTT ATGAGAAATT TTTGTAATTT TATTAGATTC AAGGCACGCC
     301  AAAAAAAAGC GGTTGAGTGA AATACCAATG TTTCCAACTT CACCTGTTTC GTTTTTGCTC
     361  GCCCTCGTTT TAAAATTCAT TCCGCTATAA TTGATTTTCA ATTTTTGATT TCTATTTGTT
     421  TATTACTTTA TGATATTATT CAGTTATATT TCGCCCAATT CATCAAACAG AAAACAAAGT
```

```
     481  GGCGGGGAAT CAAGGAGAAA TGAATGGCAA TTTTGGCCAA TATATTAACA ACTGTGACAA
     541  TTTGGATGAG CAAATCAGTG TTACAGATTT CATTAAAACT CTGAAGAAAG TTAGTTTTAA
     601  ATTAAAAATG CAAATTACCG TATTTCCTCT ATCATTCTTT TTTACAAGAC TAATCGTCTC
     661  GTATGGGCGC AAGACTATTA AAACTTGAAT TATTAATAGA GATAACGAAA CTTTTTTACT
     721  TTTACATACA TTTTTCGTGG CTTTAGGTTG AATCGTCCGA CACCCTGAAT GGTCGAGAAA
     781  GTGTTTCATC AAGTTCATCA ACCGACAAAC TTTCAAAATT GCCGGAAGAT GGTCGTAGAG
     841  TTGATAAAAA GAACAGTGCA AACGATTTGC TTCAAACAAT GCGCAAACGA TTTGTTCACC
     901  GAGCGACGGA GCCAATATTC CACGCCTTAC CAGAGGCGCA GGTAAGCAGG TGTGTGCCTC
     961  CTCTAACTGC CTTCTATTTG TCGCATGTTC TTGATGTTGT TGTCTTTTTC AAATGGAATT
    1021  TGAAACTTGA ATCTAAAACT TATTTTTGTC TATTCAAAAC CACACACCAA TAGGAAAACT
    1081  GGGATATTTG CCAATTTTTT GCCAATTTCG GTGAAACTCA ATGTAGAACA ACTTTTGCAT
    1141  TCAAACCGTA AAGTTTAATG TTTTCCGATT AAGAAACTTC AGATTATTGT CTCTAATTTT
    1201  TATATTTTAT TCCTGTTATG AATATATTTA AGACTCAAAA TTTTCGAATA ATTTTTTTTT
    1261  CAAATTTAGT ACCAACTTGA GCAAAAACAC ATCATAGTCT CATTTGTATT TCAGCAATGA
    1321  CTTCATTAAA AAGTGATATT CAATTGAATT CGGTTTCTTC AACCAAAACT GTGAAGTAAA
    1381  AACAGTTTGC TTTATGATTT TTGATACATA TTTGTTCTAG ATAGTATTAA TTTTCAAAAA
    1441  CATTCCAGAG CTTACCATTA GATGTTAGTT ATTTAGTGTG CCTATAGTAA AACACTTAGT
    1501  TCATTTACCA AACAAAACTC CACCTTTAAG TACCAAAATC TCAATCAGTA CAAATTTCCT
    1561  GATAAAGTTT GAAAATATAT TCATAATTTT ACTTACCATC GGTCATTTTT TAATTGGAAA
    1621  CTCTTAAAAT TACTGTACTG AAGCTTGTAA TGAAAAACCT GTTTCCAATT TCCCTCATTT
    1681  CCTGGTCCAC GAGTGAAATT CCGATTAAAA CTATCAAAAC ATCTCAATAA ACTTGTTTTC
    1741  TGTTTCAGGA ACGTCGACTC GTCAGCGGAA GGGACTGCGG CACTTCCTCC TCGAGTATCA
    1801  GCCTCGGCAG GAAGCCGAAA GTTTTCGAAG GAGGATGCGG TGCCAACGAC GACCTCCTAA
    1861  AAATGCCAAG TTTTGAGTCG ACGTTAGACC GGAAAACTGA TTCTGATCCA ACACTTTCTA
    1921  AAACTCATAA ATACCACGTC TCGTAATCTT TCCTATTCTT AGACAATCAT CTAGTCTACG
    1981  ATTTCCATAT TTCTCGAACA AACAGGCTAA TATACAATAA ATGCATATGT TCATGAATAT
    2041  TTAATATAAA TTTCGCTTTT TATTATGTTC TCTCCATTTT CTCTCCGGCT CCACATGCTT
    2101  CTAAATACCA CTTTTATTCT CAATTGTACT TAAAAATACT ACGTTCTTTT CATAAACGTA
    2161  CTTTGTCTGT TACATTACAA TTGGTTCCTT TTGCATTTGC ATCTGCATCT GCCATTCGTT
    2221  TTTTTCATCT TTTTTGCAAT GTTAGATCGT TGCTGTCGGA AAGGAAAATG AGATATGAGT
    2281  GACAGAAGAG GGATGTAGTG GACACAAGAG GGTGAGAGAA ACCAGTTTGT GGCCGGTACT
    2341  CTAAACGCAC CGGACATCTT CTTCGAATTC TACAAAGAGA CGTCATCTTT TTCAGCGGCT
    2401  GGCTACAAAA TCGACTGACA TGATTGTTCT TTCCAATTCT CATTTT

//

LOCUS          29     3449 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      29
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS
join(500..586,1706..1750,2155..2285,2330..2407,2714..2800,2859..2949)
                 /translation="MKPHLCDQPMNLTRRACCQLVLFLNDTERVPPCSTSMLPSTTS
                 HKPPIYLVEPVEESEDHLTFFVGIFMLMVIALAFSIGLSKMTKPKVSISRYNTSNSS
                 YDILMWLSSNPTNRSSMLLLLRSMCRTPTTPDSSVSVVSSNIPLLPSYQCASQSTTA
                 NTPPPPYSATSHLVA"
ORIGIN
       1  GGAAATCAGA AGAAAAAATA ATGCAATTTA AAGTTGGAAA GTGCAAAAAT TTCAGTGTCA
      61  AAACTAACCG AATCAAAAAT GACCTTTTCG TGCCAAGACC AGAACTTTCC AAAATACCAA
     121  TAATTGAATA AAACGTTTAG TTTCCTCGTG CCACCCCGGC GGGCGGGTGT CACTAATTGT
     181  CAGTGGCCTG GTGTGACATC AAAAAGTTTG ATTTCATTGT GGCCGTCGTG GGGTCCTCCC
     241  ATTTCTCTTC CTCTCTTTGC CAACACCACA AACGGCGTTT TTTTTCATTT TCTGCGACCG
     301  CTTCTTCCTT TTTTTTATCC CACCGAACGT TATTTTTTTT TGCATAACTT CTTCCAGTTT
     361  TTTTAGTAGT ATGAAGACAA ATTGTTGTGT CTTCTTGTCC CCCCATCCTT CCCGCCGAAA
     421  AACATAATAA AATAATTCAA TTACCGTCTA GCCGCCGCTT CTTGTCTTAT TTTGAAGAAA
     481  ATGAGCCCAG TTTTACAGAA TGAAGCCTCA TTTATGTGAT CAACCAATGA ATCTTACTCG
```

```
     541  GCGAGCATGC TGTCAACTTG TACTATTTTT GAATGATACT GAACGGGTGA GTATATTATA
     601  ATTGGTATTT GTAAAACGAT TTTTAAGTAA TAGAATGTAA TAGAAAAACT GATTCTTACA
     661  TTTTTTTTTT CATTTTTGAA GGAAAAACGT TTTAGATAGG CACAAAAATG CATATTTGCT
     721  CAAAAATACG GTACCTAGGG TCTCGACGCG ACACATTTTA AAAAGAATAC TGTAATGTTT
     781  TTAATCGAGT TTTCATATTT TTTTACGAAA AAATTTATTT AATTATTTAA AAACATATAA
     841  ATCGCTATAA AAAGCTGCGA AAAATCCTTG AAATTCCGAA AAAACAAAAG TTTGAAATTA
     901  CAGTAATTTT TAATGACGTT TGGTTTTTTG CACAAATCTC AAAACTTAGC GTCTGGATAA
     961  TACAAATCTG CGTTTACTGG GTCAACATCT GGTTTTACGA AAAAATGCGG AAAATATGTG
    1021  GTAGGCAGGC ATAGGAACGC CTGGAGATTT AATTTCTTGT CTATCAGTGT TGTTTTATTT
    1081  GTCTGAATAT GTTTTTTTTC CAAAAACTTT CGCGTAAAGT ATGGTGCATT GTCATTACTG
    1141  TACACACCAA TGTGATGCAA CATATCCATA TTTTACGCGA AAGTTATCTA GAAAACAATG
    1201  GTTACAGGCA CAGTAATAGG CATTTAAGTT TACAGGTGTG TCAATGTCTG CCTGCCTGCC
    1261  AACCGCCTAT TTTCTGCATT TTGGCGTAAA ATCTGGTTTT TTATGGTGTT ATTTTTTCCT
    1321  AAAAAATGCC AACAAAGTAC AAATTGGACA TTCAAATGCG GAACTTTCTC CAGATTTATA
    1381  TGTTTACTAT AAACACAATA AACTTTCAAA TCCACCGAGA CTACTGTAAT TATCATGTTT
    1441  TGGTGCACAC AAAACAAATG CCTGTAGTCA TATCGCGCTG ACCCTTCTCT GAGCCCCCGT
    1501  CGAGCTTCTC TACTCTCTTC TCCACATACA CACTCTCTCT ATCCCAACGA ATAAGGAGAG
    1561  TATTTATCAT TTGTGAGATT TCCCTGTCAG GGGAATCCCC GTGGGATCAT CCCCACATCT
    1621  TCTCAAATCT CCCCCGATTT TTTGCTTTTT TGTGTTGGAA CAATCATAAT TGATGGGTGG
    1681  TTTTTTGACG TGTATTTGGT TTCAGGTGCC GCCTTGCTCC ACATCAATGC TCCCATCAAC
    1741  TACGAGTCAT GTAAGTTTGG CCCCCATCTT TTGCTAGGTG GCCTAGAAAA AATGAAATCT
    1801  CGGGCACCCT TTGAAGCTGA CCCAAAAAAT TATCTCAAAA GGTGGTTTTC ATAGTCTCGT
    1861  CAAACCGATT TTAAAAGGGT AGCCGATTTT TTTAACAGTT TTTTTTTAAT TCTTGATTTT
    1921  GTATTATTTT TTGCCGACTA TTTTGAAATT AAACATTTTA AAAAGTGAAA CAACTTGAAT
    1981  TTTTAGTTTT CTACAGAATA TACATCGAGG TTTTATGTAT TTGAAAAATG AAAAATAACA
    2041  GTTCAATTTA AATTTTGAGT TTAAAAAGCT TTAAAGCATG CATTTTGGCT GGAAAAATGG
    2101  TTGAAAAATG AAAAGAGCCC TTTCCCTCGT GGAGTACAAT TCAAATTTTT CCAGAAACCA
    2161  CCGATATATT TAGTGGAACC TGTTGAAGAA TCCGAGGATC ATTTGACATT TTTCGTTGGC
    2221  ATTTTTATGC TTATGGTTAT TGCATTAGCA TTCTCAATTG GGTTGAGCAA AATGACAAAG
    2281  CCAAAGTAAG ATTTTTTGAA GATTCTGAAA ATCAAATACT TTTTTTCAGA GTATCAATAA
    2341  GTCGATATAA TACTTCAAAC TCTTCGTATG ACATATTAAT GTGGCTTTCG AGTAATCCAA
    2401  CAAATAGGTG AGTTTAGAAC AATAAAGCCA AAAATGTCTA TGAGGCCTAA AAGGCCTATA
    2461  AGGCTTATAA GGCCTATCAA GCCTAAAAGA TCTATGAGGC CTATGAGGTC TACAAAACCG
    2521  GTAAAGTCTA TATGGTCTAA AAGGTCTATT AGGCCTAGAA TCTTCAGATG GTTTATTGGG
    2581  TCTGGGGCTA TGAGGTCTAT GATATCCATA AAGCCTATAA GGCTAAAAGG TCTATAAGGC
    2641  CTATAAGTAT ATATGGTCTA TAAGATCTAA ATGGTCTATA CGGTCATCTT GAACTAGAAA
    2701  ACTTTTAATT CAGATCCTCC ATGCTCCTTC TGCTCCGATC CATGTGCCGA ACTCCAACGA
    2761  CACCGGACAG CAGTGTCTCA GTGGTATCCT CAAATATCCC GTATGTTTTC TATTTTCTTC
    2821  TCTTCTATTT TCCTGTTTAA TCTGATTAAG AACTTCAGAC TGCTCCCATC TTACCAATGT
    2881  GCCAGTCAAT CCACCACAGC CAACACACCA CCTCCTCCCT ATTCAGCCAC GTCACATCTG
    2941  GTCGCTTGAT GCCTTTGATT TTTGAAATTT TGTTTTCTAC GGGTTCCCCC TGTCTGTTTC
    3001  GCTTTGAATT ACCAAGTCCT TATTAAAGTT ATAAAACAAT TTCCAATTGC CCTGTAGTTT
    3061  GATCTTTAAA AAATGGGCGG GAAAGAAACA CAAACCACAT CGGATTGAAA TCCGATATTT
    3121  GAGGATACTC TTCATCTGCG CACCATCACC ACCGATTGGT TTAGCTCTTT GAGGTTTTTG
    3181  GTGTTCTAAT GTGGGAAAAA ACGAGTGGAA TTGAATGACA ATGACGTTGA AAATCAAACA
    3241  ACTGAATACT CCGAACATGT TTTGTTTTCT ACCAGTAGAG CTGAGAGAAG TGCCGAAAAA
    3301  TTCTCACTTT TTCGTCGTTT TTTTTTTGAA AAATTATATT AAACATATCG AGAAATATTG
    3361  GTACTTCTTG AAAATAACAT TATTTTCCCT TAATTGCCGG TTTGTCGATT TTCTGGAAAT
    3421  TTTAATTTTC CGCAAATTGC CGATTTGCG

//

LOCUS          30     1789 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      30
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
```

```
     CDS             complement(join(500..627,675..860,1016..1136,1182..1289))
                     /translation="MGDGANVFNTFSFNKWADGKREEEIDKESRRRRAEQIKRLKER
                     LRKERALREAQGDGDLDLDALIEALQRGEVINIPQQTDVIVTEPIEETSKRKRGVPD
                     NLSDDYIFNEINLYANRSIGLNNIAEPVDIAPVPLQKLDPLSFRYNRRNQNNEMIAN
                     EEQIVQHRPNNFKFNRMIQDIQK"
ORIGIN
        1  TAAATATGAA CTCGCGGGCG CGGTAATCTT TTTTAGGTTG GCGCGAGTTT TTGCAATGCA
       61  AACGCGCTCC ACCGCAAATC TGAATTTTCG AACATTTTTC GCCAATTTTT ATCCTTATTT
      121  TTTTATTTTT TGCAATATTT CGCTATTTTT ATTTGTTTCT CCCACCAAAA GCGCACATTT
      181  TTAAATCAAA ATACGATTAA ATAAAAGTAA TATAACTTTT TAAATAAGTC CAAATAACCT
      241  TTAAGCAAAG TAAGGGTCGA ATCTGAAACT ATTTCTTGAA AACAACTTTT TTGTCTCATT
      301  CATCGAAGAT TGGTTGTATT CGGAAGTGAG AGGTTGACAA GTTTTTTAAA ATCCATTTTA
      361  TTAATTATTC CGAAATACAC ATTTAGTTCC AATAGTTTCG AAATTTAAAA CGTTTGAGTC
      421  ATCCTTCATT TTTTCTGCCT AGTTGCTGTG ATCATTTGGA GAGAGTATAA GAATTTATTA
      481  ATCATTTCGA AAAAAAAAAT CATTTTTGAA TATCCTGTAT CATTCTATTA AACTTGAAGT
      541  TATTCGGCCG GTGTTGAACA ATTTGCTCCT CATTGGCGAT CATTTCGTTA TTCTGATTTC
      601  TTCGATTATA ACGGAAGGAG AGTGGATCTG AATTTACAGT AGATTTTAAA TTGGTTTTTC
      661  TTGAATTTCC TCACCTAATT TTTGTAGTGG AACAGGCGCT ATGTCCACCG GCTCTGCAAT
      721  ATTGTTCAAG CCGATCGAGC GATTTGCGTA GAGATTTATT TCATTAAAGA TATAATCATC
      781  TGACAAATTA TCTGGCACCC CTCGTTTTCT TTTAGAAGTT TCCTCGATAG GTTCGGTTAC
      841  AATTACATCG GTTTGTTGAG CTGAAATAGG ATCCAAATTA TCAAATATCC TTCTCTATTC
      901  ATGTTATAAA TCTTGGTTGT AAGAACAGTT AACATAATTA GTAACAATTA GAAACCTGAT
      961  GTGATTAAGA TTCTATCACT GTTTTTAAAT AGTAAAATAA ATAATTTTGA CATACGTATA
     1021  TTGATAACTT CTCCTCTTTG TAACGCCTCA ATCAATGCAT CCAAATCAAG ATCTCCATCA
     1081  CCTTGAGCCT CACGCAGTGC TCTCTCCTTT CGAAGCCTTT CCTTCAGTCG TTTAATCTGT
     1141  CATTAAGCCA CTAGTATACA CCTTGGAGAG CCACCACAAA CCTGCTCGGC TCGTCGCCGT
     1201  CTACTTTCCT TATCTATCTC CTCTTCACGC TTGCCGTCTG CCCATTTGTT GAAGCTGAAT
     1261  GTGTTGAACA CATTCGCACC GTCTCCCATT GTACTGTTCT GTAAATATAT AAACTGAGCT
     1321  AATGAGAAGA ATACGTACTA GAGCACCACG CTTTTGCGAA AAATGAGAAC CAAAAGCTCG
     1381  ATTGTGCACT CACTTTTGCA TATTTTATTA AAGCTTTTGA CGTCCTTTTC TCTTTTGTTC
     1441  GCCTGTTTGA AATTTTTGAA AACCAGGAAA CCATGATGAT TTTAAATGTT TAATGATGTT
     1501  CTTCTAACAG CATGCCCACG CGTTTTGCCT GAAAATTCGA ATCAAATTAA TTTTCAGCCT
     1561  TTTTATTGTT TAAATATCAG AAAAAAACTG GCTGTTCACT AATCACGTGA TGCCTGTGTT
     1621  TTCATCCTCC CTTCTACACT GAAAATATTC TATTTTAAGG ACATTTTCAG TCATTTTCCG
     1681  GCGCTGTTCC CTCTGTCGTC GCCACGTCCA GTACACCGAA ACCGACGACA AATCGATGAG
     1741  AGCTCTACGA CCGCAGGCGC GATGGATTGT GCTATTGCAG TGAGAGAGA

//

LOCUS          31     1668 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      31
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS             join(500..553,611..702,760..1168)
                     /translation="MFLRLSRPHHHDTINILTSTYQSLVNSARILFVEKGYQAVSID
                     EISGKAACYKQQLIMIDAYITTKTDLTNGWSALESIFEHYLDYIIDNNKNLIPIQEV
                     MPIIGWNELEKISLEYITGKVNAIVSKLIQENQLKAYDDDVLKNLLNGWFMHIAIHA
                     KNLKELADKKGQFIAIYRGFLLSLKDK"
ORIGIN
        1  GGCTGCAACC CTACTAGCTC AAAATTTATT CACACATGGT TACGCTTTGG GGAAATTATG
       61  AGGGGATCTC TCAGTGCATT GCCTCCAATT CCCATAATTT ATTACGCCGA TAATAACTTG
      121  GTGTAACCTT AAAAATGTAC TTAAATCGAC GTGTAAAAGA TTGTTGGGAA TCAAATTGAT
      181  ATTTTAATGC GATCTCAAGG ATAGTTTTTT TCGTCAACCT CAACTCAACA GCCGCTTTCG
      241  TCAAACGACG AGCACGAATA TAGCTAGCCA GTGTGACCCC TGTTACTTTT TTGAACAGCC
      301  GCTGAAAATA CCACTTGGTA TAACCCGCTT TATTCGCCAC ATCATCAAGC AGTAAAGACT
      361  GATCTAAATT ATGTTCAATC CATAATAGAA CATCTTTGAT AACCGTTGTA TGAAACTGCT
```

```
     421  TATCATCATA TCTTAATTGG ATGTTATTAG CTTTATTTTG ATAGCGAGAA TGCTGTTCAA
     481  TATACATAAA ATAACCTAAA TGTTCTTAAG ATTGTCACGA CCACATCATC ATGATACCAT
     541  AAACATACTG ACGGTATGTT ATTTTAAATC TATCATGGAA AATAAAAATC ATCAACAAGA
     601  AAATTTTAAG AGTACCTATC AATCACTGGT TAACTCAGCA CGAATATTGT TTGTTGAAAA
     661  AGGCTATCAA GCTGTTTCAA TAGATGAGAT CTCGGGAAAA GCGTTGGTGA CCAAAGGTGC
     721  CTTTTATCAT CACTTTAAAA ATAAAAAACA ATTACTCAGT GCCTGTTATA AGCAGCAATT
     781  AATTATGATT GATGCCTACA TCACAACAAA AACTGATTTA ACAAATGGTT GGTCTGCCTT
     841  AGAAAGTATA TTTGAACATT ATCTTGATTA TATTATTGAT AATAATAAAA ACCTTATCCC
     901  TATCCAAGAA GTGATGCCTA TCATTGGTTG GAATGAACTT GAAAAAATTA GCCTTGAATA
     961  CATTACTGGT AAGGTAAACG CCATTGTCAG CAAATTGATC CAAGAGAACC AACTTAAAGC
    1021  TTATGATGAT GATGTGCTTA AAAACTTACT CAATGGCTGG TTTATGCATA TCGCAATACA
    1081  TGCGAAAAAC CTAAAAGAGC TTGCCGATAA AAAAGGCCAA TTTATTGCTA TTTACCGCGG
    1141  CTTTTTATTG AGCTTGAAAG ATAAATAAAA TAGATAGGTT TTATTTGAAG CTAAATCTTC
    1201  TTTATCGTAA AAAATGCCCT CTTGGGTTAT CAAGAGGGTC ATTATATTTC GCGGAATAAC
    1261  ATCATTTGGT GACGAAATAA CTAAGCACTT GTCTCCTGTT TACTCCCCTG AGCTTGAGGG
    1321  GTTAACATGA AGGTCATCGA TAGCAGGATA ATAATACAGT AAAACGCTAA ACCAATAATC
    1381  CAAATCCAGC CATCCCAAAT TGGTAGTGAA TGATTATAAA TAACAGCAAA CAGTAATGGG
    1441  CCAATAACAC CGGTTGCATT GGTAAGGCTC ACCAATAATC CCTGTAAAGC ACCTTGCTGA
    1501  TGACTCTTTG TTTGGATAGA CATCACTCCC TGTAATGCAG GTAAAGCGAT CCCACCACCA
    1561  GCCAATAAAA TTAAAACAGG GAAAACTAAC CAACCTTCAG ATATAAACGC TAAAAAGGCA
    1621  AATGCACTAC TATCTGCAAT AAATCCGAGC AGTACTGCCG TTTTTTCG

//

LOCUS          32     2375 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      32
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS          join(500..590,640..724,1287..1344,1397..1522,1678..1875)
                  /translation="MIYSIVFGILIMMESTVLGCPPQPNGNCYYGQYMLSNSQGGFQ
                  NQQCGSAINRDWCAVSYQPATQQANFGCASNFPINVGGSVCGEGAGFVNQQGCSMVN
                  TPNGPCYFCCCKGGSCNNPQIFAREASKLPINQQQMQTYNPNGYNPYSRQNYYPIPW
                  YNTASQPSNPFFLIVIILIIFLILSYSA"
ORIGIN
       1  AAAACACAAG TTATCTTCTA TGCGTCTAAA TTTACAAAAC TCGTCCAACC TAGTAATAGG
      61  TTTTTTTCCA AAATTTTGAA TCTCCATAAC TTTTTTGAAA AATTTTCAAA ACGTCGAAAA
     121  AGTCTTCAAT TTCTGCTACT CCTCCTTTTA AAAAATTGGA AAACTTTCCG ATTTTTTCTG
     181  TAGATGCTTT AAAATATATC AACTCTTGGA GCCTACTTGA ATATTGAAAC TACTGTAATC
     241  GTATCAAAAT GTACTACATT TGCAAAATGC ATTTGATCAA TTTAATTTTT TTCGTTCAAT
     301  TTTCTCCAAA AAAATGGAAT AGTAAAAACG TCATAGAGTA CATATACAAA CATTCCATTC
     361  ATTCTATACA AAACCAATTG AGGGCCGAGA AATTGAAAAA GCATACAAAA AAGATGCAGG
     421  CACTCCTTGA TTTCACCATT TACATATCAC CTTCGACAAT ATTCATATTT TCTGAAACTT
     481  TTTTATTGGA AAGAGGATAA TGATTTATTC AATTGTTTTT GGTATTCTTA TTATGATGGA
     541  GTCAACCGTT TTAGGATGTC CACCACAGCC AAATGGAAAT TGTTATTACG GTAATTGATT
     601  AGGTTATTAC TGAAATAGCA TGAATTAAGC TAATTTTAGG TCAGTACATG CTTTCAAATA
     661  GTCAAGGAGG TTTTCAGAAT CAGCAGTGCG GGTCGGCGAT CAATAGGGAT TGGTGTGCAG
     721  TATCGTATGT TTTTTTTTGT TAGAAAAAAT TGAAAAATGT ATAATTGTGT TGTGTTAAGA
     781  AAAACTAATT TTTGGTCATT TTTTGGTCGG GTGTACCTGT CTGCTTACCG AATATTGTAG
     841  ATCTAAGTCA GAAAATTCAA ATATATACTT TTAAAAACAT GTTTTATTGA ATACCACTTT
     901  TCCCATAAAC TTTGAGCTCT TCACCCTTTT GAAGACTAAT TTTGGATACC CCCTCTCACT
     961  TGTATGAACT GGTATTCATT ATTTGCACAC AAACATACAA CTCCAGAATG AGGTACTTGA
    1021  TGTCGTTGAC TCTACAATTC ACTTTCAATT ATTCTCATCA TTCATTTCAA TATGTTTACT
    1081  CCTTCTTCAT ACTACAATCA ACCATCATTT AATTCGTTTT CTTCAGCCTA TCCGAGGTGA
    1141  TTTGGATCAT CTTTTTTGAT GATAAAACCC AACATTGGGT CAAAAGAGAA TTTAATATTT
    1201  TAGTGAAAAC TGAACCAAAC TTATATATTC AATAAATATA AAACATTTGG TAGATCACTT
    1261  TAAAATCCAA ATTTATTTTT CTACAGCTAC CAACCAGCAA CTCAGCAAGC AAACTTTGGG
```

```
    1321  TGTGCATCCA ATTTTCCCAT AAATGTGAGC ACTAAGCTTT TAGTCGGAAA AGTGAGCTAC
    1381  CGTAAAACTA TTTTAGGTCG GTGGATCAGT TTGTGGTGAA GGTGCTGGAT TTGTGAATCA
    1441  ACAAGGTTGT AGTATGGTCA ATACACCAAA TGGACCATGT TACTTTTGTT GTTGTAAAGG
    1501  GGGAAGTTGT AATAATCCTC AGGTACGTTA TTTTTTCAGT ATGAGAATGG TTATGTATGT
    1561  AGTGTCTTAA ATGGAATTCG AAAAATGTAC TGTTAAAAAC TACTGAAAAT AAGAATTTGA
    1621  TTTTCACTTG CTATAACCCC AAGTTTGAAT TTTTCAATTT CCTTATTTTA ATTACAGATA
    1681  TTTGCTCGAG AAGCTTCGAA ACTTCCAATA AATCAGCAAC AAATGCAAAC ATACAATCCA
    1741  AATGGATATA ATCCATATTC AAGACAAAAT TATTATCCAA TTCCTTGGTA TAATACAGCT
    1801  TCTCAACCTT CCAATCCATT TTTCCTTATT GTCATTATTC TTATTATATT TTTAATTCTC
    1861  TCGTATTCTG CTTAAAATCA TCGTTTTTTT TTTCTAAAGA TCTCTTTTTT TTATTTTATC
    1921  CCCAAATCTT GATGAGATTC TCCCATTTAT TGGTAATTAA TTTTCAATTC AATTGTTTTA
    1981  AAAACCATTC ATTATGTACT AAATTTCGAT AATGCTACAG TACAGAATGA AGATGAACAG
    2041  AAAAATTGCA GTTTCAATTG ATATAACTCG GAAACAAAAC AAATGCCATG CAAATGAGAA
    2101  AAAAATTTCA CAGACAAGTC TCACAGTGAG AGGAGGAAAG AATCAACAAG GGATGATGAT
    2161  GATGATGAAA AATTGAATTT CTCAAGTGAG AAAAACTATT GAAACCGGTG GGTTTTTGTA
    2221  ACTTTCGAGG AATTTTGTGC TTAACGATGG GCGTAAAAAG AATACATTGA AATGCTAGGG
    2281  AATCGGGGTT TCATAAGCTA AATCTTTTCA AAAACATATT TTAGAAAGTA AACAACAAAA
    2341  GGAAATAAAA ATGAACAGAT CAAATAAATT AGCAA

//

LOCUS          33     2204 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      33
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             complement(join(500..724,940..1078,1346..1611,1666..1704))
                     /translation="MSSNTQNYEAFETTSIVYSSIMSAISLLIIVLNWKLSRLLISR
                     RHHPNIQKLRKTLIKGREAVKNGKLTPIETVILALEAGMNLEDVRLEFAHVATREEI
                     LESVFGKLNPKFLARLTASGRFMETGYANEKIFLDLWKSESYRELVNLAQRSLCPTQ
                     SFHPSYIADLIDPLAPPATTVCDTKPSDPAPAAKSKRKTIIDKMANAILNEERPLLC
                     SGNNDDAH"
ORIGIN
       1  CTTTCATGAG AAGAGTGTCT CGAATTGAAG ATATTCCAGT GAAAGAGGTT CTAACAGATC
      61  TTCTCCATCT TCATGTCAAC TATGAACTAT TGGATGTTGC CACGTATGCA CTGGAATTCA
     121  TGTCATCCAC TCAGTTGGAT TATATCAGAG ATCAACTATA TTTATATCTG GAGAAGATTC
     181  GACCAAGTGC AGTATCTCTA GTTGACTCTT TCCAGATTAG TGATATGCAA TTGAGATCTG
     241  TAAGTTTTTG GAATAATTTT TCTATGAAAA ACATGTTTTA GGTTCTTGGC CGCCGGGATG
     301  GTAATGTCTA CGAGAATCTC TTCAAGTGGG CAAAGTCGAG TCCCTTGAAC AAATCTGATG
     361  TTCTACCGTC AGTCGATAAG TACTTGATGC CTATGATGGA AAAGGCGAAA CTTTAAGTTT
     421  CTGTTCTCAT CTGGAATTCT TTATCACATT CTTTATACAT CCATTTCTCT ATTAATAAAA
     481  CAATAAAAAC AGTTCACAAC TAATGAGCAT CATCATTATT TCCCGAACAG AGCAGTGGTC
     541  TTTCCTCATT CAAAATCGCG TTGGCCATCT TGTCGATTAT CGTCTTTCTC TTTGACTTTG
     601  CAGCTGGTGC TGGATCCGAT GGTTTAGTAT CACAAACAGT TGTTGCTGGA GGTGCAAGTG
     661  GATCTATTAG ATCAGCAATA TACGATGGAT GGAAGGACTG AGTTGGGCAC AGGGATCGTT
     721  GAGCCTGTAA ATGTGGAATT TAACTTCTTT ATAGTGTAGT TTATGGAAAA CCCTGGAAGG
     781  TCTAATAGTA AGTCTATGAA GGTTTTTATT TTGAAAAGTG AGAGCTTGCA TATCTCCAAA
     841  AATTGTAGTT ATCAAAAAGT TGCCAACTAT CAAATTCTAG ATTTTTTTCA ACTAACAAAT
     901  TTTCCTAAAA ACCTTCTTGT ACTTAAATCG TCAACTCACC AAGTTAACAA GTTCCCGATA
     961  GCTCTCACTT TTCCACAAAT CCAAAAAGAT CTTCTCGTTT GCGTAACCAG TTTCCATGAA
    1021  TCTTCCACTT GCGGTTAGAC GGGCAAGGAA TTTCGGATTC AGCTTTCCAA ATACACTTCT
    1081  GGAAAATGGA ATGTGTTACT TGGAAATATT AAAACCTACA AAAATTTTAC TACGTTAAAA
    1141  TACGCGCCAG GCAGCATACC ACTGACGCAA TTCTAGTGCC TTGACGCGTT TCCAGTGCTT
    1201  TGGAGCAAAT CTAGTGCCTT GGGGCATTTT CTAGTGCCTT GACGAACTTC TAGTGCCTTG
    1261  ACGCAATTCC AGTGCCTTGA CGAATTTTTA GTGCACTGGC GCATTTCTAG TGCCCTGGCG
    1321  CATTTCTTGT GCATCAAAAC TCAACTCTAA AATCTCCTCT CGTGTTGCCA CGTGAGCAAA
    1381  CTCGAGCCGA ACATCTTCCA AATTCATTCC CGCTTCCAAT GCCAAAATAA CTGTCTCAAT
```

```
    1441  CGGCGTCAAC TTTCCATTCT TAACCGCTTC CCGTCCTTTT ATCAGTGTCT TACGAAGCTT
    1501  CTGTATATTC GGATGATGTC GTCGAGAAAT CAGAAGTCGA GATAGCTTCC AGTTTAAAAC
    1561  TATTATCAAG AGAGATATAG CAGACATTAT CGACGAGTAA ACTATTGAGG TCTGCAAGAA
    1621  AAATCGATAA ATTATCAGAA ACTCAAGAAA AATCAATAAA CTCACCGTTT CGAACGCCTC
    1681  ATAATTTTGT GTGTTTGAGC TCATTTGTTT TGTGAAAAAT TACCGGAAAA CTATGGTAAT
    1741  TTCATAGAAA TGATTGAAAT TTAAACGTTT GAACTGATAA ATAATTTTTT TAAAACAATT
    1801  ATTGCGAAAT ATAAGTTTAC ACGTGAATAT TTTTATTTAT TTTTCCCAAA CATGAATTTT
    1861  TAATTAAGAG ACCCGTGAAG AAAATCAATA ATTCACTGAA AAATCAATAT TTGCTGATCT
    1921  TTCCGTTTAG ATCCGTTTTC AGCTTTATCC ATTCCGGCGG GTAGAAGTCT TCCGGGTGCA
    1981  TTTCACCAGC AACCTGAGAG GATTACTGTA GTTTTTGGTT GCTGGCGCTT TTCTCGTGCC
    2041  TACCGTACCC CATCTTTTGA CTCCAAAATA TCCCTATAAT ATGTAGTTGC AGTTCGTTTT
    2101  GCCAAGTACG ACAACCACCA TCCGAATGTA GAAGATGGCG CTGGAAAATT GTGTAGTTTG
    2161  TAGTCATCCA GTTGTAGTTT GTAGCTTACC AGTTATTAGT GTGA

//

LOCUS          34     2270 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      34
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             complement(join(500..718,767..857,914..1187,1680..1770))
                     /translation="MHIYTYSFFIFGLLTITCTGCIPTIPPEDVGFVERIIDQLEAL
                     PDPPQSINNVLPAIQDFRNELVKANSFLISKTSEVAESPYQPFKETAAVFAGCQINQ
                     IKFTSPETGKKLGRTIIPPDLKCVKKSLFGCDVAHLSCKVPADSKYPLISGFAKIDK
                     ADWATTDLTATRADFKLDCDIFNRWTYISDIPGLEGRVYPTNEVSCLLVPIITEMNQ
                     VDFQEIHYDV"
ORIGIN
       1  AAAATCTTTA TAAAGAGACG AAACATGAGA GGACAACTGA AGATCAAGAT CATTGAGAGC
      61  ACCAGCTTCA AAATCATATG AAGTGATTAT ATTTAACAAT TTAAAGGAGG AGCAACGGAT
     121  TGAAATAATT GTTTGGAAAG ACTTAGCCCC GACAACTGAA TTTCCTAACG AAACTTCATA
     181  AAAGTCTCAA AAATTAATAT GTTTACACGG TACATCGAAA TTTTGGTAAA AAGGAGGAAA
     241  ATGCATGTGA ATTTTATACT TTTCGTCATG TATTTGACAT AAATAGTGAA ATAAAACTTG
     301  AGTTAATCCA AGCTACAACA CTTAAAAGTT AAGAAAAATA CAAGTGTTAG CTTAATTCAA
     361  AATCTTTTTT ATAAAAATTA ATTACATTTT CAATTTTGTA CGTTTATTGG AATGCCTTTT
     421  GGAAATCGAA TTTTTGGGAG CTAAAGCTAA ATAATTTGAA TCCGTCACTA CTCGTTTAAT
     481  TCTTTTAAAC AGTACTTTTC TAAACATCAT AATGTATTTC TTGAAAATCG ACTTGATTCA
     541  TTTCTGTGAT AATTGGCACC AAAAGACACG AAACTTCATT CGTTGGATAA ACTCTTCCTT
     601  CAAGTCCAGG AATGTCTGAT ATGTAGGTCC ATCGATTAAA AATATCACAG TCTAGCTTGA
     661  AATCTGCTCT TGTTGCAGTC AAATCTGTTG TTGCCCAATC AGCTTTATCG ATTTTTGCCT
     721  GAACATGTTG TTGTTTCACG TTAACGTCTA TTTGTTTGTA TCTTACAAAT CCAGAAATGA
     781  GTGGATACTT GGAATCAGCT GGAACTTTAC ATGATAAGTG GGCAACGTCA CACCCGAATA
     841  ATGACTTTTT GACGCACCTA TCAAGAAATT TTAAATAAAC GCAACATTTA CAGAAAAGAA
     901  AATAAAAACT AACTTGAGAT CTGGTGGAAT TATCGTCCTT CCCAACTTTT TCCCCGTTTC
     961  CGGAGATGTG AATTTAATTT GATTTATTTG ACATCCGGCA AATACTGCAG CGGTTTCTTT
    1021  GAATGGTTGA TAAGGTGATT CGGCTACTTC GGAGGTTTTA GAGATCAAAA AACTGTTAGC
    1081  TTTTACCAGT TCGTTTCTAA AATCTTGGAT GGCTGGTAGT ACATTGTTTA TTGACTGTGG
    1141  AGGGTCTGGC AGGGCTTCAA GTTGATCAAT AATTCGTTCA ACAAAACCTG AAAATATGAA
    1201  ACATGTGAAC AAATAATTCG ACATTTTATC AATATTTTTT TGTTTGTTTG AAAACTAAAA
    1261  ATGACAGTAA TAACCGAGCA TAACAAAGTT AAAAAAATTG TTGCATTTTT AAATAATTTT
    1321  AACAGTTTTT GAAATTTTCC AAAAAAAAGT GATACAACTG GTAATTGAAT TGCCGAAAAT
    1381  AATTTATTTC GTTGTCACAA CTAAACTTAA AAACTTATAG GGTTTTTTCT CATTTTTGTC
    1441  CACAGATAAA CGCATTGCTA GCAAAAGTAG AGTAATCCTT TGCGCGGAAA TTCAAATATA
    1501  ATTTGACCAC TTTTTTTATT TTTGCAAAAT ATAAAAAAAA TATTTTATCT AAAAGTTAGG
    1561  AAACTTGTAT TAACTTTTTA ATAAACAAAG TCATCAAACG AGGATACTTT TTCAAAATCT
    1621  GAGTTCAAAT AACGAATTTT CTTTTATCAC TATGTATACG TTTCCTAGTT TTTACCTACC
    1681  AACGTCCTCT GGTGGTATAG TAGGTATACA TCCGGTACAT GTGATAGTTA ATAATCCAAA
```

```
    1741  GATAAAAAAC GAATACGTGT AGATATGCAT AGACGGTCAA AATTTGAAAT CCGCCATGAG
    1801  GCAAGACCAC TACGATTTGG CGGCAACGCT TGAGTGATCT AAATTTGATT GACGACAAAC
    1861  GCCCTCATAA TTAAAAATAT AAGCGTTGTG ATTAATTATA ATCCTGAAGA TGATGAGAAA
    1921  TCGTTTCTTG CTTTTTCACA ATGAAATAGT TTTGAGTTTT GAATAGGTTT TTCAATAACA
    1981  AAACAATATC CTATAAAAAC AATTATTCAT TTTTTCAATA ATGAATGAAA GAAAGAAGAC
    2041  AAAAAGACAA AAAGACAAAA AAGCAATAAA CCAAGAAAAT AGTTGATATT AGAATTTCGA
    2101  ACAATATTAA ACATCAGAAA AGATCAGTTT GAAGTGAACA CTTTGGAAGA TGCAGTTCTT
    2161  TTTCAAGTAG ATTTAATTTC TAACAGGAAG GCAAATGAAA TGCAGAAGCG AAAAGTAGAG
    2221  GAAACATGGA CCGGAAATAT AAGTTGACAA TTGAAAAAAG CATAGGTTTC

//

LOCUS          35     3710 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      35
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS
complement(join(500..733,1416..1534,1626..1723,1786..1853,2597..2739,3183..3210)
)
                 /translation="MLTFQDLSLNVSGYAPMLFWSAISPLIRLVFLFFTFATWHTAH
                 AIKCYSCANDFIVWQWRHFFLKRNYDISTSDPECTSRDYISDLTQSCHSTCFVFYLN
                 GTNKETGHTKVLGVGRGCSASFLTDDQHLHLGLGAHSRPSHVGEYLPHDFDQFDITE
                 HWCFCATEKCNSEDCFSSPFGSREYASSYIAKRLQYSSSYSHNQWRYRNTGSLCTLS
                 FVFMALLLVLYTICL"
ORIGIN
       1  GGCGTGGTTG ACGATGCAAC TCGACAAAAA GTTCAAGATG CTCTCCTAGA AGCATTGGCT
      61  AAAATCAGCT CTGTAACAGA ACCATCAGTT CCAGAAGGAT CCATAATCTC ATCGCAGAAT
     121  GCAAATGAAG ACTCCAAGAA GTTGCAGTTC AACTCTGAAG TTTTAGAGTT AGCAAGTCAA
     181  ATTCTTGTGG AAAAGTATGG AGTTGAAATA CTAGATGCAG CAAACGTAGT GCTCCAGGAG
     241  CTTCAATCAG AAACTACCAC CGCTGAAAGC AAGTGACAAA GATTCAACTT TTTTGAATTA
     301  AAGTAGTTTG TCTGATGCTG TTAATGAATA GATTCACGAT TGTAAACTGT ATGTTGAATT
     361  TTTAAAAAAA GTAAAATTTT TAAATCGGCA GTTTGCAGTT TGCCAATTTG CCAGAACTTT
     421  TTGAAATCCT TTTTAACCTC AAAAAGTTAA AAGAAAAGCC ATTTGAAAAG TTTTTAATTC
     481  CGATTTATTT CAGTCAATAT TACAGACAAA TTGTGTACAA GACAAGTAAC AATGCCATAA
     541  ATACAAATGA CAAAGTACAT AAGGAACCCG TATTTCGATA TCTCCATTGA TTATGTGAAT
     601  AAGAAGAAGA GTACTGTAGT CGTTTGGCAA TGTACGAAGA CGCGTATTCT CGTGAGCCAA
     661  ATGGTGATGA AAAACAGTCT TCTGAGTTAC ATTTCTCCGT GGCACAAAAA CACCAATGTT
     721  CCGTTATGTC GAACTATAAT TATTGTTAAT TGAGTTTTCA TTTTCAAATT CAACTTTTAA
     781  ATATTTTTGA CAAAGTTTAC CTCTTTTTCA ACTTGAAAAA CTTAAAATTA AACAAAAGAT
     841  TTATGGAAGA ATACAGATAT TTATAGAAGT TTATAGGAGT TTTTTCATGT AGAAAATTTA
     901  CAGATAGACA TGTATAGAAA AGGTATAGGC ATTTATAAAA TTTTATAGAA AATTTATAGA
     961  CTTTTTTCAG AAGTTTACAG TTTTTTTTTA AATTTTCTTG AAAATTCTTA AAATTTATAG
    1021  AAAATGTATA GAAATTTATT GAAGTTTATA GAATTTTTTT CATATAGAAA TGTAGTAGAT
    1081  AATTATAGAA ATGTATAGAA AAGGTATAGA CATTTATAGA AATTTATAGA ATTTTTTAGA
    1141  AGTTGACAGA TTTTTTTCTG AAACATGTTT CAGAAAAGTT TTGCAGAAAA TTTATACAAT
    1201  TTTTTAGAAG TTTACAGAAT TTTTATTTAT TTTTTAAGCT ATTAATTAAA GTTACAGTAT
    1261  TGACCAATAT TTAATTCAGT TTTGAGTATG CGATAATCTT TCGAGTCTCT CAAATGAGTT
    1321  TTACAGCAGA ACATCAAAGA AAAACCAAGC ACGAATAGTC TATTGTGTAC TTCTTTTGCT
    1381  ACAAACTTTT CTGAATAATA AAAAAGTTGT GTTACCTGAT CGAAATCATG CGGTAGGTAT
    1441  TCTCCAACAT GAGACGGCCG GGAATGAGCG CCTAGCCCCA AATGAAGATG CTGATCATCT
    1501  GTTAGAAAAC TTGCAGAACA TCCTCGGCCA ACACCTGGAA TATTAAAAAA TTGTTTTCAC
    1561  AGCTCATGGA AAATGGAAAT TTCAAGAATT GTTTCGCAAA AAAATCGCAT AACAAATAAA
    1621  CGCACCGAGA ACTTTTGTGT GTCCAGTTTC CTTATTGGTA CCATTCAGGT AAAATACAAA
    1681  ACACGTTGAA TGGCAACTTT GTGTGAGATC GGAGATATAG TCTCTGAAAC CAGTGAAACT
    1741  TCTGATTAAC ATGATTAGAA AGAGGGTGGT CGCCACGTGA AATACCTGCT TGTACATTCA
    1801  GGATCACTCG TTGATATGTC GTAGTTTCGT TTTAGGAAGA AGTGACGCCA TTGCTGAATT
```

```
    1861  ATTCACTTTT TTTGAAACAT CTTACTAAGT TTTGTAGCAA AACCTATGTA TGAGTAATTT
    1921  TAAAAAATTA TCCCCTTTTC CAAAATTTTT TTTTTCAACA TAAATTTTCC TTTCAGTTTT
    1981  CGGTATTTAT GTTTTTTTTC TGAATCCTCG AATTTTCTTT TTTTTTCTCT TGCTCTCCTC
    2041  GGCCTAAGAC TAAGCCCGAG CCTAAGCCTA AGACTAAGCA TAAGCCTAAA AACAGGAAAA
    2101  TCCTAAGCCT AAATAGCTAA CGCTCGCCAC TGACGCCACG CTGACGCCTA ATCCTAACCC
    2161  TGAGCCAAAA CCTAAACCTA AGCCTAATCC TAAGACTATG ACCAAGCCTA AGCCTAATAA
    2221  CCCTAAATAG CGTCAGTGCT AACGCTCCCC CACTGACGCC AAGCCTATGC CTATGCCTAA
    2281  GCTTAGACCT AGAGCTAAGC CAAAGCTGAA GTCTAAGCCA AAGCCTATCC GCACTCCAAA
    2341  ACATGTTCAA TCGTGTTTAA AAATTACAGT GAATTTTTTT AAAAACAATT TTAGGCAATG
    2401  ATGGTTTTTG AGCAAAAAAA AAATTTGACA ATGAAAAAGT AAATGAGTAT TTAAAAAATG
    2461  ATAATGGCTT TTACTATTCT TAAGGTGAAA TATTACAGTT TGGGCCTGAC AAAATATGCT
    2521  TTGATCTATT CAGGTTTACG ACAAACATTA ATATTGCAGA AAATGTAATA TTTTATAAAA
    2581  TTTTTCAGAA ACTTACCCAG ACAATAAAAT CATTTGCACA TGAGTAACAT TTTATTGCGT
    2641  GAGCCGTATG CCACGTGGCA AACGTGAAAA ATAAAAAGAC AAGCCGGATC AGCGGAGATA
    2701  TTGCGCTCCA AAAGAGCATT GGAGCATAGC CAGAGACGTC TGAAAAAGAA AAAAAGAGAA
    2761  ATGCACTTTT TGTGAATTCG AGTGTTAATT TTGCAAAAAG TTTAGAGTGT TCGAACTTGA
    2821  TGTAGAATAA TACAATATTT TGGAAAAGCC TGAATAAAAG AAGGATGGAT AAGGCTAAGA
    2881  GAAATAAAAA CCTAGCGATT AAGTAAGGCG AGTAGTAGGA GTAGGAAGCA CTGGGAGAGA
    2941  GGGGGGGGGG CGGCTGAGAG AAGAGTATAA TTGCATTTTG GGCTCTTCTC ATCAACGCCT
    3001  TTTTGTCGAC TGCCTGTGAT ATGGATTCAG CGAAAATAAA AGAAACACGG ATTTCGTTTT
    3061  TGTAAATATT GAAAACTTAG ATAAATAACT TAAAATCACG ACTGTAGAGC ACCTGATACG
    3121  CGGAAATCGA AAAAAGGTAC GCCGTTTTTA GTAAGAGATT ATTCAGTGTT CAAATGATAG
    3181  ACTTAAGCTT AGATCTTGAA AAGTTAACAT TTTGCAGCAT TATATGTACA TTTCAAGCGT
    3241  TCATTATTAG TTTTTTTAAA ATTTTTATTT CTTTTTTTCA CTATTTGTGT GTGTGTGTGT
    3301  TTGTGTGTTT TTTTTGTGAA TTGTTGAATT ATCTGAAACT TTTTTTCTCT CTCTTTCTCC
    3361  AACTGCCCAG CCTATGCCTA AGCCTAAACC AAAGCCTGAG CCTACGCCTA AGCCTAAGCC
    3421  AAAACCAAAG CCAAAGCCAA AGCCAAAGCC AAAGCTTCAG CCTGAGTCTA AGCCTCAGCC
    3481  AAAGCCTAAG CCTAAGCCTA AGCCTAAGCC TAAAACTAAT ACTAAGACTA AGCCTGAGCC
    3541  TAAGCCTTGG CCTAATCGCG TTGATCGGCC AACAGGAACC CCCTTAAAAA TTTAAAAAAA
    3601  AAACCGGAAA ATCCTAAGCC TAAATAGCGT CAGTGGCGAG TGTAAGGCTA CGCCCAAGCC
    3661  AAGCTTGAGG CAACGCCTAA GCCTAAGCTT GAGCCTACGC CTAAGCCTAA

//

LOCUS          36     2290 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      36
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..595,668..910,1368..1601,1647..1790)
                     /translation="MPTCMKTNSQKVIFTKNGAVATRKIEDPTSFTMSYNCVNNYLA
                     DASCDPQRMTVTCCQSENGRNFVSFAISEVSRRTATELVDSKYFEAAFNSADNRKSL
                     RNACEKAKRFYMEHILYHNPFPPKQRFSCVIEKSKKTKKEVSNKRSDEFLVRTRRTT
                     SSISVLEGTMSLTSMSSHSQLSTTTPSTPINLKQASPHPSHVPNDSQMVIQKAPTPA
                     PAKTPKKGASDEQRLMNTFNTSEY"
ORIGIN
       1  AGACAAGTTA TTAGAAGTTA TTACACTACA AAAACTATCA TAAAATACAT TCGTGTAGTT
      61  TGAGGGGAAA GGAAGTATAT CAAAACATGG TAAGTTAGTT ATGGATTGAA CCCCCCGACA
     121  ACCTAAATAT TCATGACTTT TACACAACTG TTTTTGTTGC CTCAGCTCCT GCTTAAGTTT
     181  TAAAAAAACT TGGATTGAAC TGATTATGAG CATTGTAGTC ACCAACGAAG ATCTCATTAG
     241  ATGTTGCAGT CTCTACTGAA GCTGCTGAAA AATTTAGATT TATTTAGTTT TAGGATATTT
     301  TAGGACATAT TTTAGGAAAT TGACAAAAAC ATATTTTCTA CGAGCTTTAT ACTTCAATTG
     361  CGTTTTTTTT TTTGGAAAAT TTGGCAATTT ACCAAAAATT TGAAAAATAT TAAATTATCC
     421  TATTTAAATT TTCACGTTTT GCTCAACCGC TTCTGGTTCG TTTTTTAAAA ATTTATTTCT
     481  GTACTTACGA TTTCTAAAAA TGCCAACTTG TATGAAGACT AATTCACAAA AAGTGATTTT
     541  TACAAAAAAT GGAGCTGTGG CAACTCGAAA AATTGAAGAT CCTACTTCTT TCACTGTGAG
     601  TTAGACGGAT TTTGAAACGT TTATTTTGCA AACTGAAAAT TCAAAAATCA AATGGATTCT
```

```
     661  TGTTCAGATG TCGTATAATT GTGTGAACAA TTATCTTGCT GATGCATCTT GTGATCCTCA
     721  AAGAATGACT GTAACATGTT GCCAATCGGA AAATGGAAGG AATTTTGTTT CATTTGCAAT
     781  CAGTGAAGTT TCTCGTCGAA CTGCTACAGA ATTAGTCGAT TCAAAATATT TTGAAGCTGC
     841  GTTCAATTCG GCTGATAATC GGAAATCGCT TCGTAATGCT TGTGAAAAAG CAAAACGATT
     901  TTATATGGAG GTACTTCGAT CTTTTCAAGA TTCAAAGAAC AAAGATCAAA GAAAACAATT
     961  TTGTAATAAT TATCGTATGT CTAATAATTA AAATTGGCAA AAAAGGTCAT CTAGATTATA
    1021  CCTGCACGGT GTAGTCAGTT TGGAATTTTA TGTGAGTACA GTAAATAGGT TCGGTGCGAG
    1081  ATATCTTACA TGTGTACATA CTCATATAGT TCCCCAGCTG ATAGACCAGA AATTGCGAGA
    1141  ACTTTATATC TTAAAGATGT GCCATGCTTT TCTATTGGTG GCTCATAATA AAGTCGCTCC
    1201  TTTCTAAAAT AAAATACCCG TGCCTCCTGG AAGTCAATTC TTCTGAATAC TTAAGCAACG
    1261  TGCTACTTTG ATGTGAATAT CTTTTGTAAA AAATAATATT GAGAATGAAT AATAGTCTAC
    1321  AGATTAGCTT CCAATTCATG CACGTATCTC TAAATTCTAA TTTCCAGCAC ATTCTCTACC
    1381  ACAATCCATT TCCCCCGAAA CAACGTTTCT CGTGTGTTAT TGAAAAATCA AAGAAGACAA
    1441  AGAAAGAGGT ATCAAATAAA AGATCTGATG AATTCTTGGT CAGAACCAGA AGAACTACAT
    1501  CAAGTATATC AGTATTGGAA GGTACAATGT CTTTGACTTC AATGTCTTCA CATAGTCAAC
    1561  TCTCGACGAC AACGCCATCG ACTCCGATTA ACTTGAAACA GGTGGGTCAC TGTAGTATCA
    1621  AATGAAGTTC CTATATTTGA ATTCAGGCAA GCCCTCATCC TAGTCACGTT CCGAACGATT
    1681  CTCAAATGGT TATTCAAAAA GCTCCAACTC CAGCTCCAGC AAAAACACCA AAGAAAGGAG
    1741  CATCAGATGA GCAACGACTT ATGAACACAT TTAACACTTC TGAATACTAG GACCATTTTT
    1801  ATTAGAAGAA AAATAAATTA TGAAGAGATG GAATGAATGA CATTGCACAA GCAGACTAAT
    1861  CACAGAGTTA ACAGGGGGAA ATGGTATAAA AATCTTTGGA AATTTGATAG AAAAGTTTAT
    1921  TAGATGCACT AAAAATTTCT AGCAATGCAG ATTTGATCCC AAGTTGTTGA AAATAAATTA
    1981  CACTATGAGT ATCCCGTACA GTTCGAAAAG AAAATAGAAA ATGACTAATG TTTGGAAAAA
    2041  AAATATTCTA GAAGCTACAA TTCTGAAATA GTTCGATCTT TTTCGTTTTC TTTTTGCTCT
    2101  TCTGTAAAGA AATGACGTTT TCTTCGTTTT GCTGGTAACT CCATCGCGGC TACTTGTAGA
    2161  GGCCATCCAC CACCGATGAC ACCAGCTTGA ACAGAAGCTC CTGTTACGAC GGCAAGTTCT
    2221  GGGTCGACAC CGTAATTGGC AGACTTTCTG AAAAAGTTAA TCTAATTGTT ATCCAGATGC
    2281  GAAATTTTCA

//

LOCUS          37     4237 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      37
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..635,1036..1118,1398..1625,2417..2703,3674..3737)
                     /translation="MASQPELIMVDEQVVAYEVEIDSFDVKYDEEEHDGQGTQDEPF
                     SHDEPSTSGYHHHYQFPNDVDPNDVYLFDEETDQIHQLDPNQLKNNEEIDDVEYIDQ
                     SVPSTSSMMTSLPSTVAPVQPNTYYRRKSGGPTATGNEKPNYRPLAFQTTYVNSVTQ
                     YGRRVVTKRIHESIDPTSNSINEESEILQKIQMEIEILTKENDLLKKDYEAVTKRVD
                     SVNMVFNEAKARQKMAMEEKRQLQRDLTRAKCERTKLSKQLGHTQHRYQYQ"
ORIGIN
       1  CTATGGCATC CACATGAAAA AAGCCGAATG ATTACTGCTG GATGGGATGG TCTCATTAAA
      61  ATGTGGAATT AATTGTTGAT TTTTCCCTTT TTTCCCAAAA AAAAATTATT ATTTTCCCAA
     121  ATTTTGGCTC TTTTTCTCAC GAGTTTTTGT CCCATCTCTC ACATTTCCCT GTTCATTAAA
     181  AAAATTTTTT TTTTCGTAAT AATGTTTTAC TTTCTCCTGT TTCAATTATT TATTTTGTAT
     241  CTTTGTTCAT TTATTTGTTT TTTCTTCCAA TTTTAAACCC TATTCTTACA GAACATTTCC
     301  GACACACGGA AATTGATAAA ATTATTTTTA TTTTTTAGCC TGAAATTTCA GTTTACCCAA
     361  TTAATGGTTA TCGACCAGAT TTTATTTTAA TTTTTTAAAG GAAAATTAGT TTTTTTTTCT
     421  ATGAAAAAAA GTTATATATT TTTTTGGAAA TTTATAACTC TTAAAAAAAT TCAATTTTTT
     481  CAAATAAATA AAATTTCAGA TGGCTTCTCA ACCGGAGCTC ATAATGGTTG ACGAGCAAGT
     541  CGTCGCTTAT GAAGTAGAAA TTGATAGTTT TGATGTAAAA TATGATGAAG AGGAACATGA
     601  TGGTCAAGGG ACACAAGATG AACCATTTTC TCATGGTACG GAACAGTTTT ACGCTGAAAA
     661  ATTCCAGAAT TCCAAAAAAT GAAACCTAAA ATAGTGATAA AAAGGCGTTT TGAATATTAA
     721  ATTGAAGAAA AAAATCAGCA AAAATTGTTC AAAATCAAGA ATTTTAACGG AAAAGTGTAA
     781  AATCTTCTCC ACGGGGAGTA CACATGCTTC GTAAATCGAC ATATGGTCAA TTTTAAAGTT
```

```
 841  TTGAAAATTG AAATGCCGGC AAAAAATCTT TTCTTGTTTT TTTTTCGCAA AAAATTCAAT
 901  TTTCGAAAAA ATAATTATAG AAAATTGCAT TTTTTGACCG AAAAGTCAAT AAAAATAACA
 961  GAAAAAATCG ATAAACCGTT GAAAAATTTT TTTTTAATTC AAAAATTCAG AAATTCTTAA
1021  AATTCAAATT TCCAGATGAG CCAAGCACCA GCGGTTATCA CCATCACTAC CAATTTCCCA
1081  ATGACGTGGA TCCAAATGAT GTTTATTTAT TCGATGAGGT ATCAATTATC CGAAATTTGG
1141  CGATTTTTGA GCCAAAACTA CGGTACCCGG TCTCGACACG ACAATTTTTG TTAAATTAAA
1201  AAAGGTGTGC GCCTTTGAAG GTTACTGTAG TTTCGAACTT TTGCTGATTT TTCATATTTT
1261  TTCGTTGAAA ACAAAAGTAT TTATTTGTTG AAAATCAGAA AATATTATCT TCGCGTCGAG
1321  ACCTATTACC ATTCTATTTT TGCCGCAAAA AACAAAATTT CCTTTAAAAA AAAGCTAATT
1381  TTTCCAAGTT TTTCCAGGAA ACTGATCAAA TTCATCAGCT CGACCCGAAT CAACTCAAAA
1441  ATAATGAAGA AATTGACGAT GTCGAATATA TTGATCAATC TGTGCCTTCC ACGTCATCAA
1501  TGATGACGTC ACTGCCGTCA ACGGTGGCTC CAGTTCAGCC AAATACGTAT TACAGACGGA
1561  AATCTGGAGG CCCAACTGCA ACTGGAAATG AAAAACCGAA TTATAGGCCG TTGGCGTTCC
1621  AAACGGTTCG TAAAATAAAA AAAATGTCCA TGTGTCGATT TACGCAGCTT CTCGTGTACT
1681  CCTCGAGGAA GAAATTCGCC AGTTTTCTGA ATTTTATCAT TAAAATTTCA TGATCTCCTC
1741  GGGGAGTACA CGAGAAGCTG CGTAAATCGA CATGGAGACC AAATATTCGA TTTATCGAGG
1801  TTTTTTCCTG AGAAAATCGA AGTAAAAATG TTTTAAAAAA TGAGAGAAAA CTATAAATTT
1861  ATCGTAAAAG TATATATTTC AATGAGAAAA AGTTGTATTT TCCAAAAAAA TAAAAATTAT
1921  AGGAAATTAT GAAAAAAATT TCGTTTTTCC AAAAAAATTT TTAAAAGGAG AAAACTAACT
1981  AGTAAGTTTA TCAAAAAAAG TTTATTTCGG TTAAAAAAAA TTATAAAAGA CAATTTTTAA
2041  AAAATAAATA AATTATTGGA AAATACAAAA AATTTTGGAT TTTTCAAAAA AAAAAGCCGA
2101  AAAATTGAAA AAAAAATTGA AGAAAATAGG AAAAACTATA AATGTATCAG AAATAATGAG
2161  GATTTTTATT TACAAATAAG GAAATTATAG AAAATTATGA AAAAAAATCG ATTTTCAATT
2221  AAAAATAAAA GTCTAGAAAA CTATAAATTC ATAAAAAAAA GAAATATATT TCACAATCAG
2281  TGAGAGAATG TTTTTTTTTT AAATTTTAAT TTAAAAAAAG TTTTTGAAAT TATAAAAAAA
2341  TTTGAATTTC ATCTTAAAAA TAGTTTTAGA AAAATTGAAA AAAATCGGTG TTAATAAATT
2401  TAAAAAAAAA TTGCAGACAT ATGTGAATTC TGTGACCCAA TATGGCCGCC GAGTAGTGAC
2461  AAAACGAATT CACGAGTCAA TTGATCCAAC ATCAAATTCA ATTAATGAAG AATCGGAAAT
2521  TCTGCAAAAA ATTCAAATGG AAATTGAAAT TTTGACGAAA GAGAATGATT TGTTGAAAAA
2581  GGATTATGAG GCAGTCACAA AACGAGTTGA TTCAGTGAAT ATGGTATTTA ATGAGGCAAA
2641  AGCACGTCAG AAAATGGCAA TGGAAGAGAA ACGGCAATTA CAGAGAGATT TGACTAGAGC
2701  AAAGTATGAA AGATTGTAGA ACTTATTGAG AAAATTGCGA GAAAAACGAG CTATTTGGTC
2761  CAGAAAAGTG TTGAAAAAAT GTTTAAAAAT GACTTGGAAA TCCGAGAATT CATGGAAATT
2821  GCGATGAAGA TATGCAGTTT TTTGTTGAAA ATTGAACCTA AATTTAAAAA AAAATTAATG
2881  ATTTTTTCAC TTTTGGAACA AATTTTTGAT GACAGATTTT CCTCACTTTG AACAAAAAAA
2941  CAAACTTTTT CAAAATGATT TTTCTTCAAT TTTCTGGAAA ATTGGGGCAA AACGAGAAAT
3001  TTCTTTAGAA AATTTTAAAC ATAGATTTTT TTCAAATTTG AATTTGTATT TATTTTTTAA
3061  TTGACTTTTG AAAGAAATTT CAAATTATTT TGCTAAAAAT TTGAACAAAT TTCTTTCTTT
3121  TGTCTTCAAT TTTTGGTGTT TTTTCTAAAT AATTATAGGA TTTCTCAATT TTGAAAAAAA
3181  AAATTGTGTT TAAAAATTAA ATTAATTGTT TTAAATTTTG AAAAAAAAAA TTTAAAATCT
3241  GAAAAAAAAA GATTTTCCAA CAAGAAAATT TGAAACAAAA ATTGATAATT TTTTCATTTA
3301  AAATAATTTA AAAACTTGTT TCGTGCTTTT TCCTTAAAAA TTTATTTTTT TCAATTTTAA
3361  GAAAAAAACT TTATTATTTT TCTCCAAATT TTACATATAT ATGTTTCAAT CTAATTTTTT
3421  TTTTCAAATA ATATTAATTT ATTCAAAATT ATTCAATCAA TTGGAATTTC TTTTGCAAAA
3481  TTTTTAGAAA AATCAAAAAA AAAATCTAGA AAAGCATATA ATATTTTTTT AATCAGCATA
3541  TGTGATTAAA AAAATTTTTT TTTGGTTTTT AAGCAAAAAA TCTGGAAAAA AACCCTTCTT
3601  GGCTTTTTGT AATCTGAATT TTTGTGTTTT TTTTTCATTT TTTTTTTTTA AATCCTAATA
3661  AATGTATTTC CAGATGTGAA CGAACGAAAC TATCGAAGCA ACTGGGACAT ACACAGCATC
3721  GTTATCAATA TCAATGATTC TGATTTTGGA TTTAATTTCG ATTTTTTTCT TTTTCTTTCA
3781  AAAACTGCTA ATTTTTTTTC CATTTGATCT CAAATTATCA TATCCATCTA ATCATTTCTA
3841  TGGGTTATTT TGTTAATTTT CTCATCGTTT TCTTGAAGTT TCTGTTTATT AACAAAATGT
3901  TTTCTAAAAT CTTTTTTATA TAATTTGTTT TGAAAATTTA CCAGTTTCCT TGAACTGTAA
3961  TTATTTCTTA TTTTGTGAAA ATTGGTTTTT TTTAAAGATT TTATAAAAAC TGCCCAAAGT
4021  TCAGATGCTC TCTGCAGTCA GATTTCCATT TCTTGAATCT TCTCTTCGAT TCTCAGCTGC
4081  CGCTTTCTCA ACAGTACCTG ATTTTAGTAA GGATTTGGTG CTTTCTGGCT GAAAAGCTGG
4141  ATTTTGTGTC ATTCGTTTTG AATTTACCAT TCAATTTGAC AATTTAATTA TTTTTTTCTA
4201  ATTTTTAGTT AAAAATCTGA AAATCGGTAA AATATAT
```

```
//

LOCUS          38     2187 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      38
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..574,623..859,963..1241,1289..1401,1570..1687)
                     /translation="METITAFSSTRCAVTNAVYIVNSDWIIMKLPKEFKEITGLQVW
                     RERFSRLLPPQDICTEQESLNWIQDCINHCIADFKKAAHCLRIANQHIPFESSESAF
                     RKVENSATRTLFQNVENPEWLFVPTAQCKIMLPVSFLVEGDAIEVFVAQVRNQYEPR
                     PLYGTHYHYEPEDHSEDFEAFRSIVFMWIRKWNPVDDDQMRKVANKLKKFCNDSPLP
                     KLFNQTYSNQNDTDGTLNSPDQIDRNFSIQLEMLRNKLFVDHVIRFDILGHNPFNTV
                     TV"
ORIGIN
       1  TATCTCAAGT GAACTTGAAC TGAGAGAAAG GGAGAGGGAT GGGGGTTAAT TACTGCAAAG
      61  ACTTTTTTCT GCTTATCAAT CACTAGGAGT GAGTAAACAC TGAAATTGAT CTACAGCTCA
     121  CAAGAACAGA GCAAAAAATA GGACAACACC TTCTTACCAA CCTAATATTT CCAAGTATAA
     181  ATTATCCTAT TTCATTCTGT CGAGAAAAAT AGTTAAAAAG TAGAAAATTT GTGGAAAATA
     241  TTCATAACAT CAAGTAAACA GAGTTGTAAA AACCTCAATA TACGCACGTG TTGTGCCTTT
     301  GTCGTACATT CCTCGTAGGT TCCTGGGAAT ATACGACTAA CGCCCAATAT TTGGCTTGAG
     361  AGGCGCACCA TGCGGGGAAT GTACGAGGAA AATGCGAGGA ATGTACGATG CGAATTGCCG
     421  AGCTAAACGT GCTTTTTAAC AATAACCTCA CTGGCGCTGA TCTACCTTTA ATTTAGCATA
     481  ACGTAGGGTT TGCAGGTAAA TGGAAACGAT TACGGCTTTT TCAAGCACTA GATGTGCCGT
     541  CACTAATGCT GTTTACATCG TTAACAGCGA TTGGGTATTT TGCTTTGAAA ACGTTGTTGA
     601  TTTAATCTTG TTAATTTTTT AGATAATCAT GAAATTACCA AAAGAGTTTA AGGAGATTAC
     661  GGGATTGCAA GTTTGGAGAG AACGGTTTTC CCGATTATTG CCACCACAAG ATATTTGTAC
     721  TGAACAAGAG TCTTTGAACT GGATTCAGGA TTGTATAAAT CATTGCATTG CCGATTTTAA
     781  AAAAGCTGCG CATTGCTTAA GAATTGCAAA TCAACATATA CCTTTTGAAT CAAGTGAAAG
     841  CGCATTCAGG AAAGTTGAGG TTTGTCTAAA AGTTTTTGTC CAAATTTTTA ATTAAGAATT
     901  TGGCTTTAAT TTGAACTTAA TTTTGTCAAG ACCGGTTACT ACAAATCAGT TTTTAATTAC
     961  AGAATTCTGC AACGAGAACC TTGTTTCAAA ATGTAGAAAA TCCAGAATGG CTCTTTGTTC
    1021  CCACTGCGCA ATGTAAAATC ATGCTACCAG TATCTTTTCT TGTGGAAGGT GATGCAATAG
    1081  AGGTTTTTGT TGCTCAAGTC CGGAACCAAT ATGAGCCACG CCCGTTGTAC GGAACTCACT
    1141  ATCATTATGA GCCCGAAGAC CATTCTGAAG ATTTTGAGGC GTTCAGATCG ATTGTTTTTA
    1201  TGTGGATCAG AAAGTGGAAC CCAGTCGATG ATGATCAAAT GGTAGGGTGA TAAGTTCATA
    1261  TTTCAAAGTA TGCTTTTCAT TGTTTCAGCG TAAAGTTGCC AATAAACTCA AGAAATTCTG
    1321  CAATGATTCG CCATTGCCAA AACTGTTCAA TCAGACGTAT AGCAATCAAA ACGACACAGA
    1381  TGGAACATTA AATTCTCCAG AGTTTGTATA TTTGAATTTA GAGCATATTT TGTTTGTTAA
    1441  ACTGTATTTA TTTAAAAGTT CTAAAATTGA TAAGTGGTAA ATAAAATAAT CAGAGGGTAA
    1501  AAAATGTTAA GAAACGAGTG GTTTTTAAGC TACTCTTAAA ACAAAACGGT TAAGCTTACA
    1561  TTTTTACAGC CAAATTGATC GCAACTTCTC AATACAACTG GAAATGCTGC GAAACAAGCT
    1621  GTTCGTCGAT CATGTGATAA GATTTGACAT TCTCGGGCAC AATCCATTTA ATACTGTGAC
    1681  TGTTTAATTT TTCTCATCGA AATTTCTAAG TGGCTCCACG ATGAGCGAGA AGAAGACACT
    1741  AATTGTACAC CATCACTTTT CTATCTTTCT TTTTTACACT TAAACACTTG ATTCTTTCTT
    1801  AAAAATTATA TGTTTTTCTC ACTTCTTAAT ATCAAAATTT CATGATTGGT CGGCAATCGA
    1861  CAGCTTTCTT TAAAAAAAAG AAATTAGCAA TCTTTCAGTT TTTTTTTTTA AAGAAACCCC
    1921  TTTAAAACAT TTACTCTACA TAAATTGTTA GCAACCCCTT CCCTGACAAT TGGCGATCCC
    1981  GTCCAAAAGT TCAAGGTTCC TGTTTTCATT TTTTTTTTAA AGAATATATT CATTACCAGA
    2041  TTTTCACTAA TAACTTGCAG ATTGATTTAT AAAATAGAAT TATTAAACAA CTTCAATCTT
    2101  TGAGAGAATT CAGTAATACG CAAATGTTTG GTTTGATTTT TTTATTACAA ATACATTGTG
    2161  GAACTGAACT AGTTCATTGG TTTCAAA

//

LOCUS          39     3092 bp     DNA    linear PLN 25-FEB-2004
```

```
ACCESSION      39
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             complement(join(500..715,1063..1540,1817..1917,2479..2592))
                     /translation="MLVVIFCFLNILFSVTSFDLEDDDDSWKELTFSIRTDGNGTIE
                     SAHIQWTDRFDDSFAYAYTNENRTDLKFLYDKPRPELFVSKNFIVDAFTQFMDFFQN
                     RTKISRISVTLPSLTDDLKKITKRKNSVKTKPAIDNFRFNSYDSLLYNELLRLSTGI
                     PKQLKISFFYYPERTEKVLDGTILNHPKVREITELDLRTVESDINDAQLHQLECESI
                     SLRSSTVTNAGINRILKEWQDGLRPTRYLKLVNDDLQKAEIMDGVSKRTLQRGVWEI
                     KNRFNASATVTVSRVPTNLYDSKPTFEFAVS"
ORIGIN
       1  TCATTGAAAA ACTCCTTCTC ACTCATTTTC CGCGCATTTT CCGCCTTCTA TTTTCCCGTA
      61  AGACTGCCGA AGAGAGACCC CCGCGGTTCC CACCGTCAAA TGACAGCCTC AAATGACCAT
     121  CATCACAGCC GAGTAATGAG TTCTCTTCTC GTGTGTTCGC TCTTTACCAA TCGATCACAT
     181  TTGACCATAT GTATGTGTGC AACTACGATA GTTTAGGTGA AGGTGCCACC GGTATTTCGA
     241  ATGAGACATG TCTTTTTTTA AGTAAAATGT TCTGAAATAA TTTAATATGG GGTTTAAGCA
     301  ACAAAGTTTC AAAATTGAGA TTTTTGAGCT CCAACGGATG CTCAGAGCAA AACAAGCGGC
     361  GTGCTATCGG TGCTCTCGGA ACATGTGTTA TGACAGGCTA CTTATCAAGT TTATTTCATT
     421  AGACAAGCTA TATGCATTTA ATGGTTCAAT AGATAAATAA AAAAATATAA CATTGATATT
     481  TATTTTTTAC CATGATTTTT TAAGAAACCG CAAACTCAAA CGTCGGCTTT GAATCATATA
     541  GATTAGTCGG CACACGGCTT ACGGTTACAG TAGCCGAGGC GTTGAATCGA TTCTTAATTT
     601  CCCAAACGCC ACGTTGCAAT GTTCGTTTCG AGACTCCATC CATAATTTCA GCCTTTTGTA
     661  AATCATCATT CACTAGTTTC AAATAACGAG TCGGTCGTAA TCCATCTTGC CATTCCTTAA
     721  AATAGAAAAA GTTCGTATGA TTGCATGTTT TAACAGGGGC AACGATCGCA TGAAGATGTT
     781  ATCGTGATAC TTTTTAAAAA GTCACAGAAA TTTTGATGAT TTATTGATAT TACATCAAAA
     841  AGAATGACAC AATTTGCAGA AATTCAGAAA TATCACAGTA CATTAGAGAA ATTGGGAAAT
     901  AAAGAAAAAT GGTAGCCTAC ATGATAACAT CTCTAATTAC ATTTTTTCAT TCATATTCGC
     961  AAATAGTTCA AAAGATTCAA TTCATGAAGT GGTGAAGAAA AATCATGTGA TTGAAATCAA
    1021  AATACAACAT TTTTATTTGA CAGAGTCTAG ATACATGTTT ACCTTTAAAA TTCGGTTAAT
    1081  TCCTGCATTG GTTACCGTGG AAGATCTCAG AGAAATAGAT TCACATTCCA ATTGATGAAG
    1141  TTGAGCATCA TTAATATCAC TTTCAACAGT ACGAAGATCC AGTTCAGTGA TTTCTCGAAC
    1201  CTTGGGATGA TTCAAAATTG TACCGTCCAG AACTTTCTCA GTCCGCTCAG GATAGTAGAA
    1261  AAATGAAATT TTCAACTGTT TCGGGATTCC GGTAGATAAC CTCAAAAGTT CGTTGTACAA
    1321  CAATGAATCA TACGAGTTGA AACGAAAATT ATCAATTGCT GGTTTTGTTT TAACGGAATT
    1381  TTTTCGTTTT GTGATCTTCT TGAGATCATC TGTCAATGAT GGAAGCGTCA CACTAATTCT
    1441  GGAAATTTTT GTTCGATTCT GGAAAAAATC CATGAATTGA GTGAATGCGT CCACGATGAA
    1501  GTTTTTGCTG ACGAATAGTT CAGGGCGTGG CTTGTCATAT CTGAAATTAT ATTAATTTTC
    1561  CAACTTGAAA TTTTGGCGAC CACTGCAGTA GCAGTTGGTG TTTTTGACTT CATGCAAATA
    1621  TCATGCCATT AGAAAATATC TAGAATTTCC ACAATTTACC AACAACGCCT AGAATTATAA
    1681  TTTTTGACTC ATACTTCTCA CAATCAGAAC AATTTTCTTT TCGAGTCTAG GAAAATTGTA
    1741  GGTGTTTGGA TAATACAATC TCTGCTCATT TATGACTAAC ACGTCGTTGC TCAAATTCTT
    1801  ACAAAAACAT TCTTACAAAA ACTTCAAATC TGTCCGATTT TCGTTCGTGT AGGCATAGGC
    1861  AAATGAGTCA TCAAAACGAT CAGTCCATTG GATATGAGCA CTTTCGATAG TACCATTCTG
    1921  AATAATATAA CAAGTAAGAA TAAACACTTA TTTATTAAAT ACAAATATTT TAAGAAATCT
    1981  TGCTGATTAG CGATTCCGTT TGTAGTATGT TTAAGAATTC AACAGATCAA AAACAAAATT
    2041  AGCGCGAAGT GTTAGGAAAG TAGCCTCGCA AATAAATATG TAGATTGTCA TCTACGTTTG
    2101  ATGGAAAAAC GTGAAAATGG CACGAACGAT TGTAAGATTA TTAAACCGCT TTTTGAGGAT
    2161  GAAATCAACT TTCTGAAAAC AGAAGGTTTT TGGTCGAATA CTAAAATAAC TTCATGAGCC
    2221  TAAATGTGCA CCTACTGCCG AAACCCAAGT TAGGCAAAAA CTCTTTTGGC AAAAACTGGG
    2281  TAAGAAGCTG GAAAGATCTA AACTATGCTT ATTAGTTGTG TCGGAATTCA GAATAGGTTT
    2341  ACGCATAAAT TTAATAAAAA GGTTGTAGGT AAGCACGCCG TCAACACCGG AAAGAAAAAC
    2401  TAGTTCCGAT TTTTTTCATT TAGTGTGAGC ATAAAAGTTC TAAGGATCTA GGTGAAGGAA
    2461  ATGGTTTTTC ACACTTACAC CATCTGTTCT TATTGAGAAA GTCAGTTCTT TCCATGAATC
    2521  ATCGTCGTCC TCAAGATCAA AAGATGTCAC CGAAAATAGA ATATTCAAGA AACAAAATAT
    2581  GACCACCAAC ATTTTCCTTT CTTTGTTGCA AATGTTCGTT TCCTTTTATT CCTGTTTTTT
    2641  TGAAGCATGC TCATGACCTC TAGAGCACTA GAAATGAAAT ATGATATTCT ACTTTAAAAT
```

```
    2701  TTTTATTTTT AAATTTTCAG ATATCAATTT GGGACTTTTG AAACAAATGA AGGATACTTG
    2761  AGAAGTGGCA TGTACATATG ATTTGAAAAA AGGGAAATGT TTTAGAAGAT TTTGATGTAC
    2821  TCTATTAAAT TCTTCCTTTC TAGGTACCTG ACAAACTGGC GTCGCCAGTT TTGTAGGCAC
    2881  GAAGGTTGAC ATTTTCTTAC CTCCATGGAA AACTTAGTTT ATATTTTCCC ATTCAAAAGA
    2941  AAATAATAGA ATATATCAAA ATCCATTACC AATACGATGC GAATATAGGA AACAATTGGA
    3001  AGTAGAAATG GAAAAAATAG GAAATGTTGA ATGGGATCAC AATGTAATGA AATCACAAAA
    3061  ACATAGTAAA AATGGAAGAA AATTGATTAT CT

//

LOCUS          40     2236 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      40
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(500..802,878..1072,1260..1490,1539..1736)
                     /translation="MLETALESVSTCCQLTGLGVFYTYILSFVLAFDLSLTAALYIA
                     SLGAVLVFLSANYQFDNKKFEPKTFIVYGLFAQSFCSASLATCFFAESGDESKSDKS
                     LMTQIIPKISIILFFLAFYLTHASSMFYIKILCRSGNCTDSASTLSWVVTIILIGFK
                     VVDYEDFSKSLSITMIAFGFLYLSLAFVINDRLPPAKSLFEQEIAQSFGFQSVKNIL
                     KYEDHRTSETDGILTFYIKNLTSDKIAIDISFSNPGDQLDGHFSRKYIRGKGHVKLN
                     VKRGTIMEKGDRMFIFYARSIKDLESINDSGFLTIYF"
ORIGIN
       1  CAATTCTCGC GTTCGGAATT CTCTTCCTAA ATCTGAACAT TTGTGCCATA ATAATTTTCT
      61  GATCGGGTTC CGAAATACTG TGTTTATCAA TAACTTGGAA TCGAGTACTA CTGCTTGTAG
     121  AACGGCGGCC ACCTGATTCA AGATCCTAAA AATAAAATTT TCCATCAGTT TTCAAAAGGT
     181  TTCTCTCACT TTATTCACAG CCATCGACAT TTTTCTGAAT GTGTTCAGCA GTTTTCCACG
     241  TGGCACTGCA AATTAATTGC TTTTTTTTTA AGAGGAGGCT TTAAATGAAC AAACCATCTT
     301  TCTTCTGACT CCCCTGGTCG GCTAGTTTTG GAGTGGCACT GAAATTGTAT CAAATTGAGG
     361  TTTTTTGTAG TTTTCAGAAG CATGGATAAT CCTTTCGAAA AATTGAATAA TACTAAGAAA
     421  TCATATAAAT ACAATTTTAG TCTCATAAAG TCAGAATCTA GATATTTTAT TTATTTTATT
     481  TTTTCAATTC TCCATAAAAA TGCTAGAAAC TGCTTTGGAA AGTGTCTCAA CATGTTGTCA
     541  ATTAACTGGT CTTGGTGTAT TTTACACTTA TATTTTGTCA TTTGTATTGG CATTTGATCT
     601  TTCTTTGACC GCAGCTTTAT ATATTGCTTC ACTTGGTGCT GTTCTTGTTT TTCTTTCGGC
     661  GAATTATCAA TTTGATAATA AAAAATTCGA GCCTAAAACT TTCATAGTTT ATGGATTATT
     721  TGCGCAGAGC TTTTGTTCAG CTTCTTTGGC TACATGCTTT TTTGCAGAAT CAGGAGATGA
     781  ATCGAAAAGT GATAAATCAT TGGTAGGTGT AAATTGGAAG ACGAAATTTA AAATTGTCCT
     841  GTTGTGAACT ACAGGAACCT TCACTAATTT TTTCCAGATG ACACAAATAA TTCCAAAAAT
     901  CTCAATCATA TTATTCTTCC TTGCTTTTTA TTTAACTCAT GCTTCATCAA TGTTTTACAT
     961  CAAAATTCTG TGCCGTTCTG GAAACTGTAC TGATAGTGCT TCAACTCTCA GTTGGGTTGT
    1021  TACAATTATT TTAATTGGTT TTAAAGTTGT CGACTATGAA GATTTTTCAA AGGTTTGTGT
    1081  TTTTTTTTGG AAAGTGACTT GTGTGAAATG TTGCTTTGAA TCGATCAATA CTACACAAAG
    1141  AAAATGCTGA AACATTTGGT TTTTAGAAAA TATATTTTTA AAGCAAAATT CCCTTTATTC
    1201  CTTCTCTGGG TGGATTTTAA CTTTTCTGGA TTTTCTGGAT TAACTTTCTG GTTTTGTAGA
    1261  GTCTTAGCAT CACAATGATT GCTTTTGGAT TCTTGTATTT AAGTCTAGCA TTTGTCATAA
    1321  ACGATCGATT ACCTCCAGCG AAAAGCCTTT TTGAGCAAGA AATTGCACAA AGTTTTGGAT
    1381  TTCAATCAGT TAAAAATATT CTCAAATATG AAGACCATCG CACATCGGAA ACAGATGGAA
    1441  TTTTAACTTT TTATATCAAG AATCTTACAA GCGATAAAAT AGCGATTGAT GTGAGTTTTT
    1501  TTTAATGTAA AAAGAGTTTG ATATTACAAA TATTATAGAT AAGCTTCTCA AATCCAGGAG
    1561  ATCAGTTAGA TGGACATTTT TCAAGGAAGT ATATTCGTGG AAAAGGTCAT GTGAAACTTA
    1621  ATGTAAAGCG AGGAACTATA ATGGAAAAAG GAGATCGAAT GTTCATTTTT TATGCAAGAT
    1681  CTATAAAAGA TTTGGAAAGT ATCAATGATT CGGGATTTCT AACTATTTAT TTTTGATTTA
    1741  TTGGGTTTAT ATGATCTGAA TGTTGATATA AAAAATTTTA AAACTTCGAT TTAAAAAATT
    1801  GTCTCAATAA AAAAATTTCA AATGAGAATC TATATCTCAC CTATCAACAT ATTCTGCCAA
    1861  ATCCTCATCA TCATTGATCG GGGCTCCATT TTTCCTGATA GATGACGTCA CATCGAAACC
    1921  AGAAGCCTCA TTTCTAAAGA TTATCACTCC AACTTGTTTA TCAAATGCAT TTGATAGCAT
    1981  TCCATAATAT TCTGATATTG ATTCTTTACT CAGCTTCTCC CATCCAGTTT TGTATCCCAA
```

```
    2041  CATTAAAATA TTCGGACGGA GACGGCCAAG CTGCGGTCTG AAAATTTTGG TTGTGTTGGT
    2101  TAATCTATTC ATATTTATAG CACGAGATAA TTCGCACCGT GACCTTTTTT GGAAAACCCA
    2161  CAATAAAACT ATCATTTTGG AAATTTTATA GTTAATCTAG CACTTGGCAC ATTCGAATAT
    2221  TTTACCTGCA ACAAAA

//

LOCUS          41     14183 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      41
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS
complement(join(500..691,1197..1328,1565..1720,2195..2521,3016..3094,4878..4964,
13634..13683))
                /translation="MKRRGPEIPPSFSRGGRKLPITRIPHFPVIKQGWVILTTRGSS
                IEEMSSDTKYKYAVVRIPETSVVDFFSLIVAKGYSDVSFVASLSQCNLRDDPNDDQP
                TTSSNSVKESINDDESNSENNKLSPPRRQSNPHSSLPAISSSTVKNEPTDSWTPSAL
                SNDPTPDLLSATVPAELLTNLFAKTKSTEPKPQQLFGFQASGVDFDLSNNEWHENLR
                LPNGNGTEKYHPYGGNSKNDSPLQTRMKGWQREYIKEVIKDSHYPTEEELRDIEQKC
                DLSRKQILRFIAKRLTNPNRKPRVNHHDEKRKEQEERDSLADPDDDMINDNEAVTNL
                HHILNSLQETTA"
ORIGIN
       1  TTATAGAAGA GCACAGCTTT TCCGTCTTTC TTTTGGCTTT TGAAGAGGGC ACAGAATAAT
      61  TTGTAAGAAA ATGGTGGTTT TGAACTAAAT ATTTATAATT TTTCGAAAAA TGCGTACTTT
     121  TTTGTGAACT TTCAGTTCAT GTTTTGTATT TCCAAATTCA TCACATTGAC GTTTGCCAGA
     181  GCTTTTACCA AAGAAAAACC ACATTAACTA AGTAAGCCTT ATCATGTTTC AAACATAGGT
     241  TTTAGTCTTA GACTTATAGA GTTAGATTCG AGGATAAGCT CTGGTAGGAG TAATGCAAAA
     301  TTAAATTAAT ACAAGAATTC AAAACACTCT TCTTCTGAAC AATTTCCTCC TGTCCCCCTT
     361  CACTTACCAT TTCAACACAA ATTTCAAACA ATCAATCGAA GAAACCAAAC ACTTTTTCAT
     421  TTTTATATAT GCAAAGGCAT TATGCAAGTT TTAATAAAGA ATTAAAATCA AATTCAGTAG
     481  TTAAAAACAA ATAGTTGTGT CAAGCCGTCG TCTCTTGCAG TGAATTCAAA ATATGATGTA
     541  GATTCGTTAC TGCCTCATTA TCATTGATCA TATCATCATC CGGATCGGCC AGCGAATCCC
     601  GTTCTTCCTG CTCTTTTCGC TTTTCGTCAT GGTGATTTAC ACGTGGTTTT CGGTTTGGAT
     661  TGGTTAACCG TTTTGCGATG AATCGAAGAA TCTGAAATAT ATTTGAATTT TTATAAAATT
     721  AAGAACTTCT TACTAGGGAA TATTGTTAAT TGGGCAATTA CTAGGAAGAC AAAATAAACT
     781  TTAAATGAAA GACCAGGCGT CTTCCTAGCA ATTTCCCAAT TAACAATATT CCCTAGTTAG
     841  TTAAATCAAA AATCGTGAAA TATTAAAAAC CAAATACCAT ATTTCTTCTA TTAATATTAC
     901  ACCCGAATGC TTCACTTTGA ATCACTTCGT CGTTAGCATT TCAATAAATA TTAAGTTACA
     961  AACTATAGAA AATGTTTTGT TAGTTGACAA GTTTTCAATA TAACAAACGA CAACTGAGAT
    1021  GTACGCTGTC AAGGTTTGAC AATGGAAGTG CAATACTTAT AGAAAAAACG GTTGCGAAAA
    1081  TAAGGGAGAT CTAGGATCAT TTTTCATTGC TCCTCCTGGG GGAGTTTTTC TCCAAAAAAT
    1141  GAGCCTCAAT AATTGAGGCT AGACCGTTTT TTGGGTTACT GTACTGGGAT GCTCACCTGT
    1201  TTCCTACTCA AATCACATTT CTGTTCAATA TCTCGAAGTT CTTCTTCAGT TGGATAATGA
    1261  CTATCTTTAA TCACTTCTTT AATATATTCC CTTTGCCATC CTTTCATTCT TGTTTGAAGT
    1321  GGTGAGTCCT GAAAATTGGA AAACTTATGA AACTTTGCAA ACATTAAAAT AGAAATTGGA
    1381  AATCAATTCT TCAACCCCCT CGATCAAAAT AAATACTCTA ATATATAATT TCTGTATCAG
    1441  AAAAAGCTTC AAAGTTTTCA AGACTGATCT ATAAAAACAA TTATTAATAG AAATGGAAAT
    1501  TTTGCTAAAA ATACCGTACC CCCGGGGTGA ACTTGACACA AAAACTATTT AGCTTCAAAC
    1561  TTACATTCTT ACTATTCCCT CCGTACGGGT GATACTTTTC AGTCCCATTT CCATTTGGTA
    1621  ATCTCAAATT CTCATGCCAT TCATTATTTG ATAAATCGAA ATCAACGCCT GATGCCTGGA
    1681  ATCCGAATAA TTGTTGAGGT TTTGGTTCGG TAGATTTCGT CTGTAAGAGA AAACAAAAAA
    1741  TGTGAGTAGC GACAAGGACA AAATCAAGGG TTTTTAGAGA TCAAATCAAC ATTGATTAGG
    1801  TTTTGTGGAG TTATACATTT AATTAATGAT TACATGGGAT AAGGTTGGTA ACTAGCTTAT
    1861  CTTTGACCCT AGTTTCAGAG CTGGGAGCTT TAAAAGTACT CCAACATACT TTTTCCTATA
    1921  GATGTTTAAA ATTTTCTATG AATTTTAATT TATCTGTAAC AACTTTGGGA TTTCCAACCA
    1981  GTAAATATCA TGAAACATAT TAAGAAGTTA GTGCGATTCC CTTATATTTC TAACACATTT
```

2041 TTTGATAATA TATGTCTATA TTGGAATGAA TTCATCAAAA TCGAGAACAG GAGACAAAAT
2101 ATATACTGTA GTTAGTAGTG GTTTGTATTG TTAGATCAAA CATTTGAACT CATACAAATT
2161 GATAAAAGAA TAAAAAAGAT TAGACAGCAC TTACCTTAGC AAACAAGTTG GTGAGTAATT
2221 CAGCAGGAAC TGTTGCTGAT AAAAGATCTG GAGTTGGATC ATTAGAAAGA GCCGATGGCG
2281 TCCAAGAATC AGTTGGCTCA TTTTTAACAG TAGACGATGA GATGGCTGGA AGTGATGAGT
2341 GAGGATTCGA TTGTCTTCGT GGTGGAGATA ATTTATTATT TTCTGAATTT GATTCATCGT
2401 CATTTATTGA TTCTTTTACA GAATTCGAAG AAGTCGTTGG TTGATCATCA TTTGGATCAT
2461 CACGCAAGTT ACATTGTGAT AGTGATGCAA CAAATGATAC ATCTGAATAT CCTTTCGCAA
2521 CCTGTAAACA TTTTTGAATT TATAAAAAAA CTGTTTTTTG TTTAACTTCA GTTCTATCAG
2581 ATAAGAGTTC ATACTAAACA ACTTTAGTAG GTAATCCATC ACGTTGATGC CAGTCAAGGG
2641 CGCCTCCTCA AGCATCCTCC TTGAGCCCTA CACAGTTCCA CTAGTTGTTC ACAGACGAAT
2701 CAACTTTTGC TATATATTTT TTTGTTATTT TCTAGTTTTA TCATAATGCA TGATACATCC
2761 CAAATACAAG ATTTCTACGG AAATCTTCCT TGGTCTGACC AAATTTTTTT TTCTAACTTA
2821 TTTGTATATC GAGTTGGGAC AATTTAATAC AATAACCCCA ACAATTTTAG TGAATAAATT
2881 GAATTGAAAT TTTGATCAAG ATTCACAATA AGTGTGACAA CCTTCCTCAA CAGCTATCTA
2941 TAATTCGCCT CCACTGAATC TTCTCCTCTT TAAACCTACA AGTCGTATAT CCGTAAAGTT
3001 TTTTTGCAAA CTTACAATAA GGGAAAAGAA ATCAACAACC GAAGTCTCTG GTATCCTCAC
3061 AACTGCATAT TTGTATTTTG TATCCGAAGA CATTCTGAAA AAAAGTTTAG ATGTTTGATT
3121 CGGAATGAAT GAATGCAAAG AATAAATGCT TGAAGAGTTG AGAGAAACGA CGAAAAGGCT
3181 CTTTGGGGTG ATAGGGAGCA GGTGAAGGTA GTAGATAAGC AAATAAGCGC AGTTTATTGG
3241 TTTAATTTAG GAATTGAAAA AGGCAATGAT AATGATTAGT GTAGTTGATC AGTATTAGGG
3301 GAATAAAAAT CACAAAACTT TATGAATTTG ATGACATTTT TTCAATAGTT GCGTCAATTG
3361 CACAGAGCTT CTAACTAAAA TTAATGTTTT TTTGGAAACT AGGTCTATGA ATAATGTTTA
3421 ATTGAAAATC AGAGCAAGTT TAATTGTTGC AATTTTCCAT AAAACTTACA GAAATTTCAA
3481 CCTACATACA AGAGGAAGAT GATTAATACA TCAAGAGAGT ACGGAATTAG GTATATCATT
3541 GGACACTTGT TTCTGGAACA TTGAAAATTG ACGAATTCCT TTTTCTTTTA ACTACCAGAT
3601 TTTAAAATTC AGATCGATCA AATGAAAAGT TTTAAAATGC AGTCTTTGTT TTAGGCGTGC
3661 ATCGGATTTA TCAAACTTCA GCAATACCAC GAAGGTTGAG CAATGCCAAT ACACATGGAA
3721 GATGCTCCAA TTGAGCATAA ATTTTTACTG TAAGCAAAAA CAGGAATAAA ACGCCATTTC
3781 CCCAAAGTTC TAGAGATTGT ACAAGAGCTT AATCAACAAT AGTTTAAATT ATTTACGGTG
3841 CCATAATTGT TTGTTTTTGT TTTGTAGTTC AAAATTAGGG GTTTCAACTT TGTTTCTTAC
3901 ATTTCAATTG TTCAATCTAT TCATTATCGA CATTTCTTTG CCAATTTGTG ACAAGGCATT
3961 AAATAGAATT AGTTGATTAC TCTCTTATGA GACGCAGGTG GTTGGCAACA CTTACATCCA
4021 ATTTTATTTC AATTTTACAG TTTCTTTTTC ATTTTCATTT CAAATTGGAT GATGTTTGTG
4081 AAAATGTTCA TCGTCAAAAA TTAAAAATAT TTTAAGCTAC AGGACGGATT TCCCTCGATA
4141 GGTCCTATTT TCAAAAAAAA AAAATTAACT GAGCGTCTGA AAAAAAAAAA AAAAAAAAAA
4201 AAATTAACTG AGCGAACTGA AAAAAAATTA ACTGAGCAAT GATTTTAAAT CCAATATTGG
4261 ATGATTCCTA ATACAAAATT AGATGTGAAA AAGAGTCACA ATATAAAAGA CAACAGTTGG
4321 TTGAATGAAA GGATAGAGAA AGAAAGAAGG GACACAGGCA AAGAATACTA ATTTTGAATG
4381 AAAAGAGAAT AGAGAGCGGC TGGTGGTGGA AGGGCTCCAA AAGTGTTCGA TTTGATGACG
4441 ACAAGTGTGT CTTCAAACAG TACGCACACA TAAAGAAATG GACGAAAAAA GGAAGAAAAG
4501 ATGAAGAAGA AGAAGAATGA CACTTTTTAT GATTGGTACT TGCTATGGGA ATATGGAATT
4561 TATTCTCTAT CAATCAAACA AAAATTATTG ACTTTAGAGA GACAAGTATT CACAAAAGAA
4621 AATGAGCAGA TATAATACGG TAGTAGTGAT TTAAACAAAA AGTCGGTAAA AATCTACAAG
4681 TTTACGCTTT ATATATATAC CAGAACATTT AAATGCGCTT CAGAAAATGA AGAATGCTCA
4741 ATGATTATTA GATTGAAGTA CTTTTCAATG AGAACTAAGG TGAAGGTATC AATTTCGATT
4801 CAGCAGAATA TGAAAGAGCA GTCTTTAAAA AATAATCAAG ATGCGATTCT TAAATTTTTT
4861 TAGCAATTTC ATGTTACTCC TCAATTGAGG AACCTCGGGT AGTCAGAATG ACCCAACCCT
4921 GTTTGATCAC TGGAAAGTGT GGAATACGTG TTATTGGTAA TTTGCTGGAT TTTTAGATGT
4981 TCTACTAGAA ATGCACGGAA AAATGCTTCT GAATAAAATA GTACCATGCA GACGACTCGT
5041 TGCGAGCAAG AACCCTTTTG GTGCGTGGCT GTTCAAATTT TGGCACAAAA CCACGCGTTA
5101 TTGTTCATTT GCGAACAGGA ACAGATAGGG GGACGGAGCG ATGCAAATGT GAATAACGTT
5161 CACAATACCC CCCGCAAGTA CAAAAAAAGA GAGAACGAAA GGAGAAAGAG AGAAAACATC
5221 CTAAACATCG TTCCGTGTTA TAGAGATGTT ACTTTATGTG CCATAAATGT GCGCACTGTA
5281 TTCTTTGAAA CTCTACAGAA TTTTGGGACA AAATGGGGTT TTGCAGAAGC ATACCGATTT
5341 GCGAGGTGTA CGCAGCCATC ACAACTCAAA ATTGTTTAGG TTAGTAGAGA AACAAAAATT
5401 GTTTATGTAT TGTGTGACAG GTGTAGAATA TTTGAAAACG TTCACTATTC TAGAACCATT

5461 AGACCCTCAC TTTATTGATT TTTGTTGAAA AATGCAAATG GTGTAGGGAA ATTGGGAAGT
5521 TGTTAGAAAA ATTAACTCAA AGTTGCTCCA GATTCTTTAA TTCTTCAAAT TTGAGATTGC
5581 AATTTTCGAC AACACTTGCT GAAATGTTGG GTAAAACTTC AAGAAAATTG ATAATGAATT
5641 GATGAAAAAA GAGCCAACAA TTGCAACTTT TGGAAAATTT GGAAAAATGA TTGTAAAATG
5701 AGCAAACGCA CGGTTAAAAT CTCAAAATTA AGGAGGCAAT TTTCTAATAA ACAGGAAAAG
5761 CAAGATCGAT TTCGTGAAGT TTTCAAAATG CAATACTAAC AGAAGAAAAA CGGTAATCAA
5821 ATTTCGATCC TAATATTCTT CTACGGAACT TATGTTTTCA CGTACAGTTT GGTACAGTAC
5881 AGTCTGGAAA TTCATAATGT TGAAAGTAAT GATATCTGAT GATCAGTCGT GCTCTCTCTC
5941 TGGATCTATC TCGCGGAATT CTACATTCAA AATAAGGATT TTCAAAAATC CTATAGCAAG
6001 GAAAAACTGA TTTTATGACT AGATGACATG AAACTTACTA GGGAATGGTG TTACCCCTGC
6061 AATCGCTAGG AGGACAAAAC AAAGTGAAAA TGGAAGATCA GATCTATGGC TACAAGAATC
6121 AGTTTTTGGA ATCTGCTTAA CTCGGTCCCA AAGCTGAAAT TTTTGGCCTT AAAAATTAGT
6181 GCTTTTTTTG CTCAAAATGC CGTTTTTCGT GTGTAGAAGT GGCTACATAA AAATTTGGCA
6241 CTGCCGCCGC CGTTCGGATG GCCGATTGGT GAACTCTCGT GCTCATTAAT CTAAAGGTCA
6301 CGGGTTCGTT CCCCACTAGG TATTAATTTT TTTTTTGAAT TTTTCCTCTA GCATGAAATA
6361 TGCACAAAAA CACGGTTTCT TCCGTTTTTG GGTAACTTCG GAATAAATTT TTTGGTTCCT
6421 GGGAATCATA AAAGTAATTT GGAAACTTTT TGGAAGTTTT CTTTTTTTGA TAAATTTTCA
6481 CTTTTCTATG CACATTTTTT TCTTCAAAAA TAATTTTTCA AAGTTTTTTC AAAAAGTGTT
6541 TGGTTTGAAA GAGCAGCATT AGTTTGGATA TTTCTACGAT TTTGGAGGAT CAAATGTCAT
6601 TTTTAGGAGA AAACCACTGT TGTTTTATCT AGAGTTTTTC ACGAAAAGGA CCTACTTTCA
6661 AAGACTTCCA GCTCCGTTGA AGAGTAAAAT AAAAAAAAGT TGTCAACTGC AAAGTTATTT
6721 TAAAAGTACC AATAAACAAA TTGTCAGTTG AAGTTACTCT GCTAAATTTT ATGGGTAGAG
6781 AGATATGGAC ATTAGTAATT AAGTAAGTTA AAGCTTCACT TCGATTAGCT ATTTCCGTCT
6841 ACCCATAAAA TTTAGCAGAA TAACTTCAAC TGACAATTTG TTTATTGGTA CTTTTAAAAT
6901 AACTTTGCAG TTGACAACTT TTTTTTATTT TACTCGTCAA CGGAGCTGGA AGTCTTTGAA
6961 AGTAGGTCCT TTTCGTGAAA AACTCTAGAT AAAACAACAG TGGTTTTCTC CTAAAAATGA
7021 CATTTGATCC TCCAAAATCG TAGAAATATC CAAACTAATG CTGCTCTTTC AAACCAAACA
7081 CTTTTTGAAA AAATTTTGAA AAATTATTTT TGAAGAAAAA AATGTGCATA CATCAGGTGG
7141 GAGAAAAGGG TGGGACAGGT GGGAGAAAAG TGAAAATTTA TCAAAAAAAG AAAACTTCCA
7201 AAAAGTTTCC AAATTACTTT TATGATTCCC AGGAACCAAA AAATTTATTC CGAAGTTACC
7261 CAAAAACGGA AGAAACCGTG TTTTTGTGCA TATTTCATGC TAGAGGAAAA ATTCAAAAAA
7321 AAAATTAATA CCTAGTGGGG AACGAACCCG TGACCTTTAG ATTAATGAGC ACGAGAGTTC
7381 ACCAATCGGC CATTCGAACG GCGGCTGCAG TGCCAAACTT TTATGTAGCC ACTTCTACAC
7441 ACGAAAAACG GCATTTTGAG CAAAAAAAGC ACTAATTTTT AAGGCCAAAA ATTTCAGCTT
7501 TGGGACCGAG TTAAGCAGAT TCCAAAAACT GATTCTTGTA GCCATAGATC TGATCTTCCA
7561 TTTTCACTTT GTTTTGTCCT CCTAGCGATT GCAGGGGTAA CACCATTCCC TAGTTAGCAT
7621 AAAAGTTTTG TGTGAGTTAT CGGATTAAGA ATTCGATTGC TCACAAGATT TATTAGTTTA
7681 TACTGGTAAG GAAAGATTCT AATTGAATTC TTGGTAGAAA AAGAGTAAAA TGGAAATGAT
7741 TGATAAGGCA GATTTGCACT TTTTGTTTAA TTTTAGGGTT TGACGTTTTT AGAAAACTTG
7801 AGAAATGTGA TTCGGTGTAG GACAGTAGGA GAAGGAAAAG TTGGGCTTAA AAAAACGTTT
7861 AGTTGTAACT TTGGCTGTAA TTTTTTTGTA TTTCTTTTCC CCTGACAGTA AGAGTTATCT
7921 AGAGGAGATT AGCCCTAAGA ATGAGCGACA TGATTACGGT AGCGAGAAAG TCGCACCAAA
7981 AGAAAATGTC AGACATTTTG ATGAACTGGA AGTTGAAAAG TGGTGATGTT TGGGAGTGAG
8041 AGATAGAGAG AGACCTCTTA CTACCCTTTT TCCACTTGCC TATACACTAA AACGTTAGAG
8101 TAGTTGGCCA TAACACTCTC AACGAGTATT TATTCGAACA AGAAGAGGTG GTAGTTGGAC
8161 GTCGTTGCAG AAGTAGAAGA AGTAGAAGAG AAAAGAGAAA GAAGGTCTCT AAGTCTCTCT
8221 ATCTTTTTCC CCTATTCCAC TTGTATGCCC ATTGCTGGTA ACGACTGCAG TTTAGAGGGC
8281 TTTAAAACGT TTGGGACACA CACTTTGACC ACCATCCAAC ACTATTTAAG TGAAGAGAGT
8341 TAAATGGAGA AAAGGAATGA GGGAAAACGA TTATTTGAAG ATTTGATGGA AATGTTGGAA
8401 AAGAAAATTA CATTAAGGAT ATCACACTAA TTCTGAATCT AGTCAGTTAA TGTGCCAATA
8461 TCAATTTTAA AACTGTTTAG GTGTTCATTT TCTCAAAGAA AAAAAATTAG ATAGTTCATT
8521 GTTTATAAAC CGTGTTCCAC CAGACACTTA TCTAAAAGGG TTTTATATTA CGGTATTCAA
8581 TTATGGGGAT AGTCGATTAT AGATAATAAA TTCTTCCTCG AGGCGTGACG GCGGAAGCTT
8641 TTCCCGCCTC TCTTCGTACC CTAATGTAAG CCATGTGAAC TCCGGTAGAA CGTTTTAGAC
8701 AACTTTTTAA ATTTATAAAA AGTTATTATC GCTTGATTGC AGAACATTTT TCATGAAAAT
8761 ATATCTTCAG ATCGAGCTCT GGCTATAGAG TATGGAGAAT GTGTGGGTCT TATGTTTCCA
8821 GTTAAATGAA ATCCTTAAAA GTTTGAACTA GTAAAACACG CAACCCAGAA ACATGAACAT

```
 8881  GAAGGAAAGA GCATAAATAA AAGTAAAAAG AGTTGAGAGA AGGAGTGGAA AAGGTTTTTG
 8941  GACGCACACA TAGACATATA AACTTTTGAA AATTGGGTAG AAGTGGAATC GTTTCGAAAA
 9001  TTCTTGGAAG TGATACGGAT TCTGCAGACT TGTGAGGTTT AACTTAAATC TAGTTTTATA
 9061  CCAATTTCAG TTTTCTGGAC CATTATACAA ATTGTTTCGA AATGTTTAAA GTCAGAAAAA
 9121  ATTCTGAAAT GACATGTCGG TAAATAGAGT AACAAGTAAC GACTAACACT AGTTGAACAG
 9181  TAAGAGACTA CAGGTGTACA GGTAGACAGA AGGGTGATCG GGATTACGAA CAAGGATTAA
 9241  GGCGGGTGGG AGAGGATAGA AGAACGAAAA AGAAGAGAAG AAGAGGGTAA TAGTAGTACA
 9301  GTAAGGGATT GAAGAGAGAG GGAGAAGGCC ATTATGAACA CAGAGAGAGG CCTTATGACC
 9361  AATAGATAGG GCAATGAAGA GAATGGGGTT GCTAGGGAAT AGAAAAATGA AACAAGGAGA
 9421  GTATGTTTAC AGCTGTGTTA TGGATATAAA TTGAGTTTAA AGTTGAATAG GGAAAAATCT
 9481  GAAGAAAATG GATAATGAAT GAGCAAAGGA CATTTCTAGA AATTGCTTGG TTTTTAATTC
 9541  ACACAACTAA TTGTGAGCAT TTTTTAAGAT TGGATAATAG CTTTTTCGAT TATTATTAAT
 9601  TAATGCTATG TTTATACTTT AACAAAACGA GATTTTTAAC TTGAACTAGA CTAGGTTAAA
 9661  ATATGAGCGG CAAAAAGATC CTAAAAATTA GGCTTGGTAT GGAAATTTCC AAAATTTTAT
 9721  GACACTTGTA GTAAATAATT AACAATTAGA AACTATGGGA AGACTTTTTA AAAAATAAAA
 9781  TGGAAATAAA TGTTCAAAAA CACAATGTCT TATTGACTTT TCCACATTCC CACATTTTCA
 9841  CAACTTAATA GGTACTTGGC ACATATAATA TTCAATCAGT TATAAAACTA AGGGATTTTT
 9901  AATATATTCC AATTCCAACC TATGAAATTG GTCTTTAACT TTATGAAATA TAGGTTTGAA
 9961  AACTCAACTG TAACTTCTCC CCTTCAAAAC TTTGGAACCG CCCGCTTAAA CGAGTTGGAT
10021  AATCCGGATA AGATATTTAT ATGTATATTT AAAATGTCAT TTCATTTTTA GACATCTATT
10081  TCGCCATCTC CGTCATTTTG TGGTTTTCTG GTTGCAATTG ATATTGCAAA GTTAAAAATT
10141  GTTACTCATA TACATAGGTA GACAACAGAG CCTCTTTATT TCAGAGATAA GTACTAGATA
10201  AGTGAAGTAA TCTAATAAGT GCATACTACT TGGTCACGTA TATTTGGGAA ACCATTGAGA
10261  AATTTGCATT CCTTATTCTA TTCTGACCTA GTTTTCTCCA AACTCCCCTT ACTTTTCTCT
10321  TATTCTCTTA TAATTTCCAC CAGTTTCTCA CATTTATATA GACAACCCTT AATCGAATAA
10381  ACTTCACAAG TAGGAAATTT GAAGAAAATA AAGAAGAGCA TTATCTTGAG TTGAGTAATA
10441  TTTTGGGGTA GTTTTTGAAG TCCGGGCTTT TTGGTGTAAA AACGTGGAAC TCCAAAGAGC
10501  ATTTTTGGAA ATAGATTTTT GTCACAAAAA GATTTGGGTA GCTTGAATAT CAGTGAGTTC
10561  TAAAACCAAC ACTGCCACTA ATAAATAAGA GATATATAAA TCGTAGAGTA GTGTTGTATT
10621  GAAAGAAGAA AAAAGAGTTT CTTGTGTGTG TGCTTCTCAC AAATTAAAGG TGTGGGTGGT
10681  GGGATCCCTT TGGCTCCTCG TGCTCCACCT TGTGCACTTT CAATATTGAC GAGAGATTCT
10741  TCTGACAAGA GAATCCCCCC TTCTCTATTG TGTTAAACTT TTTCTATTGG TCCCTCGTCT
10801  TTCAATTTAC GAGCAGTTTA CGAGAATTCT CTTATCTATT TCTGATCAGC TGATTATTAT
10861  CTCTTATCAG TTCATGATCA CTATTCCTTT AATCTTCATT TACTATGTAT TTTTTAAGTG
10921  AAAGAATGAA TTGGATTTGA GTGAGGATTG AAAAATAAAT TCTATTATTT TCTAAAATTT
10981  AATGGAACTG GAGATTACCA AAATACTAGT AAGAGCTTAG CTCATAGGTG AAAAGTTGTA
11041  TTTTTGGGAA AAAATATAAC ATTCAGATTT GAATCTCGGT ACTGGTATTT TCTACCGTAT
11101  TCCCTTATCA GTCTTTTAGC CGTCTATTAT TTGATTGCTC ATATCTCATT TCATTTTGAA
11161  TATTTTGCAA AATGTTTCAC TAACAAGCTG AAGAACTGAA TAACAAATGT TTTGTCAGTT
11221  AACAACTTTT TTATACGCCT AAAGAAAATA AAGATATAAG CTGAGCAATG AATGTTGAGC
11281  AATGAGGTGC AACTCTGAAC ATAAAATAAA AAATTGAAAT ATTTGCAGAA GGATTGTCAC
11341  CGTAAGAAAA CTAAAAAAGA CCTTCCCTCC TCCAACATTC TCCATTTATT TTCCTTATTC
11401  AACCAGCTTT GCCTTCAATG TCCGTAAGCC ATCCTTAAGC GGCTTCCCTT AGGGGCGGAT
11461  TTATGTTGTC GTTTAGGCGA AATTATTCTC TATGTCTATG CGTAGAGAAC AATTTAATGG
11521  AAAATAAAAG CGATGTACTA CGTCAAATAA GAGCACCTGG ACAGGCTTCT TCTAAGCTTT
11581  TGGAAATTTA ATAGTGAAAA AGGGTAAGGG CAAGAGACGA AAAAATAAAG CATTATGATA
11641  ATGGATTTGT GGGGTCAAAG AGAATTTAGA TATTCGATTT TCGGGAAGAA AGGAATGAGG
11701  AAATTGGGAA ATTGAAAATG AAATTTGAGA CAAAATTGAT AAATCTTTGA TGGATTTTAT
11761  GAATTAAAAA AGTTTTTAAA CTCTATTTCA AAGCTGCAGA GTAATGCAAA AGTTTCAAAA
11821  CATTAGAAGT TTAAGACATA TTCACACTGT TTTCTGTTAT TTTTCACATA CTAAAATAAA
11881  ATAAACATAT AAAATAAGGG AAAGTCGGTA GAATATTCAA TAATCTTTAA GATAAAGTCA
11941  GAAATTCAGA GAGAAAATTG TGGTTAGTGA TCATAAGCCA GAAGAAGAAA CTGGAGCTTG
12001  AGATGTTTGG GAAAAAAATG AAAGAGAAAG AGAAAGTAGT GGGACACTCC AATCAATTTC
12061  ACACGCTTTA AGAATAGAGT TTAGGGTTGA ATGATTCGAA ATTTTAAATT GAAAAGATTT
12121  CATACGAAAT TTATAAATTG GTATATGTAG CTGAATATTA GACTTGACAG TGCAAACTCT
12181  AAAAAAGCAA TTGAAACAAG TTGTTCCAAT GTTTTTGACT AAAATTGGTT GTCAAAAAAT
12241  TTGTTTGATA AAAAACATTT GAGAAGTGGT CAAAAGTTGC ATGAAAAGTT GTCAATTGCC
```

```
   12301  AAAAATACGT ACAGCCAGGA AAGATTTGGC AATTTATGAA ATCAACTTCA AATTTTTCAT
   12361  TTAAATTTTG GGCTAATGAG TTGATAAAAT TTTGCCCTGT TTAAACGATA CACTTCAAGG
   12421  AAAATTGGAA ATCATACCAC AACAAGTCAT TTTCTAAATT TTCATAAATT TTTGTTACTA
   12481  TGAAAAGAAC AACGAACAAA CATTTTCTAG ATACATTTTA ATTGTATTTA GTTTTCCAAT
   12541  GAATACATAC ACAACACATA TATAATTAGC ATGAAAAAGC AGAAAAGTGG TGATCTTTCA
   12601  AAGTTATAGT TCGTACGAGT ATTCATCATT GAAAACTAAG ACATTGGCAT TTAGAACGAA
   12661  AACGTGTGAT TTACTTTATT ATTCCAAGGT TTCACTTTAG AGAAAAAAGA GAGAACAAAA
   12721  CAAAAAATGT CTTAGAGATA GAAATATAGG GAAATGAATA TGTTAACCGA ATATAATCTG
   12781  AAAGTATCAA ATAAATAAAA AATAGAAGCA TTTTTGTCAG GAGCATTATG AGAGCAACAT
   12841  TCAAACAGTA TGTTTATGAG ATTTGAAAGT GAAAAGATTA TATAAAGAGA GTCTCTAAAA
   12901  AATCAATATA AACATAGAGA ACAGATGGAG GGCTGCCGTA AGGAAAGTCT GTGTCCTTTC
   12961  CCTTAGCGTC CTCCTTCTCT TTTTCACTAA GCCCTAGACA CACTAAGGGA AAGAGAGAGA
   13021  GCATTAAGTG GAGAGACACA GGCATTACTT GGGACACTAC TCTAAACATG TTATATTAGT
   13081  CTTTAGAGAA ACAGAGGAGA AGAGTAGATG GATATACAAT GGGCCGACGG GGAATTCTGA
   13141  ACAAGCACAT TAGTGTGATT GTGGTGTGGT TCGGAGAAGT TTGTTTTAGA GAGAAAAGAT
   13201  ACGGGTAATA AGAACATCTT GGACAATCGA AGTCTCGAGA ACAAATTTCA AGGAACGAAA
   13261  ATTAGTTTTA AAACTGCAAA AGAACCATGT TACAGTTTTA AAGGCAAGAT AGCTCAAAAA
   13321  AGTAATGTCA TTGAATTTCT GCAAATCTAT TTTTGACATC TGTTAGGAAA GTTGCTTGCC
   13381  GTCCGTCGTT GGAAAATTAA ATTCTCATCT GAAAATTATT CAGATAATTC TAACCCAAAA
   13441  ATTAATGCAA ACCAAGTTCC TTGGGTTCTT GGAGTTCAAA AAATTGATCT CCAAGTTTGA
   13501  ACCTTCAAAC AACTTTTTCA TATAAACTAT AAAGTCTAGT AAATCTATCG ATTTGTTAAT
   13561  GGATTAATTT CTTATTAGCA TCCATTGGAA TAAAATCGAT AATTGGTGTC GAGAGAGGGG
   13621  ATAAAGGGGT CACCGTCCAC CACGCGAGAA AGAGGGTGGG ATTTCAGGTC CTCTTCTTTT
   13681  CATTCTGTTT ATCAACGATC TATTAATCGA CCTAGCTCCC TCCATAAATA TTTCTTGTTT
   13741  TGCTGATGAC GTAAAGATCT ATCACACTGA CCCTATTACT ATCCAAAATT CCATAGATTT
   13801  CATAGTTATG TGGTCCAAGA AAAATAAACT TTCGCTTGCA CCAACTAAAT CGGCTTTACT
   13861  AGCACTTGGC ACTCGGAACA AAAATCATTC TTATTTTGTG GACGGTGTCC CCATTCCCCC
   13921  TTCATCCACT GTTCGTGACT TAGGTCTCAT TACTGATTGC AAACTTAAGT TTGATCCCCA
   13981  CATTGCTAAA GTTAGCTGCC TAGCCTTGCT CAGAGCAAAA CAAATTCTCA AGGCATTCTC
   14041  CTCCAATACT CCTAAGTTTT ATGCCCATCT CTTTAAGACT TACGTGGCCC CTAGTCTCAA
   14101  TTACTGCTCC GAAATCTATT CCCCTCCACC TAAATCTTTG CTCTCAGCTA AACTTGAACA
   14161  ACCTCTAAGA CACTTTACCA AGC

//

LOCUS          42     2208 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      42
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS          500..1708
                  /translation="MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTLTLTE
                  LGRNLPTKARTKHNIKRIDRLLGNRHLHKERLAVYRWHASFICSGNTMPIVLVDWSD
                  IREQKRLMVLRASVALHGRSVTLYEKAFPLSEQCSKKAHDQFLADLASILPSNTTPL
                  IVSDAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAENWKPISNLHDMSSSHSKTL
                  GYKRLTKSNPISCQILLYKSRSKGRKNQRSTRTHCHHPSPKIYSASAKEPWILATNL
                  PVEIRTPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALML
                  QLTCWLAGVHAQKQGWDKHFQANTVRNRNVLSTVRLGMEVLRHSGYTITREDLLVAA
                  TLLAQNLFTHGYALGKL"
ORIGIN
        1  TTAAAAACAC CATTGAAATC CACTGTCTTA CCATGCAATC CACGCAAATC TCAGCTTGCG
       61  TGACCATAAA GTTGTAAAAC AACTCATTCA AACATTTTTG CGCATTTTTA AAGTGATTTT
      121  ATTTTTATTC GGGAACCTAG AATCATTGTC CGCACTTTTT GGAAATTTTT ATTTTTTTCA
      181  TTTTTGCTCA AAATTTCTTG ATAAACTCAA AGCAAAAAAT TCAAATAATT TCATTTTTCT
      241  AAACAATTAT GAAATTGCTA TGTTGTTGTT CAGAAATGTA TGAAACGTAC ATTACACAAG
      301  TTTTAACTCT CTATTCGCAA GAAACCGTCG AAATGATCTA CATCTCACGA ACTTTGTGCA
      361  AAATATTTAA CCAACTTTGA AGTTGCATAA CTCTGATGAA TCCCCTAATG ATTTTTATCA
```

```
     421  AAATCATTAA GTTAAGGTAG ATACACATCT TGTCATATGA TCAAATGGTT TCGCCAAAAA
     481  TCAATAATCA GACAACAAAA TGTGCGAACT CGATATTTTA CACGACTCTC TTTACCAATT
     541  CTGCCCCGAA TTACACTTAA AACGACTCAA CAGCTTAACG TTGGCTTGCC ACGCCTTACT
     601  TGACTGTAAA ACTCTCACTC TTACCGAACT TGGCCGTAAC CTGCCAACCA AAGCGAGAAC
     661  AAAACATAAC ATCAAACGAA TCGACCGATT GTTAGGTAAT CGTCACCTCC ACAAAGAGCG
     721  ACTCGCTGTA TACCGTTGGC ATGCTAGCTT TATCTGTTCG GGCAATACGA TGCCCATTGT
     781  ACTTGTTGAC TGGTCTGATA TCCGTGAGCA AAAACGGCTT ATGGTATTGC GAGCTTCAGT
     841  CGCACTACAC GGTCGTTCTG TTACTCTTTA TGAGAAAGCG TTCCCGCTTT CAGAGCAATG
     901  TTCAAAGAAA GCTCATGACC AATTTCTAGC CGACCTTGCG AGCATTCTAC CGAGTAACAC
     961  CACACCGCTC ATTGTCAGTG ATGCTGGCTT TAAAGTGCCA TGGTATAAAT CCGTTGAGAA
    1021  GCTGGGTTGG TACTGGTTAA GTCGAGTAAG AGGAAAAGTA CAATATGCAG ACCTAGGAGC
    1081  GGAAAACTGG AAACCTATCA GCAACTTACA TGATATGTCA TCTAGTCACT CAAAGACTTT
    1141  AGGCTATAAG AGGCTGACTA AAAGCAATCC AATCTCATGC CAAATTCTAT TGTATAAATC
    1201  TCGCTCTAAA GGCCGAAAAA ATCAGCGCTC GACACGGACT CATTGTCACC ACCCGTCACC
    1261  TAAAATCTAC TCAGCGTCGG CAAAGGAGCC ATGGATTCTA GCAACTAACT TACCTGTTGA
    1321  AATTCGAACA CCCAAACAAC TTGTTAATAT CTATTCGAAG CGAATGCAGA TTGAAGAAAC
    1381  CTTCCGAGAC TTGAAAAGTC CTGCCTACGG ACTAGGCCTA CGCCATAGCC GAACGAGCAG
    1441  CTCAGAGCGT TTTGATATCA TGCTGCTAAT CGCCCTGATG CTTCAACTAA CATGTTGGCT
    1501  TGCGGGCGTT CATGCTCAGA AACAAGGTTG GGACAAGCAC TTCCAGGCTA ACACAGTCAG
    1561  AAATCGAAAC GTACTCTCAA CAGTTCGCTT AGGCATGGAA GTTTTGCGGC ATTCTGGCTA
    1621  CACAATAACA AGGGAAGACT TACTCGTGGC TGCAACCCTA CTAGCTCAAA ATTTATTCAC
    1681  ACATGGTTAC GCTTTGGGGA AATTATGAGG GGATCTCTCA GTGCATTGCC TCCAATTCCC
    1741  ATAATTTATT ACGCCGATAA TAACTTGGTG TAACCTTAAA AATGTACTTA AATCGACGTG
    1801  TAAAAGATTG TTGGGAATCA AATTGATATT TTAATGCGAT CTCAAGGATA GTTTTTTTCG
    1861  TCAACCTCAA CTCAACAGCC GCTTTCGTCA AACGACGAGC ACGAATATAG CTAGCCAGTG
    1921  TGACCCCTGT TACTTTTTTG AACAGCCGCT GAAAATACCA CTTGGTATAA CCCGCTTTAT
    1981  TCGCCACATC ATCAAGCAGT AAAGACTGAT CTAAATTATG TTCAATCCAT AATAGAACAT
    2041  CTTTGATAAC CGTTGTATGA AACTGCTTAT CATCATATCT TAATTGGATG TTATTAGCTT
    2101  TATTTTGATA GCGAGAATGC TGTTCAATAT ACATAAAATA ACCTAAATGT TCTTAAGATT
    2161  GTCACGACCA CATCATCATG ATACCATAAA CATACTGACG GTATGTTA

//

LOCUS          43     2208 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      43
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS          complement(500..1708)
                  /translation="MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTLTLTE
                  LGRNLPTKARTKHNIKRIDRLLGNRHLHKERLAVYRWHASFICSGNTMPIVLVDWSD
                  IREQKRLMVLRASVALHGRSVTLYEKAFPLSEQCSKKAHDQFLADLASILPSNTTPL
                  IVSDAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAENWKPISNLHDMSSSHSKTL
                  GYKRLTKSNPISCQILLYKSRSKGRKNQRSTRTHCHHPSPKIYSASAKEPWILATNL
                  PVEIRTPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRTSSSERFDIMLLIALML
                  QLTCWLAGVHAQKQGWDKHFQANTVRNRNVLSTVRLGMEVLRHSGYTITREDLLVAA
                  TLLAQNLFTHGYALGKL"
ORIGIN
       1  GTAATGCAGG TAAAGCGATC CCACCACCAG CCAATAAAAT TAAAACAGGG AAAACTAACC
      61  AACCTTCAGA TATAAACGCT AAAAAGGCAA ATGCACTACT ATCTGCAATA AATCCGAGCA
     121  GTACTGCCGT TTTTTCGCCC CATTTAGTGG CTATTCTTCC TGCCACAAAG GCTTGGAATA
     181  CTGAGTGTAA AAGACCAAGA CCCGCTAATG AAAAGCCAAC CATCATGCTA TTCCATCCAA
     241  AACGATTTTC GGTAAATAGC ACCCACACCG TTGCGGGAAT TTGGCCTATC AATTGCGCTG
     301  AAAAATAAAT AATCAACAAA ATGGGCATCG TTTTAAATAA AGTGATGTAT ACCGAATTCG
     361  ATTGCGTCTC AACCCCTACT TCGGTATCTG TATTATCACG TGTATTTTTG GTTTCACGGA
     421  ACCAAAAACAT AACCACAAGG AAAGTGACAA TATTTAGCAA CGCAGCGATA AAAAAGGGAC
     481  TATGCGCTGA GAGATCCCCT CATAATTTCC CCAAAGCGTA ACCATGTGTG AATAAATTTT
```

```
     541  GAGCTAGTAG GGTTGCAGCC ACGAGTAAGT CTTCCCTTGT TATTGTGTAG CCAGAATGCC
     601  GCAAAACTTC CATGCCTAAG CGAACTGTTG AGAGTACGTT TCGATTTCTG ACTGTGTTAG
     661  CCTGGAAGTG CTTGTCCCAA CCTTGTTTCT GAGCATGAAC GCCCGCAAGC CAACATGTTA
     721  GTTGAAGCAT CAGGGCGATT AGCAGCATGA TATCAAAACG CTCTGAGCTG CTCGTTCGGC
     781  TATGGCGTAG GCCTAGTCCG TAGGCAGGAC TTTTCAAGTC TCGGAAGGTT TCTTCAATCT
     841  GCATTCGCTT CGAATAGATA TTAACAAGTT GTTTGGGTGT TCGAATTTCA ACAGGTAAGT
     901  TAGTTGCTAG AATCCATGGC TCCTTTGCCG ACGCTGAGTA GATTTTAGGT GACGGGTGGT
     961  GACAATGAGT CCGTGTCGAG CGCTGATTTT TTCGGCCTTT AGAGCGAGAT TTATACAATA
    1021  GAATTTGGCA TGAGATTGGA TTGCTTTTAG TCAGCCTCTT ATAGCCTAAA GTCTTTGAGT
    1081  GACTAGATGA CATATCATGT AAGTTGCTGA TAGGTTTCCA GTTTTCCGCT CCTAGGTCTG
    1141  CATATTGTAC TTTTCCTCTT ACTCGACTTA ACCAGTACCA ACCCAGCTTC TCAACGGATT
    1201  TATACCATGG CACTTTAAAG CCAGCATCAC TGACAATGAG CGGTGTGGTG TTACTCGGTA
    1261  GAATGCTCGC AAGGTCGGCT AGAAATTGGT CATGAGCTTT CTTTGAACAT TGCTCTGAAA
    1321  GCGGGAACGC TTTCTCATAA AGAGTAACAG AACGACCGTG TAGTGCGACT GAAGCTCGCA
    1381  ATACCATAAG CCGTTTTTGC TCACGGATAT CAGACCAGTC AACAAGTACA ATGGGCATCG
    1441  TATTGCCCGA ACAGATAAAG CTAGCATGCC AACGGTATAC AGCGAGTCGC TCTTTGTGGA
    1501  GGTGACGATT ACCTAACAAT CGGTCGATTC GTTTGATGTT ATGTTTTGTT CTCGCTTTGG
    1561  TTGGCAGGTT ACGGCCAAGT TCGGTAAGAG TGAGAGTTTT ACAGTCAAGT AAGGCGTGGC
    1621  AAGCCAACGT TAAGCTGTTG AGTCGTTTTA AGTGTAATTC GGGGCAGAAT TGGTAAAGAG
    1681  AGTCGTGTAA AATATCGAGT TCGCACATTT TGTTGTCTGA TTATTGATTT TTGGCGAAAC
    1741  CATTTGATCA TATGACAAGA TGTGTATCTA CCTTAACTTA ATGATTTTGA TAAAAATCAT
    1801  TAGGGGATTC ATCAGTGCAT AACTTCGTTG AGATAAATTA TTTTGAAAAA TGATCAACTA
    1861  ATATAACTAA AACTTTGTTC ACTGTTATTC AACAAACATT TTGTTAGTTG ATCATTTTTC
    1921  AAAATAATTT ATCTCAACGA AGTTATGCAA CTTCAAAGTT GGTTAAATAT TTTGCACAAA
    1981  GTTCGTGAGA TGTAGATCAT TTCGACGGTT TCTTGAGAAT AGAGAGTTAA AACTTGTGTA
    2041  ATGTACGTTT CATACATTTC TGAACAACAA CATAGCAATT TCATAATTGT TTAGAAAAAT
    2101  GAAATTATTT GAATTTTTTG CTTTGAGTTT ATCAAGAAAT TTTGAGCAAA AATGAAAAAA
    2161  ATAAAAATTT CCAAAAAGTG CGGACAATGA TTCTAGGTTC CCGAATAA

//

LOCUS          44     12591 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      44
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.elegans
FEATURES            CHR  ---
     CDS
complement(join(161..330,2644..2705,2759..2928,4062..4587,5700..5825,7326..7553,
10017..10277,11776..11957,12080..12091))
                /translation="MWSTGSKLRQQVDLSTCTLFWLDESVFCADGCLFLSPSFPQQQ
                PPTLYLCFRPYTAFLKWIRILRSRTIYQDVPPPMLQGTISVGEPNSQVSILSIEIDK
                FRSELLKSDMFYSAQVSLNGVKIGVSNAFAPAGNKGPNEMPVVVIDSKFVIPCIPTC
                STNIQLSMNSHSSIGKRGRPCGVTVTICLNENNESICQSQADSTGFVFRAQRNRCPV
                LPLDRYKPLLDMIRECPSSILSWPSQVLPSHLKQFMYTCISHLYALNPQFMSQVIRR
                IINDILVTSTAEDVFRKDSLATGIITQCLRHLFKTPFDEFLLENAQFSQSLKQQNLD
                SGGGAVELLVNFVDQRLLTIPLATRLLTIAAECAGCRFGDEQHEQHLIKRTLSALLI
                LRVLNPIIFSTLNTGIGSQIAKLVQISANSAASQTPTDSLTPAASTIRRMFDRIIFL
                VNQQQPAESPEDSEIGEVHTEWISCLTYMIGHSLTIRSATSSEDVHLPLPVLELIRL
                HHLLISLILEIKSYVVASSASHHAIFNLKFYKATVGEQRESGFQTQDIDPAKPVFWS
                RCTMFKMFFLWAKGTTNRLKFS"
ORIGIN
       1  TTTTTGCCCG ATTTTTCAGC CATCGTTTGA AATTACTGGA GATTGAAAAA TAGCAATTTT
      61  TAGAGTTTGG AGATCAATTT AGAGTCCTCG TCTTTTTTTT GGGCAAAAAA AATTCTTCTG
     121  AAAATGTTCG AAACCACTAT ATTCTTAAAA ATGGCAATTT TTATGAAAAT TTTAATCGAT
     181  TAGTAGTTCC TTTAGCCCAC AGGAAGAACA TTTTAAACAT GGTGCATCGT GACCAAAATA
     241  CTGGCTTGGC GGGGTCAATG TCCTGAGTTT GGAATCCGCT CTCACGCTGC TCGCCCACAG
     301  TTGCCTTGTA AAATTTCAAG TTGAAAATAG CTGAATTCCA TAAATTTAAG TTTTTAACAA
```

```
 361  GAAAACGTTT TCAAGCAAAA AACAACGACA CTCCGATCTT CCGCGTGTTG TTTAGCCTTA
 421  AGAGTTTACA GCGAATTCAA AAAAACAAAT TTTCACTTCA AAAAAAAATT GTGTATTTTT
 481  TTTATTCCTC GAAAGTGTCT GTTCAAGTTT TTAAAAAATT ATAAATTTAA AATTTAAAGA
 541  ATTTTGTTTA TTTAGAATTG AAAAATTCGA AGAGCAAAAA AAAAGACATT CGAGAATAAA
 601  TATAAAACAC AAGGTTTGAG TGGGTATTCT TTTTTTTAAT ACCGGTTTTG CAGGAGAATT
 661  GAATTGAACG GATGATAAAA CTAGGCACAC ACACACACAA AAAACACACA CAAAAATGCT
 721  CATTAAATGA CAATGAAAAA GAGCAAGAAT TGAATGAGAT TTTCGAAAAA AGAGAAATGG
 781  AACAGGGATG GATACACTAC TACTATATAG AGATATAAAG ACTATATATT GTACATATAT
 841  AATGGAAAGA TTATTGGAGG GTGCAAATTT TTTGGCGGGA AAGCTGAGCA ATTTTCGAGT
 901  GTATTTTCGC AAAGAAATTC GATAAATTTT CTGGGTAAAA ATTTGAAAAT TTCCAGAAAA
 961  ATCAATTTTT TCAGTTATTT TTTTTAACTC AATAATATTT TGAGAAGTTT CTTTTTTTAA
1021  AGCGAGGTTT TCAACGCCTA TTTGAAAAAT CGAAGAATTT TTTTTTGGAT TGCAAAAAAA
1081  CATCACTTGG AATTTTCAAA TATAACGTTG AAGATCCTAG ATTTTTTAAA AAAAGAAATG
1141  CCAAGAGTTA GAAACAATAA AAAAAACATA CAAATTTGAA TCAAAAAATA AATTGAAATT
1201  TTCGAAAATT TTTTTGAAAA CTGAAAATTT TGATAATTTC CTGGAAAATC AAATGTTTGA
1261  AAGAAATACA AAAATATCGA TTTTTCGAGT TATTTTTTTG TTGTATTTTC AAAGCGAGGC
1321  CTCGCTTTAA AAAAAATCGA TATTTTAGTA TTGCTCTCAA AATTTTTTTG ATTTTCCAGG
1381  AAATTATCAA AATTTTTAGT TTTCAAAAAA ATTGTCGAAA TTTTCAATTT ATTTTTTGAA
1441  TAGAATTTGT TTTTTTAAAA ATTATTTCTA ACTCTTGGCA TTTCCTTTTT TAAAAATAGA
1501  GGAAAATCGA TAATTCCAAG TGATTTTTTG CAATCCAAAA AAAATTCTTC GAAAAAAATC
1561  GATAATTTTC AGTATGCAGT GAAATTCTTG GAAAAATCGA TAATTTTCAG TATGCAGTGA
1621  AATTCTTGGA AAAATCGATA ATTTTCAGTA TGCAGTGAAA TTCTTGGAAA AATCGATAAT
1681  TTTCAGTATG CAGTGAAATT CAATGAATTT TTGGCGAAAA ATATCGAATT TCGAGTGAAT
1741  TTCCAGAATT ATAAGTGATT TTTAGAAATT CATAAATTTT CGGTTAATTT TTCGGAAAAA
1801  ATCAGAAATT TTTGAGAAAA TAATTCAATA ATGTTCAATG TTTTTAAACA TTTTTCCGAA
1861  AAAATTATTT TAATTTTCGT TATCCAGAAA ATATTTTTGA ATGCAAAAAA TTGTACCGAC
1921  ACGACGGTTC CAGGATGCTT CATTGCGATT TGATCTACGT AGATCTACAA AAAATGCGGC
1981  AGAGTTCCCA ACTGATTTCA CATGGTTAAG AACGTGCTGA CGTCACATTT TTTGAGCAAA
2041  AAATTCCCGC ATTTTTTGTA GATCAAACCG TAATGGGACA CTTAGACACC ACGTGTACAG
2101  AGCAGCAATT TTTTTTTGCA AAACTTTTTT TCCGCGATTC CAGGAACCGT CCAGCTTGAA
2161  AATTTCTTGA AAATCTCGAA AAATGCTCAG TTTTCACGGG CGGGACGGGG CGGTCATCTT
2221  TGAGATGACG TGGTGGTGGT AGATTATTTG ACAGAAAAAG GATAAATATA TATATATATG
2281  TATATATAAA CTGGGGCGAA ATGAGGAACC CGATCAGATC AGGGGGGCTG GCGGGGTAAA
2341  AAAAAGAAAG TAAAGCAGGG CATGTGTGAT GAGATGTCAA AGGTTGGAAA TGGGTGCAAG
2401  AATTTTACAT GAAATTGGAA TTTCAAAATT TTGTGCCGCG CGCGAGTTTC GATTTTCTTG
2461  AACTGCGGGG GCGTCAACGG AAATTTAGAA AAAAGTTATT AGGAAACGGG CGGAGGAGGA
2521  GGGTAAAACG ATGCACATAT ATATGTAGAT TTCACGTGGT GGTGTCAGAG TGTCGGCTTG
2581  ATCTACGTTG ATCTACAAAA AAATTTTTGG ACAAAAAATT GCCGCATTTT TTGTAGATCA
2641  AACCGTGATG GCTCGCAGAG CTCGCCACCA CGTAAGATTT GATTTCTAAA ATAAGAGAAA
2701  TAAGTCTAAA TTTTCGCCTA TTTAAAAAAA AAGTCTCTTT TTTCTCTGTG GGCTCTACAA
2761  ATGATGAAGC CGAATGAGCT CAAGAACTGG TAACGGAAGA TGCACATCTT CTGATGAAGT
2821  GGCACTGCGA ATAGTTAGCG AGTGTCCGAT CATATATGTA AGGCAGGATA TCCATTCGGT
2881  ATGAACTTCT CCGATTTCCG AGTCTTCCGG TGATTCAGCG GGTTGTTGCT GCAACAAAAA
2941  AAAATCAAAA ATCTATGTTT ATTAGAGCGT ACTTGCTTTT TTTTTATTAG AGCGCATCAG
3001  TCAATTTTTT TAGTCCCAGA AATTTCCGAA AAATTCGGAA AATTCCATAT TATAGGAACA
3061  CAAAATTCTG AGAATGCGTA TTATGCAACA TTTTTGAAGC GCAAAATATC TCGAAGCGAA
3121  AACTACAGTA ATTCCTTTAA TGACTACTGT GTCGATTTAC GGGACCTTGA TTAAAACATA
3181  TCTTTTAATC GAGCCCGTAA ATCGACACAA GCGCTACAGT AGTCATTTAA AGGGTTACTG
3241  TAGTTTTCGC TACGAGATAT TTTCGGCAGA GTTCTTGGTT TGTAAACCAC GCCTACATTC
3301  TGCTTTTTTA GCAAATTTTG TGGCGCAAAT ATCCAAAGGA GCACAATATC CGAGACGTAA
3361  ATTCGTGGAA ATTTTTGAGC ATTTTTTCTA AGGGTCACCA CAACGACTGA GGAAAACGTT
3421  TCGGGAGACG CGAAAATTCG GGAATTTGTG GGCGGGGCTT GGAAAATCGC GAGTCACCAC
3481  AAATTTCGGA GGATAACATT TTTGGAGACA AAAATTCCTG GAAATTTGTG AGGTTTTTCG
3541  GTGATTTTTC AAAAATGTCA ATTTTTTTAG AGTCACCACA AACTCTGGGG AAACCAAAAA
3601  TTTCCGGACA TTTATGGGCG GGGCTTTGTT TTGCATACCC CGCCCACTCA AAAAGCTCGA
3661  ATTCTGTGTC CACGCGGAAA ATGGTTTCCC AAGATTTGTG ACCAGTTTCG AGTCTCGTTC
3721  ATAAAATGAG TCACCACAAA TTTCCGGAGA CGCGAAAGTT CGGAAATTTG TGGGCGGGGC
```

3781 TTGGTTTCAC AGTTTTCGAC ATTTTTTAAA GTTGTTTTTT TAATTTTTCG GCATTTTTCA
3841 ACATTTTTCT TCATTTTTCA ACGTTTTTCG ACATTTTGCG ATCGAAAAAA ATGTCGGAAA
3901 ATGTCAAAAA ATGTCGAAAA AATGTCAAAA AAATGTCGAA AAATGCAAAG AATGCATACG
3961 TTTTTAAAAA CATTTTTCTA CATCACATTT GCTGGGTAAT TTTTGGACTA TTTTTCGGCG
4021 CTTTTTCAGG ATTTCCAACT TTCCGGAGGG TCTTTTCCTA CCTGATTAAC CAGAAAAATA
4081 ATTCTATCAA ACATTCTTCG AATCGTGCTT GCCGCCGGTG TCAGTGAGTC TGTCGGCGTC
4141 TGTGAAGCCG CTGAATTCGC AGAAATCTGA ACTAATTTCG CAATTTGACT GCCGATTCCC
4201 GTGTTCAGAG TGGAGAAAAT AATTGGATTA AGCACTCGAA GTATGAGCAG AGCACTTAAA
4261 GTCCTCTTTA TCAAATGCTG TTCATGTTGT TCATCTCCAA ATCGGCAACC AGCACACTCG
4321 GCGGCAATTG TCAAAAGCCG CGTGGCAAGT GGAATAGTGA GTAGCCGCTG ATCTACAAAA
4381 TTGACGAGAA GCTCAACGGC TCCTCCTCCA CTGTCTAGAT TCTGTTGTTT CAGAGATTGT
4441 GAGAATTGTG CATTTTCGAG AAGAAATTCG TCGAATGGAG TTTTGAATAA ATGTCGGAGG
4501 CATTGAGTGA TAATTCCAGT GGCAAGGGAA TCTTTTCGAA AAACATCTTC AGCTGTTGAG
4561 GTTACCAGTA TATCATTGAT TATTCGTCTG TAAGAGAAAT TTTCAGGAAT TTTTTAGTAT
4621 CATTGGAATT TTCTACATAC TGATAATAGA TTTTCTTCAC GAGTTTCTGA AATTTCCTGA
4681 TTTTTTTTCT GATATATTTT TAGGTTTCGC TGAATATTTT AAATTTATCT AAAGATTTTC
4741 TGGAATTTTT TTTTAATTTT TAAAGATTTT TCAGAACTTT CAGGGACTTT TTTTAAGGTT
4801 TTAATTTATC GGGCTTAAAA AATATTTTTT CTAAAATTTT CCAATAAAAA TTTAAAAATT
4861 AAAAATAAAT TTATCTAAAT TTTTTTTGTT GTTGTAATTT TTAGAATTTT TCGAATTTTT
4921 CTGAGAAATT CTAAATTTTC GACAGAATTT TATGGAATTA TTTAAGATTT TTCAGAATTT
4981 TTAGAGATAA TTTTTAACGT TCTAATTTAT TGGGTTAAAA AAAATCTTTC TCTCAAACGT
5041 TCGGAAATAA TTTTACATTT TCTGTGATTT TCTAAAACTT TCCAAAAAAT TGTCGAATTT
5101 TTTGGAACAT TCTAGATAAT TTTGATTTTT TTTATTCCTC AAATTGAATA AAATGATGAA
5161 TAAAATAAAT AAATAAATAA ATAAATAAAT GCAATGTTCT GTGTTTTCTG AAACATCTTG
5221 AATTTTTCTG AAATATTACG GTCATAGGGA AAAAAATTCC GGAATTTTCA AAATTAGTTT
5281 ATGAATTTTT TTTTTGAAA AATATTCTAA ATTTCCGGAC TTAGCGAAAT GTTTGTGCTA
5341 CTTTATATAT TGTTAGGATT AGCTGATTTT CGAAAATTTT CTATAACTTT CTAGTATTTT
5401 TTTAATATAC TTTAAATTTA TCTAAATATG TTTTTGTTGT AATTTTTGGT TTCGAATTTT
5461 CTGGTATTTT TGGAATTTTC CGAATTTTTC TGAGAAATTC TAAATTTTCG GCGATTTTCT
5521 AGAATTTTAT GGAATTTTTA ATGATTTTTC AGAATTTTTA GAGATAATTT TTAACGTTCT
5581 AATTTATTGG GTTTAAAAAA AATTCTAAAA TTTTCGGAAA TAATCTAAAT TTTCTGTGAT
5641 TTTCTAAAAC TTTCCAAAAA ATTTCCTAAT TTTCAGTATT TTTTTCTAAA AATGCTCACC
5701 TTATAACTTG GCTCATAAAT TGTGGATTAA GCGCATATAA ATGTGAAATG CACGTATACA
5761 TAAACTGTTT CAAGTGTGAC GGAAGTACTT GTGATGGCCA CGAGAGAATC GACGACGGGC
5821 ATTCACTGCA ACAAAAATTT CATAAAATTA GCTAAGGTTT AGGCATTTTT TGCACATTTT
5881 TTGCACATTT TTAGCACATT TTTAGCATAT TTTTTGCACA TTTTTGCACA TTTTTGCACA
5941 CTTTTGCACA TTTTTAGCAC ATTTTTAGCA CATTTTTAGC ACATTTTTAG CACATTTTTA
6001 GCATATTTTT TGCACATTTT TGTACATTTT TTTGCACATT TTTGCACATT TTTAGCATAT
6061 TTTTTGCACA TTTTTGCACA TTTTTGCACA TTTTTGCACA TTTTTAGCAC ATTTTTAGCA
6121 CATTTTTAGC ACATTTTTAG CATATTTTTG CACATTTTTG TACATTTTTT TGCACATTTT
6181 TGCACATTTT TAGCATATTT TTTGCACATT TTTGCACATT TTTGCACATT TTTGCACATT
6241 TTTAGCACAT TTTTAGCACA TTTTTAGCAC ATTTTTAGCA TATTTTTGCA CATTTTTGTA
6301 CATTTTTTTG CACATTTTTT TTCACATTTT TGCACATTTT TAGCATATTT TTTGCACATT
6361 TTTGCACATT TTTGCACATT TTTGCACATT TTTAGCACAT TTTTAGCACA TTTTTAGCAC
6421 ATTTTTAGCA TATTTTTGCA CATTTTTGTA CATTTTTTTG CACATTTTTT TTCACATTTT
6481 TGCACATTTT TTGCAAAATT TTTGCTTTCC ATTTTTTGAT TTCATCATTT TTGATGTCAA
6541 AATTAAATTT TGAATATATT TATTTCCAAT AATCAAGTTA ATTTTAATGG CAACATTTTA
6601 ATATTTTCAT TTAAAATTTT CGAAATTTGT TTTTTTTGTT TGATTATTTC TAACTCTTGG
6661 CATTTCTTTG TGATTTCTTG CAATCCAAAA AAAAATCTTC GAAAAAATCG GTAATTTTCC
6721 GTATGTAGTG AAATTCTTGG AAAATCAATA TTTAAATTAC AACGAATTTT GAAAAAATCG
6781 AGTGGTTTTT TTTGCGATTG CAGTGAAATT CGATGAATTT TTGCGAAAAA TATCGAATTT
6841 TCCGTGTATT TCCAGAAAAA TCACCACATT CATTTTTAAT TCAAATTCGT ATTTTTTAAA
6901 TCTAAATTTT CATGTTAAAA CTGCATTTTT TTAAATGTTT TTGAGTGAAT TTCGAGTCTT
6961 CTCACTACAA AATTAATCAT TTTAGAGGAG TTTCAAAATT GCGATTTTTT ACCAAAATTG
7021 CCCATTTTTT CCACTTTTAA ATAGTTTTTG ATGGGTTAAA GCTAGATTTT CTGAATTCAG
7081 CATATATGAA TTACCCGTTT TCAACAAATT CAGAATTTCT TCATTTTTGC CCAATTTTTT
7141 TCTCAGTCAT TTAATTTTAC TGTCTTTTTT TAGGGCAAAA AAAATAATTT TTCTGAAAAT

```
 7201  GATCGAAACC ACTATATTCT TAAACAAGGG CAATTTTTTA ATGTTTTTGA GCGAATTTTG
 7261  AGTCCTCTAA CTCCAAGTTC GAACAATTTC CAATTTATCG CGTCGAGACC CCCCCCCCCC
 7321  CGTACCGTAT CATATCCAAC AACGGTTTAT ATCGATCCAA TGGTAGAACG GGACACCGAT
 7381  TTCTCTGTGC ACGAAATACG AATCCAGTTG AATCGGCTTG TGATTGGCAT ATACTTTCAT
 7441  TATTCTCATT TAGACAAATT GTCACTGTCA CACCGCACGG GCGTCCTCTC TTTCCAATTG
 7501  ATGAGTGACT ATTCATTGAT AGTTGAATAT TTGTACTGCA CGTTGGAATA CATCTGCAAA
 7561  AAAATCAACC GAATATCATA ATAAAACATT CCCGAAATTT TTTGGATATA TTTTTTTTAA
 7621  ATTTTCCAGT CAAAGTTTTG GTAAATTACC ACAAACATTT TTGAAAATAA TGAGCTTTTA
 7681  AAGGTGGTGT AGTCGAATTT TTTATTGCCA TATTAGACTC AAAATTGTCT GAAAACACCG
 7741  AATTTCATAA TGAAACTTCT TGAAAACTTC TCAAAAAACA GTTATGACGG CTCAAAAAAT
 7801  GGCCTAAAAT TAGTTAAAAT TCGAAATTTG ACCGACTTGT CAAGCGGCTG GAAATTACCT
 7861  TTTTCTGAAA TTACCGTCAA ATTTTGAGTA TACAAGTCAA TTATCTTGCG TTTTCAACTA
 7921  GATTCAGGTA TTTTAAAGTC GATGGACGGC GAGATTTTTG TATTTTTTTC AGCAAATCTC
 7981  GCCGTCCATC GACTTTAAAA TACCTAAATC TAATTGGAAA CGCAAGGTAA TTATATTGTA
 8041  TACTCAAAAT GTTACGGTAA TTTCAAAAAA AAGCTAGTCT CCAGCCGCTT GACAAGTCGG
 8101  TCAAATTTCA AATTTTCACT AATTTTAGGC CATTTTTTGA GCCGTCATAA CTTTTTTTTT
 8161  GAGAAGTTTT CAAAAAGTTT CATTATGAAA TTCGGTGTTT TCAGACAATT TTGAATCTAC
 8221  CGTATATCCT CTATTAGTAA GGCGTGCAAA ACTAATTTTC GAGCATTTCT TTTTCTTTTT
 8281  TTTTCCATTA ATTTATGGCA TGATATCATC TATTTCAATA ACAACGTATG AACCAAAATG
 8341  AACGAATTTT ACGACTGATA CGCAAAAATT GTCCGAGTTG TACTCATATT TTGCCAATTT
 8401  TGACTTGTTA TACCAAGTCT GTAAGAATTT TCCTGATTTG TAAAACGATT TTATTTTGCA
 8461  AATTTTGAAT TCCTGAAAAT AGGAAACAAA TGAAGGGGTG CAAAACTAAT AGAGAATATA
 8521  CGGTAATAAC GCAATAAAAA ATTCTACTAC ACCACCTTTA ATGAAACCAA GAGTAGTTTA
 8581  GCATGTTCCG ACCCTTACAT TATTTTAATA CGAATAATTA AAATAAAATA AAAGTATTAA
 8641  AAATGTAAAA AAAATTTTGG TCGATTTCCA AAATTAAGAG TGGCAAAAAC TGAGTAATTG
 8701  CCACTTTTTG ACAGTAAATA GAAAATTTTC AAAATTGCAA AAACATTTTT TTTTAATTAA
 8761  AAATGTTTTT TTTTTGTTGA AAATTTTTCA AAACAGTAGA ATATCGATTT TTTCTGTTTT
 8821  TTCTTATAAA ATCATAACTT TCATATAACC AAAAACACTT ATTTGAAGCA ATTTTTCGGA
 8881  AAATGTACCG AAATGTTTTT TTCTATAAAT ACTTTACTTC AGAAAATTGA AGTTTTTATG
 8941  GTTTTTCTGG AAAAATTTCC AAAAAATTCT ACTTTTTTCC AGCAGAAGAA TAAACTACGA
 9001  TTTTTTTTTT GGAAATCCAG AATTTTCATT AAAAATCTGT AAAAAAAATT TTTTTTAGCC
 9061  TTTTTCCGCA ATTTTTCAAC ATTATTTTGT TTAATTTTTT CAGAAAAAAC CTAGTTTTTA
 9121  GGTATCTCTG TTTTCAAAAA AAAAATTAAA TTTTCTAAAA AGTTTCAGAA AAAATTTTTT
 9181  TTTTAAAAGA AAATCAAATT TTTAGTATAA AATTTTTGTC AAACTAATTC AAATGAAATT
 9241  TATTTATTTT TCAAAAACTC AATTACGATT TTCGGCAAAA ATATAAATTC ATGTTTTTCA
 9301  AATAATATTC AGAAAATCAC ATATTTCCTT CAGAAAAAAA ATCATATTTT TAATTTAACC
 9361  AATATACTTT TTCCAGAAAT TTTCCGGAAA AATACCTTTC AGAAAATTGA TTAATTCAAT
 9421  TTTTTCAACA TAACCTCTAG TTTTTTCTCA AAAAATTAAA AAAAAAAGAT TTAAAAAAAG
 9481  GGTTTCTGAA ATAATTTTCC ATCAGATAAT TATTTTTATT AAAAATAAAA ATCCGAACCT
 9541  ACTTTTTTCC AGCAAAAAAA ATCAATTTTC GAAATTATTT TCCGGATAAT AAAAAAAATA
 9601  ATACCCTAAA GTTTTTTTCA AAGAAAATTA ATTTTTTAAT AAAAAATTAT TTCAAATAAA
 9661  TTACCGTTTT TTTTTTCAAA AGTGAAAATT TAAAAATTTA TTATAATTAA ATTAATTTAC
 9721  TTTTTGAAAA AAAAAGTGAA ATTTGTTTGA AACAATTTTT TATTAAAAAA TTGATTTTTT
 9781  GAAATTTTAT AAAAAAACTA ATTCTGGAAA ACTGCTGGAA AAAAGTAAGT TAAGAATTTT
 9841  TTTTTAAATT TTGATTTTTG GAAATCAAAA TTTTCCTGCA AAAACCACTT TTACAGAAAA
 9901  TTAAAATTAA AAAAAAATTA TTTGATGGAA AAAAAATAAT ATTCAGAAAA TGACAAAACA
 9961  AAAAAGCTTC AGTAATTTTT CGGAAAATAA TTTCTATTTT CAAAAAAAAA ACTCACGGAA
10021  TAACAAATTT CGAGTCAATA ACAACGACAG GCATCTCATT GGGTCCTTTA TTACCAGCAG
10081  GAGCAAATGC ATTAGAAACA CCAATTTTGA CACCATTCAG TGATACTTGT GCCGAATAGA
10141  ACATATCAGA TTTTAAGAGT TCCGATCGAA ATTTATCAAT TTCAATTGAT AATATTGATA
10201  CTTGAGAGTT GGGTTCACCA ACTGATATTG TTCCTTGAAG CATTGGTGGT GGCACGTCTT
10261  GATAGATTGT TCGGCTTCTG CAATTTTAGA ATAATTTTTT TTCAAAATTC GGGAATTAAA
10321  ACATTAATTT AATTTAAAGT GTTATACATT TTATTTCTAA TTGAACATTA TTTAAAGAAA
10381  CAGTATCCTC ATCGAGGATA CTATTTCTTT AAAAGTGGAG CACTGAAATT TGAGACTTTG
10441  CGTTTTTAGA CCCAAAATGT TCCAAAACTA CCAAATTCCG TAATAAGACG TTCTGAAAAT
10501  TTCTCAAAAA AAAGTTATGG TGGTTTGAAG TTCTGGAAAA AATGGTATAT TTTTTTAGCT
10561  AAACTCTCAA ATTTTGCCAA CTTTTCAGCT GTTGCAGCAG TTGGAACTTA ATTTTTATTT
```

```
   10621  GATTACCACA CTTTATACAT ATTATCGTAG TTTATTACGC GTTTATTTCT TGATTGCCAT
   10681  GCTTTTTCGG TCGATGGGTG CACAAAAATG TATTATATTT GTGCAACCAT CGATCGAAAA
   10741  AGCATCTCAA TCAACGAAAA AACGCGTAAT AAACTACCAT AATATGTATT AAATAGTGTT
   10801  AATTAAATAA AAATCAAATT TCAACCGCTG CGACACTGAA AAGTTGTCAA AACTTGAGAT
   10861  TTTAGCTAAA AAAAATTAAA CCATTTTTTC CAGAACTTCA AACCACCACA ACTTTTTTTG
   10921  AGAAATTTTC AGAACGTCTT ATTACGGAAT TTGGTAGTTT TGGAACATTT TGGGTCCAAA
   10981  AAAGCAAAGT TTCAGATTTC GGTACTCCAC CTTTAAAGAA ATAGTATCCT GAATGAGGAA
   11041  ATTGTTTTAA ATGCAATATT CGTGCTCAAA GTTACGAATA TAACAAAAGA AAAACGGGGA
   11101  AAATCCGTTA AAATTGAGCA TTTCAAGTCA GAAAAATCGA CAGCGAAATC CTAGGCCACG
   11161  AGGATTTTTT TTCAAATTTT TTTTGTTCTG AATCGTATTT TTCCAAGGAT TTTCTTCAAT
   11221  ACCTTCTTCC CAGTTTTCAT CAACTTTTTG AGAAGGAAAA AAAAATTTCA ACAGAATTGA
   11281  GAAATACTTG TTATTTCTAC ATATTCGATG AATTGTTTTA GCAAATTAAA TATTATAGAA
   11341  AGCATGTGAA AATGTTTTAC TTTTTGAGAA ATTAAACGTG TTTTTTTTCA ATTTTCGTAT
   11401  ACTTTCATAG AATTATGCAA TGCGCATATA GATATAGATT TAAGGAGTAT GCAATTTTTA
   11461  AAAAACAATT TGAAAATTAG TCGTAAAGAA CTTCTCGAAT TAAAAAAGTG AGATTTTCTT
   11521  AGAATTTACT ACTACATATT TGGTTTTTTT TTCGATTTTA TTTTATTTGA GACACTTTAA
   11581  TAGTAGAAAC AATTCCGGTA GTTTTGTTTG AAAATGCAAT TTTTTGATTG TTGAAAAATT
   11641  TTCTGAAATT TCGGTATTTT TAACCACTTA AATTTAAAAT ACCTATCTAG AAACCGAAAC
   11701  CACCAAATTT TCTAAAATCT TGATTTTTAA AACATTCCTA CTTAATTGAG AAATTCAAAT
   11761  CAACAGAACA AAAACCTCAA TATCCGAATC CATTTGAGAA ACGCTGTGTA CGGTCGAAAG
   11821  CATAAATAAA GTGTCGGCGG CTGTTGTTGA GGAAATGATG GTGACAGAAA TAGACATCCG
   11881  TCTGCACAGA ATACAGACTC ATCCAACCAA AATAGAGTAC ACGTCGACAA ATCCACCTGC
   11941  TGCCGTAGTT TTGATCCCTG GAAAATTGAT ATAGGGAATT ATTTAAACCT GATCGGGAAT
   12001  CAATATAATA AATCTGGAAC GAAAAATAAC CAAGATTATT ATGCATATGC ATATTTGGTT
   12061  TATTATTATG TATATGTACA GTGCTCCACA TAATGATACG GCCACCCCAG AATTATGATA
   12121  GAACGCCTCT TTCGACGTGT ACTGAGAAAT AATTTTTTAT TGTGTGGAAT GATACTCTTG
   12181  AACAACATTT ATTGATTATA CAGCGAAAAT AAAATAAAGC TGTGCAAGCT AGAAAAATCA
   12241  ATTTTTCTGA TTTTTGCACT TTTTCCAATC CACAAAATGA TACGGCCACC CTTTGTTTTT
   12301  TTTCTTATTT CGTTTTCTTT GCAAGTTTTT TTTTGCTAAA CGTTAGTGTT CATGTTCATT
   12361  AGTGTTTTTA CAGCGATGGG TCGTGGAATT ACTTTAACTG ACTACGAAAA AGGAGAAATT
   12421  GATGCAAAAC TATCTCAAGG CTTGTCGCAT CGTCAGATTT CTCGTGATTT GAAACGCTCG
   12481  AGAGATATGA TCACTCGATA TGTTTCCAAT CCTGCGGTTT ATGGCACCAA AAAGTCTTCT
   12541  GGTCGCCCAC CACTTCTTTC TGATCGAGAC AAGCGGAAAA TCTGTCGTCG A

//

LOCUS          45     1203            2668 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      45          1203
VERSION        45
ORGANISM   C.elegans
FEATURES             CHR  ---
     CDS             join(1000..1053,1111..1202,1260..1668)
                     /translation="MFLRLSRPHHHDTINILTSTYQSLVNSARILFVEKGYQAVSID
                     EISGKAACYKQQLIMIDAYITTKTDLTNGWSALESIFEHYLDYIIDNNKNLIPIQEV
                     MPIIGWNELEKISLEYITGKVNAIVSKLIQENQLKAYDDDVLKNLLNGWFMHIAIHA
                     KNLKELADKKGQFIAIYRGFLLSLKDK"
ORIGIN
       1  AAGAGGCTGA CTAAAAGCAA TCCAATCTCA TGCCAAATTC TATTGTATAA ATCTCGCTCT
      61  AAAGGCCGAA AAAATCAGCG CTCGACACGG ACTCATTGTC ACCACCCGTC ACCTAAAATC
     121  TACTCAGCGT CGGCAAAGGA GCCATGGATT CTAGCAACTA ACTTACCTGT TGAAATTCGA
     181  ACACCCAAAC AACTTGTTAA TATCTATTCG AAGCGAATGC AGATTGAAGA AACCTTCCGA
     241  GACTTGAAAA GTCCTGCCTA CGGACTAGGC CTACGCCATA GCCGAACGAG CAGCTCAGAG
     301  CGTTTTGATA TCATGCTGCT AATCGCCCTG ATGCTTCAAC TAACATGTTG GCTTGCGGGC
     361  GTTCATGCTC AGAAACAAGG TTGGGACAAG CACTTCCAGG CTAACACAGT CAGAAATCGA
     421  AACGTACTCT CAACAGTTCG CTTAGGCATG GAAGTTTTGC GGCATTCTGG CTACACAATA
     481  ACAAGGGAAG ACTTACTCGT GGCTGCAACC CTACTAGCTC AAAATTTATT CACACATGGT
     541  TACGCTTTGG GGAAATTATG AGGGGATCTC TCAGTGCATT GCCTCCAATT CCCATAATTT
```

```
     601  ATTACGCCGA TAATAACTTG GTGTAACCTT AAAAATGTAC TTAAATCGAC GTGTAAAAGA
     661  TTGTTGGGAA TCAAATTGAT ATTTTAATGC GATCTCAAGG ATAGTTTTTT TCGTCAACCT
     721  CAACTCAACA GCCGCTTTCG TCAAACGACG AGCACGAATA TAGCTAGCCA GTGTGACCCC
     781  TGTTACTTTT TTGAACAGCC GCTGAAAATA CCACTTGGTA TAACCCGCTT TATTCGCCAC
     841  ATCATCAAGC AGTAAAGACT GATCTAAATT ATGTTCAATC CATAATAGAA CATCTTTGAT
     901  AACCGTTGTA TGAAACTGCT TATCATCATA TCTTAATTGG ATGTTATTAG CTTTATTTTG
     961  ATAGCGAGAA TGCTGTTCAA TATACATAAA ATAACCTAAA TGTTCTTAAG ATTGTCACGA
    1021  CCACATCATC ATGATACCAT AAACATACTG ACGGTATGTT ATTTTAAATC TATCATGGAA
    1081  AATAAAAATC ATCAACAAGA AAATTTTAAG AGTACCTATC AATCACTGGT TAACTCAGCA
    1141  CGAATATTGT TTGTTGAAAA AGGCTATCAA GCTGTTTCAA TAGATGAGAT CTCGGGAAAA
    1201  GCGTTGGTGA CCAAAGGTGC CTTTTATCAT CACTTTAAAA ATAAAAAACA ATTACTCAGT
    1261  GCCTGTTATA AGCAGCAATT AATTATGATT GATGCCTACA TCACAACAAA AACTGATTTA
    1321  ACAAATGGTT GGTCTGCCTT AGAAAGTATA TTTGAACATT ATCTTGATTA TATTATTGAT
    1381  AATAATAAAA ACCTTATCCC TATCCAAGAA GTGATGCCTA TCATTGGTTG GAATGAACTT
    1441  GAAAAAATTA GCCTTGAATA CATTACTGGT AAGGTAAACG CCATTGTCAG CAAATTGATC
    1501  CAAGAGAACC AACTTAAAGC TTATGATGAT GATGTGCTTA AAAACTTACT CAATGGCTGG
    1561  TTTATGCATA TCGCAATACA TGCGAAAAAC CTAAAAGAGC TTGCCGATAA AAAAGGCCAA
    1621  TTTATTGCTA TTTACCGCGG CTTTTTATTG AGCTTGAAAG ATAAATAAAA TAGATAGGTT
    1681  TTATTTGAAG CTAAATCTTC TTTATCGTAA AAAATGCCCT CTTGGGTTAT CAAGAGGGTC
    1741  ATTATATTTC GCGGAATAAC ATCATTTGGT GACGAAATAA CTAAGCACTT GTCTCCTGTT
    1801  TACTCCCCTG AGCTTGAGGG GTTAACATGA AGGTCATCGA TAGCAGGATA ATAATACAGT
    1861  AAAACGCTAA ACCAATAATC CAAATCCAGC CATCCCAAAT TGGTAGTGAA TGATTATAAA
    1921  TAACAGCAAA CAGTAATGGG CCAATAACAC CGGTTGCATT GGTAAGGCTC ACCAATAATC
    1981  CCTGTAAAGC ACCTTGCTGA TGACTCTTTG TTTGGATAGA CATCACTCCC TGTAATGCAG
    2041  GTAAAGCGAT CCCACCACCA GCCAATAAAA TTAAAACAGG GAAAACTAAC CAACCTTCAG
    2101  ATATAAACGC TAAAAAGGCA AATGCACTAC TATCTGCAAT AAATCCGAGC AGTACTGCCG
    2161  TTTTTTCGCC CCATTTAGTG GCTATTCTTC CTGCCACAAA GGCTTGGAAT ACTGAGTGTA
    2221  AAAGACCAAG ACCCGCTAAT GAAAAGCCAA CCATCATGCT ATTCCATCCA AAACGATTTT
    2281  CGGTAAATAG CACCCACACC GTTGCGGGAA TTTGGCCTAT CAATTGCGCT GAAAAATAAA
    2341  TAATCAACAA AATGGGCATC GTTTTAAATA AAGTGATGTA TACCGAATTC GATTGCGTCT
    2401  CAACCCCTAC TTCGGTATCT GTATTATCAC GTGTATTTTT GGTTTCACGG AACCAAAACA
    2461  TAACCACAAG GAAAGTGACA ATATTTAGCA ACGCAGCGAT AAAAAAGGGA CTATGCGCTG
    2521  AGAGATCCCC TCATAATTTC CCCAAAGCGT AACCATGTGT GAATAAATTT TGAGCTAGTA
    2581  GGGTTGCAGC CACGAGTAAG TCTTCCCTTG TTATTGTGTA GCCAGAATGC CGCAAAACTT
    2641  CCATGCCTAA GCGAACTGTT GAGAGTAC

//

LOCUS          1     1296 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      1
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          join(500..586,677..796)
                  /translation="MNSQLSCNLESARKLVDELKYHTNISRLPVSQAATELLNYVLE
                  NQGRDFLLVKPPDNPFKPKSSCMFL"
ORIGIN
       1  TCGCCTAATT AAATTATATT TTTTAAACTT TAATATAGCA AAAAAAACAA TAATAAAGTA
      61  AGTACTCATT TTCAATTAAA ATATTTCGTC CCCGATAGTC TCGAAGCAGC TGTGAGGTTG
     121  TAAAATAGAA GATCCTCAGA TTTGATAATT TATGTTGTTT GTATGCGGAG GGTGTGGTCA
     181  CAAAAAATGT TAAGATTGGT GTCTAAAATA TAGTAATTGG TCGAAACCTT TTGGTAATGT
     241  ATGGTCGTAG AATGTGCTGT AGTTATAGTT CCTGCAATTA TGTTATTTAT AAGTGCGTTT
     301  GTTGAAAAAG TGTATGCGCG TATTCGGCGT ATAACAAACG TTCGAAGTGC AATATCCAAA
     361  ACTTTGACCT AATTTAGCCG CGTACAGACG TATATGTGTA TTCCTATTGC TGCAATGGTA
     421  TAACCTATTC TACTTCTTGT AGGGTTCATT GCCTTACGGT AGGTTGTGCC TATAATTATA
     481  GGAATCAACA AGAGCAATCA TGAACTCCCA ATTATCTTGT AATTTGGAAA GCGCGAGAAA
     541  GCTTGTGGAT GAACTGAAGT ATCACACCAA CATATCAAGA CTACCGGTAA GCAAATACTT
```

```
     601  AATTATGTAT GTACTATATC TAGTTTCCTT ATTGATATGT ATTTTACCTA TATCTGATTG
     661  GCATTTTTCA TTACAGGTCT CCCAAGCCGC TACAGAGTTG CTGAACTATG TTCTAGAAAA
     721  CCAAGGGCGA GATTTCCTGC TGGTGAAGCC TCCAGATAAC CCCTTCAAAC CCAAAAGCTC
     781  ATGCATGTTT CTCTGACAAA AGCTGCATGC AGCACTCCAA ATAAACTTCT ACTTTTGGAG
     841  TGTCTAAATA GGAACTCTAT GATTCGGTGC TGAATCATTC TGCAATTGTA TTTTTAAGCA
     901  ATCGCGAAAA GTGAAAATTG TGTGCTTTTA TTCTCGGTAC AAGTCTAAGC TCTTATCTCT
     961  TCATTTAAAT CTTGTAAGTT GTAGCAATAT TCTTTTTCCT TATATGGTAC CAAACAAACC
    1021  AGTGGCCTGT GAAGTGTTTA TCAGTGATGT ATGCTTCTAT GTAAATACGT AAGGTATAGC
    1081  ATGCCTTAAA TGAACGATGG TTAAAAGTAT TCAACTTAAG TTGCACTCTT GTTTGGATGA
    1141  ATGTCAGTTT GTTTTTTCGC TCGTTGTTAG TTAAATGAAC TATGTCCTCT TCAGTTGACT
    1201  AAGGAAAATA GTACTGCTGT CTTTATTAAT TTAAATTTCT TTTACCAACG GCTCTTTCAG
    1261  TGTCTTGTAA CAAACCAATG AAATACATCA TTTATG

//

LOCUS          2     1425 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      2
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          complement(join(415..571,831..925))
                  /translation="MTDVIHDEVEIEDFEFDEETETYYYPCPCGDKFEISLEELQSG
                  EDIATCPSCSLLVKVIYDPDDFLLDEAKEVGTNESVKQTA"
ORIGIN
       1  AGAGAATTTA ACAAACACCT GTAAAACAAA AAAATCATAA CCTTACACTT ATTTTCAGAA
      61  AAATTCTCAC TGATTATTAT GTTATTTATA TTTGCTTGTT ATTTACACTG TGTTAGCAGC
     121  TGCTATTTTG GTGCAGTGGT TATGCACCTG TAACCCAGAG GTTACTGATT CAAGGCTTGT
     181  CGTTGGTACC ATTGCGGGTG TACACTGTCA TAATAAAAAA GTTTGACCCT TTAAAATGCT
     241  AGAAAAATAC AGAAACCAAA CTAACTATCT TCTAATGACT TATGACTGTA CAGATTATTC
     301  CAAGCTCATT GTGGGTGTAT GCTGTCATAA TAGGACAATA GATCAAAAAA AATGCTAGAA
     361  AAATACAGAA ATAAAACTAA ATACCTTCTA ATGACTTATG ATGTACAGAT GATTTTAAGC
     421  AGTTTGTTTC ACGCTTTCAT TCGTTCCAAC TTCTTTAGCT TCATCTAGCA AGAAATCGTC
     481  AGGATCATAA ATTACTTTAA CAAGTAATGA ACAACTTGGG CAGGTAGCAA TATCTTCACC
     541  ACTTTGTAAC TCTTCAAGTG ATATTTCAAA CCTGTAAAGG TAGTGTTACT AACATGAACT
     601  TCTGATAAAC ATATATGGTG CGAAACATTT ATGGAGATTA CTGATTTGAA ATACAAATAA
     661  GTTATTGTTC TAGTTTATAA ATTGCAAACC AAATAAAACA TATATATATA CATGGTCATT
     721  TTGACATAAA CATCAAAATA TTAATATAAA ACTCAAATTC GCCAACAAAA ACTGTATAAC
     781  AATCACATTA TCCTAGCAGG AAGCTACAAT TTAGGCACAA AAAAACATAC TTATCTCCGC
     841  AAGGACAAGG GTAATAATAC GTTTCAGTTT CTTCGTCAAA TTCAAAGTCT TCTATCTCGA
     901  CTTCGTCATG AATTACATCA GTCATTCTTA TATAAGTTTA AACACATAAG GAACTAGCCA
     961  TATATGTCTT TAAAAAAAGG GGGAAATCAA GTTATAGTAT TTACTAAAAT ACGAAAAGCA
    1021  ACTTTACTAT GGGCACGTGT TACAAAATAC GGGGAATCCC CTATGGTCAG TGTTACCTAT
    1081  AAATTTTCAC ATTTTTTCTT AACATCGTAT GCTTATTGTT GTTGCTTGTA CATATTTTGT
    1141  TGGGGAACAC TTGTTTTAAG GTATCATATT TGGGGAGCAA TAAGACACTA ATGTCTTCCA
    1201  CAACCAAACG AGTTTTTTTG AGATGTAAAT GTAATATTTT GTAATGTGTG CCCGTTTATT
    1261  CAGAGCTGTG CCCTTGTTTT ACGCTGTTGG CTCAAAATAA TCAATTCAAG GGCAAGTGCG
    1321  AATAATGTCA ATGCCGAAAC CTTACTCAAA CCGACGAAAG TACATATGAA ACTTCTTACT
    1381  TCTGTTTGCC GTTAAGTTTC TTATGTGTAT CAAATACACA CTGCA

//

LOCUS          3     2311 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      3
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          complement(join(500..584,1594..1811))
```

```
                /translation="MSRNYSKYALNVARLSSRIFCEPLPESANRGKHVDKYLRELPY
                YKDRSFTHYYPPVSNINLLLRTLRNHGLFVDEHKDFKEEMELQRNLRGKGRSLKKSL
                "
ORIGIN
       1  TATTGTGTCT TTACAATTTG AAGAAGATAT TAAAAAGTGA GAACTTGTGA GTATATATAG
      61  TGGCTGAGGG TTACAACTAA AAGATATATG AATTTTAGCA CACACTAAAG ACCAGACCCC
     121  ATCTCATGAC ACACAATCAA TAAATAAGAT ACATAAATAC AACATAAACT TGTTAGGAAG
     181  GTGGTCAAGG GAAATGAAGG GTCGGAAAAC AATGGGATGT GATTTTTGGT TGTGATATTA
     241  CTCATATATT TTATTAACTT TCAATCTTTG GCTTTAAATT TGCAGTTTGA AATATGTAAT
     301  ATTTGCAGTT TGAAAAATAT AATAGAGTCT ACACCTTGGC ACATAGGAAG CAAAACACAT
     361  GCGAGCATGG AAGCTCTACA GGCTGCATAC AAGGTTGGAG ACAGATTGCG CAATCTATTT
     421  CTTCTGTCAT TGTTTATCAT ATAATAACAA TCTGTTCGGG CAAGGGTCTT TCAATAGTTC
     481  GAGTTTGCTT AAAATAAATT TAAAGAGATT TCTTCAAAGA TCTTCCTTTT CCTCGAAGAT
     541  TTCTTTGAAG TTCCATTTCT TCTTTAAAGT CTTTATGTTC ATCTCTGTAA TCAAAAGAAA
     601  TTACATAAAT ATGGTATATC AGGAACTGTA AAAGTCTCCA AATATTATAC AGAACTTATA
     661  GTTCTGTATA GTATGCTTTT CAGATGTTTG TGTGGTATTG TGCTACGATT ATCTATCCTG
     721  TAAATCACAT TTTAACCAAC ATCTCTAAAA TGTTCACTAT ACTTTCTTAC TATTAATTTA
     781  AGAAAAGACA GGATAACTTT TCCGACAGAT GGCACTGAGT TAATTTATTA TTTTTGTGTG
     841  GAGAAAAAAT GTGTGTTTTT TTAAATTACC TACTAAATAA TAAATCAATT TCTAGTTAAC
     901  ATGGGTAATA CTGCCGCTGG TGACCAGTGA AAATTTTGGT TTCAAAATGT TCCTTACGTC
     961  TCCCTGATAC CGGCTGTATA TTTGTTTTTG AAAAATCTCC GGCGTTCGTG TTGAAAATTA
    1021  ACTATGAGAA TGCAAAAGCA TGCTTCGTCA CCGTAATTCG CCTGACTTGC TTAAGGTTAA
    1081  CGATATACTA CGTCACAATA GCGATATGAT ACGTCACAGT TGTGACATAT TACGAAAGTT
    1141  ACAGCACGTT ATAACTTTGT TTTTGCACTG TTTTATGCTA GGGAAAACCC AAAGCCCCAT
    1201  TCCTAACCCC TACTTTTAAC CCCCAACCCC TAAATCCCCC CAAAAAACTA ACAATTTACG
    1261  TAATCTGTAA CGTAGAAATC AAACGTGGTA ATTCGCATTT TATGACGTAA CAATCTAACC
    1321  TTTGTCTAAT CCATGACGGC TGGTGTGGAA AATGACGCAG CACGATTTCG TGCATTTTCA
    1381  AAGTTAATTT TCGACTTGAA CGCCCGAGTT TTTTCAAAAA AAATTTATAC AGCCGGTATC
    1441  AGGGAGACGT AAGGAACATT TTGAAACCAA AATTTTCACT GGTCACTAGC GACAGTATTA
    1501  CCGTTAACGT GTATCCTTAA AATAATGGTG ATCGTTTGGT AAAAGCAATG TACTTAGTGA
    1561  CATTTTAACT ATTTCAAAAA GCACAAAACT TACACAAAGA GTCCGTGGTT TCTTAGTGTT
    1621  CGTAGTAAAA GATTAATATT TGATACGGGC GGGTAATAAT GAGTGAATGA TCTGTCTTTA
    1681  TAGTAAGGTA ACTCCCTCAG ATATTTATCG ACGTGTTTTC CTCGATTGGC AGATTCAGGA
    1741  AGTGGTTCAC AAAATATGCG CGAGCTTAAT CTTGCTACGT TAAGTGCGTA CTTGGAATAA
    1801  TTACGACTCA TATCCTAAAA CGTTCATTTA TATCTACAAA TTGTTACGAC ACATGAAACG
    1861  AAATTACATA CAAACTTACA GGTATGGAAA ATCACAACTA ATCACCTACC AAAGTGATCT
    1921  CAGGGTATAG ATTATTGCGG CCATTTTTGA TACTTTCTAT TGGTTCTCCA ATAGATATCT
    1981  AAACATATTG GTAGGTACTC ATAGCAAATA GCATGAATTG TATAAAAAAA CAACATTTTT
    2041  TTAGAAAAAA TATTTCATAA ATAAAATCCA AGTCTTGTAT GTGGCAGAGT GGAAAGTGTC
    2101  GAAAATGACG GTCAAAACCT GACGCTATGT ATACTATATA CTTGGATCAC CTTGGTCTAA
    2161  AAATTCCCAA AGAATCAACT AATTTAAAAA AAAAAGTAAA TTACTTTTAA GGAGTTCGTT
    2221  GAAACATGGA TCAAGAAACA GCAAAAGCTT TATTTAAAGA AGGAGCATGC TTATTGCTGC
    2281  AAGGAGTTCC TGAAGGGACA GAGTTTGGGA T

//

LOCUS          4     1308 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      4
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS          500..808
                  /translation="MAGKVFRSFMPLFDRVLVQRFAPETTTKGGIVLPEKSAGKVLR
                  ATVVATGPGVEDKDGKLKPVTVGPGDEVLLPEYGGTKVTLGDEEFHLFRDGDILGKF
                  SN"
ORIGIN
       1  AACCTGATTG TTTTCTGTCC CAGAGGGTTC CAGTCCATTC TCTCGGCCAT ATTGAAGAAT
```

```
       61  ATAAGTTTTC CGACCATATA ATGTCCAAAC CTTTGTAAAC AAACAATGAT ATAGGATCTC
      121  TATAAAATAG CGAAAAAAGG TGACGCACCG AAGCGTGAAA AACGTGGATG ACTGCGCTCA
      181  AAAAAGTGGT GCCAAAAGAT TACCTTTACT GTCTGTATTG TGAGTTTAAA TACTTTAAAA
      241  TTGTATTTTT AAAGCAAATA GACTAAACAA ATATAAAATC ATAGTATACA ACTGAGTTCT
      301  GAAGAAATAT TGCGTTTTAT ATCAAAGTAC CGCGATACGA AAATACCGTA GTGCATCACT
      361  TTTTAAATTG GGGAAGTACC AATCCGAGAG ACTTCTAGAA TTGCTACAGC GTTGCGTCAG
      421  CGTGTAGTTA CTATAGGACT TGTAGTGGTA ATTGCAGTAG TTTAATTTCA CCCCAAAACG
      481  CAAATTATAA AATACAAAAA TGGCCGGGAA AGTTTTTAGA AGCTTTATGC CTTTATTTGA
      541  CCGGGTTTTG GTTCAGCGTT TCGCCCCAGA AACTACCACG AAAGGCGGGA TAGTACTGCC
      601  CGAGAAAAGC GCCGGCAAAG TCCTTCGCGC AACTGTAGTA GCTACCGGTC CTGGTGTTGA
      661  AGATAAAGAC GGAAAATTGA AGCCGGTAAC TGTCGGGCCT GGGGATGAAG TTTTGCTCCC
      721  AGAATACGGA GGAACAAAAG TGACGCTCGG GGATGAAGAA TTTCATCTTT TCCGTGACGG
      781  AGATATCTTG GGAAAATTTT CAAACTGAAA ACACTATTTG CTCTAAAAAA TCGCGTGCTT
      841  AACCAAATCT CTCTGCTCAA AACCTGCCAA TTTCAGACCT AAAGTCATTT GCGGAGTTAT
      901  TTATAATTTC AGTGTTGTTT GGCTTTTTCC ATTTCTTCAT AGTACCAAAA CTGCAGTGTA
      961  TGTTTATATA ATTACACTAT GATCTAATAA AATCGTGTTT GGTGTATTGT GAACTTCATG
     1021  CATACTTTAC GACAACTGCA ATTTAAATAA ACTGGATGTT TTCTACTTAG ACTGTTCGTA
     1081  CGCCTTAAGT TTTTTTGCAA TAAAAATAAT TGACAACAAC AATCATCAAT CTGTAAGCAG
     1141  AAAAAAATGC AGGTAATGCT GTAATAGTAG GGCACAAACT TGAAATTTTT ATGAACGAAA
     1201  GATTTTTATT GGAAGAAATT TCTTAATTTA GTTGTGTATG AGTATTTTGG ATAGAAATCC
     1261  AAATTTTAGG GCATTTTTTC GGAATATGGC ATTTGTTTAA CCAAGAAT

//

LOCUS          5     1338 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      5
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          complement(500..838)
                  /translation="MAGNSLQNGMDEQLLSFEKLDRSSPDLWPEQLSVTDFLSGLKS
                  PASPAVKWGGQLDDDIDMLQEFGNLTTTNLMEKVRELQNLAYQLGLDESREMTRGKL
                  LNILDKPKTNQR"
ORIGIN
       1  TTTATTATTA ATGCCATTTA AAATATAAAA ACAAAGCTTG CACAGTGCTT ATTCAATTGA
      61  ATTGTCACCA AATATATTTA GAATAAACCT CCATCTCACC AGTTACACCC GCCTTTACGA
     121  CTCTTTGAAC GATCGCTTTC ATTTTAAAAT CGACAAATCT AAATATATCA ACTTTTATTT
     181  AATTCATTTT TCATTAAAAA TTCTTCAAAA TAAAAAAATA CTTGTTTAAC CAACAATATT
     241  TAACAATGAA TGTAAAGCGT GTGATAGATT TATATATTTT CACATAAAAT ATTCTTTTAC
     301  AATTCAATCT ATTTTTTTAA TAAATACATT TTCTGCTTGA ATTCGCAATA AATGTTGTGT
     361  AATGAACAGG TTTACATAGA AAATAAAAAC CTCACAAATG AAATCACTAA AATACCCTCA
     421  TTGATTGAAC ATGTACCACA TTTCAACACG TGTTGTAACA CAGGGAAGTA TTACATTACA
     481  ACAAAACCAG GGAAATATTT TATCGTTGGT TTGTTTTTGG TTTGTCCAAA ATATTTAAAA
     541  GTTTTCCCCG GGTCATCTCA CGTGACTCAT CTAATCCAAG TTGGTACGCC AAATTTTGTA
     601  ATTCCCGAAC TTTTTCCATG AGATTGGTTG TTGTTAGGTT CCCAAACTCT TGTAGCATAT
     661  CTATGTCGTC ATCCAGTTGT CCCCCCCACT TCACAGCGGG GGAAGCGGGA CTTTTTAACC
     721  CGGATAAAAA GTCCGTTACA GATAATTGTT CCGGCCACAG ATCGGGTGAC GAACGATCCA
     781  ATTTCTCAAA ACTTAGAAGT TGTTCATCCA TGCCGTTTTG TAGTGAGTTA CCTGCCATGT
     841  TAGCGCAACA AAAAAACGAA AAGAATTTCT CAAAAACACG TCATGTGACG TCAGATCCTT
     901  GAGATCTTAA ATTCCAATCA GGGAATCTCC CCGCAGTAGT TTCGACAAAA GTTCCAGACA
     961  ACTCGACACT AATGAGTCAC CTCACCCGAT GTTGTAACAT GTAAATGTTT TGCGCCGAAA
    1021  ATGAATAAAA TTCACACCTA TTTAAAATTC ACGTATTTTT GGGTATTTAT AAAATAAACA
    1081  CAAAGTAAAA ATGCACCTTT AAAAATTTCA TGCATATAAA ATTTAAATAT TCCACTATTA
    1141  AGTAATGTTT AAAAATATTC ACCCTTGAAA TACTTATGAC CGAATAAAGG AATTACCCAA
    1201  ATTTATAAAT AATTTTTTTA GTAAATAAAG AAAAAAAATC ATCCGGCCAT GTAAGTACCC
    1261  AGTAATATAT TATTCTAAAT CAAATATAAA TTAATATTTA TTTCAAATGT TAATAACTCA
    1321  AAGTTTAATA ATGTTGGC
```

```
//

LOCUS          6     1395 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      6
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS          500..895
                  /translation="MAEKLEVLEACLEQFIENVRQLGVIVTDFQPQGQEVLNQKINS
                  TITCLKDIENCKASVSDIHVPLEVFDYIDSGRNPQHYTKDCIEKSLMKNEALKGKID
                  VYGKFRQELMTQIGKVYPQQIIEYKQARGEL"
ORIGIN
       1  ATAAACATCC ATGTAAAGTA AGAAGTGATC TTTTTAATTT TTTAGTGATT TTTTGTTATA
      61  TTTTTAATTT TAAGAAGAAT TGTATAAAAG CATTCGGGAA AAACAAAAAT ATTGTCATTT
     121  TAAAATTGAA AGGGATGTGT AAAAGTTTTT GGATTTACAA CATAGATATC GTGTATTTTT
     181  ATAAAAGAAA GAGCAATGGC GCGGTGTTCG ACGTTTGATA TACAAAACTA CATGTCAGTA
     241  TAGATGTTTT TAAGTAGTTT TAATATTACG AAACTGATAC TGTAGTGCCG TACTGGTGAA
     301  GCCCTTGCAG ATTAGGCTAG CCAGTTATAA TGGTTTAAAT GGTAAAGGGC CATTGGGGCC
     361  TTGACCTGTT GACCAGATAC CGTAGAAGGG AGATTCTTTA TTAGCACCTA GTAATGTTTT
     421  ACTAATTTCC ATAGAAATCT AACAAACTGA GGTATTAAGT TGCTTTAAAC TATAACAAAC
     481  TCCTTCATGT ATAACCTAAA TGGCTGAAAA ATTAGAAGTC CTTGAAGCGT GTCTTGAACA
     541  ATTCATTGAA AACGTGCGGC AACTTGGTGT AATCGTTACT GACTTCCAAC CACAGGGGCA
     601  GGAGGTGCTT AATCAGAAAA TCAACTCAAC TATTACATGT CTTAAGGATA TTGAGAACTG
     661  TAAAGCCTCA GTTTCAGACA TCCACGTGCC GCTTGAGGTG TTTGATTACA TTGACTCTGG
     721  TAGAAACCCC CAGCACTACA CAAAAGACTG CATCGAGAAG TCGTTAATGA AAAACGAAGC
     781  CTTGAAAGGA AAAATCGACG TTTACGGAAA ATTTCGGCAG GAATTAATGA CTCAAATTGG
     841  AAAGGTTTAT CCACAACAAA TTATCGAGTA TAAGCAAGCA AGGGGGGAAC TCTGATAGTT
     901  CTTTCTAAGC TTAAAACATC AACATTGCAA ATTGTTCTGC GCATAACATT ATTTTTTACA
     961  TCTTATTTTT TAAGTTTCAA TCATGGATGA TAAAAAATCA ATTTCATTTG GATTTAAAAA
    1021  GAAAATCCAG ACCAAGGTGG ATGAAAGTAA AGTTGTAAAA GATGCTCCTG AAGAGAGCAG
    1081  CACCAAGGAC TTTCTCACAG AGGTTGATGC TAAAAGTTTG CAGAAGGTGA CTCCAAAGGA
    1141  ATTGGTAATC CCTCTCATAT CAAAGAATAG ATATGTGATG CCAAAAAAAG AAAAACTGCA
    1201  AACAAATGAG ATAACTGCAG GAATTGAAGA TGAAGCAGCA AAGGAACTAA TGGCAGAATC
    1261  AAAACAGTTT TTGGATCAAG CTCCTACGGA AGAACAAAGC AATGAAGATT TTGAAAACAT
    1321  GGCTGTTCCT CTTCTTTTAA GAAATAAAGT CCCGGATGGG TATGAGGATG ACACAGTATT
    1381  GGATGTTTCG ATCCG

//

LOCUS          7     1753 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      7
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS          join(500..703,772..882,1173..1253)
                  /translation="MELSLKNELLASFAWYTVIVVLKMMIMSIICVMHRIRNKAFVN
                  EEDTKAFGSSVKGKLATVTNASVERVRRCHLNDIENIIPFVIIGFAYILTDPAVETA
                  IFHFQPSRFVFCMSGYMATLSMAFRTLAALY"
ORIGIN
       1  CCGTAGCCGC AATGAGCTAT TGCTTTTTAC TATCGTCAAA AAAACAAATC ACATACCATA
      61  TATTCCACAA ATTGCGGCGG GTAATTGGCT TAACCAGGTT CTAGTGTTTT AACTATCCGG
     121  CATTATACCG TGTATTTCTG TTACCTGATA GCCGACGGGT TTTGCTTACA TATCCAGAAG
     181  CCAGCGAAGC GTAACAAAGT TTGTCACGGG AAAGAAGCGC TCTGTGTTAA CTTTAGTTCA
     241  GTTAACCGCT CTGCAAATAA GTGTTGAGCA AGTGTATTAA CAGATTCGAC GTACAGAATT
     301  ATTAAGGCAT GGCAACTTGT GTGTAGTCTG TATTATGTGT ATGACACGAG TCATACACGT
     361  TAATTACTGT ACGGTACAAG TTAGATGCGT CAAATATAAC AGAATAGAGT TAAATGGAAA
```

```
     421  ACTTGGTTAA GATGTTTCGT AGTTTAACAA ATTGGCTTTG GATTTTTAGT GAAAAGTAGA
     481  CGATTGTTAC TGCTTTGCCA TGGAACTTAG TTTGAAAAAC GAGCTTCTTG CCAGCTTTGC
     541  TTGGTACACA GTAATAGTTG TGTTGAAAAT GATGATAATG TCGATTATCT GTGTAATGCA
     601  TCGAATTCGC AATAAAGCAT TTGTAAATGA AGAAGATACA AAAGCGTTTG GTTCATCCGT
     661  AAAAGGCAAA CTTGCAACAG TTACAAATGC TTCAGTCGAA AGGGTATGTA GTGTGCTTCG
     721  TTAAAAAGCT TCATTTATCG TTTGTTGTTG TTTAATTATT TCCAATTCTA GGTTCGACGT
     781  TGCCACTTAA ACGATATTGA AAATATTATT CCGTTTGTTA TCATTGGTTT TGCTTATATA
     841  CTGACAGATC CTGCCGTAGA AACTGCTATA TTTCATTTCC AGGTAATGTT AAACTTTGTG
     901  CATTTAGAAC TGTAAACGTT GGTCTTTTGT GTGCTGTATA AGAACTTCGT GTATGTACTG
     961  TCTATTATTA GAACGTGCAA AACAAAGAAC TCTTAGTGTT TACACGATAA GCACCGCAAC
    1021  GCTATACAGT TCATAAAAAA CTAACTATAA TTGTTGTTTT AGCAAATTTA CATGAATTCC
    1081  TATAACAATC AATAAAAATG GAGGTTATTT TTCAGCTCTT CACCGGCTCG AGGATTTCCC
    1141  ATACTTTCTG TTACTTGATT CCAATTCCAC AGCCATCTAG ATTCGTATTC TGTATGTCAG
    1201  GATATATGGC TACTCTATCA ATGGCATTCC GCACCCTGGC TGCACTCTAT TAATTTTTAT
    1261  GCAAAAACGG AAATTACAAA TTTATGAAAT TGATTTTCCA ACTGCTACAA ATGGATTTAA
    1321  TTGACTTTTT ATTACAAAAT GCTATGTAAT TAAATGTATG TCATGCTATT GTGGAGATAC
    1381  AATACGTTTG TATTTTGTTG TTTCTGACTA GCCCACATAT TACTATTCCA TTACAAACAT
    1441  CGTATTATAT CGATTCAAAC CTGTCAAAAC TGGAACAACA TGAAACATAT GGTAAAAAAC
    1501  ACTTTCGTTT TTAGATTTTT AAATCCATCT TTATTCGCTG TATAATCGAG AAAAATAAAC
    1561  ATTGCAAAAT TCGAATCATT TTTCAGTTGG TGGTGCTAGA AGTGACATTG GCATAAATAC
    1621  TTCTGTTTCT TGCTTTCTTT GTGCTGACTC GACGCCGTCT TTGACGAAAC TATCCTTGGA
    1681  AGAAAAAGCT AATGTACCGT CTATTATTTT GGATGAATCG GCATGAATGT CTGTGAACAC
    1741  GAAATTTCCT GAT

//

LOCUS          8     1419 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      8
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          500..919
                  /translation="MSVDIRLDQNVEKVPEVDLHHIPCKIEANETANISKFFCPKSV
                  KTESKETLSASFRGRPLDGKNITLPDGYIGAVITEKQRPFSEEENRELDVTGKFKVF
                  TKWYLGSQFAGGNSMDKACLTWTSNLASVIHDPVEPKHD"
ORIGIN
       1  TTCCCCCACC CTAACAAAAA AAATTAGGTT TGTAGGAATT TTGAAAAAAT GTTAGTTGTT
      61  TACTCGCAGT ATCATCGTCA GCTTCGGCTT CGGCAACCAT TTCTGAATTT GCGATCTACA
     121  AACATAGATA CCTGCTTTAA ATCTATGTTA ACAACGAAAC TGTTATGCTT GAGTAGCCAG
     181  CGAGTAAAAT AAACGAAATT GAACTTTGAC CTAGAAAAGC ACACGAGGTG AAGCTGAATC
     241  GTTGTAAAGA TCGGATTTGT TTTGTTGTGG TACAGCTCGT AGATGTTAAT ATTGCTAGAG
     301  GGATGTATTG CTGTAAGTAA CCTCGGTGGC TTTCATAATG AAAGCGAACG CCTTTCATCA
     361  ACAGAGTAGC TTGTTCGTCT GCTATGTTTA ACAAAGTTTG TAATAAGTGA TAAGAACTTC
     421  GGTGCTAGTG AATTGTCATA ACAACATGTT TCTTGTTTCC ATAGAATGCA TTTGCTTTAT
     481  TTTTGATAAG AATAAAATAA TGAGTGTTGA CATACGTTTA GATCAAAACG TTGAAAAAGT
     541  ACCTGAAGTG GATTTACATC ATATTCCATG TAAAATTGAA GCCAACGAAA CTGCAAATAT
     601  TTCAAAGTTC TTTTGCCCCA AAAGTGTTAA GACAGAGAGT AAAGAAACAC TGTCAGCTTC
     661  TTTCAGGGGG AGACCTTTAG ATGGAAAGAA TATTACTCTT CCTGATGGTT ATATTGGAGC
     721  AGTAATAACG GAAAAACAGA GGCCATTTTC TGAGGAGGAA AACCGGGAAT TGGATGTTAC
     781  TGGGAAGTTT AAAGTATTCA CAAAGTGGTA CCTTGGTTCA CAGTTTGCAG GTGGAAATTC
     841  AATGGACAAA GCTTGCTTAA CATGGACCAG CAATCTTGCT TCTGTGATTC ATGATCCTGT
     901  TGAACCAAAA CATGATTGAA GACCTGGCCA CCTCAGTAAA ACTAAATTCC AGTATCAAAG
     961  CATTCCTTGT GATTTCCTTT TTAAGGCAAA CTTCTTTAAA ACATGTTTGC TAAAGCTATA
    1021  ACTCTAAAAT GTATTTTATT ATTCCCTAAC AATGTAACAT TTACATCACA TACAAATCTT
    1081  CATTTATTTT TGAAGGTTTT TTATTGATAT ATATAATTGT AGTATATTGT TTACTAAAGA
    1141  CGTTTCTGTA ATTATTTGTG AGATCATCTT TCAAGAATTC ATAATGGAGC TTCATAACTT
    1201  GGATTTAATA ATATTCAGTG AAACAATACG GTTGGTGTTT TCTAACATAC TTTCATAACC
```

```
    1261  TATACTCTCT ATGACTGTTT GTTGCTACAT AAAATAATAA TTAAATAATA TCCCGATATT
    1321  TGTCCGGCGC TGTGTCAAAG TGGTTAGCCC GCTTGCATTT AACTTGGAGG TAATGAGTTC
    1381  AAGCCTCGTC GCTGCTACCA ATTTTGACAT ATGTGTCCT

//

LOCUS          9     1934 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      9
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          join(500..653,846..1007,1295..1434)
                  /translation="MTEVTSRARLHCRIAYSLFGIAIGVSVFVIWLTVFDNRKETYA
                  TVTYGGLSGVLAAWNLILHIMHYQDLWRTWLKGLRLFIGGGLLFLGITGISFVTFLV
                  LAIYLNQSDLVRSYYITAIWCALSAKWSFLLFWYSRNYRNEFADLTALLEF"
ORIGIN
       1  TATAGGACGC TTAATAAGCT AAGTTCTGAG TACACATATG TCAATCGCGT TGGAATATTA
      61  GCTTTAAGAA AATTTCTAAG TTGTAGTAAA GGGTTAATTA ATATTATGTT AACTAGTTTA
     121  ATGCGTGCTC ACTCATTAGA GCTTTGTAGA ATTAAAAAAA TGTTTTGAAA TCATGTTAAG
     181  TATACTGTAA GGATTGCGTC CATAATATTG TATCATGTCA GCGTTTCTTT TTGCCATATA
     241  CATTTGAATA GGCATTAGAT TATCTTTCTA GTAGCCTATC TATGCTTTTA CGATTTACGT
     301  AACCTTGTTT ACCAAATCGT CACGAATGAG TTAAATCACG TGATCAAATA CAAAAGCATT
     361  TCATCAGCAT TTCCCCATAG CGAGATAGTT CCGTAGTATT TACGGCCATG TAAGCATTGG
     421  TTAAAACGAA TTTTACTTAT TATTTGTAGT GTTACTATTG TTGTTTAGCT GAATAATCTG
     481  GATGGACTGA ACTATCGAGA TGACAGAAGT AACGTCACGG GCTAGACTTC ATTGTAGAAT
     541  CGCTTATTCT TTGTTTGGAA TAGCAATTGG TGTTTCTGTG TTCGTGATCT GGTTGACGGT
     601  GTTTGATAAT CGAAAGGAAA CGTATGCAAC AGTTACATAT GGTGGTTTAT CAGGTATTGA
     661  AAATGTTTAT ATTATATAAT GAATAGCTCA AAAGCAATTT AGTTTTCTGT ATGCTGTATT
     721  TGTAATTGTA ATGATGATAT ATTTGTGCTT TGTTTTTATA TTATGAGGGC AAAAGTAATT
     781  TAGTTTTATG TATGCTGTAT TTGTTATGGT AGTGACTAAA GATTTATGTT TTGATCTACA
     841  TATAGGTGTG CTGGCAGCAT GGAATCTAAT TCTCCACATC ATGCATTACC AAGATTTATG
     901  GAGAACTTGG CTTAAAGGGT TAAGATTGTT TATTGGCGGT GGACTCTTAT TCTTGGGCAT
     961  CACAGGAATT TCTTTTGTTA CATTTTTAGT TCTTGCCATT TATCTAAGTA AGCTATTATT
    1021  TTATGGCAAC TGTTTTATTG TAAGTGTGAA TGAAATTTCC TGGTGTACTT GTTAAAGTCT
    1081  GTGCACATAT ATTATATGAT TTACCACCAA TTCAGCTAGG TCAAGTTCGT GTATTTCTAG
    1141  CTGTTTAGAA TGCCAGTTTG CTTTAGCCTT TTGTATACAT GATTCGAAAA AGAATGGTTT
    1201  GTGCATATGG ATACACCAGT GGGCAGTGAT GATCGTTGCC TTATGATATG TTTTATATCT
    1261  AGTATAAGTA ACTAATTAGG TTTATTATTT ACAGATCAAA GCGACTTGGT ACGGAGTTAT
    1321  TACATCACAG CTATATGGTG CGCTTTGAGT GCCAAATGGT CTTTTCTATT GTTTTGGTAT
    1381  TCTAGAAACT ACAGAAACGA ATTCGCAGAT TTAACTGCAT TACTCGAGTT CTAAAACAAA
    1441  TTAATTTATT TATTTTTGTT CCAATCATAA GAGAAGTACC AACTGAAATA TATGTGTTTG
    1501  TTGATTGTAT TTTCACAGTA TTGTGACTAA AGATTAATAT GTTCATTGGT GCAAGTTAAA
    1561  AATTATATTT TTTCTGCATT GCTTGTTACT TTAAGCTGTT GCGTTGAGTC AGGCACCTCT
    1621  GTGGTGAAAA GTAGTGTATT TCATGCTCGT TAAATTTATT TTTGTTATTT AAGAGTGGCT
    1681  GAAAATAATA AATTGCAGAG GGTATAACAT TTTCCATTCA AAATTGCACA TTTTTTGCTG
    1741  GATAAGTCCA TGCTATACGA TCAAACAATT TTTAAGCAAT CTTCTACCTT TAAATTCATG
    1801  AAGTAGCAGT ATTTTTTAGG TTGTTTTCAA AAGTACTGAT AACCTATTAA CGTGCATATT
    1861  GCTTTCAACT GCATTAATAA AATTGGTCTA ATAAAACATT GATGACATGT TTTTGGTTGA
    1921  AAGTATAATC TACT

//

LOCUS          10     1677 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      10
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
```

```
     CDS             join(500..709,863..973,1040..1177)
                     /translation="MSAFTLENEAFSGLVFYGGLCLAKTAAMSMVTAYHRITKGAMP
                     TEEDTRMMTKDPEKMKKMLVPNDDVERSRKAHLNDLENVVPFVLLGLLYVSTNPPAD
                     SAILHFKIFFWARVLHSVAYICKLRSPSRGVGFFLGFATTISMLYRTISAVY"
ORIGIN
        1  CAGTTGCTAA TGTGTGCTTC TAAAATTCTT TTTCCTTGTG TTTGCTCTGT CTTATAAATT
       61  GATAATACGT TGTGGGAATG CCGTAAGAAG AAATTAGAGT TCCTATTCAA GGTTTCAGAT
      121  GAAGACCAAT TTAGTTGTTT TTTTATTCTC TCTATTCGCC GTTAGTTTCA ATGGTCATGC
      181  TCTACTAGTC TACTACTTTA CTGTTGGACG TTCTGCTTAA ATAGCAAGCA TTGTAGTGAG
      241  AAGACTATGT AACCAATGAA TGCTGCTCAG AAACAGTAGG TTGATTATTA GCAGAGTGGT
      301  AGGACCAAAT AGAACTTTGT TTGAGCAAGG TTTGAGTGGG ACGCACAGGC TAAAGTGCTG
      361  CTAGTCACGT ACGGTCTGTT GTACCATATA AAGGTACCGG TAGCTGTGGT TTGAGGTGGT
      421  TTAGATTTTA TAGGACGAAA TTTAGTAAAA TCTGTTCAAA TACGTCTCTT TTTTGTATAC
      481  TAGAACGTAG CTACATAAAA TGTCAGCTTT TACATTGGAA AACGAGGCTT TTTCTGGGTT
      541  AGTTTTCTAC GGCGGGTTAT GCCTTGCAAA GACAGCAGCA ATGTCCATGG TTACTGCATA
      601  CCATCGCATC ACCAAAGGAG CAATGCCAAC GGAAGAAGAC ACGAGGATGA TGACCAAAGA
      661  TCCGGAAAAG ATGAAGAAAA TGTTGGTTCC AAATGATGAT GTTGAAAGGG TGTGTGTGCA
      721  GTATGGTTTA TATTAGGAAA TCCGAGATTT TAAATGTAAA CTACAAAGTA TTGTATAGTA
      781  TTTGCAAATC TTTCAATTCA AAATTGTCCT AGTTGCAAGT TTTCAAAATT TTTATTGTGA
      841  CGTAACAATT CACCTCATGC AGTCACGAAA GGCGCATCTG AACGACTTGG AAAACGTCGT
      901  ACCGTTTGTT CTACTTGGTT TACTCTACGT ATCTACGAAC CCACCGGCAG ATTCTGCTAT
      961  TTTACATTTC AAGGTAACAC ATTAACGAAG TGGAGGTTTC TTGTTTGAAT TACGCGCCTT
     1021  CAACGTACCT CCATTACAGA TCTTTTTCTG GGCTCGTGTT TTGCACTCAG TGGCTTACAT
     1081  ATGCAAGCTT CGGTCGCCGT CCAGGGGTGT AGGTTTCTTT TTGGGATTTG CAACAACTAT
     1141  CTCCATGTTG TACCGCACCA TCTCTGCTGT TTATTAAACT GAAAGCTCTC TTACACGTCC
     1201  GAACCATGTT ATCGTGTTTA ATTAAAGCAA AACAGCAATT TTGACACATA TAAAAAGGGT
     1261  GCAATTTTTA CACCAAAAAA AAGAACAAAT AAGAAATGTG CAGTTTTTAC ACAATGTTGT
     1321  GCTACGAAAA ATTAAGAAGG TATTTAGTGT AGAAATGCAT CCATATGTTA TTGTCGTGCG
     1381  AACAATTATT AAATACAATA TCACATAGAA AACATTCTGT TTACTTTTCC CGTAGATTTA
     1441  TATCATTCAA ACAAATCCAA ATAATTATGT GGACTTAACA ATTAGGCTAC ACCTACAGCT
     1501  CAATTGGATT TTTCCACTTT GCCCAGCAAT CCTCCAGCTA TTTACCAAAT TAATTATCAT
     1561  GACTGTATAT TTTCACTAAA AAACTTGGCT TAAATCATAT CGAACACGGC CTACTCAATG
     1621  TATTGAGTAG TCCATGTATC GAAATAAATA GATTTTATAA TATGTATACT AGAAGCC

//

LOCUS          11     1503 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      11
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS             500..1003
                     /translation="MHCVFVLLVLVSVNYKVSAIPLHASCSIQWEFPSTVCDDVRTK
                     LVNQIKTWSGKCTHSGEKCLYKLIRNDDLKVTATHTTPVKNYVDSLTFTFRTSQKIR
                     FADDSTWCVMKGYSRSNTWYAVLDHGTNYCNLHNLITGSGLNNTKGYEENTNDSICT
                     QYTSADCTKY"
ORIGIN
        1  TGTAGAAATT GTTAAAACTG CTTTGTTTTT TTGAGTAGCC TATAATTTAT TGTTTGTATA
       61  ATATTCTTAC TTAGAGCTGT GGGTCGGTTG GGAAATGTAT ACGTCATAAC TGAAGGATAA
      121  TGGGTTCACA ACATATAGTC AGTAAAACAC TGTTATTAAA CCAATGTCCA AACCAGGGGG
      181  TAATAATATC AAGTCCTGAT GCTGCTACCA TTGTGAGCGT ATTTGTTCTT GTGCAAGACC
      241  TAAACAACGG TCATTGCACC AACCACTGAT TGTTGACATT TCATTTGTCT TATTTCCTCC
      301  CCGTCAAATA AAAACTACTC AACCGCGAAG AACGTTATTG TCATGTGCCT GTGACGTAAT
      361  AAACCTCATA GTTGTAAACA ATGTTTACCG ATGCAACGTC GAGACGGAGA TTCCTCGTTC
      421  TGCAACCGTG TACATTAAAA TAGTTTTTCA TCTCGTAACT ATACATACTT TTTACAGCGT
      481  TAATAGCTGT AGCTTTATAA TGCATTGCGT ATTTGTGTTA TTGGTTCTGG TGTCAGTTAA
      541  CTACAAAGTA TCGGCGATCC CACTTCATGC TTCTTGTTCC ATTCAATGGG AGTTCCCGTC
```

```
     601  AACAGTATGT GATGATGTCC GAACCAAGCT AGTCAACCAA ATCAAAACAT GGTCGGGCAA
     661  ATGTACACAT AGCGGAGAAA AATGCCTTTA TAAATTGATT AGGAACGATG ATCTAAAAGT
     721  TACTGCAACC CATACGACCC CAGTTAAGAA TTATGTGGAT AGTTTGACTT TTACATTTCG
     781  CACGTCGCAA AAAATACGTT TCGCCGACGA TTCTACGTGG TGTGTAATGA AAGGTTACTC
     841  AAGGTCAAAT ACGTGGTACG CTGTGCTGGA TCACGGAACA AACTATTGCA ATCTACATAA
     901  CTTGATCACT GGTAGCGGGC TTAACAATAC AAAAGGGTAC GAGGAGAACA CAAACGACAG
     961  CATATGCACT CAATACACAT CCGCAGACTG CACGAAATAT TGAAATGTTA ATATTATGTT
    1021  GTTATTTGCA AATAAAATAC AAGTTTGAAA ATTTCTTTTC GTGTAAACTT TTACTGACTG
    1081  TTTAGTATGC TGCATTGCAC GTGACATTGT CTTGTGAACT GCAAATAAAA TATTGATGCA
    1141  ATAAAAAAAT GTTGGTAACC AAAACAGTTA TTTTTAAGCT ATAAATATAA TGTCTCTATC
    1201  GATTTGACTC AGCGTAGGGT ATTTCCCATT GGAAAAAAAT GGCGCGAATT TTGCAAATAA
    1261  TTTAAGTTGT TTTGTTGTTA AAATATTCAA TAAAAGTCGT GAAAGTATGC CAAGGAAGCG
    1321  AAAAAGTGAC GAAATTGAAG CGAAGGCCAC AATTTCAAAG AAACAGAAGA ACGTAGAACA
    1381  AGATACGGGA AGCGAAAAAG AGCGAATCGA AAAGTTGTAC GGCATTTCGT TTTCGAGAGA
    1441  TTTTTTCGCG TTTTGGAAAT TCTGCGTGGA GTTAAAACCG GAAAACCCCT TAAGTATGTA
    1501  TGA

//

LOCUS          12     2046 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      12
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          join(500..699,926..1179,1446..1546)
                  /translation="MSENIDTLCLTMLELVAEIHQKKLEQEKHMQEGYLCLAKARYS
                  MGVGRVSSLQYPNEITPTVSISCREKENNKFPSFTVTQNLQDEETQPPTSAVRRRKK
                  TKKDDHNEAKQAVEEVDLLSQDKLHAQVNPLLWFGVLVPTPLRQAQNCFRQATEVSC
                  QLCTLKEELNYKTRLYKEVMKPTETKG"
ORIGIN
       1  GAGAATGGCT TACCAGCACT GCTGGCTTAC AAGAACAATG TTATCATTGG AAACTTTGTT
      61  GCTCTATCTG AAACACTGGG GGAAGATTTC TATGCAACTG ACGTTGAATC ATTTTTACAT
     121  AGCTACGGAA CTCTCCCAAC TTCAAATGAG TCGTTAAGCG ATGACTGACA ATCTCTGTTA
     181  TTATTTTTTT CAGTATCTTT ATGTACTGTT ATCTGTTACA GAGTTGCATG AATGGATAAT
     241  GTTTATAATT TTCCATTTAT TAATATTTTT TTGACCATGA ATTGCTATTG TCATTCGTCG
     301  TAATGTGTTG TTATGTTTCT GTGCTATATT GTATCGCACT GTACCCGGAC AAGAGCAGTT
     361  TTTAAAATAT GTATCTTTAT GCACAAACAT AAAGTTACAA CTGTATTATT CAGTAATGCA
     421  TTGAATGCTT TTTTATTTTT AAAAACATGA AATCTATTTA CACTGCTTTG GATGTCTTTC
     481  CGAAGATAGA AAAACAGATA TGTCTGAAAA CATAGACACA TTATGTTTAA CCATGCTGGA
     541  GTTAGTGGCT GAAATTCACC AGAAGAAACT GGAGCAAGAG AAACACATGC AGGAAGGTTA
     601  CTTGTGTCTG GCTAAAGCCC GATATTCCAT GGGTGTGGGC CGGGTTTCAT CACTGCAGTA
     661  CCCAAATGAA ATCACACCCA CAGTCAGTAT TAGTTGCAGG TAAAGGCAGC ATTTTCTCCT
     721  AATTTAATGT AAATAGGTAA AGGCAGCATT TTCTCCTAAT TTAATGTAAA TATCTGTAAG
     781  CATGTTTTAA CAACCGTTGG TTTGACGCTA TGTTCCGCCT TTTTTTTTTG TATCAAAGAG
     841  ACAAATAAAG TTATTATTTT TGTACTGTGA TTTTACTTTG TATACTTTCT AACATAAGTT
     901  TTAGACTAAT TTGTTGATAC TACAGGGAAA AAGAAAACAA CAAATTTCCT TCCTTTACTG
     961  TCACACAAAA CCTTCAAGAT GAAGAAACCC AACCCCCCAC TTCTGCTGTG AGAAGAAGAA
    1021  AAAAGACAAA AAAAGATGAT CATAATGAAG CCAAACAAGC AGTAGAAGAA GTTGATCTTT
    1081  TATCACAAGA TAAACTGCAT GCACAAGTCA ACCCATTGCT TTGGTTTGGT GTTCTTGTAC
    1141  CTACACCATT GCGACAAGCC CAAAATTGTT TCAGACAAGG TAAAGTTTGG GCTTTTGACT
    1201  TGTTTTTTTT TAATGTATTG TACTGTGCGG AAAATGGATA CTGTTAGCAC AAAGTATACA
    1261  ATGACTCCTA ACCATGTTTT CAACAATTAC CATAGCTATT TTATAGTCTG GGAATAGATT
    1321  AAATTGTTCT TTAAATATTG TTTGTTTACC ACAAAATGGC ACGATATGAG AAAGTAAAAA
    1381  CATGGACCAT TTTTCATTAC TCTAAACTAC TATATGTGTT GATTTAGGTT AATATTGTTT
    1441  TACAGCAACT GAAGTCAGCT GTCAGCTTTG TACTTTAAAG GAAGAATTGA ATTACAAAAC
    1501  TCGTTTATAC AAAGAAGTAA TGAAGCCAAC TGAGACAAAA GGATAATCAC AGAATCTAAT
    1561  TGACATCAAA TCATATGCTA TCTTTATAAA GCTGGCATGT AATCATTTTC AATCACATAA
```

```
    1621  TCCTTTGTTT AATGTACCAC AGAATCAATG GATTTAATCA CGCAAATGCA ATAGAAGAGG
    1681  ATTTGCTGGA ACATATGGCG GGATATTCTA ACCAAGAAAC ATTCAATGTA ATCATCTGTC
    1741  ATATTGCATT GATTTTTTAT AGAATTAGTT TTTTGATGAC TTGTAATCAA GCTGGGAATT
    1801  CCACAGTTGA TTTCTCACAG TCAATAATCA AGTAGATGGG AATGTTTTTT TGTTTATTTT
    1861  AAAAATACGA TGATCGTTTC AAAAGTAACA AGAAAAATTT TCTATCTTTA TTTTCGCCTT
    1921  GATTTACTGT CAATGAAGTT GCAATACAGC ACTGAAAAAC TAAGTGGATT TTCAGGTGTG
    1981  TATTATTGTA AAAAAAAATA CCAACAAGAT TGACACTATT GTTTCATCTG TTTGAAACGC
    2041  ATCGAG

//

LOCUS          13     1695 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      13
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          complement(500..1195)
                  /translation="MAWRCHGKSNLDLINNLRKNGIITHDAVYEGMKLTDRKFYVSS
                  GAYNDSPQSIGYQATISAPHMHAAALEALHDQLTRSENPTALDVGSGSGYLTACFAR
                  MMGDNGRAYGIEHIPELVNKSIENVNRDDSTLITSGRVTLKKGDGRLGYDPDNRKTE
                  LYDAIHVGAAASQVPRPLLEQLKKGGRLILPVGEPGTTQVFEQWDKNNVGELTKKEL
                  MKVVYIPLTDKRKQWPG"
ORIGIN
       1  ACATCCATTC ATTCAACTAA ATATTCATAA ACTGCCAAAA AGCGACAAAG CATAAAAGTC
      61  AGGATTTTAA ATTATCGACA TTAGCCAAAT ATACATTTGC TGCAACAAAA ATAATTCAGC
     121  ACTGTAATAC ATACGCTATG GGTGAGTAAT TAGTTCATGC ATAAATTAAA TAAACCAATA
     181  TACTTAAAAT TAGCAACTGT AGCTAGAGTT CATAATAATA TATTGTTTTG CTGAACTTAT
     241  ATAGGCCTAT ACCAAATATA CTCTTAACTT TACTCCAACC TAAAAGAACT TTAATATTTA
     301  TCACACAAAA TAAAGACTTA AATAATAGAA TTGATAAATA GGGCAATTGT TATGACATCA
     361  AAGCAAGACA GTTGAGTTCT GGAACCTGAC ACTCCACTTC TATGACATCA CAATCCAATG
     421  TTTGACATAA TAAAGACTGT TTGGTTTACA AAACCAATAT TGGTTTTTCG ATTTAACTTC
     481  TTGTATATGA GTGGACCTTT CAACCTGGCC ACTGTTTCCG CTTATCAGTA AGTGGAATGT
     541  AAACAACTTT CATCAATTCT TTCTTGGTAA GTTCACCAAC ATTATTTTTG TCCCATTGCT
     601  CAAATACTTG TGTTGTACCA GGCTCTCCCA CGGGCAAAAT AAGTCGGCCA CCTTTTTTCA
     661  ATTGTTCAAG TAAAGGCCTA GGTACTTGAG AAGCAGCCGC TCCGACGTGT ATTGCATCAT
     721  AAAGCTCTGT CTTTCTATTG TCCGGGTCAT AACCCAGTCT CCCATCCCCC TTTTTCAAAG
     781  TAACCCTCCC AGAAGTAATC AAAGTTGAGT CATCTCTGTT TACATTTTCA ATGGATTTAT
     841  TCACAAGTTC TGGAATATGT TCAATCCCAT AAGCTCTCCC ATTATCTCCC ATCATTCTCG
     901  CAAAACAAGC CGTTAAATAT CCGCTACCAG ACCCAACATC TAATGCTGTT GGATTCTCAC
     961  TTCTGGTAAG TTGATCATGT AACGCCTCTA AAGCTGCAGC ATGCATGTGG GGTGCACTTA
    1021  TAGTAGCCTG GTATCCAATA GACTGTGGAC TGTCATTATA TGCCCCACTT GAAACATAAA
    1081  ATTTCCGGTC CGTCAACTTC ATACCTTCAT AAACTGCATC ATGGGTAATA ATTCCATTTT
    1141  TTCTGAGGTT GTTAATTAAA TCCAAGTTAC TTTTTCCATG ACAGCGCCAC GCCATTGTAC
    1201  AAATTATGAA TAACACCAAC CTTATATAAA ATAAAAACGT CATTTTAGGA TTTTAAACAA
    1261  ATTAACTTAT GTGCCAAATA TTCAGTTCTA TGTGTAATTC CTAAAACAGA AAATAAAACA
    1321  AATTTTATGC AGTTCAGGCC AAGAAAACAC AGTTACACAT ACAAAACACG AGCCAATCTT
    1381  TCTTAAGCAG AATACACTTA ATGATGTATA TACAAAAATA CAATATACCC AAATTTCCTT
    1441  TTCTTACATT TAACAGTTTT TTTTCCGATA CACAAGTCAG ATAAACCTTA TTGTGCACGT
    1501  TACCTGTTAC CACGCGTTGT CCAAATATGC GACGAATCTG CACTCGCTGT TACAGTGTAG
    1561  TTTGTAAACA GCAGTCGGAA TGTCTGTCAC ACGACTCCTG TTGCGATCTC TTCTAGCTCT
    1621  ACAGCGGCGC TCGCTACTAC TATGACGTCA TCAGAATACT AACAGCGTGA CGCCGAACTA
    1681  AATTTGATGA CGTCG

//

LOCUS          14     1704 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      14
```

```
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS             complement(500..1204)
                     /translation="MTESPDTEGTLSFFQVASDEELSDSDCSENEEAVKSEEVVPEK
                     QRSSVLPSPLCVLADAKTPEFVTSHLSSDLKWDKMVKKAPYIAPKEFKPWENDLPAF
                     ESKSNVKDSKPDIALVKQIPSSMARPVEVKPTPVEPKKKEIVDHAISWSKMYKDDVE
                     DRRAYKRPRGEDEDEELRGFCHTDLDAKKVAFHSKNNQSTTHHDKEKRKRNLGQANR
                     EKMFVEDEKRVLRQEFAENF"
ORIGIN
        1  CCAAATTATT ACATCATAAT TACCTCCGGT GCCTTGGACT TTCAAATCGT TTTGATTTCC
       61  TTCGTATGTG AACAAAGCCC TGAAACAAAT CAATTATTAC ATCATAATGA ATTGAGGTGT
      121  TAGGTGTACG AGGTAATTAA ACTAGAGCAA TAAAACGAAC TGCAAACCTG GTAAAAAAAA
      181  ACCTAACTTA CCAATCAGTG TCTGTGTTGT CATTTAAAAC ATCTTTGTAG GCCTCAACTA
      241  GCTTGGCGTT ATGTTTAAAA AAATTTACAC TCATATTTAT AATATATATA TTGTACAAAA
      301  CACTCTTCTT CCTGTGCAAT GGAAAGGTAT TCTATAAAAT ATGATCATTT TGTTTAAAAG
      361  TTTAACAAAA CTATATCATA AACGGCAAAC ACCTAGTTTT TTGATTAGTT GGCAAACAAT
      421  TTAAAATAAA ATCCTTTTCA AATAAAGAAA AGGCTAAAAG TAATAAATGG AAAAAATATC
      481  AACAGTTATT AATGCAAGTT TAAAAATTCT CTGCAAATTC TTGTCTTAAA ACTCTTTTTT
      541  CATCTTCAAC AAACATTTTT TCTCTATTCG CTTGTCCTAA ATTTCTTTTC CGCTTTTCCT
      601  TATCGTGGTG AGTTGTAGAT TGGTTGTTTT TACTGTGGAA AGCAACCTTT TTCGCATCCA
      661  AGTCTGTGTG ACAAAATCCT CTTAACTCTT CGTCTTCATC TTCTCCCCGA GGACGTTTAT
      721  ATGCTCTTCT GTCTTCAACA TCATCTTTAT ACATTTTGGA CCACGAAATA GCATGGTCAA
      781  CAATCTCCTT TTTCTTTGGT TCAACAGGGG TGGGTTTAAC CTCAACAGGT CGAGCCATAC
      841  TCGAAGGTAT TTGTTTCACA AGAGCTATGT CCGGTTTAGA GTCTTTAACA TTTGATTTTG
      901  ACTCAAACGC AGGTAGATCA TTTTCCCATG GTTTAAATTC CTTAGGAGCG ATATATGGAG
      961  CTTTTTTAAC CATTTTGTCC CACTTCAAAT CTGATGACAA ATGACTTGTG ACAAATTCTG
     1021  GAGTTTTTGC ATCTGCCAGG ACACATAAAG GGGAGGGAAG AACAGATGAT CTTTGTTTTT
     1081  CTGGAACAAC CTCTTCGGAC TTGACAGCTT CTTCATTTTC CGAACAATCA GAGTCGGATA
     1141  ATTCTTCATC ACTTGCCACT TGGAAGAAGG ATAAGGTGCC TTCAGTATCT GGACTTTCAG
     1201  TCATTATAAA CTACCCGTCT GGCCGTCTAT ATGCAAATAT CCTATCAATA TCAACTAGAA
     1261  CATTTTGCAT ATATTGATTT ATTAATATGA ATTTATTTGA GAAATGTTTA AAAACATAAC
     1321  TGTAACCTTA ATTTATATAC TACATTTCAA AGTAGCGTGA TAATATAGTT ATCTTCTGAC
     1381  ACCAGACATT ATTATACACA TACACATATA TAGTTATAAA ACATAAATGC TAAGAGTTAA
     1441  AACGCGGATT TCCTTGTCCA ACTTTAATTT TATACGCCAA ACCTATGCTT GGAATATACC
     1501  CACTTTCAGT TTTTTGTTTA CTAATTAAAT CATTGACTCA GTGACTGACT ACTCACTAGG
     1561  GCAGTTTCTG CCGTTTAAAA CTCTGTTTGA CACTTACCGG GCAACTTCTT GATGTTCGTA
     1621  CGAAAGTTTA TAACCTCTTA TTGCTTAGAG GGTTACGCAT GGACGATTTA AAATTATTTC
     1681  AACCCCGCAA TAACCAAAGC CTAA

//

LOCUS          15     2592 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      15
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS             join(500..563,842..993,1220..1369,1578..1734,1881..2092)
                     /translation="MNSPCLRTALSGIYIRHIGALGSYTNIPKRSRYFWEARPECRE
                     RKLLMQITLPRLKQAVDPTQFVQEVEEEKLYGLDVILHKEVQQIFNENELIAVCLYS
                     EMPSDNFYQVRYELQQHGMSMEKYPEHVMEFTLKQTPYENMLELFRGETAVIFGEAN
                     VKQMISAVSKNNYLLLLGGLVQGRYMSLKELIKFSNLAGLDSSRAELGFLLNTAVSS
                     TKNMMQKPTNQLVHSLQQYIEQCSNASNEN"
ORIGIN
        1  AAACAATAAC TGTATGAATG CATCGTGAAT TTTATGTTTT TGTACCCATT GTGCTTCATT
       61  CTTTTAGTAA TCAAGTTGTC GTGTACTATC TGAATACTTT TCATATATTT ACTACTAAAA
      121  AAGCAATGCT TATTAACCTT GGTTTCAATC CCAAATTTTG GATTGGGTTA CCCATTGTAA
```

```
     181  TGTTTCATTA AACCCTTTAT CTGTTAACAC AAAAGTTAAA TCATTTTCTG TATGCTGACA
     241  AAAAGGTCAA TTAAGTGTTA TTGCTAAAAT CGAGTGACAA AATTTTATTA TGTATATTTT
     301  TTTGATTATT TGGTATGGTC GGTAAAAGGC TTTAAAGCAT TGGAGTAAAT GAATATCATT
     361  CTTTTTTTCT TATATATGGA AATATAAACT TCTTTTTGCA TCTTCTTAAA AATGACCTTT
     421  GCAGTGATAC ATGAGCACTG ATAATGTGGC ACCTTTTATT TTAACATTTT ATGTTTAGAT
     481  CACCAACAAT AAGTTTAGGA TGAATTCTCC ATGTTTAAGA ACAGCTCTAA GTGGAATATA
     541  CATTCGTCAT ATTGGAGCTT TAGGTATGTA AACTGTTTTT ATTAAGATAT GTTCCTATGG
     601  GTGCATGTAA GATTATTAAT GGTCATTAGG GATGTCAGAA CATCTTATAA GTTATGATTT
     661  AACATGTGTT CCCCAGCAGG TTTTATAGGT GTCTCTAAAG TTTTAAAAAT TATAAAATTC
     721  AATTAAATGA TTACAATACT AACAGGCATT TGAGTATAAC ACTTTCTGAA ATATTACTCC
     781  CCATAAAAAT GCAAGTTTAT GGTAGATTTG TTGCATTCTT TCACATAAAT TTCAATGGCA
     841  GGATCGTACA CAAATATTCC AAAGAGAAGT AGGTACTTTT GGGAAGCAAG GCCAGAATGC
     901  CGAGAACGAA AACTCTTAAT GCAAATTACT TTGCCTCGTC TCAAACAAGC AGTGGACCCC
     961  ACGCAGTTTG TTCAAGAAGT AGAAGAGGAA AAAGTTAGTT ATTTTAATGT ATTTAACCTC
    1021  TCTCTTGGTC ATGGCTTATG TTAAATATAC CAGTAAGTTT TTGATAACAG AACAGGGTCC
    1081  AGTATTTGCT GACCCAAATT TAAGGTGATA GTCCTCTAAG GCCACAATTT AATTTATTGT
    1141  TATAATTCGT TCATTAACAT TTTGCAACCA ATGATTGCAC TAAAGCAATA ATAATTATTA
    1201  TAAGCCTTTT CTTCAAAAGC TTTACGGACT TGATGTGATC CTGCATAAAG AAGTTCAGCA
    1261  AATATTCAAT GAAAATGAAC TTATTGCTGT GTGTTTATAT TCTGAAATGC CCAGTGACAA
    1321  TTTTTACCAA GTCCGATATG AATTGCAGCA ACATGGAATG AGCATGGAGG TATTTTGCAA
    1381  AGTTTGCTAT GATAATCTGA TTATATTGTA CATGACATTT GTTTCTGTGA CACAGGCCGT
    1441  GGGTGCCCCA TTTCATACAC TGCATTTTAT TTTGCATAAC CTTCTGACAA TAAGTAGGCA
    1501  ATATTTATTT ATTCATTTAT TTATTTAACA CATTGATAGT TTTATTTGTT TTGAAAACAT
    1561  ACTTTTGTAT TATGCAGAAA TATCCTGAAC ATGTGATGGA GTTCACATTG AAACAAACAC
    1621  CATATGAAAA TATGTTGGAG CTTTTTCGGG GGGAAACAGC AGTCATATTT GGAGAAGCAA
    1681  ACGTCAAACA AATGATTAGT GCAGTTAGCA AGAATAATTA CCTTCTTTTA CTTGGTTTGT
    1741  ATTTTTATCA TTTATAAACG GATGATTTAT AGTGTATTGG ATTTTATTGT TTACTAAAGA
    1801  TGTATTAGAT ATTGTTTCAT TTTGCACAAC TTTCTTCTAT TTTATTTAAC ATTTCTTAGT
    1861  AATTATTTCT ATTTTTATAG GTGGTTTAGT TCAAGGAAGA TACATGTCTT TAAAAGAATT
    1921  GATAAAATTT TCAAACTTAG CTGGTTTGGA TTCATCTAGA GCGGAATTGG GCTTTTTACT
    1981  GAACACAGCA GTTAGTTCAA CTAAAAATAT GATGCAGAAA CCAACTAATC AACTTGTTCA
    2041  TTCATTACAA CAATATATTG AGCAGTGTTC TAATGCTTCA AATGAAAATT AAATTTTATG
    2101  TCCGCGGTTT AATAAATATT AATTCTTTTA CAACTTGTTT CCTGAATTCT AACAAACGTT
    2161  TGACCAAAGG TCACATTTAA TAATGCTTTA ATCTTTCAGC TTTTACCAAT CAATTTAATT
    2221  GTAGATTTCA TTCTAAGTCT TGCTGGTCAG GTTAGTACTA CTGTATATAG CTCTGTGAAC
    2281  TGGGCTTCAT ATGAAATGTC ATAAACTAAG GGACACAACA TATCAGAGAT CAAATTTTGT
    2341  GTTAGTTTTT GGACTTCTAT TGTTTTCAAT ATTAAAAGCT GTAATACACG TTTTAATGAA
    2401  GAATCATAAA CATGCGTACC TTATATACTT TCTAGTGTGA ACAACTTGTT TTATGTCAGA
    2461  CAAAGTAAAA TACTTTCCAT TGGTAGTTGT AAGGGCCAAC TACTATTTCT AAATATCACG
    2521  TACTTACCTA AATTAAAATG ACGTCTGGCT AGAGAAATGA ACGTATGTTT TGTATTGATT
    2581  ACTCATGAAA AG

//

LOCUS          16     1968 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      16
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS          complement(join(500..856,917..1155,1315..1468))
                  /translation="MADKEYDGSAGIEASAAKILTRAITLEKEARLQAALICYQEGI
                  DLLFDALKGCKDSAKKKHFRDKLEEYMTRAEQVKTRVNHLKAAGKYHDQTKIEENST
                  GHSYASLFQPYFDEFVREVWVEDAYIRSYHQVNNFLRFCEFVVTRCKNVRAIHLKTT
                  HDDYGNQDQQISNLESIRASLKSREIKLVIMFSSSLHDREIRLDTGWILKIGRGLDY
                  FKKAEKFSLGFFDMDMRPCHSTTVDVFHSNHTRER"
ORIGIN
       1  ATTATGCATA ATTAAAACTT TAACATTACA GGAGCCTATT CAAGAACGAC ACAAGGAATT
```

```
      61  GTTAAAAATT CGTGACTATC TGAAGAAAAA GGTTGAAGAT TTGGAACGGC AGCTCGCAGC
     121  AAAACAGCTT TCCCCTTCGG GCGGACCAAA ACTAACTGTT CATTCAGATG CAGAGCCTTT
     181  GGCTGTTTAG TAATGCTTAT GCACGCAATA TTTAACAAAT GTTTCCATAG TATTTGTCCT
     241  AACTGTTCAT CTTTATTAGA GAGTGTGTAT TGTTCTTTAA ATGTTGATCT TTTTATAATC
     301  ATATGTTCAG AATTATGTAT GTTTAAAACT GTAGCTTTGC ACAGCAGCCA TGTGTTTGCT
     361  TGGTGGTCAT TAATTAAAGT GATACTTGTT ATTTTACTGA AGGCTCAAAA TTAAATGATG
     421  TATGTATGTA ACCATTTGAT TTTGAGGGAT TGGGACAGTG TTTAAATTTT GTTTGGACTG
     481  TACAGTATTG TTTCAAGTGC TATCTTTCTC GAGTGTGATT TGAATGAAAA ACATCCACAG
     541  TGGTAGAATG GCAGGGCCGC ATATCCATAT CGAAGAACCC TAATGAAAAT TTTTCGGCTT
     601  TTTTAAAGTA ATCAAGACCA CGGCCTATTT TCAAGATCCA TCCGGTATCC AACCGAATTT
     661  CCCTGTCATG CAGGGAGGAA GAAAACATAA TTACTAGTTT TATTTCACGG GATTTGAGAC
     721  TGGCCCTAAT GCTTTCTAAA TTGGATATCT GTTGGTCTTG ATTTCCATAG TCATCATGTG
     781  TAGTTTTTAG GTGAATAGCT CTGACATTTT TGCACCTTGT CACAACGAAC TCACAAAACC
     841  TTAGAAAATT GTTAACCTGC AACATAAAGT ATTTAGACTA TAGTAATTTA AAAAGTGATT
     901  GCACAAAATA GCGAACCTGA TGATAGCTTC TTATATATGC ATCTTCAACC CATACTTCTC
     961  GAACAAACTC ATCGAAATAT GGTTGAAATA AAGACGCATA ACTATGTCCG GTAGAATTCT
    1021  CTTCAATCTT TGTTTGGTCA TGATATTTCC CAGCAGCTTT AAGGTGGTTT ACACGAGTCT
    1081  TTACCTGTTC AGCCCTTGTC ATATACTCTT CTAGTTTGTC TCGAAAGTGC TTCTTTTTAG
    1141  CACTGTCTTT ACAGCCTATA AAATAAAATG GAATATAATT TAATGGTTTC AAACCCAATA
    1201  AAACTTTAGA CGAGACATAC AACTTGTCTT TCTGCTATAA AATATTTACC AGATAGCCTT
    1261  TACTTTTACA TTTATTTATT TACCTTGACC AATTAGTAAT TCCTGTACAC GTACCTTTCA
    1321  GTGCATCAAA CAACAAATCG ATTCCTTCTT GGTAACAAAT TAACGCCGCC TGTAGTCGTG
    1381  CTTCCTTTTC CAGTGTAATT GCTCTCGTTA ATATTTTTGC AGCGGAGGCT TCAATACCAG
    1441  CTGAACCATC GTATTCTTTG TCCGCCATAC GCTTTTAATT TCGACGGATT GAACATTGAA
    1501  GGGTTAAAAA CAAAATTGAG ATTCAGGATA TGACGTCAAA TGACAACGGC TGCGTCGAGA
    1561  CTAAAGTCAC ATGACCAGCA TCGTTTTTCA TTGCTTACTG CTTTGTTCGC TGGTGACAGT
    1621  ATTGTGTCGT CATGCTGTGT TTTTCGTACT TTAACGTTCG AATTTATGAA GTTAAGTTAT
    1681  TAATTTATAA TGCTTATGTT TTCCAGCGGT AGAATGTGCA TGAGTAGTCT ATAATTTTGT
    1741  TAGTTTTCGT GGACCAAGGC CCTTTTTCGT CAATAAAAGT GCCCGCGGTC CAGTTTCAAT
    1801  ATGGACGACG TTATTGAAGA ATATTCGTCG TCATTGCAAG ATTTAACCTT CAATTCCAAG
    1861  CCACATATCA ACATGCTAAC CATCCTGGCC GAAGAAAACA AAGCCCATGC TGAGAGTGTC
    1921  GTCAAAGTTA TAGAGGCCCA GTTGTATAAG GTTAATTGGT TATTTTAA

//

LOCUS          17     1761 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      17
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS          500..1261
                  /translation="MGSWRSYLTSGILICISRFPFWEALSTMMLLAEIEQKSKRKLN
                  IPTYLVYIDAAVLILCGTFMSFQASARRSASLIAATHLAVRTLVQFHFHKMKSEVMI
                  HILLRNVGVIASYLLLASGIGTTKAKHRSLMDLSVFAIGTLYAAIAGLLWSSKFEQT
                  FVKRSLPQQVTRQVSLKLLYSFLEICLAICALLFATQRKGLQQAKKFALIFWLFLMV
                  PSDFLFSRMDSDIETWSHLRLMASSFCTLGTIVVLNFVN"
ORIGIN
       1  ATCACGTATG TTGTAATCTA ACTTGTGTGC TTTACTGCTC ATGCATATCT GACGTAAAAC
      61  AGTGCTGCCA TACGAATTTT AATATCGTAT AAATCATGCT AAAAACGAAG AACACAAGGT
     121  CTTATAACGC AATATTTATT TTCAAACAAT AAAATACATG CAATCGTATT ATTTATAAAA
     181  GAAGCTGCAA AAATAGTATT TTAAACAGAC ACAAAGACAG TTAAATTTTT GTAGTTGGCA
     241  ACACGTAATA AAAAGTTCAA ATTGAACTTT GACCCCTACA ACATGCCGTT AGTATACAAA
     301  CACTGTTTTG TAAACAATCT GATTTTGACC TTGTTATATC AAGGTTTTGT TTAATATGTT
     361  TGCCAACATT GTGGAGTTGA CAGACAAATT AATTATCAAA ATTGTTACCC AAAGCAATTA
     421  TAGAAATGTA ACTATATATA CATTTAACAT TTATTTATAA ATATAGGTTT TGTGCTAGTA
     481  AATATTATAA ACACACAAAA TGGGAAGTTG GCGATCGTAT TTAACTTCTG GCATATTAAT
     541  TTGTATTTCA CGTTTTCCAT TTTGGGAAGC ATTATCAACG ATGATGTTGC TGGCTGAAAT
```

```
     601  CGAGCAAAAA TCAAAGAGAA AATTAAACAT TCCTACATAT TTGGTGTATA TTGATGCAGC
     661  TGTTTTAATT CTTTGTGGAA CATTTATGTC ATTTCAAGCA TCTGCACGAC GCTCTGCTTC
     721  TTTGATTGCA GCAACTCACC TTGCAGTAAG GACATTGGTT CAGTTTCATT TCCACAAGAT
     781  GAAATCTGAA GTTATGATTC ATATCCTGTT GCGTAATGTT GGAGTCATTG CATCCTACTT
     841  GCTGCTAGCA TCTGGTATCG GCACAACAAA AGCAAAACAC AGATCTCTGA TGGATTTGTC
     901  TGTATTTGCT ATCGGTACAT TATATGCAGC AATAGCAGGG TTGCTGTGGT CGTCTAAATT
     961  CGAGCAAACA TTTGTAAAAA GATCTCTCCC TCAACAAGTA ACCCGGCAGG TCTCTCTAAA
    1021  GCTTCTTTAC AGCTTCCTTG AAATATGTTT GGCAATATGT GCTCTGCTGT TTGCAACACA
    1081  ACGCAAAGGG TTGCAGCAGG CCAAAAAGTT TGCCCTAATT TTTTGGTTGT TTCTTATGGT
    1141  TCCATCAGAC TTTCTCTTTT CACGAATGGA CAGTGATATT GAGACATGGT CACATTTAAG
    1201  GCTAATGGCT TCTAGCTTCT GTACCTTGGG GACAATTGTT GTATTAAATT TCGTCAACTA
    1261  GACACTGACG ATTTTATTTA AATAATCAAG AAAAAGTTTG GTTCATTGCA AAAAAAAGAA
    1321  TATTATTTTG CAGCTGTGTG TTTGGTGCTC CAGATCTTTT AAAAGAAAAA TCGTATGCAA
    1381  TGTAATCTTT AGTTGAAAAC TTTTATGTGT AGTATGTATG CAAATGTGAA GCAAAGTATT
    1441  GATGCAACCG TTTTTATTCA AGGGCCTTCT GTAGAAATGT TATGGGTTCA ATTAAAAAGT
    1501  TAAATTTCGA GGTATGTGGT AAGGTACAGG GTGTATTTTT TCGTAAACAT ACAAAGAAAA
    1561  CTTGCGATAA ACATGGTGTG GCTGGCTGGG TCTTGAACAC AAAGGGTGGA ACTGTCCAAG
    1621  GGTTTCTGGA AGGATCACAA CAGAGTGTAT CCAAAGTGTA AGCATCAGCT TGGATGTGTG
    1681  TTTTGCTTGC CTTCATATTA TATTTAAACA TATTTAATAA AAACTGTCCT ACTTTAATAA
    1741  TATCTGCAAA TTTACGTGTT T

//

LOCUS          18     1860 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      18
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS          500..1360
                  /translation="MSNSSYKSLLRNLPRVNLQNLKANPYAEVKHSPNRGPHAIGLK
                  NGVRPNIKPRIGYVQDKTPFQKRIPKYGYNREQHLKRQYFPLTLWQLQRHVDLGRID
                  PTKPVDITAISNGRVLSMVGSESNYYGIYLLEQGANIFKAKLDIEVQIVDELALATV
                  EKAGGTVSTAFYDRRSFVALCNPVEYFLMGKPIHKRLLPPQELVTYYTDPKVRGYLS
                  DPAKIWEERMKLAKKYGYDLPELSEEQKHLMVEGKKDPRQVFNGLMPGSIVNLADKT
                  VLVPENEHFKQHAAM"
ORIGIN
       1  TTCCCGTCCA ACAAATTAGA CGATTTACGT AAATTGTCAT TTCTTGTTTA TGATTTCCGT
      61  ACACGAGCAC GAAATACAGT CATGCAACAG TCTCCATTTA CGCTGCAACA TGCCGCAATC
     121  ACCGGCGGGG AATAACCCTA TTAAACTCGG TTGACATCGG AGTTTTTCGG AAAGTAAAAC
     181  AAATTTACAT AAATATGGAG CCATAAATAC TACAACTGTA TAAACAAATT AACAAATATT
     241  TAAACAAATA TGATTTACTA TTTAAATAGT AGAGTAGTAT TTTGCTTACC GTGTGGATAA
     301  ATATATATTT TTTATCACTT TGTGTAAAAA TGGTAACACT GTAATGGTGT AGATTTTGTC
     361  AAAATTAGAA TAACCATGCG AGTTTAAAGA TTATGTTTTG AGTGCGGTGC GTGTTTCATA
     421  GCGGTGTTGT TATCTGTTAG ATTTGTGTTT CAGGACGCTA TCTGCTTTAT AGGTACAACA
     481  GCTTTGAACC AATGAGCAAA TGTCAAATTC ATCGTATAAA TCTTTGTTGC GTAATCTTCC
     541  AAGGGTAAAT TTGCAAAATT TGAAAGCAAA CCCTTACGCC GAAGTGAAAC ATTCTCCCAA
     601  TCGCGGACCC CATGCAATAG GGCTTAAGAA TGGAGTACGA CCAAATATCA AACCAAGGAT
     661  TGGGTACGTT CAAGATAAAA CGCCGTTTCA AAAAAGAATC CCTAAATACG GATACAACAG
     721  AGAACAGCAT TTAAAACGAC AGTACTTCCC ATTGACGTTG TGGCAATTGC AACGTCATGT
     781  TGACCTTGGA AGAATTGACC CCACTAAGCC AGTGGATATC ACTGCTATTT CTAATGGCCG
     841  AGTGTTAAGC ATGGTTGGTA GTGAGTCTAA TTATTATGGA ATATACTTAC TTGAGCAGGG
     901  TGCCAATATT TTTAAAGCAA AACTTGATAT AGAAGTTCAA ATTGTAGATG AACTGGCTCT
     961  TGCCACGGTT GAAAAAGCTG GTGGTACTGT TTCTACTGCA TTCTATGACC GCAGAAGCTT
    1021  TGTGGCTCTG TGTAACCCTG TAGAGTATTT TCTTATGGGA AAACCAATCC ATAAAAGGCT
    1081  GCTACCACCA CAGGAGTTGG TAACTTACTA TACAGATCCC AAAGTGAGAG GCTATTTGAG
    1141  TGATCCAGCA AAAATATGGG AGGAAAGGAT GAAACTAGCT AAAAAATATG GATATGATTT
    1201  GCCTGAACTT TCTGAGGAAC AGAAACATCT GATGGTTGAA GGAAAAAAAG ATCCAAGACA
```

```
    1261  AGTGTTCAAT GGGTTAATGC CGGGATCGAT TGTAAATTTG GCAGATAAAA CAGTTTTGGT
    1321  GCCTGAAAAT GAACATTTCA AGCAACATGC AGCGATGTAG CAGTTACTAT TTTGTTTTAA
    1381  ATTTTACATG TTGTGAAAAC TTGAAATCTG TTCGCATAAT TAATATTAGG TATGTTTCAA
    1441  ACCGTGAACT ATTTTGTGTA AATATATATC CTTTTTTCAG AATAATTAGG ACAATATGAC
    1501  TGAGGATGTG GTTAAAGTAA AGAAACGTAA ACATACAAAA GATGAAACTC AAGAGGAAGA
    1561  GGAAGCTCAA AAGGAAGTAA AGGAAAAAAC AAAGAAAAAA AAGAAGTCAA AAAACAAGAT
    1621  GCAAAAGACC GAAGAAAATG AAATTTCACC CGAAAAATCC ATTGAACCGA CCCCTGCAGC
    1681  CAAGCCTGAA GAAGAAAAAA ATAAAGAAGA TAAGCCCAAG GAAGAAAATA CCGAAATTAT
    1741  AAAGGAAACA AACAAAGAAA ATTCAGAGAC TGATGGAAAA AAGAAGAAGA AGCGAAAGAA
    1801  GAAGAAACAG CTGCCTCCAC TTCCTAATCC AGATCCAGGA GCTTTTGAAG CCATGTTAAC
    1861

//

LOCUS          19     3928 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      19
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS
join(500..611,672..736,843..926,1234..1290,1355..1401,1572..1634,1693..1739,2095
..2176,2236..2278,2431..2512,2674..2761,3284..3428)
                 /translation="MEDWLHCNACFVRIIHGPVSFSVTNCAHIFCNKCIERVANSTC
                 TVCNKKCSAYTLLHNKMDPTTANYFKDPGELAQKAMNMFKLSEDVRKFQATQRKSLF
                 KAIVNKAEQKEKLLQTAVATMSETKRKNSELKFIIKNKNNLSSVSPKSHSYMAMQYS
                 PDCSPPSSPNRNISTRKTPPHYYSSVPCTPNKTSVDGHQHRTLLSGGYTVSPNKEKP
                 RMSLRSPPVGGKIQTPPDLRLHHSSPDGIVSKQASLYDCSGINEPHRKTKTRGSSIK
                 VMSRQLRIETAESSANTPESLKKICPVMKSHKD"
ORIGIN
       1  GTAAACATAC ATGAGCCAGT CTGCAAAGTA TGGCAGTTCC TGACGAGATC TCGCGCGTGC
      61  AAGACGGTTA TTTTTCGCGT TTGCGCGTGC TATCATTAAA AATCTGCGTG CAAAGCATGT
     121  ACATTTTCAA GCCACGTTAG GAAATCTTAC TTAATTTAAT ACGTATTTTT TTCTTTGTTT
     181  AATTAATTGG CGTGCGTAGG ACAAATTGTA GCGTGCATAA CCAAAAGTCT GTAAAAAAAT
     241  CGCGCGTGCA CCTGCCATAC TCTGCAGTCT GGTACATGAG TCAGTTTACC GTCGAAAGCC
     301  CGCAGTTGCT TAATCTGTAA CTAACGTTTG GAAGGATTCT TCTCTAGTGG TGTTTATTCC
     361  AGTGGTTAAG AATAGTCTTA GGTGGTATGG TAGCAATAAT GTACAGTGTA TAGGCAATTT
     421  CTAGTCCTAC TGCTTACCCG ATAACTGAGT TTAAAGCATA TTGGATGAGA TTGTAAAGGT
     481  CATATAGAAG AGAACGAAAA TGGAAGATTG GTTGCACTGT AATGCTTGCT TTGTAAGGAT
     541  TATCCATGGA CCAGTTTCTT TTTCTGTGAC AAACTGTGCT CATATATTCT GCAACAAGTG
     601  CATTGAAAGG GGTTAGTACT ATTTGCATAC TACATACCTT ATAGATGTGT TTTGTTTTAC
     661  TTTCATTGCA GTTGCCAACT CCACTTGTAC TGTGTGCAAC AAAAAGTGTT CAGCATACAC
     721  CTTACTTCAT AACAAGGTAA AGGAAATATT TTAAAACGTT TAGTGGCATC ATTTACTTTT
     781  AACAAGTTTA CCTAACATAG CATGTTCAAA AATCATTATA ATGTATATAA TTGTTACACC
     841  AGATGGATCC CACCACAGCA AACTATTTCA AAGACCCAGG AGAACTTGCC CAAAAAGCCA
     901  TGAATATGTT CAAACTATCC GAAGATGTGG GTTTTTAACT TTACAGTATA ATATATTTAC
     961  TTTACCTCTG GCCTCTGTCT CTTTATTTTG TTGTATATGG ATGAGTTTCT AACGACAGGG
    1021  TTCAGACAAG AGTTGGTTCG TTACTGCAAG AACCTTCCAT TCTATAAGAA AAAATAAGTT
    1081  ATATATATTA TAACTTTGCT GTTAACATGT TGAATTTGGT TATGCAATGT ATGTAAACTT
    1141  CAGTTAAGTG CAGTTTAAAG TATGAATATA AAACTCCTGT GGCCTATACT ATACTTTTAT
    1201  ACTAACCGAA TAATTTTACA TATTTTTAAA CAGGTTCGAA AATTCCAAGC TACCCAAAGG
    1261  AAATCTTTAT TCAAAGCGAT TGTTAACAAG GTAATGTATG TTAAAATTTT GATATATTGA
    1321  TAAAAGGCAA ACTGACACAT CTGCTTATTA TCAGGCCGAA CAAAAAGAGA AGTTACTGCA
    1381  GACAGCAGTT GCAACAATGT CGTAAGTATT AGGGAAAATA ACCGCAGTTT TTTCGCAAAG
    1441  AAAATTGATT TGTTTCACAG CTTGTGACAA AAATAAAACA CCCAGTAGTC TAAACTTTGC
    1501  AGGAGTTTTT GAAATACTAA GGAAACTTTT AGTTTATTTG CAATGCAAAA ATATCAATTA
    1561  TTTTCCTTTA GAGAAACCAA AAGAAAGAAT TCAGAGTTAA AGTTTATCAT CAAGAATAAA
    1621  AACAACTTAT CAAGGTAATA TCTGTGAATT TAATCTACAA TACATTAACA TAGAGTTTCT
```

```
    1681  TACATATTTC AGTGTCTCAC CAAAAAGTCA CTCTTATATG GCTATGCAAT ATTCACCAGG
    1741  TTAGTAATGT ATTCACTTAT TTGGTCCTGC ATCAATTGAC AATTTATTTA TTCCTGCGTG
    1801  GCAGGGCAAG AACAGTTGTT AGAATACGAC TGTTCTGTTT TATAAACCTC ATGTCCAATT
    1861  ACCATGTATC TATGTAACTT TGTGGATGAT TTTTTCGTAA ATAATTTTCC AACAGCTTAT
    1921  CCGGAATGAA AAAACAGTGT ATTCCCGACA TTTCGTCTGC CATTCGGCAA GACTTTAGTA
    1981  GGAAATGTGT TTTTTTTATT GTTTTTTTTT TAATCTAATG TAATAAGAAA CCTAGTCAAA
    2041  AGGTCTATAA AATGGGTGTT ATGGATTTGT TTTAACATTT ACTGTTTCTT ACAGATTGTT
    2101  CTCCCCCTTC ATCTCCCAAT AGAAACATAT CCACTCGTAA AACACCACCA CATTACTACA
    2161  GCAGTGTACC ATGCACGTAT GTACTCATAT TCCATGTTCA ATATATACAG GTTACTTTAT
    2221  CTCCGTAATT TTCAGGCCAA ACAAAACTTC TGTAGATGGC CACCAGCACA GAACGCTTGT
    2281  AAGTTATTTG TATTAACTCT TATAACCTTT TTTCACATAA TCTGCTGCGT TTAGGAATAG
    2341  CAGCAGAATG TTATTGGACA TACACTTTAT AACACTATCT ACGTTTATTA ATCTGTTACC
    2401  ATAATCACAC TAACACTTCA AAATGTTTAG CTTTCTGGTG GTTATACAGT GTCACCTAAT
    2461  AAGGAAAAGC CAAGAATGTC TCTTCGTTCT CCACCTGTTG GTGGAAAAAT ACGTAAGATT
    2521  ATAAACTATG CTTTTAGTAG TCATACTTTT GCTTGGTTTT ACCCTTAGCG TATTTTAACG
    2581  ACTGTCATTT TTACTTCTAA TTCCATTTTC TTTGTTATTT CTACAAGATA AATGAAAGTT
    2641  GTATTTTATC TTATAATGAA ACCTGTATTC TAGAAACACC TCCAGATTTA CGTCTACACC
    2701  ACAGCTCACC AGATGGAATA GTTTCAAAGC AGGCCAGCTT GTATGATTGT TCTGGAATAA
    2761  AGTTGGTTGT TTTTAACTAA CTCAATTGTT TTTACAGACA TTCAATGAAA CAAAGCACTT
    2821  ACTTCTCAAA GTTCCAAACG CCTAGTAAGT TTTTCTGTTA TTTTCCACTC ACAAGAAATA
    2881  TCTATTCTCT ATATATGGAT AAGACTCCCA GCAAGGCATA ATATGCAATT GAAGCTGTGT
    2941  GACCTTGGGC AAAATAATAC TTTAGCCAAA ACCCTGTAGT TCCGAACTGA TTATTTTAAG
    3001  CGTCAGCCAT TAATTATATA AAAAATGTAA CAAAGGGTAT GAAGCAGAAC ACCCATGTTG
    3061  TAACGACTTT AGTTTCCCGC CAAGCGAGGA TAAACAAGGA ACATCATTTT TAACGAGATA
    3121  CTTTTTTCTT GTAGAATCTC CCATGCATGG AACCACCCCA CATAGTACAC ACAGCACTCC
    3181  ACAGTCGAGG TGAGGGTTTA TTACTAAAAT AAATTTGCCA AACAGATAAA ATACATTGTC
    3241  AGATGTGGAA CTCCCAGAGT GAAAGCCATT ACTAACTATG GAGTGAGCCT CATCGAAAAA
    3301  CGAAGACTCG TGGGTCATCA ATAAAAGTTA TGAGTCGTCA ATTAAGAATA GAAACTGCCG
    3361  AATCATCCGC AAACACTCCA GAGTCATTGA AGAAAATTTG CCCAGTCATG AAGAGCCATA
    3421  AAGATTGATG ATGTAACAAT AAAACATTCG ATTTACATTC TGATTCAAAA ACCTAATTTC
    3481  GAGGTGTTTT TCGTTGCGAA ACCAAGCGTT GTCGCAGGTT TCTTACTTCA TCCATTAGGT
    3541  GTCGCTCTTC AGTCCTTAAG TTTCCCCTGC CACGGTCGAG TGACCCTGCG TGAACATATT
    3601  TTATCGCTTG TTGTTTTAAA AGTTAATTAA ATGTTGCCGT TCTTACATTG AACTGAGTCC
    3661  TCTATATCTG ATACATCGTT CAAACACATC TTGTTGCTCA CAGCAGTCAT ACGATCTTCT
    3721  GCAGCGCTAC GATCTATAAG ACAGATATAT TAATATTTAC ATAGCAATAT GGGAAGAACA
    3781  TTAATGCTAT TATGGTATAA TGAGAAGAGC ATTGCTAACA TGCCATTTGT AAAAATTGAT
    3841  AAACGATAAT TAATAAACAG CCTGAAATTC ATGACATACA TTTAATTGAC AATTTTTTTA
    3901  CTGAATTAGA AAATATTTAG TGTTTACT

//

LOCUS          20     2180 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      20
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS             join(500..586,655..758,912..1680)
                     /translation="MIRLLYVVCCFLILQTSKASEGDRCVQYQGCLQPCQHQCKDGI
                     FDKNQTRYMLLLGWDCREECKYTCMWKTVEAYEAANVRVPQFHGKWPFVRIIGVQEP
                     ASVLFSILNGISNIWAYKQYYSAVSSNAPLYTTTTIHAILAANAWLWSSVFHARDFP
                     WTEKLDYFCATSLVLYSFYLSIHRLSYELHGHNVHILRWIAGNLIGAFYLGHISYLT
                     FQSFDYGYNMKANVAVGLMNSITWLSLCFRKRKKHLHVKKMAAAIIMTNLFLMFELS
                     DFPPVWWTFDAHSIWHFLTIPMPLLVYSFLKDESLKLQKEYETKQKFP"
ORIGIN
       1  AGGGAAAGGA GGACGGGAGA CATCAAAAAT CAATGGGGCT TCAGGACCAG GCATTGCGTG
      61  CGCCATTGTA GCACGGTCTG CTGGATTGTT TCAAAATAAC AAACAGACTT GTGCTTGTGA
     121  TGGAGTATCT GTGTGGGACG AAAGGAATGT ACCTTTAGCA GGTGCAGAGA GAAGCAAGCA
```

```
     181  TAAACAGAAA TTGTAGTTTA TTGTTTTAAA TTTAATTCCT CATCATTGTG AAATGTTTCT
     241  TTGATTAGTG TTGTACATCC GCAAAGACCC CAACAGTTTA ATAATACCTG AGCTTAATTA
     301  GCATCCTTGG CACTTAGCAA TGCATATAGG CAGTGGTACA GCTCAACCCT TCTGTTTACC
     361  AAAATTTGTC AATCTATTAT ATCGCTTATC GTTTCTATTT GTTCTTTTTA AAGTTGGCTG
     421  TGCCACTGCA TTTTTAGGGG TGTTCCTTTG TAAAATCATG TTGTTTCATT TAAAGGTACT
     481  ATATATATAT ATGTATGCAA TGATACGACT ATTGTATGTA GTTTGCTGCT TTTTGATTTT
     541  ACAAACTTCA AAGGCATCAG AAGGTGATAG ATGCGTACAG TACCAGGTTA GCTGATAATT
     601  TAATTTAAAT GGGGATGAAA TGGAAGTTTT TGATTATAAT GTATCTGTAT TTAGGGATGC
     661  TTACAACCAT GCCAGCATCA ATGCAAAGAT GGAATCTTTG ATAAGAACCA AACCAGATAC
     721  ATGTTGTTGT TGGGATGGGA TTGCAGAGAG GAATGCAAGT TAGTGTTGAT TCATAATATG
     781  GTTGAATAAA CTTGTATTAT AGTGATATTT GGATCCTGAC TTTTAAACAA ATTTTCTTTC
     841  TGAATTTATG CCTGAATGTA TCTGATGATT AAAATAGCTT TGATATTTTT TAACAAAAAG
     901  TAATTCCCCA GATACACTTG CATGTGGAAG ACTGTAGAAG CTTATGAAGC TGCAAATGTC
     961  AGAGTTCCAC AGTTTCATGG TAAATGGCCA TTTGTAAGAA TAATTGGTGT CCAGGAACCG
    1021  GCGTCTGTTC TGTTTTCTAT ACTGAATGGA ATATCCAACA TATGGGCTTA CAAACAATAT
    1081  TATTCAGCTG TTTCATCCAA TGCACCATTG TACACAACCA CCACAATCCA TGCAATACTT
    1141  GCTGCAAATG CATGGCTATG GTCCAGTGTT TTTCATGCAA GAGATTTTCC ATGGACTGAA
    1201  AAGTTAGATT ACTTTTGTGC TACATCTTTG GTTTTATACT CATTCTATCT TTCTATTCAT
    1261  CGCTTATCAT ATGAGCTTCA TGGACACAAT GTGCATATTT TAAGGTGGAT AGCTGGAAAT
    1321  TTGATTGGTG CCTTTTACCT TGGCCACATA TCTTACTTAA CCTTTCAAAG TTTTGATTAC
    1381  GGTTACAACA TGAAAGCAAA TGTAGCAGTT GGTCTCATGA ACTCAATTAC TTGGTTAAGT
    1441  CTCTGCTTTC GAAAAAGAAA AAAACACCTA CATGTTAAAA AGATGGCAGC TGCAATAATA
    1501  ATGACCAATC TATTTCTGAT GTTTGAGCTT TCCGATTTCC CACCAGTATG GTGGACATTC
    1561  GATGCACATT CTATATGGCA CTTCTTAACT ATTCCAATGC CATTGCTTGT GTATTCATTT
    1621  TTAAAGGATG AAAGTTTAAA ATTGCAAAAA GAGTATGAAA CAAAACAAAA GTTTCCATGA
    1681  CCTACCTTTT TTACATTTTT ACAAACACTG TAGTACTGTA CTTGCTATAA TTATGCAGAT
    1741  TTTTATAAAC AAATATATTT TATTCAAGTT TTATTGTTAC CATAACCAAT GCAAATAAAC
    1801  ATCAGTCTGC TGTGAGCAAT CTACTAAGAT GCTGTGCAAA ACAAACACCA AACAACGCAA
    1861  AAAAATATAT CAAATTACCC TATACTGTTA CAATAAAACT GGTTTTGAAT CATCTTCACT
    1921  TATAATATTC AATATTCTGC CACCTATCAA ATTGAACTCA ACAAATTTTA GGTAAATGAA
    1981  AAATGTTCAC TTACACTATT TTGCCACCAA ACTTAAACAA ATATTAGGTA AATTACGTCA
    2041  CTCAATCTTG GTAATGCAAA ACAAAACAGT CAGTCTCTAG TTTTACAGAA AAAATCTTAT
    2101  ATGACCACCA TCTACTCCTC CGTAATCACC ACCTAATACT GCATACCATA ACCATTTACA
    2161  CTTAACTATA CAGAGAGTTG

//

LOCUS          21     2779 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      21
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS
complement(join(500..747,1012..1395,1470..1649,1831..1989,2177..2279))
                 /translation="MKCLMILALAVLATNAMENHRQLSAVFEGSEQACSDTCDDVSV
                 WCGHDRMKAFCNHTNYVGIIVREKCALSCGTCRLCPNVEPRSCGEEVDTCQDTNKFC
                 ATYLAPAYCNNHLLSFLVQRVCKRSCGKCPVNNCARSSVSNQIPCTEEFCHKKRLFS
                 KKLVFEKVFLIDCADMDAECQYWAEEMNYCSPASLYKTWMETRCRKSCKKCVKCGES
                 ELIAPPEAVEGLLVAPARPLTEATTELPTTSEIPDTTTEPPTTVEAVTAAIEVFTEV
                 PRGAVGCDIQCRNGSCSGQAENSICVCDKSAPGVLSPRCGGQCEANCARPDDFIGQW
                 EVFCTGEIVCTCHPTGDVEAGIPGGALCR"
ORIGIN
       1  GTTATACGCG GAACAGGGAG TACTTATTGC ATAACGGTGA ATATACTATC GAACGTACGA
      61  CTTCTGCAGG CATAAGACCT ACATCGAATA TAACACGTAG TCTTTACTTC ATTGTGGCCA
     121  ATACAAATGG AAATGACCGT AAATTCACTC ACCACAGTTT GCGTATTCTA TATCATTAAC
     181  ACTGTTCTAC TGTTTATCTT GGACTACAAA TAAACAATAG GTAAATGCTG TTTAAAACAC
     241  GATCAAGATA TTTTGATATT ATGAGCTAAA GGTGCCCCAT ATTCCCCCAC TCTACTATAT
```

```
     301  ATTTTTGGAA ATATTAGGCA ATATGTGCTT AGGTGTCCCA CGATGTTGAT AGAACACCAT
     361  ATTGTATATT TGAAATTCTT TTATTAAAAA CGAACTGCTA TCGGTATAAT AAAAAAACGA
     421  ATCCTTTTAA TAAATACACT ATATCACGTA AAAAATTTCG TTGCACAATA TCGAGCAAAT
     481  GATACATAGT TTGCAAAAAT TATCTGCACA ACGCTCCTCC TGGTATCCCG GCTTCAACAT
     541  CGCCTGTCGG GTGGCAAGTG CACACAATTT CTCCGGTACA AAACACTTCC CACTGCCCGA
     601  TGAAGTCGTC GGGTCTTGCA CAATTGGCTT CGCACTGACC TCCGCACCTT GGTGATAAGA
     661  CTCCAGGAGC TGACTTGTCA CAGACGCAAA TCGAGTTCTC TGCCTGCCCT GAGCATGATC
     721  CGTTGCGACA TTGGATGTCA CAGCCGACTG TTGAAAATCA AAGACGTATA AGAAAAAATA
     781  TTGTAAATAT AGGTATTGGA CATAAAAAGG TGCAGAAGGC TTGTCATTTT TTGCCCTTTT
     841  TACAAACTAC AACAGTAAAG TTGCATTCAT TGCATTTTAC TCCCAACGTA TGTGTAATTA
     901  GTTTATTTAC ACTATTTGTA ACTGTACGTA TATAGATTGT TCCCGTTGAA GTAAGATGTG
     961  CCTTTAAATT TAAAAATCGG ACGTGTTTAG TGACAAAAAT TCTTTGCTTA CCTGCACCTC
    1021  TTGGAACTTC AGTAAAAACC TCAATCGCAG CGGTTACTGC TTCAACTGTT GTTGGGGGCT
    1081  CAGTGGTTGT GTCTGGTATT TCACTTGTAG TTGGCAACTC AGTAGTTGCT TCAGTCAAAG
    1141  GTCGAGCTGG TGCTACCAAA AGTCCCTCTA CCGCCTCAGG TGGTGCGATA AGTTCACTTT
    1201  CTCCACATTT TACGCACTTC TTGCAACTTT TTCTGCACCT TGTTTCCATC CATGTTTTAT
    1261  AAAGAGAAGC TGGACTGCAG TAGTTCATCT CTTCGGCCCA ATATTGACAC TCGGCGTCCA
    1321  TGTCTGCACA ATCTATGAGA AATACTTTTT CAAAAACAAG CTTTTTCGAA AACAATCGTT
    1381  TCTTATGACA AAATTCTAGA ATTAAACTAA AGTGTATATG TTTACGGGAC GCTTTTATTT
    1441  GTTTAGTTCT ATTTTTAATA TTCACGTACC TTCAGTGCAT GGTATTTGGT TGGATACGGA
    1501  CGACCTTGCG CAGTTATTAA CCGGACATTT TCCACATGAC CTCTTGCAAA CTCTTTGCAC
    1561  GAGGAATGAC AACAGATGGT TGTTGCAGTA GGCAGGGGCG AGATACGTTG CACAGAATTT
    1621  ATTCGTGTCT TGACAGGTGT CTACTTCTTC TATATTCAAA TAAATCTTGA TTAAATGTCA
    1681  CTACTAAGAT CGTTATGTTA AATTAGTAGT AACAGTTAGC GCATACACGT CACCAGCTAT
    1741  ATTTAAAATC AGCACAGCAA ATAGATTGTA TCTCGAATGC AAAACAGTCT ACATAAAAGT
    1801  AAATGCAACT CAAGCTTATG CCATAATTAC CTCCGCATGA TCTAGGTTCA ACATTAGGAC
    1861  AAAGTCGACA GGTACCGCAT GATAAGGCAC ATTTCTCACG TACGATAATC CCGACGTAAT
    1921  TTGTGTGGTT GCAAAACGCT TTCATTCTGT CGTGGCCACA CCATACAGAA ACATCGTCAC
    1981  ATGTATCACC TGAAAAATAC ATGTTGTTAA AATTAACAGT TCATCGTGTA CGTAATCGTG
    2041  CGTATTAGCC ATTCGGGTAC AGTTAAACGA ATACCGTGTT ATCGCTAGTA ACCTCATGAC
    2101  GTCAATGATC ACATACGAAC CTATTGTGAT ATATCATAAA CCATAACTTA TTCACTGTTT
    2161  AGTCAGTATA ACTTACTGCA AGCTTGTTCT GAGCCTTCAA ACACCGCTGA TAGTTGTCGA
    2221  TGGTTTTCCA TTGCATTGGT TGCGAGGACA GCTAAGGCGA GAATCATTAA ACACTTCATC
    2281  TTCTATTTCG TTTGTACTTG TTGCTTCTTT AAATTCCGTG ATGTCTTTCT GCTCTTCTTT
    2341  ATATATCTTC TTGTTGTGTT TGAATTGTGG ATGCCTTGAC TATGGATCTC ATCTTTATAT
    2401  ATACTGTTAT TGTGTCCAAA CCTTTTTCCT CATATTTCCA TGAAGTATTG TTTATTAGAA
    2461  CACTATACAC ATGGCAGTAA TAATCACTGG GTAGGAGTTG TATATAATTG TCAAATAACG
    2521  CACGATATAA ATTCGTTGTA CACATAGTGA GCTGTATGAA TAATTACAAG GAACGCAATC
    2581  TCGCTGAATG GGGGAACGGT ATACACATAA TGATGGAGTA TTAGAATCTA TTAGACGTTA
    2641  ACGACTTGAT AATTACACGT TGAACAGACT CACAAACGGT ACGTTCGTAT AAAGTTTGTT
    2701  TAGTCGCTAT TTACACGGAA GTTTAACGTA CCACATATAC GTAGGTTGAT TTGTTAAACG
    2761  TCACCAACGT AATTCCAAT

//

LOCUS          22     6045 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      22
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
     CDS
join(500..670,1044..1210,1843..1978,2203..2375,2589..2733,3144..3211,4737..4844,
4908..5015,5428..5545)
                 /translation="MAETMLHPRPGMATILKSDNNAGGNYVEQIEFNRSLRQTKFPT
                 IRGPEGNYDYPSFQGRLKESKSFIKWNKHSQAPHQGYPLTPHRDGHDALDPTSGYLG
                 AGTDIDLKTNIKKISSMNQLSDKIQLGTPTSKPPASGRPQTPPRAVAPEHGGKWASN
                 QVSDVSLRAKLGGWTSNNAVKPKNINDSGNLVVGTFTFKPGNWRDEAAKYYMFSSST
```

QTSNRMVDWDKRLPPKLLPCPSTLEKTPDPISQNMNIWPKRFESDPEFWQTIGRSWD
WFQFRNKNHGNRPYSYGCVGAYNFEDKDHPHEEYRPLTSLRTPKPRYTDTAHRPNIP
HYKGCVHWTATHSANNTFPAPQPPSTAHSHRKLPTSPNTSIHSKKGPLSRMVTTVPP
KNPFNQVEDQQD"

ORIGIN

```
       1  TTTCGCAATC ATGTGTGATT GTAAAAAAGC CTAAAGTACA ATCATGAAAT TCAGAGCTGC
      61  GCAGTACTTT CTATTGCAAA AGTTCGTTAC TTATGCACAA TTACCAAAGA GACTCTAATT
     121  TCGCAAAAAT AATCTGGCAA CACTGAAGTG GATTCAGGCT TTTCACAGCA GTCAGGGTTT
     181  TAGACATATG CAGATGCCCA TGTGTAACAT AGATAATAGA TATCATACAT AGTAAATAGT
     241  ATCATGCTAT TACTATGGTG GTAATCTGGT ACCATTAATA ATTCCTAAGA CAGGCAATTG
     301  TTGTGACATT TTCACACATT CAACTTAATT TCTGAACCAC TTAATTTTTC TCCAAAGAAA
     361  TTCTAAAGCG GGTACAGATT TTGTTGTCCA ACCTCAAATT TTACATTTAA AATGAAATTT
     421  TGGTTGCATC ACGACATTGT TCAATCCTAT TTATTTTCTT CCCAATCTCA ATTCTCAAGT
     481  TTAGGATTAT CCTTTAACAA TGGCTGAAAC GATGCTTCAC CCACGGCCTG GAATGGCAAC
     541  AATATTAAAA AGTGATAATA ATGCTGGTGG TAACTATGTT GAACAAATTG AATTTAATCG
     601  GTCGCTAAGA CAAACAAAAT TTCCAACAAT TCGGGGACCA GAAGGGAATT ATGATTATCC
     661  AAGTTTTCAG GTAATACATC TTAATGGAAT CATGATATGA ACTTTACTGT ATTCGCTTGA
     721  TAATCAAAAT TTATCTGGCT CACATAATGT TAATATTTTG TAGCAGAAAT GAAATTCAAA
     781  AAGCATGAAT GTGTATATTT ATCAGATAAA TATGTTATAA CAATCGTACT TTGTTTTAAT
     841  TAAAACATAG TAATATGCAT TTTAGTGCCT AGCGGCTTAT GCATAATAAT CTTAAACTTG
     901  TTTAGTTCGG TAGCACCAGA TGTTTACCCT ATTTTTTTAC ACAAAAAACG TTCAACTCTA
     961  TGAGCATCAG ATCATGGTCA GATTTGCAGA ATTTATTACA GTCACTTTGT TAAGTTCTTG
    1021  ACATGTATGT TGTTGTTGTA CAGGGCAGAT TGAAAGAAAG CAAAAGTTTC ATTAAATGGA
    1081  ATAAGCATTC ACAAGCTCCA CACCAAGGGT ACCCTCTTAC CCCTCACAGA GATGGTCATG
    1141  ATGCTCTGGA CCCAACATCG GGATATCTTG GTGCTGGAAC AGACATTGAT TTAAAAACAA
    1201  ATATTAAAAA GTCAGTTAAT GAAGTAATAA CAGTTATAAC ACATGTGTTG TAATGCACAT
    1261  TACACAAATA CTAATATTTA AAATGTTGTT AATTATTTTG CCCAAAATGC ATACTCCCAA
    1321  AATGGTAGCA ATGACAAGCC TTGAACTCAT AACCGCGTCT TACTATTATA ATAGAAAAAT
    1381  AAAAAAGGTC TTTGAATAAA AATCGTGCGC TCAAGTGAGC AAGATCCATG CTAGTGCTGC
    1441  AAATATTGGT GACAAGTGAA AACAGTAACG TTTTAAATGG AAAAATGAAA AGTATAGATT
    1501  CTACTCTGTA AAGACTATTT GTACACCAAT CATCAAAAGT TGACTTGAGT TTTTCTTTAT
    1561  TATTCCTGTT TTGGATTGTA CAAAATTTAC TATGAGAATT TGAATCTTCA TTTGTCTCAA
    1621  ATTCGCCTTC AGTGGTAAAC AAAAATTTGG GCACGTCTGT GGTGTAATTG TTTGTTTTAT
    1681  GTCCAACAGT TATTCTCATT TTATGCGATG ACATTTTCCA CATACCAACA CAGGTAAAGA
    1741  GCTTTCCTCT CTTGTTAGTT TTACTCTTAT GTTCCATCTT TATTATTCTT AAAAATATTC
    1801  CGCACACCAG ACTAAGTAAT TGTACCATTG CATATTGTTC AGAATCTCGA GTATGAACCA
    1861  ATTAAGTGAT AAGATTCAAC TTGGAACACC AACCAGCAAA CCTCCCGCGA GTGGAAGACC
    1921  TCAAACTCCA CCTCGTGCTG TTGCACCAGA ACATGGGGGA AAATGGGCAA GTAATCAGGT
    1981  AACCAATCAA AAGATAAAGT TGAGTATTTT TCTAATTGGA ATTGTGTAAA TGGCATCATT
    2041  AACTGTTTCT AAACGATTTC ATTTGTTGCA AAATGTTTTG GTTTGCTTAT CTCTATTATC
    2101  TAGTAACTTA TATCAAGAAT ATTGCACTAA CCACATTAAG AGTAGCTTTG ATGCTGTTAT
    2161  GAGTTTTCAA GATCAAATTT ATAATCCATT GTTGAAACAA AGGTGAGTGA TGTAAGTCTT
    2221  CGTGCGAAGT TGGGTGGCTG GACGAGTAAC AATGCAGTAA AGCCAAAAAA TATTAATGAT
    2281  TCGGGAAATC TTGTGGTTGG AACTTTTACC TTCAAACCTG GAAATTGGAG AGATGAGGCG
    2341  GCAAAATATT ATATGTTCAG CAGCAGTACA CAAACGTAAT GTGTGTGTTT ATAATATAAT
    2401  TATTTGAAAA TTATTGTAAT GTTAGAAAAA AGTCAAACAA ATATACTGGA TTTTCCCAGT
    2461  GGCAAAGCCA AGGAGTGCTT GGGTTTGTGA TCAAAATTGC TCCCATTATA AATTAAAATG
    2521  GCAGTTGCAA TACTATTTTT TGTTTCAGGA GAGTTTAGTT GAAGTTTAAT ACTTGTTTTA
    2581  ATACTTAGGA GTAACAGAAT GGTAGATTGG GACAAAAGGT TGCCGCCCAA ATTGCTTCCA
    2641  TGTCCAAGTA CCCTTGAGAA GACACCAGAT CCGATCTCCC AAAATATGAA CATTTGGCCA
    2701  AAAAGATTTG AAAGTGACCC TGAATTCTGG CAGGTTCGAT TATCATAGAA TGTGTTAAAC
    2761  ATGTAAATTT GTTTGGTGTA CGATTGTACA AATAAGGTCG CTTTCATTAT TTGTCCCTGT
    2821  TAAAAATAGA AACTCATAAC GGTCAAAACA GAGTTACAAA CAAAATATTT GATTTTTGTG
    2881  TTTTTGTTTC AGAGTACATG TGTGGGAACA TTGCAATTAT TGTTACCAAA ATACACTCAG
    2941  AATCTTATTT GGGATATTTC TTGATTAGAG AATGGTTTAT TATATACTAT TGGTTTATTA
    3001  TATACTATTA TGCTTACATA TGTCTTATAT GGTTTGATAT CAAGTGTAAT AGGTTTTTAT
    3061  ACCTATTCCT GACAATGTAT CTTTCCATGA CACATAAAAC AAAAAAGGTT TGTTCCCCAA
```

```
    3121  AATTACTTTC TTATGTATTG CAGACAATTG GAAGGTCATG GGATTGGTTT CAGTTTCGAA
    3181  ATAAGAACCA TGGAAACCGA CCATACAGCT AGTAAGTGTT GGTGGAATAA TTATAAATTT
    3241  TTAAGGATTA TGTTGTTATT TTTGCAGCAC AAGTCATTAC AGGAAGAGTG GACACATTCC
    3301  TTGTTATGGG TAACTATGTT TCACGCAATG CTTTGTAGTT TAAATTTAAG TTGATTTAAT
    3361  TAGGTCAGTT TTTATCTTTT TATAATTTAA CTGATAAAAC AGTTGGTTAT TATTACAGTG
    3421  TATGCAAAAC AGTTTTAGTT TCTTAAAGCA GTTATTACAT AATAAATGCC TGGAAGAGAT
    3481  TTTATGCAAA TATTTCGGAG GTTCTAATTT GTTTGAATTC TGATTTTGCA AATTTTTAGG
    3541  GCAGAGTAAT TATTCTGTTA AAAGAAATAA AAATTTGTGT AAGTTTGACA TAGAGAGTAT
    3601  ATAAACAAGT GGTGAGAAAG CTATAATTTA ACAAGAGAGC CTAGCTCTAA TGATTAAATG
    3661  AATGTAATGA ATAAATGTAA TCAATAAATG TAACTTACTT TATCCTCGGT GGCCGGTAAA
    3721  CGGCAGTCGT TATAACACGG GTTTAACACC TTGGCCAGCT AACGAGTTAC CATGTATGTT
    3781  ACTTTGTGGG TGATAATTTT TTTGGATTTT TTTTATGTAT GACAGTATGA CTGATAATTT
    3841  GGACAACCCA TTAGTGACCA CTGGGTTGGA GCAATTGCCG TTAAGTGTTT TGCCCAAGGA
    3901  CACATGCGCT CACAATGGTA GCAGCGATGG GCCTTGAACC CATTACCTCT GGGTTACAGG
    3961  CAGGCGTGTT AACAACTATG CCATGGCGCC GGATTGAGCA CTCGTTATAA TGTTAACAAC
    4021  TATGCCATGG CGCCGGATTG AGCACTCGTT ATAATTTTTT CACACTCCTT AATTCCCAGA
    4081  CATGCAGAGA GGTATAAGCT ATGCTTCAAC AGATGGTAGA AGGGGAAATG ATGACTTAAT
    4141  CGGTTAAGCC TTAAACGATA AGCCATTGGG AACAGATTAT AATGTCCTGA GTTCGATACT
    4201  TGCCCAACAA AACAAAAACT TTGTATTTAT TTTCTTTCAA TAACAAATGG AAAGTTATTA
    4261  TTCTCAAAGC CAAACACAGC TAAACGATAC CAATTACTTA GTTACTAACA GGTTTCTAAA
    4321  ACATTAAATT TTACTTTGCT TGCAAGCATT ATATTCATTA ATTCCTCATT AATTAGGTGT
    4381  ACACTATGGT TTATAACTAT GCTCAGTTAG AAATAGTGTA TCAAGTGCAA TCCATTGACT
    4441  GGGTTCAAGG TACATTTTGG ATTTTGCGCT TTTAACCAAA CAGAAAATGA ATCGTTAATT
    4501  ACTTGCTAAT TATGGGGTTG CATATCTGAA CACTTGTTTT AAAATTATTG AATTTATATT
    4561  TTGATTTAAT TTTAGTCAAG AGTTTAAAAA ATAAAACCTA GATTTAAGCA TTGAAATATA
    4621  TGAGAAAAAG CCTATAGGAT CACTTACATG TTAAAAAAAG GGTATTTTTA AATTTAATAA
    4681  AAATCTTTGC TCTTTTATTT ACAATTACCT TATTGTTATT TTGTGTATTG TTTTAGCGGT
    4741  TGCGTTGGAG CTTACAACTT TGAAGATAAG GACCATCCAC ATGAGGAATA TCGTCCACTT
    4801  ACAAGTTTAA GAACTCCCAA ACCAAGATAC ACTGACACAG CACAGTAGGT TTTTGCCATA
    4861  GCATGTTTAT GATGTGACAA TGACATTACC ATATTTTTTG TATACAGCCG TCCCAATATA
    4921  CCCCACTACA AAGGATGTGT GCATTGGACA GCAACACATA GTGCTAACAA CACTTTCCCG
    4981  GCACCACAAC CTCCTTCTAC AGCTCATTCA CACAGGTCAC TTTGCATTAA TGTTTTTATT
    5041  CTTAATACTA TGGCTGCATT TCCAGAAACA ATTTCAGAAA AGTTAATTAT TATTAACAAC
    5101  TTTGTATTCG TGACATCACA TATAGCAGTG ACGAACCCAG GGACAAATGA GGCGATTATT
    5161  CCATTTCTTT AACAGTTGAA TGTGAGGAAG CTAGTGTCGA ATTTATGTTG CGAGTTCACA
    5221  CAAAAATTTA TATCTACCAG ACATATGGAA TGACTTATAT TTCATAGTTA ACTCAAAGAT
    5281  CATCCTATGG CATACCTTTT TAGATCGAGT ATTATAATAA ATGTTAAATT TTATAATCGG
    5341  ACTGAACAAT GTTGTTTAGA TTTACAATAT TATCTTTTAA ATACTTAACA TTTTCAGCAG
    5401  TAGCATTTAA AACTCATGTT CTTACAGAAA ACTTCCAACC TCTCCCAATA CTTCTATTCA
    5461  TTCAAAGAAA GGCCCTCTAT CAAGAATGGT CACAACTGTT CCACCAAAGA ACCCGTTCAA
    5521  TCAAGTTGAA GATCAGCAAG ACTAGATTCT AAAGTTTACA AATCAAGATG AATGTACATT
    5581  TTTATATTGT GAATCTTTCC GATGACCAAA CTTTTGTTTG AATGTTTACA CGTTAAATAT
    5641  ATTTTGATCA AAGAATACTG TGGTTTTAAC AAGCTCTCTT GTATATGTTT TCAATAGCAG
    5701  AGTGCTCTGT TTCAAAGCCG CTCTGTTTCA GCATAGATAC CTTACAAGGT ATCTATGGTT
    5761  TCAGTGAAAG GTATTTCTAT GGTTCTGCCG TATTTTGGAT GTTTAGTTAT CATCTGGATA
    5821  GACACGTTGA CTGGACACGA CGGCATTTAA TTTTCCTTAC GTCGAAGATG TTGCGTAAGT
    5881  TACTTGCGGT TAATAATAAA ACATGAACGA TTCTTTGTCG AAAAGAAACA TTTATCGGAA
    5941  TGTACGGAAA ATAGCTTTGC TGTGCGTTGC TTGGTAAGTG GTGCCTTAGT TTTATAACTA
    6001  GATATTATAA TTCAACACGT ACCGGAGCAC TTCTTACCAG TGCCT

//

LOCUS          23     2223 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      23
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
```

```
     CDS             complement(500..1723)
                     /translation="MSEIILKYDNKSSTKTLVEFSNGTPKPDSEFQCSLYADKRLQN
                     PKFKNNKVLVVQGSHLKYVGDNYEDRHANPTASRYVLMEVDEKKQVGKIFETEYFKL
                     RPIIGSEDIDIDEEREKTEKQLEQDERSFIERQDDLTEVFGSHRRKRAMVTRLRNKV
                     KDSTDAVQTAVEAIMENADTSLLLDEEGALDTMIIPCCKTAETVEDVYPLTNIFPDH
                     VSRCLIEEAQKFIHPTGAEITEWKNQGEGLGSRFIVEHAKNPALAIVQNSQIHTEYA
                     CCLQLMVCLIRLYKVKSSEYRAKTVFTDLAEPVRDWVLSEFFLDGCNRNRKMPTRMK
                     DKVLAYAIVLAWHMDSFSTDGDMIKEAFSISQKKLSTHVRALGGNVNLLSKNGFKRT
                     SVILKLPLQFPNTAFKGRGQGH"
ORIGIN
        1  GTTTCTATTT TATGCAATTA ATTTTTAATC TGTTCTAGAA GTTCGAGTTC ATTTTTTGTG
       61  TTAGCAACTG TGTGAATTTT ATAACGAAGC ATTATATTTA TACGTTTCAC TATTATGGCA
      121  TAGTCTTAGG ACTACTTAAA AATGCTTAAG AGCTACTGCT TGTTTTGTAT CTTAAATTAT
      181  TGTTTATTTA TTTACAGAAA ACTCAGTAAG ATAGAACCTT GTTTATTTGG ATGTTAAAAA
      241  TCTGGTGTTT GATTCGAAGA CAATGAAAAA ACAATTAATA ATGATTTACA ATTTGGCCTT
      301  CCAAGACAGT TCCAAAAAGA GAACATACAT TTTACTTTTA CATTGAATGA AAGTTATGAA
      361  ACAAGTTAAT TTAAACGTAA TATTCTATTC TATAATAACA TAAGATATAC TGAAATGCAG
      421  TATGACTGAC CTCTTATGTA TATATTAAAC AAAAACAATG GCAGTTAATA GTAAAATAAG
      481  GATTCTGTTT CCCCATGTGC TAATGGCCTT GTCCTCTGCC CTTAAAAGCA GTATTTGGAA
      541  ACTGTAAAGG CAATTTCAGG ATAACACTTG TACGTTTAAA GCCATTCTTA GATAACAAGT
      601  TTACATTTCC ACCTAGTGCT CGAACATGGG TTGATAATTT TTTTTGTGAA ATTGAAAAAG
      661  CCTCCTTAAT CATGTCCCCA TCAGTGCTAA AGGAATCCAT ATGCCAGGCC AGCACAATAG
      721  CATAAGCCAA CACCTTATCT TTCATTCGGG TAGGCATTTT TCGATTTCTG TTACAACCAT
      781  CCAAAAAGAA TTCAGAGAGT ACCCAGTCTC GAACTGGTTC AGCTAGATCA GTAAATACTG
      841  TCTTTGCACG ATATTCACTT GATTTTACTT TATACAATCT GATGAGACAC ACCATGAGTT
      901  GTAAGCAGCA AGCATACTCA GTGTGAATTT GAGAGTTTTG CACGATTGCA AGTGCAGGAT
      961  TCTTCGCATG TTCAACAATA AATCGAGATC CCAACCCTTC CCCTTGATTT TTCCATTCTG
     1021  TAATTTCTGC TCCCGTTGGG TGAATAAATT TTTGAGCTTC CTCAATCAAG CAGCGACTTA
     1081  CATGATCAGG GAAGATATTT GTTAACGGGT AAACATCTTC TACGGTCTCT GCAGTTTTAC
     1141  AACATGGAAT TATCATTGTG TCCAAAGCAC CTTCTTCATC TAAAAGTAAA GAGGTGTCCG
     1201  CATTTTCCAT GATTGCTTCA ACCGCAGTTT GAACAGCATC TGTGCTGTCT TTTACTTTGT
     1261  TCCTTAATCT CGTTACCATC GCACGTTTTC TTCTGTGGCT TCCAAACACT TCAGTCAGGT
     1321  CATCTTGCCT CTCGATAAAG CTACGTTCAT CTTGCTCGAG CTGTTTTTCT GTCTTCTCTC
     1381  TTTCTTCATC AATATCGATG TCCTCACTTC CAATAATGGG TCTCAATTTA AAATACTCAG
     1441  TTTCAAAAAT CTTACCCACT TGTTTTTTTT CGTCAACTTC CATAAGAACA TATCTTGATG
     1501  CAGTTGGATT GGCATGCCTA TCTTCATAAT TGTCTCCAAC ATATTTTAAA TGTGATCCTT
     1561  GCACAACGAG CACTTTGTTG TTTTTGAATT TTGGGTTCTG CAAACGTTTA TCCGCATATA
     1621  GAGAGCATTG AAACTCTGAA TCAGGTTTGG GTGTGCCATT GCTGAACTCT ACAAGAGTTT
     1681  TTGTACTAGA TTTATTATCA TATTTCAGTA TAATTTCACT CATTCTGTCT GGGGGAACAA
     1741  ACTTTATAAG TTTTTTTAAA AGAAACACAT TTGTGTAGCT TTTTTTTAAT ATTGCATAAA
     1801  TATTTTATAT TTTCAATTAT GCATAAGTAT TAATATAGCA AAGGCAATAT GCAGCTAAGG
     1861  GTTGTGTCTT GTGTGTTATA ATGACTACAC TGGCACTGAT GCAAACTTTA CAGTATAATA
     1921  CAATGACTGG TTTGTAATTG CTGTGTCAAT ATGGGTTGTA GTAAACTCAA ATTAAAACAA
     1981  TTCGTAAACA TTTTAAAACA ACGCGTTCAA TTTCTTCAAA TGTAAGTTGA ATTTTCTACC
     2041  AAACTCGTAG AAAACACGTG CGTCGCACTG CACTCTACTT TATGGAAAGC ACCGTCATAG
     2101  TTTGGTCACG TGTCTAAACG CTGACTAACG GTAGGTTACT AAGAGTCATC ACTTTTCGCC
     2161  AGCGATCGGC GTGAATGTTG TGCAGTACAA TATTTAGGAA AATATTAAGA ATAAACTGCC
     2221  GTT

//

LOCUS          24     6075 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      24
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES            CHR  ---
```

```
     CDS
join(500..539,792..922,1251..1295,1521..1620,2011..2120,2474..2575,2963..3106,32
55..3425,3696..3741,4000..4119,4453..4612,5029..5119,5561..5575)
                 /translation="MGSIWDLRGIVLWAFTACYSKQYTQTFVSNSNFDNPERWVGHG
                 VPCAEDSIVFNEDQVVSIFVQTNASIKELYFPLNGEFIFADGVVWSADESLSPCVGD
                 DHGDHVLTYNSDPYNWFDIGNWQGSDENGDELSSKDLLDIEKIPCDFDTAAFPLKQS
                 WMVSMPEDINLHRLVIWNQDFDSATFSNYRTTGRGRYQFTGSGVISAERAICSDVTG
                 CACRAGDSLQSLICGMVTCYDVSCSLPISAAGQCCQLCGAVLTISYTQAYRESALQQ
                 LMVDLDQNVQYTGVHHILSKIDSYSRKLEISEPFVQVLITDENTGTSTGSNAKDYAT
                 ALYNNILASADSIGVSEITMTSSGEGISPQTGMQSGEIAGLVIGLLIAIILVAVLVI
                 YIRKRTSSYSLRFSPFPKKSNDNYGNPLFDIDMNEVNIT"
ORIGIN
       1  TTAAAACATT TTCAGTCGTT TTCTGTTCAA TGTTTTGCTA TGCCTGGTAT AAATTAACCC
      61  TTAAGTCACT TCATGGGTGC AAATTTTGTT GTGTTTGTTA GTTACAGTGG GTGGGGGAAG
     121  ATGGGACACT GTTTTGTTCA ATTTTCTTGT CCCATTTGGT AGTAAACAAA GAAAATGCAA
     181  AGAATTATAA AACCATATTC TCATGACTCC CATAGACCGT TGTTAATTGT TTAAAACACG
     241  TTCAGGATAT TTGAATATCA TGTGCCAAAT GTGTCCCATC TTCCCCCAAA GTATTGTATC
     301  TGTTCTGCTA AAGTCATTTT ATTGCTGGTG TTTGCAAAAC TACTTTAGTT TAGAGAGTGA
     361  CCCTTATATG CATAATACAC ATATTAGTAC ACAGCGATAT CGTAGGGTGT AGATAAAATA
     421  TTTACAGAAT TCTACAACCT TAGCCCGGCG ACTGTATATA TAAGAGCGTT GTTAATAATT
     481  AAATTCATGA TCAGAAAATA TGGGATCTAT ATGGGATCTA CGTGGTATAG TTCTATGGGG
     541  TAAGATGGGA CACCTTTAGC AAGTAATATC CGAATATATT GAATTTTAAA CAATTCTATA
     601  TGGGTATCAT GAGGATACTG TTTTATAACT GATTGTTCTT TGTTTACTAC CAACTAGGAA
     661  GAGAATATAG AATGAAAAAT TGTTCCATCT TCCCCTACCC TAATAGATGA CGGCATCCTA
     721  TCTTTAGGTG GTAACACCAT ACAAAAAATA TATGCACAGT TCATAAATTT AAGCGGTGTT
     781  TGTTCCTACA GCATTTACAG CGTGTTACTC GAAGCAATAT ACACAAACGT TTGTTTCCAA
     841  TTCCAACTTC GATAACCCTG AAAGATGGGT CGGACACGGC GTGCCGTGTG CAGAGGACAG
     901  CATCGTATTT AATGAAGACC AGGTAGTTAA AATTAGTCAT TTAAACAAAT CTTCTGCCAT
     961  TTAAATTTTT AGGTAAGTTT TTCTAATTGT CTTTTAGCAG CTTATCTGGT ATAAAACGTC
    1021  TACTGCAGTC TTGCCGTACT ACAGTACTGT GGGGGAGATG GGACACATTT ATCACATATT
    1081  ATTTAAATAT CCTGATCGTT TTTTAAACAA TTAACAATGA TCTGTGGGAG TCGTGAGGGT
    1141  ACGATTTTAT AATTCTTTGG AATGTTTTTT GATTACTATC AAATTGGATG AAAAAATATA
    1201  ATAAAAAGGT GTCCCATCTT TCCCCACCCT ACCATGTGTA TTTGTTTCAG GTTGTGTCTA
    1261  TATTTGTTCA GACCAACGCC AGCATAAAAG AATTGGTAAG AAGTTTTATT TGCTCATTCT
    1321  AGAGCAGCAG CATTTTGTGA ATATTGGAAG CTACGACCAC GCTTTTCAAA TGACACTTAA
    1381  AACTTTTTCT TTCATAGCGT AGTTTATCAA CTGTTGTTTT GCTGTTGATT CCCTCCTCTT
    1441  TGTTTTTTTA AATGGCATGA CCTCGGCTAT CGTTTTTAAA GGATAAATAA AAATTATGTT
    1501  ATATATTATG TTTTAAACAG TATTTTCCTT TGAACGGCGA GTTTATATTC GCTGATGGAG
    1561  TTGTATGGAG TGCAGATGAA AGCCTTTCAC CATGTGTGGG TGACGATCAC GGCGATCATG
    1621  GTGAGTTCGT GATATCGTGA TAATGTGAAA AAAATATTTT CAGTAGTTTT TTCAAAAGCG
    1681  CGCGAACGAG TGTCGTTTTT TGCTAATTCA AATCTCTTCA CCGGTTTAAT TAATACTATC
    1741  TTCGATATTT TTGACAGGTA TTTATTTAAA AACTATATTA TAGTAGGTTG GGAAAGATGG
    1801  GAATAATATA TAAACAACCT GATCCTGTTT TAAACAATTA ACAACGATCT ATCGAAGTCG
    1861  TGAGGATACG GTTTTATAAT TATTTGAATG TTCTTGGTTT ACTACTAAAT GGGATGAGAA
    1921  AATAGAATGA AAAGGTGTCT AAACTTCCCC CACCCTACAA TATTTAAGTT GTTATGTTTA
    1981  TAAGTAGTTA AATTATGTTT TTTTTTATAG TTTTGACTTA CAATTCTGAT CCTTACAATT
    2041  GGTTTGACAT CGGTAACTGG CAGGGTAGCG ACGAAAACGG AGACGAACTG TCAAGCAAAG
    2101  ATTTATTAGA CATTGAAAAG GTATATTTGA AATATTTTTC ATAAAACGCC TAAAGTGCGC
    2161  AAATATATCA TGTGTGTTTT GAAAATTCAA TGTATAGTTT GCCAAAAATT TGCTCCAACT
    2221  TAGCGTTTTT TATTTGGTTG TTCGAAATGT TAGCCATATA AAATAAAAAA TATGATACTC
    2281  CCAAAAACTA ATTACCCACA AAGTTAAATG CGTGAAACCT CGTAAGCGGG CATGAATTGT
    2341  ATGAAACAGA ATGTGTTATA ACGACTATCA TTGCCCCGCA CGCCAAGATA AATAAGTTAC
    2401  ACTCATTCCA TTCACGCCCT TTGCACTGTA TAGGCCTATA ATGTAATTTA CAATTGTTTG
    2461  CAATTAATTT TAGATTCCGT GTGATTTCGA CACTGCAGCA TTTCCACTTA AGCAATCATG
    2521  GATGGTTTCG ATGCCAGAAG ATATTAATCT ACACCGGTTA GTTATTTGGA ATCAGGTACG
    2581  TTTTTCAATT TTTTTGCCAT AAATTTAGGC TGTTTTGTAA TCAGAGAAAA GTTATATAAC
    2641  TATAAAACCT TATCCCTGCG GCTCTTATAG ACCGTGGTTA ACTGTTTAAA ACACAATCAG
```

2701 GGTATTTGGA TATTATATTT TGGAGGTTTC CTATCTTCCC CTACCGACTA TTCATTGGAA
2761 TATGACCCAT ATAAGCTAAT TTATCCTCTA TTGTGGGTTT TGCAGATACG CTTAAATTTT
2821 AAAATTTCAC CGAAGAAGGT TTTTACCCAT AGCATGTTTT ATAAATACTG TCGTTTTTTG
2881 CCAGTTCTCA ATTTCTACAT TGTTTGAATT TAAATTAATT TGATCTTCTA TTAAAAAAAT
2941 GAAAATTAAG TTTGTGTTTC AGGATTTCGA CAGCGCCACA TTTTCCAATT ATCGAACAAC
3001 CGGTAGAGGG CGCTACCAGT TCACAGGCAG TGGTGTAATA TCTGCTGAGC GAGCCATATG
3061 CAGTGACGTA ACTGGATGTG CATGTCGTGC GGGCGACAGT CTTCAGGTTT TATTGCAAAA
3121 TGACTTAATT TTGTTGTAAA ATGTTTTAAT TTCGTTCTAC AATTACTTAA TTGTGTTGTA
3181 AAATTGATTA TTTTTTGTAA AATTACTTAA TTAAAAGTTA CTCAGTTTTA CAAAATTACC
3241 GTTTTAAAAA ACAGAGTCTA ATTTGCGGAA TGGTCACGTG TTATGACGTC AGCTGTTCTC
3301 TTCCGATCTC TGCAGCCGGC CAGTGTTGCC AGCTGTGTGG AGCAGTGTTG ACCATTTCTT
3361 ACACACAAGC TTATAGGGAA AGCGCATTGC AACAGCTAAT GGTCGACTTG GACCAAAACG
3421 TTCAGGTAAC ACATATATGG TACTGTGGGG TAAGATGGAT ACTTTTAACA CATACTGTTC
3481 CATATACTAA TCATGTTTTA AACAATTACA AACGCTTTTT ATGGTCGCGA GGCTATACAG
3541 TTTGTAATTA TTTAAACTTT ACCTACCAAA TGGGACAAAA AATAAAATGT TTACTAAATG
3601 GGTTAAGAAA AATTAATGTA AGCATATATA CCACGTTACC GGGCCTACTA TGAATAAAAT
3661 ATAAAGATAA AAAATTAACT AATTTGTGGT TTTAGTACAC AGGTGTGCAT CATATACTCT
3721 CAAAAATTGA CTCCTACTCA AGTAAGTATA AATTCAGCTT ATATTGCGGA ACTATGGTGG
3781 GGTGGGCGAA GATGGGACAC TTTATCACTG AATATTCAAA TATGCTGATC GTGTTTTAAA
3841 CAATAATGGT CTATGGGAGT CGTGGGAATA CGGTTTTATA ATTTTTTGAA TATTCTATGT
3901 TTACTACCAA ATGAGACGAG AAAATAGCAT GAAAAGCTGC CCTATCTTTC CCTATCCTAC
3961 TAAACCCAAA ACCGTTTTTT TTTATACACC TCGCTGCAGG AAAATTGGAA ATTTCGGAGC
4021 CTTTTGTTCA GGTTTTAATA ACGGACGAAA ACACGGGGAC TTCTACTGGT TCTAATGCGA
4081 AGGATTACGC AACTGCATTG TACAATAATA TACTAGCTTG TAAGACTTAG TAATTTTTTA
4141 TTTTTCTAAG ACTTATAACG GTTTAATTTA ATTTTTTTCT AAGACTTCAA TTTTTTTTTT
4201 AAGACTTTAC ATTTAATTTT TTTTTCTAAA ACCTAATATT TTTTTCCAAG ACCTAATATT
4261 TTGTTTTAAT TTTTTTTCCA AGATTTAACA TTTCATTCTA TTTTTTTTCT AAGACTTAAT
4321 ACTTTTATTT TTTCTCTAAG ATTTAATATT TTATTTCATT TTTTGTAAGA CTTAATATTT
4381 TATTTTAACT TTTTTGTGAG ACTTGATATT TAACTTTATT TTTTTCCAAG ACTTAAAAAT
4441 CATATTTACC AGCTGCAGAT TCAATCGGAG TTTCCGAAAT CACAATGACG TCATCAGGAG
4501 AAGGGATTTC CCCGCAGACG GGAATGCAAA GCGGGGAAAT TGCTGGACTC GTCATAGGGT
4561 TACTGATTGC TATAATATTG GTCGCCGTTT TGGTGATTTA TATTCGTAAA AGGTATATGT
4621 ATATATATAT ATAGTAGTGT GGGGAAAGAT GGGACACCTT TAGTACACTT TATTTATATA
4681 TCCTGATTGT GCTTAAAATA TTTAACAACG GTCTATGAGA GTTGCTAGGA TACGGTTTTA
4741 TAATTCTTTG AATGTTCTTT GTTTACTACC AAATGCTGGA ACAAAAAAAA AATAGAATGA
4801 GAATATGTCC CGTCTTCCCC CACCTTACTA TAGTAATGAT AAAATAGTGT TTTTTTGTGT
4861 ACGATGTCAA ATATTAGGGC GACTTTATTT TTTTTTAAAT ACGCAAGGCT ATTTTAAACG
4921 AATATTGAAA ATAAATTTTG GTGTTAACTA CCTGAAAATA AAAGATGGTT GTTGAAACTA
4981 TGACAAACTA GACTTTAATA CCATAAAAAT ATTAATTTAA TTTTTCAGAA CATCATCATA
5041 TTCACTTCGA TTCAGTCCAT TCCCGAAGAA GAGCAACGAC AATTATGGAA ATCCTCTGTT
5101 TGACATCGAC ATGAATGAGG TGGGAGTAAA ATTATCTGAT AGGAATAAAA AAATATTTGT
5161 GAGGAGTAAA AATTATCCGC GACGAAAAAA TCGATTTAAA TTATATAGTA GGCTACTGTA
5221 GGCTAAAGTG AATAACTTTA GCACATAGCA GTAAAATATC TTGATAGTGA AATGGCGTTT
5281 TATCTCCGTA CAAATAATTA AACACACTCA CTCACAAATT TGAAATAATT TATCTGAAAT
5341 AAAAATGTAG AATATTTTCT CAATGTAATT AGAATCGCGA AAGCTAAACA CAAAAAAGTC
5401 ATGATAAAAA AATGACAAAA CATAAAAAAA TTGATAAAAC ATGAAATACA AATTGGACCC
5461 AAAACATGGT CAAAGCCCAA CTCTATAAAT CAGGACCTCC TTATATACCC ATACAAAATA
5521 AAACCACTAA ACTTCCAACT TGTCATTCAA AATCTCCCAG GTCAATATCA CTTAATAAGC
5581 CCAACCTATC AAATTAGCCC TTATAGCGTA ACCGGTTTTA TACATTGGTC TTTAGCTACC
5641 TTATAGCGTA ACCGGTTTTA TACATTGGTC TACTACCTTA TAGCGTAACC GGTTTTTTAT
5701 ACATTGGTCT TTAGCGACCT CATTTCCTTA TAAGTGTACC CCATTTAAAT TTCACTTACA
5761 GCTACCTATT TCTGTTAAAA CCTACCCCCA TTTAAAATAC ACTTTCCACC GTAACTCTAC
5821 TTCATTAATT CTAGTTATAA ATTTCTGACA CGAGTTTAGA AAAAGTTTTA TTAAAGTTAA
5881 AATCAACAGA AATAAAACGG TGGTATATTT TTTTTTAATT TAACACAAGT TAAGTTTTAT
5941 TAAATTTTAC ATTTGGTGGT ATATTTTTTC AAATTTAACA CAAGTGAAGT TTTATTAAAT
6001 TTAAAATCAA CAAATATAAA ACGGTGTTTA ATCCAATTTT TTAAACAGTT TTATTAAAAT
6061 CTTTCATTAG CAAAA

```
//

LOCUS          25     6098 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      25
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS             join(500..647,1387..1525,3402..3614,4229..4380,4799..5598)
                     /translation="MSVTDPLLKSLKYLIEEFYSVDEKVTDESPYLEQFCITVEKCF
                     TKGMKPGQWLIPKDNGDSYRRKRLWQILESFETNGPVPYSISTAVNAALSNKRVRGK
                     EGRLRYLIRHCLAHSSLHVLVEFLMTRRKELRDIYMVGDKENILTMLNVFGGGSSIM
                     IHEQLSEQLRSLLLTCSTSLSFHLDLTNSFFLDETWKLPQTKSAEFVPTNELGIEMI
                     YVQGHGIISRIKAGSVAAETPSLGVGDLLVEMNGHLLFPAAAGGEKDAKCVWRECKS
                     RSKPCMLTVVKPRSPVTGGIYPPLKAILTEYHEWMDIYYKTRQQKLEDKNLRDEIKA
                     DVMSPLNENEKIPVLDPTAAENGSGGSYKVTYIGEVLTGEDGGVDQIESVIRQALNK
                     RSDRTSFNVTITLTQIGVETGIGNHSQVEPLFNHKYADISSCGKIVTDDFHFCYIAG
                     DTFCTLSKQFVGYVFRAETLAQSRAVLNGIYQGFKRTTWFM"
ORIGIN
       1  ACGAGATAAT TTATAACGAG CTGTTAAGAG GAAACTGAAG CAGCTTCGCA ACAGGAAATC
      61  AGTGGCACAG TGCGTTCTAA CAACGATGTG GGTGTTTGGT GGTTACAATG GTTTGTTTGC
     121  TTGTGTGAGT TGTGAGGTTA CTTGGTGGTA ATACATAAAT AGAAAGATAT CTTACAGAAA
     181  TTTGAATTAT AATATGAATT TAAATCGTGA ATAATTTGTC TGCCAGCAGC ACCATATGTT
     241  TATACAGTAT ATAGGTTGAG CTGTATATGA TGTATTAGCT AGATTGTTGA TTATACAGTT
     301  TGTTGGTTGT CTAGTTATAT TAGCACCAGG TGGATGGCTA TTAATATTAT AATGATAATA
     361  ATTAGCATGT TTTACGTGTG ATGTTTGTTT TAAAAGTTAT TTATCTGTTT GTTTGTTTAT
     421  ACAGCTCTAA ATTACCACCA TGACTAGCAT AGTAAACCTG GAACTCAGAT ATTAAAATAA
     481  CATTAAAAGA TTAACCAGTA TGTCGGTCAC TGACCCTCTA TTGAAAAGTT TGAAATATTT
     541  GATCGAGGAA TTTTATTCGG TTGATGAAAA AGTAACCGAT GAAAGTCCGT ATTTGGAACA
     601  GTTTTGTATA ACTGTGGAGA AGTGTTTTAC CAAGGGAATG AAACCAGGTG AGGGTTTTAT
     661  ACAAGGGTGT ATTGAACTGT GTGTTTGCTG CACACTGATG CGAATGAATG TAACTTTTTT
     721  ATTTATCCTT GTGTGACTGG AAAAACAGTC GTCATAATTC TCGCTTGCAA GTTACCATGT
     781  ATGTCAATTT GAGGGTGATT GTTTTCTGTA TGGCTGACAA TTTAGATAAC CCATGAGTGT
     841  CCACTGGGTT GAAACAATTA TTGTCTTGTT AAAAAGGACA CATGCTTTTA CAACGAGAAC
     901  AGCCACGACC TTGAACAAGT TACCCCTTTA AAAAGTGGCA GGCGAGCTAA CCACTGTTCC
     961  ATGATGTGCA CTTTAATAAA ATCTGTGTAT TTTTTGTGCA AAAGGATCAG TTTTCTCAAA
    1021  TAACATGTTG TTACACCAAA GTAATATATC CAGGAAATGA TTAATTGCAT GGTAGTGTGT
    1081  TGTGTTGCTA ATTTGGGATG AATAACAATT TTAATGCGAA AAACGTGAAT CTTGCTATAT
    1141  CTGGTTATGA CTATTTTGCA GAGTTCTCAG TTATATTTAT TTCAAGTGTT ACAATATGAT
    1201  TTATCAATTC TTAAAATGTT ACTATTTATT TCCACGTGTT TTTTGAACGA CCAATTTAAT
    1261  TTTTCTAGTT AGATTTTTCC AAGTTGTCAA CTATATTTCT GTTTTTACAA TGTTGCGAAG
    1321  TGCAGACTAT ATTTCCCAGT GTTTCCAGAT ACATTCTTGT TTTTAAGTGT TGCGAATTAT
    1381  TTCAAGGACA ATGGTTGATT CCGAAAGACA ATGGGGATTC ATACAGAAGA AAAAGGTTGT
    1441  GGCAGATTCT AGAATCCTTC GAAACCAACG GACCCGTCCC ATATTCGATA TCAACTGCTG
    1501  TTAATGCTGC GTTGAGCAAT AAAAGGTGAC GGAATAATAA TCTATATGGG TGAGGACGGT
    1561  GTGACAAGCG TTGGGACCTG GGATTGAGAC CACAGTTGGT CCATTACCCC AACATTGTAT
    1621  ATGCAGGTTC TATTCTATAT GGGTGAGTTT CTACAGTGAG GCATGCCATG GGACCTGGGA
    1681  TTTAGACCAG AGTTAGCCCA TTACCCCAAC ATTGTATATG TACAGTCTGT TGCATGGGTG
    1741  AGGTCCTCAG TAAGGCATGC CATAGGGCCG GAGATTTTAG CCAACACTTA AAGTGCCACA
    1801  AGCCATTCAG TGACCACTGT GTATTGTACA AATGTTCAAA GTACCTAAAG ATACTTACTA
    1861  AGCCATGGTG GTGAAATAAA CAATGCAATG ACAATTAAGA AGTTGAAACA TTATTGTTCA
    1921  ATCATAACAC AAGTTGTTTT CACAGAAGTT AAATAATAAA CATTAAGAAT ATGCCAAGTT
    1981  GAAACAATAA TAAACATACT GTAGATAAAC CTCATGCTTA CTATCCAGTT ATACCATGGC
    2041  ACACATGGTG TATTCATACA CCTAACGCTC ACTGTCAAGT TATAACATGG GTGTCTTCTT
    2101  ATACCATGGC ACTGGCACAC ATGGTGTATT CATACACCTC ACGCTCACTG TCAAGTTATA
    2161  ACATGGGTGT CTTCTTATAC GCCAGGCACA CATGGTGTAT TCATACACCT CATGCTCATT
    2221  GTCAAGTTAT AACATGGGTG TCTTCTTATA CACCCAGCAC ACATGGTGTA TTCATACACC
```

2281 TCACGCTCAC TGTCAAGTTA TAACATGGGT GTCTTCTTAT ACCATGGCAC TGGCACACAT
2341 GGTGTATTCA TACACCTAAC GCTCACTGTC AAGTTATAAC ATGGGTGTCT TCTTATACAG
2401 CCACTGGCAC ACATGGTGTA TTCATACACC TCATGCTCAT TGTCAAGTTA TAACATGGGT
2461 GTCTTCTTAT ACATCCTAGC ACACATGGTG TATTCATACA CCTCACGCTC ATTGTCAAGT
2521 TATAACATGG GTGTCTTCTT ATACCATGGC ACTGGCACAC ATGGTGTATT CATACACCCA
2581 TGCTCACTGT CAAGTTATAA CATGGGTGTC TTCTTATACC ATGGCACTGG CACACATGGT
2641 GTATTCATAC ACCTAACGCT CACTGTCAAG TTATAACATG GGTGTCTTCT TATACCATGG
2701 CACTGGCACA CATGGTGTAT TCATACACCT AACGCTCACT GTCAAGTTAT AACATGGGTG
2761 TCTTCTTATA CCATGGCACT GGCACACATG GTGTATTCAT ACACCTAACG CTCACTGTCA
2821 AGTTATAACA TTAAGGTGTC AGTTTTTAAT TTTTTTTTGT GACATAATGT GTTTTACTAT
2881 CTTAGTATTC AATGAGGTTC CCTATGTTTA ACCTATTTTG ATGGGTGTAA CTGTATTTTA
2941 ACAATGTCAC CCCAGATTTT ACCCTGCTGT TAGGTGTCTA TACAAACAAG CAGAGCCAAA
3001 GTATATTGCA TTCACCACAT TAATCAATGA GGTTCCCTAT GTTTGACAAC TATGAGTGTA
3061 ACTGTATTTT AACAATGTTA CCCCAGATTT TACCCTGCTG TTAGGTGTCT ATACAAACAA
3121 TAAGAGCCAA AGTATATTGC ATTCACCACA TTAATCAATG AGGTTCCCTA TGTTTGACAA
3181 CTATGAATGT AACTGTATTT TAACAACGTC ACCCCAGATT TTACCCTGCC GTTAGGTGTC
3241 CATACAAACA AGCAGATCCA AAGTATATTG CATTCACCAC ATTAATCAAT GAGGTTCTCT
3301 ATGTTTGACA ACTATGAGTG TAATTACATT CTAACAATTT CTCACCACAT TTTGTTAAGT
3361 TTTAAAAATC TAACACCAAT AAATCGACCC CAACCACGCA GGGTACGGGG CAAGGAGGGC
3421 CGACTACGAT ATCTCATCCG ACATTGCCTC GCACACAGCA GCTTACATGT GTTGGTGGAA
3481 TTCCTTATGA CAAGAAGGAA GGAACTGAGA GATATTTACA TGGTGGGGGA CAAGGAAAAT
3541 ATATTGACCA TGCTCAATGT GTTTGGGGGA GGTAGCTCCA TCATGATACA TGAACAACTA
3601 AGTGAGCAGC TCAGGTAAGA TTAAGGGATC GTGTATTGGA ATATTTTTAA TATGTTGATA
3661 CCCAGTGTAC CCCATTGTAT ATGTTTGAGG TATTCTGGCG ATATGTCGGC AAATAAGACT
3721 AGATTTCCCC TATTGTTTAT CGTATATATT GGACAATGAC CAAGCTGTCA AGGGATGATC
3781 TTCCATCCAT TCTATAAACT CCTTGTCCCC ATACCTTCCA TAAAGGGACA AGCTTTTTGC
3841 ACCAAGCTGT TTAACACATA GATATGTTAG ACAAAGATGT AATACACAGC AATTAACAAT
3901 GCTGCAATCA GTCTATTGCC CCAGTTGAAA CCACTCGGCA AATTCAAGAC TCAGTGCACC
3961 ACTAATACTA ATGCTATGAA TGTAACTTGC TTCATCCACG CATGGCCAGG AAATGGCAGT
4021 TGTTATAATT CTAACACCCC GATTTACAGG TCCCTTCTAT TAACATGTCC CAACAATATA
4081 TCACACATTC TTATGTGAAA AGACTGGTGG CTATAACTGA ACCCCATACA CCATTAAAGC
4141 TTGGGATTCT TAACCCTACA CAGAACTTGC TTTATCCTCG CGCGGCTGGA AAATGACAAT
4201 TACTATAACA CTAACCCCCC ATCTACAGGT CCCTCCTACT AACATGTTCC ACCTCTCTTT
4261 CCTTCCATCT TGACCTCACT AATAGTTTCT TCCTGGATGA AACTTGGAAA CTTCCACAAA
4321 CTAAATCAGC CGAGTTTGTT CCAACCAATG AACTCGGGAT AGAGATGATA TATGTACAAG
4381 GTATGGGACA ATCACATGTT TGAAAAATTG TTCATAGTCT TTGATCCATC TCACTTACAA
4441 CAACATTTAT GGCAGTGGTT CCCAAACTGT GGCACGCGAG CACTCATTAG GTTTACACCA
4501 AACAATCGCG ATACGTTGCT TTACGCGTAC AAATTATTCT TTGCAAACAT ATTTATACTT
4561 TAATTGTGCG CATACGTGGA AATGTATTTT AAACACCTGC AGGCTAAATA ATCAAGGGGA
4621 CAATCGATAT TATGATGCAA ATAATCGCCG TCGACGCCTT GTGAATTGCT ATTATGTCGT
4681 AATCAATTGC CAGGTCAAAC AATCGCTATT ATGATGCAAT GAATCCACGT GTGAATTCCT
4741 ACTATTACGT AACAAACATC ACATTCATTC ATTCATTTAC TCAAGTTATA TGTTACAGGT
4801 CATGGAATCA TCAGTCGAAT TAAAGCTGGC AGTGTAGCAG CGGAAACCCC AAGCCTTGGT
4861 GTCGGGGATC TCCTGGTTGA GATGAACGGA CACCTCTTGT TCCCTGCTGC TGCTGGTGGT
4921 GAGAAAGACG CCAAGTGCGT GTGGAGGGAA TGTAAAAGTC GCAGCAAGCC TTGTATGTTG
4981 ACGGTGGTCA AACCGCGATC ACCCGTCACT GGAGGCATAT ATCCCCCACT GAAAGCGATC
5041 CTCACAGAGT ACCATGAATG GATGGACATC TATTATAAAA CGCGGCAACA AAAACTTGAG
5101 GATAAAAACT TGAGGGATGA AATTAAAGCC GACGTGATGT CGCCACTGAA TGAGAACGAG
5161 AAGATACCGG TCCTTGATCC GACAGCTGCA GAAAATGGGT CCGGGGGATC GTATAAAGTC
5221 ACGTATATTG GTGAAGTATT GACAGGGGAG GATGGGGGGG TGGATCAGAT TGAAAGCGTG
5281 ATTCGGCAGG CTTTAAATAA AAGATCGGAC AGAACTTCCT TTAATGTGAC GATCACACTT
5341 ACACAGATAG GGGTGGAGAC GGGCATAGGA AACCACAGCC AAGTCGAACC ACTGTTTAAT
5401 CATAAATACG CTGATATATC AAGCTGTGGG AAGATTGTAA CAGACGATTT CCACTTCTGC
5461 TACATAGCTG GCGACACCTT CTGTACATTG AGCAAGCAGT TTGTGGGATA TGTGTTTAGA
5521 GCTGAAACAT TGGCTCAATC AAGAGCTGTA TTGAATGGTA TCTACCAAGG GTTTAAACGC
5581 ACTACATGGT TCATGTAACG CGATAATATT CACTTAACAG TTTAATTGTG CAATAACATG
5641 TATGGTGAAA TTATAGGTGC GACTTTAAGG GGAGGGAGGG ATAGTTAGAC ACACAATCCT

```
    5701  ATGAAATAAA CTTGTTGGTT GTAATGTGTT TACTGATTAA TGCAGCGTGG ATTGTTCCAA
    5761  GGTTGGCATC TCTACCCTGC CACCCCAGGT TTGGTGCCTT TGATTGCAAT GTCTTGTTTC
    5821  GTAGCAAGTG TATACGTAAC TCAATGCTGT CACTACTTTA CTGTGTGAAT ACTTTGTACT
    5881  TTTTGCTTGA ACTAATATTA CAAAGTTAAC AAGGAGTTTA TTATTGTTGC CTGGTGTTAA
    5941  TTATTTCTAA CTTTTATCAT GAACTTTGTA ATGCTGTTGT TTTTGTTTAC ACACATAAGG
    6001  TACTGCAATA AAACAAACTT AACTATTCTA CAAAGAAGTA ATAAAAGTTA CAGCTTATTT
    6061  TAATGTCCTT GCACATTAAA TAAAGGTAAC CGTGTTCC

//

LOCUS          26     2553 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      26
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS             500..2053
                     /translation="MTFCFSSGNTGMVCPKEFVSMVVFVVFECKRKYVVRLTMQSTI
                     NYDQLKVVNSTAEDRIESRHENFMKTFTSVFVKPEPVKPGMPTKPTIAFELGHAGLN
                     RPKSRYAAVRGKITRKTEDPPTVYDDPRVHDVRLHISKHKKGLQTETAPDDLVLAKD
                     RNIGCLRMLMKKLPFERTRKEHKRVYRLLRELWPSELACLLEDSGDSHPVLRELATV
                     AALDSYGEAGLTVFGNTGLHMILRGSVRPQTLPYIRAGLSTEDDAEVFPSPTPLLRR
                     KMGKLQHGDWFGTLEKVEGREVNTKVMTSVTLEPCQFLKIAIADYKRILERIKERIQ
                     AEKTTVVQASDLYRNWPGLSLRKLAALVEWQTIPGNTVVIPEGEIAPFIFFLKKGKC
                     DVFKQVEAIKTLPNGRRKKLFRSVRMGSLREGDSIGEDSLLEKKPLSCSIVSCEDLE
                     VGVITLERLELLDPTTRSLLSQSSRPKFDDVTDDIIQENFVQQEMKDDWAELKRNVL
                     LRTINYCGIRPGYGKWAK"
ORIGIN
       1  AAGCAAGTAA AGAAACTGTT AAGAAAACAT TAGTATGTCG ATTTAAATTA GTTGTTTATT
      61  AGGTAATTTT CTAATCGTAC AACCGTAAAC TAAACACCGG TTACAGAAAA CAGGTTTCGT
     121  ATATATGTGA TTAACAAAGT TTATAAAATT AATACGTTGA CCAAAATCAA AACGATTGTG
     181  GCACTGAACA GCAAAAACTG GAGCGTGGCA TAACAGAATT AAAATTCGGT GTCGCGTGAA
     241  ATAAGGTCAC CTGAAACGAA CAAGGCACAC GTAGATCAGA GTTTCAATAA CAAATCCGCC
     301  CCTTGAAGGA TAATGGAAAC AAATGGAAAG TTTTCGCAAC AAAAGGCGGG AGCTCTTTAT
     361  TCGCGCTTAT GTGAATGTTG AAGTGTTTGT TTCGAGTCAT GATTTCAGTG TTGCCAAAAT
     421  AAAAATGTTG TAATCGTTTT TGGAAAAATA GTTGTATTTA CCTAAATTTG GAAAACATAC
     481  CTTGTTTCTT TTGATTAAGA TGACTTTTTG CTTTTCGTCT GGAAACACTG GCATGGTGTG
     541  TCCAAAAGAA TTCGTTTCGA TGGTTGTGTT CGTAGTTTTT GAGTGTAAGC GAAAGTACGT
     601  GGTAAGGTTA ACGATGCAGT CAACTATAAA CTACGATCAA CTTAAAGTTG TTAATAGTAC
     661  AGCGGAAGAC AGGATCGAAT CACGACATGA AAATTTCATG AAAACTTTCA CCAGCGTTTT
     721  TGTGAAGCCC GAGCCTGTGA AGCCGGGAAT GCCAACCAAA CCAACGATAG CGTTTGAACT
     781  GGGCCATGCT GGTTTGAATC GACCCAAATC CAGATATGCG GCGGTAAGAG GCAAAATCAC
     841  ACGAAAAACG GAGGACCCAC CTACTGTGTA CGATGATCCA AGAGTACATG ATGTTAGACT
     901  TCATATATCA AAGCATAAGA AAGGTTTGCA AACTGAAACA GCTCCGGACG ATCTGGTATT
     961  GGCAAAAGAT CGGAATATTG GGTGTCTGAG AATGCTGATG AAGAAGTTGC CTTTTGAGAG
    1021  AACACGCAAG GAACACAAAA GGGTGTATCG TCTACTACGC GAGCTGTGGC CTTCTGAACT
    1081  TGCTTGCTTG CTCGAAGATA GTGGTGACAG TCACCCAGTG CTCCGCGAAC TCGCAACAGT
    1141  GGCAGCTTTG GATAGTTACG GAGAAGCGGG ACTGACAGTG TTCGGAAACA CTGGCCTACA
    1201  TATGATTCTT CGTGGATCTG TAAGACCGCA AACTCTTCCA TATATTCGTG CAGGGTTGTC
    1261  TACTGAAGAT GACGCTGAAG TATTTCCTTC CCCCACCCCT TTGCTTCGGA GAAAAATGGG
    1321  GAAACTGCAA CATGGCGATT GGTTCGGGAC CCTTGAAAAG GTCGAAGGTC GAGAGGTCAA
    1381  CACAAAGGTC ATGACCTCCG TCACCTTAGA ACCGTGCCAA TTTTTAAAGA TCGCAATAGC
    1441  GGATTACAAG CGCATTTTGG AACGAATAAA GGAGCGAATT CAAGCCGAAA AGACAACCGT
    1501  TGTTCAAGCG AGCGACTTAT ATAGAAATTG GCCCGGATTA TCACTTCGAA AACTGGCCGC
    1561  ACTGGTTGAA TGGCAAACCA TACCTGGAAA CACTGTTGTT ATACCAGAAG GAGAAATCGC
    1621  TCCCTTCATA TTTTTCCTCA AAAAGGGTAA ATGCGACGTC TTCAAGCAAG TGGAGGCAAT
    1681  CAAGACTCTT CCGAACGGTA GAAGAAAAAA GCTCTTCCGG AGCGTGAGGA TGGGATCGTT
    1741  AAGGGAGGGA GACTCGATCG GAGAGGACAG TTTGTTGGAA AAGAAACCTT TGTCTTGTTC
```

```
    1801  CATCGTGTCT TGCGAAGATT TGGAAGTTGG TGTTATTACA TTGGAAAGAC TTGAGTTGTT
    1861  GGACCCCACC ACACGTTCCT TGCTTTCTCA AAGCTCTCGG CCAAAGTTTG ATGACGTCAC
    1921  TGATGACATC ATTCAAGAGA ACTTCGTTCA GCAGGAAATG AAAGACGATT GGGCAGAGTT
    1981  GAAACGTAAC GTGTTACTAC GAACCATAAA CTACTGTGGT ATTCGACCCG GTTACGGAAA
    2041  ATGGGCGAAA TAAAAATATA AATTCTTGTT TTTTTTAGTT TGTATAAAAA AGTGACACTT
    2101  TATATCGAAT GCGAATATTT ATTTTTCAGA AAAAATAAAT ACTGAGACTG CGTACAACGG
    2161  TTTGAAAATG CTACAACGAA AGGGATAGCT GCGTGGAATG AACTACATAG AAGTTTTAAT
    2221  CGGAAATTAT TCCAAAACAC TGTTCCAGCT ATAATTGTTA ATAAAGTCAA TAATATGTAG
    2281  ATTTTATACA AAAAGATGAG ACTTAAGGTT GGTAGCTGGC AGTGCTGACC GATAATATTA
    2341  TGTGCAGTGT TGTGTAGTTT GAGAGAAAAC TATAAAGCAA CACTGTTAAT CCACCATGCT
    2401  TTCAATTTAT CGACTCGAAC TTTGTTTGAT CTGCGCAAGT CGTCGTTTTA AATCGGATAT
    2461  ATTTGCTGAA GAACTAATGC TGCGTTCTTT CGGTTGCGAT TGAAAGCCGC TGGATCGAAC
    2521  AGCCGCTTCA TCGGCCTGTA ATGACACAAT AAA

//

LOCUS          27     7071 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      27
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.intestinalis
FEATURES             CHR  ---
     CDS             join(500..2719,6497..6571)
                     /translation="MSKFFAAESSSDTDVSSSSEDETPAPSRSSRATVQRQFLLSDD
                     EEEVQHVVRSAKEKRWEELQNIIKLMRNARKIKDMSKLMDEFDNLCKLYIKSKTVVL
                     KEGIPRFYIQNLVETEDCITELWQDKEGRKKMSKLNSKSLGALRQKLRKYLRDFEKE
                     VEDYRLHPDEEEEHVPEKHVEENDAPTFLGSPVKVAQLEKPEPLQDKDDDSDDSDWG
                     FDDDSTSDESDEGGIGARQLTADYFLKKTAGKDDEKSKKKERVQQSKSKKKTDDDWE
                     EVKGGGKVAKVTALFEKGTEITHPVVMKKLNEIILNRGRRSTVRKDQFDLLKQLRQI
                     AEENNLGIALTVRLTFDIIASLFDYNSKILDCMKSEPWAEVLKYVEEVLDLLQENPD
                     IVVAEHISQESENVSDKEAGYRLCGSPVTLVEKLHSEYIKILQTADAHSTDYVKHLR
                     EESHIVRVINKLKEYVERIGSSQDICSSYLLLIEYVYYKYDRNTAEGAVAEEKTSDE
                     EPETEESLDSTQLMSTLCQYIYTKDSSDRIRTRAILHHIYHHSLHDRWYEARDLMLM
                     SHLQETINYSDVPTQIAYNRTMVQLGLCAFRQGLMREAHNALVDIQSTGRAKELLAQ
                     GLLLQRQGERISPEQEKVEKRRQIPFHMHINLELLECVYLVSAMLLEIPYVAAHEFD
                     ARRRMISKQFHHQLRVSERQSLLGPPDNMREHVVAAAKALKMGDYEECLGFIINEKM
                     NNKVWNLFSKSVESRRGSSGVNSHGGEVAGGNSHGEK"
ORIGIN
       1  ATCTGAATTT GACTATCGCG TTATGTATCT GATTCTAGAA TATGTTTCTT TTTTGATATC
      61  ATAGTTACAA CTTCCTTTTA CGTCAGACAC CCAGAAGTCA CGTGGTTTTC ACGGCGTGTG
     121  TTCAGAAATT CACGTATTCT GTTAGTTGGT TTAATTATCA AGGTTTAGAG TACAGTAGAT
     181  GTTTATTTTA CTTCAACTAT ATGTAGAAAT AAACGTTTTA TTGCGTTAGC TGTAGTAGTT
     241  TTACCCAAAA CGTTCTTTAA ATATTACGGC GTATTGCATA TACTAACAGA AAGAATAAGA
     301  CAATGTCTTG TAGAGTGAAC TAAAACTATG TCTGGCAAAT AAAATGAATG GTTAACATAG
     361  TATGTTAGGT GGCCATATTG TTAATATTTG GCCTATCATA CCTATTATAA TTCCACCTTG
     421  GGTATTAATT GTTTTTCCAA TCACAGCACT GCAAGCCAAT ATAAGATTCT TTGCTGAGTT
     481  CTGTTGAACA TAAACAAGTA TGTCAAAGTT CTTTGCTGCT GAATCAAGCA GTGATACGGA
     541  TGTTTCATCC TCCTCGGAAG ATGAGACCCC AGCACCTTCA AGGTCTTCAA GAGCAACAGT
     601  CCAGCGACAG TTTCTTCTAA GTGATGATGA AGAAGAGGTC CAGCATGTTG TGAGGAGTGC
     661  TAAGGAGAAG AGATGGGAGG AGCTACAGAA CATCATTAAG CTGATGAGAA ACGCAAGGAA
     721  GATCAAAGAC ATGTCGAAGC TGATGGACGA GTTTGACAAT CTCTGCAAGT TGTATATCAA
     781  ATCAAAGACG GTTGTGTTGA AGGAAGGAAT TCCAAGGTTT TATATTCAAA ACCTGGTGGA
     841  GACTGAAGAC TGCATCACTG AGCTGTGGCA GGATAAGGAG GGAAGGAAAA AGATGAGCAA
     901  GTTGAATTCA AAGAGTCTTG GAGCTTTGAG GCAGAAGCTG AGGAAATATC TGAGGGATTT
     961  TGAAAAAGAA GTGGAGGATT ACAGACTGCA TCCTGATGAG GAAGAAGAGC ATGTTCCAGA
    1021  GAAGCACGTG GAGGAAAATG ATGCTCCAAC ATTTCTTGGT TCTCCTGTAA AAGTTGCGCA
    1081  ATTGGAGAAA CCCGAACCTT TGCAGGATAA AGATGACGAT AGTGATGACT CTGACTGGGG
    1141  GTTCGATGAT GATTCAACCT CAGATGAATC GGATGAGGGA GGTATCGGAG CGAGGCAACT
```

```
1201  CACTGCTGAT TATTTCCTAA AAAAGACCGC CGGAAAAGAT GATGAAAAGT CAAAGAAAAA
1261  GGAGCGAGTG CAGCAATCGA AGAGTAAAAA GAAGACCGAT GATGACTGGG AGGAGGTTAA
1321  AGGAGGTGGA AAAGTCGCTA AAGTCACAGC TCTGTTTGAG AAAGGAACTG AGATCACCCA
1381  TCCTGTGGTT ATGAAAAAGT TGAACGAAAT CATCCTGAAT CGTGGTCGAA GATCTACTGT
1441  TAGAAAAGAT CAATTTGATT TGTTGAAACA ACTGCGGCAA ATAGCTGAGG AAAATAATTT
1501  AGGAATTGCT TTAACTGTTC GCCTTACATT TGACATCATT GCTTCTTTGT TTGATTATAA
1561  TTCCAAAATA CTAGACTGCA TGAAATCTGA ACCATGGGCA GAGGTGTTAA AGTATGTGGA
1621  AGAAGTATTG GATCTACTAC AAGAAAATCC AGACATCGTG GTGGCCGAAC ATATTTCACA
1681  GGAGTCTGAG AATGTTTCTG ACAAAGAAGC AGGTTATAGA CTTTGTGGAA GTCCAGTCAC
1741  TCTTGTGGAG AAACTACATA GCGAGTATAT TAAGATCTTG CAAACTGCAG ATGCTCATTC
1801  AACTGATTAT GTGAAGCATT TACGTGAAGA AAGTCACATT GTACGTGTGA TCAACAAGCT
1861  AAAGGAATAT GTGGAAAGGA TTGGAAGCTC CCAGGATATC TGCAGCTCAT ATTTGCTGCT
1921  TATTGAATAC GTGTATTACA AGTATGATCG CAACACTGCA GAAGGTGCGG TTGCAGAAGA
1981  GAAAACTTCT GATGAGGAAC CAGAAACAGA AGAAAGCCTG GATTCAACTC AGTTAATGAG
2041  CACGCTATGT CAGTACATAT ACACAAAAGA TTCATCTGAC AGAATTCGCA CAAGAGCCAT
2101  TCTACATCAT ATCTACCATC ACTCACTACA TGATCGCTGG TATGAAGCAA GGGATTTGAT
2161  GCTCATGTCT CATCTACAAG AAACCATCAA CTACTCAGAT GTTCCAACAC AAATTGCATA
2221  CAATCGCACA ATGGTTCAGC TAGGCCTTTG TGCGTTCAGG CAAGGTTTGA TGCGAGAGGC
2281  TCACAACGCT CTAGTTGACA TACAATCTAC TGGAAGAGCA AAAGAATTAC TTGCACAAGG
2341  CCTGCTTTTG CAAAGGCAAG GAGAGAGAAT CTCACCAGAG CAAGAAAAGG TGGAAAAGCG
2401  TCGGCAAATT CCCTTTCACA TGCATATTAA CCTGGAATTG CTGGAGTGTG TATACCTGGT
2461  GTCTGCAATG CTGCTTGAGA TACCATATGT TGCTGCACAC GAGTTTGATG CAAGAAGAAG
2521  AATGATTTCC AAGCAGTTTC ACCACCAACT TCGAGTCAGT GAACGGCAAT CTTTGCTTGG
2581  GCCACCAGAT AACATGAGGG AGCATGTTGT AGCAGCCGCA AAAGCCTTGA AGATGGGTGA
2641  TTATGAGGAA TGCCTCGGTT TCATTATTAA TGAGAAAATG AACAACAAAG TGTGGAACCT
2701  GTTTTCAAAA AGCGTTGAAG TTAGAACCAT GCTTGAAACA AAANNNNNNN NNNNNNNNNN
2761  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2821  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2881  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2941  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3001  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3061  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3121  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3181  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3241  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3301  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3361  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3421  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3481  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3541  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3601  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3661  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3721  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3781  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3841  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3901  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3961  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4021  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4081  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4141  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4201  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4261  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4321  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4381  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4441  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4501  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4561  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
```

```
    4621  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    4681  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    4741  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    4801  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    4861  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    4921  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    4981  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5041  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5101  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5161  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5221  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5281  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5341  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5401  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5461  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5521  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5581  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5641  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5701  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5761  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5821  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
    5881  NNNNNNNNNN NNAGTAGCCG GGGGAATAGA CACAGAGAGG AAGTAGCCGG GGGGATGGTC
    5941  ATGGGGGGGA AGTAGCCGGG GGGATAGTCA TGGGGGGAAG TAACGGGGTG CATAGACACG
    6001  GAGGGGAAGT AGCCGGGGAG AATAGTCATG GGGGGAAGTA GCCGGGGGGA AATAGTTGCC
    6061  GAGGGGAAGT AGCCGGGGGG AATGGTCATG GGGAGAAATA GCGGGGGAAA GTCATGGGGG
    6121  GAAGTAGCCG GGGAAATAGT CATGAGGGGA AGTAGCCGGG GGGAATGGTC ATGGGGGGAA
    6181  GTAGCCGGGG GGAATAGTCA TGGGGAGAAA TAGCCGGGGG AACGGTCATG GGGGGAAGTA
    6241  GCCGGGTGAA ATAGTCATGA GGGGAAGTAG CCGGGGGGAA TGGTCATGGG GAGAAATAGC
    6301  CGGGGGAATA GTCATGGGGG GAAGTAGCCG GGTGAAATAG TCACGAGGGG AAGTAGCCGG
    6361  GGGGAATAGT CATGGGGGAA ATAGCCGGGG GAACGGTCAT GAGGGGAAGT AGCCGGGGGA
    6421  AATAGTCATG GGGGGAATAG TCACGGGGGA AAGTAGCCGG GGGGAATGGT CATGGGGAAA
    6481  TTAGCCGGGT GAATAGTCAC GGAGGGGAAG TAGCGGGGTG AATAGTCATG GAGGGGAAGT
    6541  AGCCGGGGGG AATAGTCATG GAGAGAAATA GCCGGGGGAA TGGTCATGGG GGGAAGTAGC
    6601  CGGGTGGAAT AGTCATGGGG GAAGTCGGGG TGAATAGTCA TGGGGGAAGT AGCTGGGTGA
    6661  ATAGTCACGG AGGGAAAGTA GTGGGGGGGA ATGGTCATGG GGGAAATTAG CTGGGTGAAT
    6721  AGTCACGGAG GGGAAGTAGC TGGGGAAAAT AGTCATGGAG GGGAAGTAGC CGGGGGGAAT
    6781  AGTCATGGGG GAAAGTAGCC GGGTGAATAG TCACGGAGGG GAAGTAGCCG GGGTGCACAG
    6841  AATACGGGGG GAAGTAGCCG GGGGGAATGG TCATGGGGGA AGTAGCCAGG AGGAATGGTC
    6901  ATGGGGTAAG TAGCCGGGGG GAATGGTCAT GGGGTAAGTA GCCGGGGGAA TGGTAATGGG
    6961  GGGGAAGTAG CCGGTAGGAA TGGTCATGGG GGAAGTAGCC GGGAGGAATG GTCATGGGGG
    7021  AGAAGTAGCC GGGTGCACAG AATACGGGGG GAAGTAGCCG GGGGGAATGG T

//

LOCUS          1     1466 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      1
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES            CHR  ---
     CDS          complement(join(500..508,687..786,905..966))
                  /translation="MGHNDVWNSHPKGYGQGSRHCRVCNNNHALIRKYGLNICRQCF
                  REKAADIGFVKYK"
ORIGIN
       1  TCGCCTGCAG CTGCACTCCT CCGGCACCAA CCCGCCCGGT CGGCAGCAGC CCGGCGCGCC
      61  GCACCCGCCA GGTCCGCTGA GGGCAGCTCT CGCGACTGCG CGACTCTCAC GCCGCACCCG
     121  CCACCCGCCG ATTGTAAACG CCTGGGCCTC GCCGTGCCGC CGCCAGGGCT GTCGCTGCTG
     181  CTGCTGCTCC TGCTACGCTC ACCCGCACCT GCTTGTCTGC TCCCGCCCCA GCCCGATCCA
```

```
     241  AGGGTGCCGT ACTCACCTGA TCCATTGCTT GCTGCTAGCT GTGCTGCTGC CTGGTGGTCC
     301  ACCAAGTACA ATGCACGAAG CTGCTCGGCC CTGACCTGAT GCTGCATTGC GCTGCGCTGC
     361  CTGATGTGGC CAATCTGTTT GCGGTGCTGC TCTGGCGCTA ATGCGTGTGG GGCACTGGTT
     421  GAGATTCGGG ACACGTCCGG TCAACGCCTC AGCACCTTAC AAACTGCCGT GTGAGCTAGA
     481  TACCCAACAT TCGGGCATTT TACTTGTACT GCGTGGGGCA AGGAAACAAG CACAGGGAGC
     541  ACACGGTTAG TCCAGTATGT GATACCCCGG GGCCCAGCCC AATAACGGGA GGGGGGGGCG
     601  GACCTGGGGT CGCCCGCTCC CGGCAGCCTA CCCGCCGGCC CCATACCGCG GCCCAGCCAG
     661  CCAAGCCCTC CCCGCGCAGC ACCTACCTTG ACGAAGCCGA TGTCAGCGGC CTTCTCGCGG
     721  AAGCACTGAC GGCAGATGTT CAGGCCATAC TTGCGGATGA GCGCGTGGTT GTTGTTGCAC
     781  ACACGGCTGC GGCAACACAA TCAGATATCA GCAAAATTTC GCCGTTTGAT ACAGCTCGGG
     841  ACGCTAATCG GGACGTGCGG CGTCCCCGCC TGCGCCGGCG CTGCATCGGC TGCCCCCCGC
     901  TCACCAGTGG CGCGAGCCCT GGCCGTAGCC CTTGGGGTGG CTGTTCCAGA CGTCGTTGTG
     961  GCCCATGTTT GCAGGTTACA GCTGCCTGAC CTGCACAAAA AGACCCCTCG TTTCAGGCGG
    1021  CAGTCGGAGA TTGCTAGCCA AGCCGAGCAG CGCTCAGTTC CACACCCACC AGGACAAGAG
    1081  CACGTTGGGC ACTAGGCGCG CGAGAGGGAA AGAGACTGTG CGTGAGCCCG CGTTTGGGCA
    1141  AGCGCCGACC ATGTACGGCG GCGCCCATTG CGAATTCTCC TTGTGCAGAA AGCCACCCTG
    1201  TAAATTCTCA GGATATAATG AGCTTAGTTC AATGCCCAAG AATTTTCCCT GCGCCACATA
    1261  TGTTGCCGCT GCCGCGAGAA ATGTCCCGCA ACATTTTGTG CGCTCACTTC AGCCCGGGCA
    1321  CGTTGTCAAG GTGCTTCAAT GTTTTGTCCA GTACACGTAT CCACAAACTG AATACGCGCG
    1381  GCGGCTCGGT GTCACGGCAC GCTCACAGCG CATCCCCGGG CAGGCGGTGC ACCTCTCTGG
    1441  CGTGCGTAGC AAACCGAAAG CAGCCC

//

LOCUS          2     1434 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      2
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES            CHR  ---
     CDS          complement(join(500..640,881..934))
                  /translation="MYLMYYDDDAGNRVYTLKKIAPDGTPTKSAHPARFSPDDKFSR
                  ERITCKKRFGLLPIQKPPPEL"
ORIGIN
       1  TTCACTCAGA CCAACATGAA CCCGGTCCCT TGTCGGGGGC CCGTAGCCCG AGCCTAATAC
      61  CGCTCCGCCA CCGGCCCAAA CCTCTGCGTG CCATTCCTAT CCGTCTCCTC AGGGGCCCGT
     121  CGTATCCTCC GTCGCGTCTG CAGGGACAGT GCCCTGCACG CCACCGTAAT CTGGCTGGCT
     181  ATCAAGCAGC CAACACCGAA TAAGAAGTAA AGCAATCCCA CTTGCGCTCG CGCTTCCACT
     241  GCGCGGCGGC AAGCTATCAC CCGTCAAGGG CTGGGGACCG CGGCATTCAT GCAACTAGAC
     301  AGGCGTCGCG GTGACAGAAG CACCCTCGCG CGCGAGACAG ACGTCACGAG ATGTGGAGCT
     361  GACCTGCAAC CCCGCACACT TGCCCAGTGC GACACACGCA CACACACGAA ACACACCCAC
     421  GCGTCGTCTA CCCACGCTCC TCCCTATCCC CAACATCTAG TACAATGATA CACTCCACGC
     481  CATGTCGCCG TCATGTTGCT CAAAGCTCCG GAGGGGGCTT CTGGATCGGC AGCAGGCCGA
     541  ACCGCTTCTT GCAGGTGATG CGTTCCCGAG AGAACTTGTC ATCCGGCGAG AAGCGAGCTG
     601  GGTGCGCCGA CTTGGTCGGA GTGCCGTCAG GCGCAATCTT CTGCAAGGTG CATAAACGTA
     661  GGCGGGTTAA CATGATGCGC AGCAGCAAAA GAGAAGGCGG CTCGGTAGCG CCGGCATGGC
     721  GAACCCCGTT CCAGGCGGCC ATGGCTCATC CCCATCATCA CAGCTCCATT TAGTTATCTC
     781  TACTTGATGC TTATCGCTGA CAAGGACGAA GCTGGGCTGC ATGGGTTGCG GCCGGCGGCC
     841  CGTTCTTTCG CGACACCCGT TCCCACGCCC CGCTATCCAC CTTGAGAGTG TACACGCGGT
     901  TGCCCGCATC GTCGTCGTAG TACATCAGGT ACATGGTTGC TGAGTGGCAA AGTTTAGGCG
     961  ATGGAGAGTG CTCAGTCGGC TCAGACAGCA AGCGCTAGGA ATCGGTCAAG TGATAGTAAT
    1021  CCTGAGTGCA GGTGACTGTG GCTATAGACG TAGAGCCTGA AGAAGTTTTG CTTGGCCCGA
    1081  TTTCCGTCTG GAACCCATCC TGAATGAAAC GTCACTTCCT GACCTTTCGT GAGCGTCAAG
    1141  GTCCACACAG ACCATAGATA TTTATCAGCA GCTGATATAT AGCTCAACTG GGAAGACTTA
    1201  CACACACAGC TGTCGCTCGC TCGCTCAGCC TTCCACACAA TGCCCGCAGC TGGGACTGCG
    1261  CCTGCTGATG CCATGCAAAT AGAAGAAAAC TCAGGGAGCC AGGAAGGGAA CGAGGAGGTT
    1321  CAGCAAGAGC AAAATACCGG GTCGCCGAGC AACCTTGGGA CGGAGACCGT GCTTCAACGA
    1381  AGCGCCCTGC CCCGGATGGC TCTTCTACAA GAGGCATTCG AGAAGGCGCT GGAG
```

```
//

LOCUS          3     1586 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      3
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES             CHR  ---
     CDS             join(500..601,696..770,994..1086)
                     /translation="MSYDEVEIEDMTWNEKLQAYTYPCPCGDLFQITLEELRAGEEI
                     ARCPSCSLYITVIYDPEDYENAEPPTRGTDSKVPSAAPQPSEVEVH"
ORIGIN
       1  TGCGGGCCGT GTGCCCGTGT CCGGGCTTTG GGGAAGGCAT GTGCCCAAAT TTGCTGTCCA
      61  GTCGTGCGGC CCCTCTCCTT CAATCTCTCG AACGCCATCG TTGCTAACGG TTCAGGCCAC
     121  CACTAGCAGC TCAGCAACCC GTCTGTGACG TTTGAGCTAC GTACTCGATT TCCCGCGGTT
     181  CCTTGCCACC GCTCGTCCGC ATGGGTGTGC GACCTTTCGC CTTTCTGCCC TTTCAGTACC
     241  CTTGGTCCGC CATCTAAAGA CTTATCTGTG CGAAGCGCAT ACAACTTAAA CTCTAACAGG
     301  TCCTATACTA ACTTTTCCAC CATCGCTGTT TGGGCGATTG TTCTCTGGCC TGTTCTCTGG
     361  CCGCTGACTT TGCCTGCGCT ATACGCAAAC AAGCAAACCA AACGACTTAG TCGCTGATAT
     421  TCCATTGGGC CGCCCGCCGC ATTAGGTAGC CAGTCGCTCT AGACGCGCCA GCCAGCAGTG
     481  CAGGCGCGAG CCGCTAAAAA TGAGCTATGA TGAGGTGGAG ATTGAGGACA TGACGTGGAA
     541  CGAAAAGCTG CAAGCGTACA CGTACCCGTG CCCATGCGGC GACCTCTTCC AAATTACCCT
     601  GGTAAGCGCC GCCCTTGTCG CGTTCTCTAA CCTTACCTGA TGATGGGCAA TACAGCTGTG
     661  TAAGGATGAT GATTTGGGCA TAACACACTG TCCAGGAGGA GCTGCGAGCA GGCGAGGAAA
     721  TTGCCCGATG CCCCAGCTGC AGCCTCTACA TCACGGTCAT CTATGACCCG GTGAGCCCTG
     781  CGGACAGGGT CTTGCTCTGA CTAGGGATCT AACTCAGGAG GGATGCAATA TGGGAATTGC
     841  ACTGGCGTGA GACGGGCTGC ACGGTTCAAG GAGTGGGGCT GGTATGCACG CGGGTGGCAC
     901  GCTGGACTGA CTGAGTCTTC TTGACGCTTG CGGTGCTTGG GTTCGTGATG GCTTTTTTAC
     961  TGATAGGAAC GGCCCTTTGA ATTACGTTTG CAGGAAGATT ATGAGAACGC AGAGCCACCA
    1021  ACGAGAGGAA CCGACAGTAA AGTGCCTTCG GCGGCACCTC AGCCTTCTGA GGTGGAAGTC
    1081  CACTGATGAC ACGGAGTGCC TGCTAGGCCG ACGAGTTCGG TGTAGTGGGG CGTGCTGGAT
    1141  GCATCAAGCT GCGCATCAAT GCGGCTGAGA GTTGCAATGC CTCTGCAGTT TGAGGCACTA
    1201  TGAGCTCAAC GTGCAGCACG TGGTTGCGTG CAAGGCTGCG TGCGACCCCA CTAGCGCACG
    1261  AGACGTGTCA CAGCAAGAGA CTGAAAGCAT GTTATAGCAC TTCCAAGTGG TTGCACTGGC
    1321  CTGACGCGAT TGTGCATGCT TGGACGGCGT CACATTGCGA GTTGTGGGAT AGAAATGCAG
    1381  GATGGCGAGC CGCATCACTT CTGGACGGCG GGCACTTAGC ATGGGGTCCG GAGTGATCGC
    1441  ACTGCGCATG CATTTGGCTT GAGCGACGGT TGTAGCAAGG GACGAGGTTG AGCTGAGCGA
    1501  GGAGGGGCTG TCATTTTGTT AGGCGGGGGC GCGGAGCAAT CCACCCACTG CACGCGGGTA
    1561  GGACAACATC AAAGCGAAGT ATGTAG

//

LOCUS          4     1799 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      4
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES             CHR  ---
     CDS             complement(join(500..517,868..933,1023..1114,1197..1299))
                     /translation="MAGSQSQASTLVARGGRSSSNAPTGAAGALRRRPVRGSGGSGT
                     TKAQQAAMNFYTDDTPGWKMSPVVVITMSLSFIAFVTILHVVGKFQKAA"
ORIGIN
       1  CCGGTGCTTC TCCTCTACCA TCTCACGCGG CACACAGGTG CTTTACGGCA ATGTAATGAC
      61  CGGCTGATAC TTCTCTCCAC CATTCCGTGC TTGTTCGCCC ATGCGTCCAC ACGCTTTGTG
     121  GGGCTGATCG TGTCGGTAGC CTCATGTTAC ACCAAAAGGC CAAGCGCGCC CTGCTTTCAT
     181  GGCCACACCC ACCACACCGT CCGAGCTACA GCCTCCTATC GCGTGCCAAC TATGCGTCGA
     241  CCGCCCGCCC CTGCCACCTG GTAGCGAAAT GCGCCACGCC TAGTTAATCA TCACACATGC
     301  AAGCCTATCT GCTTTACCCG ACATCCACCG TCACCACGAA CTCACATCGT GCGCACGCTT
```

```
     361  TGAAGCGTGT GCGACGTCAA CAGTCCTGTC TGTGTCTAGT GCAGCAGCAA TCCGCTGCTT
     421  AGGTCAAGGT CCAGTGCTAT CCGTTGAATT CAATGCTGCA GTTCTCATAT GTTTGCATCG
     481  CCGTCCCGCT GCAAGGTTCT TACGCGGCCT TCTGGAACTG CGGGTACAAG GGAAACGACG
     541  TGTAGCGGGA AGGCAGGTTC AGTCAAGGTC GCTTCCGCGG TGTAGTCAGA ACTATCAGTT
     601  CTAGGCGTGA GGCGGCTCTC GCACCTACGG GCCTGGCCCG TCCCATGTAT CCGGCTCCAC
     661  TCTACCCAGT ACCCACGCAG TGGCAAGCAC GAGCCCGAGG TGTTGATGGT TTTATCCCGC
     721  GCGCGCGCCC GTGTAACTAA CCGGTCCTCC GCTACGGTCC CAGCTTGCCA CCACATATTT
     781  CCTCTGCTAC CTCCCTGCTG TGCTCCAGCA TGAGCACACC ACGCCCGGGC CTTGCCCGGT
     841  GTCACCCACA CCTGTGCACC CACGCACCTT GCCAACGACG TGCAGGATGG TCACGAAGGC
     901  GATGAAGCTC AGGCTCATAG TGATCACCAC CACCTACACA TACAAGCGCG GAGCTGTTAG
     961  TTGAGAAGCT GGGCAACGAA GCCGGCAAGC AGCAGATCGA TGCCACCGTC GCGCGCCCTC
    1021  ACCGGCGACA TCTTCCAGCC GGGAGTGTCG TCCGTGTAGA AGTTCATGGC TGCTTGCTGG
    1081  GCCTTGGTCG TGCCGCTGCC ACCGCTGCCG CGAACTGCAA GGGCGAAACC CGAAACTCAG
    1141  CGACTCGGCC AAGAAACATG CGTGTTGCTG TAGCGAGACC ACCGCGCCGT TCGCACCTGG
    1201  CCGCCGGCGC AGAGCACCCG CCGCTCCAGT AGGAGCATTG GAGGAGCTGC GGCCACCGCG
    1261  GGCCACCAGA GTAGAGGCCT GAGACTGAGA ACCGGCCATT GTGATCTGGT CTTTTAACAC
    1321  AGAGGCTAGT CGAGTTGGGT GCAGCTAAAA GCCTTGGTAA TCGCTTGGTA TAATGTCTGT
    1381  AGGAACGGTG GATGGCCCAG CAGCTTTTGT GTGAGTTCGA CTGCTTCGCT ACATCGCTGT
    1441  CCCCGATTGC GCGGCCCATC CCGGGCACGG TGTCCCCCCG ACTTGCACTG TGACCAGGCG
    1501  CTTTCACGCT CAGCAGCTTT TTGATAAACT CTCAATTTTC AACGCTTTAT TCATAAATAC
    1561  TTACAGTACG CGGTATGGCG CGACAGGTGC CCGCTCGCGT GCCACACGCT ACATACCCAA
    1621  TACTGGAACC GCCAGCTGCG CCCCCCCACG CTCCCTTACC CCGAAAGCCT CCCTCCCGCT
    1681  CCTTATCCAT TCAAAGCCTT ACCAAAACCC TGGCGAGGCA CATGTCGTAG ATGCCGCCGG
    1741  TAGCGGTTGG GTGATTGATT GGAACCGACG CCTTTTTTTG CAGCGGCAGA TCAGCAGCG

//

LOCUS          5     2039 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      5
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES            CHR  ---
     CDS          join(500..574,848..904,1112..1204,1459..1539)
                  /translation="MDIAFAVVTTAGGVMGYVKKKSLPSLLGGLAFGAAYGATAWTI
                  QNHDAYVGHTVGAATSAVMATMMGLRLAKTKKVMPSGILTGLGLLGLLYHANKARQW
                  A"
ORIGIN
       1  AGCAACGTTT CTGTTTGTCG CTGCCTCCTC CGTGCAAGCC CAGCGCTTGG GCCCCGCGCG
      61  CGCATAAGCA GAAGCGGGGT CAGCCGGTCA CTGTGCGCAA CTCACGACGT CGTTATTGTG
     121  CCTTGCATCG CCCGCATATC GCACGGTGAG CTTGTCTTTC AATTCAAGGC CGTGCCCTTG
     181  GCGGGCTGAA TTCAGAGCGC TCGGCAGTGG TTCAAGCGAC CCCATAATTG CTAGCTCAAA
     241  AACATCATAT ACTCATCACA GCGACAATAC GAAGCTTGCA ACGGCCTGAG GGGAAGCTAC
     301  ATCATGCATA AGCTTTGCGC ATCAGTTCCC GATTGGTGAT CCTGAACCGG AACCCAGTAA
     361  TTACTTGCCA AACTTCACAT TTGCTTGTAG TTCTGCAGTC TGAAATTCAT CAAACGCCAA
     421  AACTACTTTG TCTTGTCGAC GCTGCAACAC CTATCGGCTA TCGCACCTCA GTAGGATGGC
     481  TCCTTCAGGA GAGGTGCACA TGGATATTGC CTTCGCCGTG GTGACTACAG CGGGCGGCGT
     541  GATGGGCTAC GTGAAGAAGA AGTCGCTGCC ATCGGTAGGT GCAGATGTCC AGGTGTGGGT
     601  ATGCTGTGGT GGACGGGTCA AGGTGCACCG CGGTTCAGGT TCGACTTCGG CATGTTAGCT
     661  GGAGTCTCAA TTGCTCGGTT TGGGGGAGCG TAGGGGTCAT GATGTCGGCC TTGGGGCAAT
     721  TGAGGCTGCG TGGCTGCGTG GCACGGCTGT GTGGTACAGT GTTGGTATGT AAAGGGGCCC
     781  AAACGCCCCG GTAGCATCAT TACATAGTAT TAATCTGTAA AGCTGTGTAC AATCTGCTTG
     841  CGCGCAGCTG CTTGGTGGCC TGGCTTTTGG TGCGGCGTAC GGGGCCACCG CCTGGACCAT
     901  CCAGGTGGGC ATCCGGTGCC GCGCCATGTG CACTGCGCTC GTGTTGGCAG GGCACGCAGC
     961  CAAGCAGCCA CGCAAGGGGG ATGGTGGGAA TGAAAGGGCC CGCTTGCAGT GCCGATCCCC
    1021  GCGCCGAGTG CCGCTCCGTG CGTCGTCAGC AGCCTTGACG TGGCAACCGC AATGTCTATC
    1081  CACGAGGTGG CTTGGTTCGT GTGTCTTGCA GAACCACGAC GCCTATGTGG GCCACACCGT
    1141  GGGTGCAGCC ACCAGCGCGG TCATGGCCAC CATGATGGGC CTGCGCCTGG CCAAGACCAA
```

```
    1201  GAAGGTAGGG GTGGGGAGGA GGGGAGCCCA AGAGTGGTGG GCTGGAGACC AGTAGCGGAG
    1261  CTAGGGGGGG CTGCGCAGGC CGGGGGAAGG GAGGGAAATG GCAGTGGGTT TGGGCAGAGG
    1321  GAAGGTGTGG AGGGATTAGA GGTAAGTCCA AGCGGCTGTG GTTTGTAGTG GCATGCCTGC
    1381  GATGCAGTGT ATGCAGGCGG TGGGTGACAG GCGCTACTAA TCCGTACCAC CACCATTCGT
    1441  TCACGTGCAT GTGTTCAGGT TATGCCGTCA GGTATCCTAA CGGGCCTGGG CCTGCTGGGG
    1501  TTGCTGTACC ACGCCAACAA GGCCCGGCAG TGGGCCTAGT CACCCGGCGA CCCGTAACAT
    1561  ACGGAAAATG TACGAGACCT GTTTGGCTGC TCCCGGCCAA GACGCGGTGT CAGGCGGTGG
    1621  CGGCGGAGAA TGCGCCCAAA CCATCTTGAT ACCGCGTACG TGTGGGAATG TGAGCGGTTG
    1681  CTCGATTCGT GGCACGCGAA GTTGAGCGCG GTGGGTGGTG GGTGTTTGCT GGGCGACGGT
    1741  GTGGTGGGGA CGGACCAAAC GTGTAGCGTT ATTAAGGAGG TGTGCTGGAA ACGGAAAGAG
    1801  TTGTGCGCGC GGAGATTACC GTGGGGGCGA GGGGCGTTGG GTGCGTCAGA CGGTCAGGGG
    1861  CCACCCGAGG GGTAACGTCT CGTTTGCGCG GGCCGTTGCG CGCAGTGGAG TGCTTCGCAA
    1921  CCGCGCCCAG CTTCGGCTGT TTGAAGTTGA AGGAATTTAT ACTACTGACA ACTCGGTTTT
    1981  TCAAGCACGT AATAAGCAAG CCAAGCGTAA TCTGCTTGGG AAAGCGGTTT GTGCGCATC

//

LOCUS          6     1543 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      6
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES             CHR  ---
     CDS             join(500..546,636..834,978..1043)
                     /translation="MINDEGQVVDLYIPRNAWTNKLITAKDHASVQINIGHLDANGV
                     YSNQYTTFAIAGNVRAMVSEGRDGRLLRGGDPSSSPQPLGEGDSALDIMWRKKQVEA
                     EAH"
ORIGIN
       1  GTAGTGCGTC GGTTTCAAGA AGCACCATGC AGGCGCCTCA CAGCTGAGCA ACTGCTGGCA
      61  GATACACCCG GCTGAGCCTC AGTCCAGGTG GCCCGCGCGA TGCCGGCTGG AGGCCATTTC
     121  TGCCATTGGT CCGAGCTGCT CGCACTCAGG CGATAGTGTA CTGTGGCAGT TTCGATGTGA
     181  GTTTCTTGGA TGCAGACCTG ATAAAGAGGC GGGCAAGTTC AAGCGCCAAC GCCGCCACTT
     241  AGTGTTGGCC AGGGGGGCGT GTTGATACGC AACCATAGTA ACGTGAGGCA TTCCTGTAGA
     301  GGTGGACGTG TGTATGTGAT GCTTAAAAGG ACGTGTAGCT TGGGACACTT GTGAGCGCGG
     361  CGGGCAAAAG TCCTGAGCGG GGGCTGCATG CGAGGCCCAT CGCGACGGGC GTCCAGCCGG
     421  GGTCAACTTC GCCTCCATCG CACTCTGGTT CTCTCTCTCC CCGCCGCACA CACGAGGTCC
     481  ATTTCAAGAC CCTTGAAAGA TGATCAACGA CGAGGGCCAG GTCGTCGACC TCTACATCCC
     541  GCGCAAGTGG TGGGTCGGCC GGTGACGGCG CTGCGCTTCT CGCTGTGGTA TTTTCTCATG
     601  GATAAATCTT ACAAGTCTAT AATTTTGTGC TGCAGCGCGT GGACCAACAA GCTCATCACC
     661  GCCAAGGACC ACGCCAGCGT GCAGATTAAC ATCGGCCACC TGGACGCCAA CGGCGTGTAC
     721  AGCAACCAGT ACACCACCTT CGCCATTGCC GGCAACGTGC GCGCCATGGT GAGCGAGGGG
     781  CGCGATGGCC GGCTTCTCCG CGGCGGCGAC CCCAGCAGCA GCCCCCAGCC GCTGGTAGCT
     841  GTTTTCCGTT TCCTAAAGAG ATACGTCGCT TTCATTTTGG AGCCGGATGG ACTCGTTAGG
     901  GCACCTAGGC GGACCGAGGC GCACATGATG CGATGCCCTG TTCTCCACTG ACGCAAATTG
     961  CCCTGGCTGC GCTGCAGGGT GAGGGCGACT CCGCTCTGGA CATCATGTGG CGGAAGAAGC
    1021  AGGTGGAGGC TGAGGCCCAC TAAGTGCTGA AGCATCCGCA GCGTGTAACC GGTGGTGTGT
    1081  GTGGAGCTCG CTTCCCCCTT CTCTTGGCCC CAGCCAGCAG GGCGCAGCAA AGCATCAGCA
    1141  TATACTCCGA AGAATGCTCT GGCAGCATCA GCGCGGAGGG GGAGCGCAGG GTGCAAGGCA
    1201  CGGGGGCACG GCAGCCGTAG GGGGCGGGGT GCGTGGCGTG GGGCGGTGGC GAGCCTGCTG
    1261  GCCGCGGCGG GGCGAGCAGG CTGCTCTGGC GCTGCGGAGA GGCCGACGGA CGGACTGGTG
    1321  GACACACTGG CGTACTGGCG CTTGCAAGCG CTCACACATG GACGAGAGCG GCGGCTTCAG
    1381  GTGTCAGCTG CGGCCGCAGG GTGCGGGGCT GGCGTCGGCG CTGCAGCCCC ACTGCTGCCA
    1441  GGCCGCAGTT CCAGCTGCAA TGCACGCAAG CGGGCACAGG CGCTAGCGGG CACAGGCCAT
    1501  GCTGCGCTGT GCCCTCTGGC GCACTGACCA GCTGCGCAGC GCA

//

LOCUS          7     1936 bp     DNA    linear PLN 25-FEB-2004
```

```
ACCESSION      7
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES             CHR  ---
     CDS             complement(join(500..591,1074..1176,1299..1436))
                     /translation="MYVEDFEAFYEQAVALYQAKPLETRYTIKFRHSEGKVVLKVTD
                     DRTCLKFKTDQQVDLRKIERITTIFFPLMACGELPAEGQDAAAAQQAPAAAATQSGG
                     AGGKKARRRG"
ORIGIN
       1  GCACAGGTCA CCCCATGGCT GGGCAATTCC TGACCCGAGT ATCAGCCACG ATACGAGTGC
      61  ATCCTGGCTG CGACAGGAGC AAGGCCGCCG CGTCGCCATG CAGCGCCAAC CAAGCCGCAC
     121  CAGCAATCCT GTTCTGTAAG ACCGGGGCCC CGCTCCGCAC CCATGACTCA CATCCTTGTC
     181  CCACCACTGT CTAACACGGT GGGGTCCAGG TGCGCAGCTC AGGCGCCCCC AAAACCGCAA
     241  TGCCACTTCT TGCGTGCGTC TGCGGGTGAA CTGCGTTCAT ACATCACTCG CGCATTCAAG
     301  CCCTGCAGCC ATAGCGCAAC GCAGGCTGCA CATTGCAGCC TGTTACCTCC TTTCCGCATG
     361  CAATCTCGAG AACCCAAACA CCACCTCGCC GACAACCCCT CACAGCGCCA ACACCTCCCC
     421  AATTCTCTAA GCAGCAAGCT CTTGCGCCAC CACACTACCT CCGCCAGCCT TCTCGGCCAT
     481  CTAAAAGAGT CTGTCTGGGT TACCCCCGCC GCCGCGCCTT CTTCCCGCCG GCTCCTCCTG
     541  ATTGCGTTGC AGCTGCCGCC GGTGCTTGTT GCGCCGCCGC CGCGTCTTGC CCTGTAACCA
     601  AGTTGACACA GGTGATGTCT CATCATGAGG GGCAGGAGGT AACAGTCGTG GGGTGGGTTG
     661  TATCAGCAAG ACACCTGCTT GGGAGCTGGC AGCACCATTG AAGTCGTACG GGCTTGGGCG
     721  AGCAAAGTCT GTGCCCCCGC TAGGTTCATG GCGTTTGTGC CCCCGCTAGG TTCATAGCAT
     781  AGGGGCCAAG CTGGTGCAGC TGTATGGGTT TTCACCCACG TAAGTCCGTC CAGTAAACAG
     841  GCACACGCTG TGTTGCCCCT CCCCACGCCA CCGCCCCGCC CCCGAAGCCT GCCACCCGCC
     901  CCGGCCTTTC ACACGAAACA CCAAGCGCCG GATGCTCCGG ACACCCAGCA TATCCCACTG
     961  CAGCAGCTAG CGAGTGTGCG CGCCGCCATC GCCTTGCCGG TGCTCAGCAC CACAGCGGAT
    1021  CGGCCGTGCC TAGGCGACCT TACTGCCCCC GTCCCACTCC GTCCATCCCT CACCCTCCGC
    1081  AGGCAGCTCT CCGCACGCCA TGAGCGGGAA GAAGATTGTT GTTATCCGCT CAATCTTTCT
    1141  GAGGTCTACC TGTTGGTCCG TCTTGAATTT TAGGCACTGC AGGCGACGCA AACATGGATC
    1201  GTCAGTGGTC CGCCGCCGGG GAAGCAGCCC ATATTGAGTG CCATCTATAC AGATATACCA
    1261  CCATTTCTAG TCCAAAAGGC CCAACAGCGC CTGCGCACCG TGCGGTCGTC CGTGACCTTC
    1321  AGCACTACCT TGCCCTCGCT GTGCCTGAAC TTGATTGTGT AACGAGTTTC TAGGGGCTTT
    1381  GCCTGATACA ATGCTACTGC TTGCTCATAG AAAGCCTCAA AATCCTCCAC GTACATTTTC
    1441  GCTCAGCGAC CGCAGCGTGG CTTCTATCAA AGGCAATTGT AAATTAAAGT TTGTGTGCCG
    1501  GGTTTGATAG AAATGTTGAT CGTCGGCGCA GATCAGCAAG TTTGCTTGGG AGCGCTCATT
    1561  CCGCAGTGTT TACAATGCCC CAAGGCCCAA GCTCTCTAGG CACATTTCTC CCACGTCGCC
    1621  AACAACTTGT AGAAGATACA CATAAATTAT CAGTGTGTAT ATCCCGAAAC AGCAAGACAT
    1681  TTAACGAGTT CCAGTCCCGA ACACCCTGTT TTCCTGCGCG TTTTCATTTA AAGATGGCCG
    1741  ATGTGGCTGC TGAGGCGGTT TCCCTGGAGC TTGAAAGCCC TTCTTCCAGC TCACAGGGTT
    1801  CGCCGGAGTC CGCAGCAAAT GGCGCAGCTA GGGCCGCCTC GCCCCGCGCG CAATCGGATG
    1861  CCAGTGGAGC TGTTATGAAG AGCAAGAAGG GCAAGGCACA GGTTTGTTCC GGGGCAGGGA
    1921  GTCCTCTCGG CAAACG

//

LOCUS          8     1389 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      8
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES             CHR  ---
     CDS             500..889
                     /translation="MSSANPRTLAGPIVQLFRDVMKLHRERLPPPMRDLGDSYARAE
                     FRSHLRGKTTMEQWKQFVREWQVYLATLRGDEQAAASADANVARVFELLSDDQRKRM
                     DQLQQELNSRTGGGGGDSDGGDAGRGPHQ"
ORIGIN
       1  GTCATACACC TAAACTCTTA CTTTGCGCCA GCAGCCCTGA ATGATGTTTG GCGGTCCTTC
      61  CAAGCCCGAG CCTCAGGGCC CCCTGACTGC TTCTCTCGCT GCCAAGAAGC TTTTCGTTTC
```

```
     121  GTTTAAGAAG GCTGCTTGGA TTTCCTCGAC CACTCTGCTC GTGCTGCTGG TGCCTCTGAT
     181  CATTGAGATG GATCGCGAGG CACAGATGGT AAGCAATCAT CAGGCCTCGT AAGGCTCTTG
     241  GCGAGGAAAA CGGGGCGTGA TGGCCGATGG ATCGGAACTC AATACCGTGC CCGCTGATTG
     301  GCTTGCAGAT TGAGCTTGAG AAGGAGCAGC TGAGCGTGCT CACCGGTGGC CCTGCCGCCG
     361  CGAAGGCGTA ATCAGCTCGA CGTGAGCATC TTTGTACTCG CTCGCTGGTT TAGGAGCCGG
     421  AGCCGCAGCG CCCCGGTCGC TCATTAGGAC TTTCGGACGT GAAAGCTGGT CGCTCAGCCA
     481  GGATCGGGAG GCAGGAAGGA TGTCGTCGGC AAACCCGCGC ACGCTTGCCG GGCCAATCGT
     541  GCAGCTCTTC CGCGACGTTA TGAAGCTGCA CCGTGAGCGG CTGCCCCCGC CTATGCGGGA
     601  CCTGGGGGAT AGCTACGCTC GCGCGGAGTT CCGCAGCCAC CTGCGCGGCA AGACGACCAT
     661  GGAGCAGTGG AAGCAGTTTG TGCGGGAGTG GCAGGTCTAC CTCGCCACGC TGCGTGGCGA
     721  TGAGCAGGCG GCCGCCAGCG CGGACGCAAA CGTGGCGCGT GTGTTTGAGC TGCTGTCTGA
     781  CGACCAGCGC AAGCGCATGG ACCAGCTGCA GCAGGAACTC AACAGCCGCA CCGGCGGAGG
     841  CGGAGGAGAC AGCGACGGTG GCGACGCCGG GCGCGGCCCT CACCAGTGAG GGGCGCGGCG
     901  GAGCGGTGGA GTGCGGGCAG CGGCTTTGGC GGGGCAGGAG GCAACAGTGT GCTTTGTTGG
     961  GCAGGATGGG ATTTCTGGGG GCCGGAAAGG GTGAGGCCGC GGCGCGGCAC GGGTCTGCGC
    1021  CCGTACTGGC GGGTTGATGC CGGTGGCAAG TTCGTGCAGT GGCGGCGGCG GCCAGTGCGT
    1081  AAGCGGGAGC TGTGGTCCGT ACTGGCACCC TGTAGCAGTG GTTTCGGGCG GTTGCGTTGG
    1141  TGTAGAAACT AGCCAGGGCT TGCCGTGCCA ATGCGAAAGC ATGCGGATGC AGGAGAGGCA
    1201  TGCGAGCGGT GTTGTGCGCT GCGAAGCTCT GCTTGCCTGG GCCTCGGTTG GACCAGACGA
    1261  AGGCATCATG CAAGAGTTTG GGTTGTTGGT CTGGTGAATG ATTACTACTG GTCTGGTTGC
    1321  GGCCATCCAA GAGGACTGGG ACGCTGCGGG TGCACATCAC AGCTGAGGCT GGGGTTGCTA
    1381  CTGTGTGTT

//

LOCUS          9     1904 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      9
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES            CHR  ---
     CDS          join(500..619,901..1074,1306..1404)
                  /translation="MQQVLIRMPTAGGAKASGPLLQQLAKSLETEGAQLASKIKGIV
                  VFKIDDAEWSLDLSEGSSGALYEGPPAEGAKPDLTLTTSDATFAQMVMGKVNPQTAF
                  LTRKLKINGSMGMAMKLAPVLEAAQPRAKM"
ORIGIN
       1  GGCGGCGACC TGGCGTTGGC CCGGCGGTGG TCCAGGGGGC CGTCTCAGGG CGTGAAGGCC
      61  GGTGGTGGCC TGCACAGCCG ACAGACAGGA CGTGCCATGC GGGGGAGAGG TCCTAGCAGG
     121  TGAAAGGGCC TTCCGGTTGT GCGCGCTTGT GATAAGCGAA GAGGTTGTGG TGCTGGCAAG
     181  TGCCTAGCAA ACCGCTTGCT ATAATAGTTG CTCGGGGTGC GTCGCTCGGA GCGGAGGTGA
     241  TGCTGTCAGG AGCGGGGGCT GGCGTGGATC TGGTGCAGTG TCGCCGCTGC ATGGGGGTCG
     301  TTGCGCTGAC GAGGACGGTT CCTGAGCAGG AACTGACAGG AACTGAGCAC ATATAGCACC
     361  ACATCCCTTC CTTACTTTTG TTTCAATATT TACAGCACTT CCAAACCTGT TGGCTCTGCT
     421  ACGCTCCTTG ACGCTGCGAC TTTGCTTCTG CACAGTCTCT CCGCCTCGCC ATCATTTGCC
     481  ACATTTCGTC CACAACACAA TGCAGCAAGT ACTGATTCGT ATGCCTACGG CGGGAGGCGC
     541  AAAGGCGTCG GGGCCGCTTC TGCAGCAACT TGCAAAGTCG CTGGAGACGG AGGGCGCGCA
     601  GCTGGCCAGC AAAATAAAGG TGCGAGCACA ATGCTGGGGA GCAAGGGGAG GTCGGGAACT
     661  GAAAAAGCGG TACCGGGCGT GCTTTCGGCA TTTTGGCGGT TTGCGACGTG GAGGACGGCA
     721  GGCTAGCTCT GCGTATAGCA GTTCCCCAAT TGACCATGCG TGGAAATTTT CAATGCTTGC
     781  ACTCTGCCCG CTATCCCATC GCCCCTGGCC GGCGGGAGCC CGAGTGCCGG CCTCCATCGC
     841  TCCGTGGAAC CGGCTTACAT GCTTGCCGGT GCCCCCCACC CGCATGATGG CTGCGCCCAG
     901  GGCATAGTGG TGTTCAAGAT CGACGACGCC GAGTGGAGCC TGGACCTGAG CGAGGGCAGC
     961  TCGGGGGCGC TGTACGAGGG GCCGCCGGCG GAGGGCGCCA AGCCCGACCT CACCCTCACC
    1021  ACCAGCGACG CCACCTTCGC GCAGATGGTC ATGGGCAAGG TCAACCCGCA GACGGTGCGT
    1081  CCGGGCGGCG GGGCCGGGGG CGGCTGATGT GCAGCTTATG AAGCGGCTGC ACGTGGGGAA
    1141  CGAGGGCAGC TGGGGAGGGC GGGCAGGGGC CGCCGGTGCT GGCACTGGTA GGGCCCTACT
    1201  GGTACCGGGC CGCCGGCCTG CAACCTACCG GTACCGGGCA CCTCGCATCC GAGTGACTCC
    1261  TCCCCTGCCG CCTTGCCCCC CCCCCCCCCC ATACACGCAC GCCAGGCCTT CCTGACGCGC
```

```
    1321  AAGCTCAAGA TCAACGGCTC CATGGGCATG GCTATGAAGC TGGCGCCCGT GCTGGAGGCG
    1381  GCGCAGCCCA GGGCCAAGAT GTGAGAGGTG GCAGCAGGGT GCAGGGCGCA GGGCACGGGG
    1441  TGGTGTGGCC GCTGACTGGC TGACAGGATG GGGCGGGAGG CGCGGAGGAG TGATGGAGCG
    1501  GCACGAGGAC GGGTGGCGCT GTAGGGGTGT GCGCTGGGAG GAGGGAGGAA GCGCCGGACA
    1561  GGGAGGGAGC GGCAGGTGTT GAGCTGCCGG CCAGGGTGAG GACGAGGGCG GGGTGTGCTG
    1621  ATGTGCTGGT GTGCTGATGT GCTGGTGTGC TGATGTGCGG GACGCGGGGC GCTGGTGGTG
    1681  CGTGAAGCGG AATAGAGAGG ATGGCAACGC GTGGACGTGT GCGGCAGCTG CGAATACGGT
    1741  ACACAATGTT ATGCTGCCCG CCGCGGCGTG TCAGGCGCGG CTGTGTGTTG CACGGCAGTT
    1801  GTGCACTGAC TGCTGACGAG CCGGCCTGCC AGGCGGCCAG CCTGCCACGC GGCCCGGCGC
    1861  TGCGACGGGA GGGCGGGGCT GGAGGGGCGC GGGACTGCGG CGGT

//

LOCUS          10     3780 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      10
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES            CHR  ---
     CDS
complement(join(500..548,984..1128,1970..2110,2972..3029,3269..3280))
                 /translation="MATQKTAVAVAYCKQGSGLIKLNGQPLELVKPEALRFKVFEPI
                 LVIGRAKVKNLDIRLRVKGGGHVSQIYALRQAIAKAVVAFYQKYVDEQSKQQIKDAL
                 LQFDRTLLVADPRRCEPKKFGGRGARARFQKSYR"
ORIGIN
       1  CTGAGCGCGG TCGAGGAGAG CTGCTCGAGG CATAGGTCAC TCTGGTGGAC TCGCACCCAC
      61  AAGCTGCAGC GGCACTTGAC GCTATTTCAC CAGTGAGTTA CACCTGGACT AGACGGCGGT
     121  TCGTCCAGAC GCGTTAAACC TCTCCCTCCC CGGCCGGCTC GCCAAGCACA CGCGCCGAGC
     181  CGAGCCGTCG CCCTACGCCG CACTGGCTCA GGCGACCGCC GGCGCCCGCC GTACCCCGCG
     241  GGCCGCGCGC GCCGCCGCCG CAGCCGCCAG CCGCTGCGAA GCGGGCCCAC GACTGCGCCA
     301  CCCCGCGCGC CGCCAGCGAC TAGCCGCCGC CGTCGCCACC GACGCCCGGC GGCCGGCGCG
     361  CCAGCCACGA CAACACCACG ACGACACCTC CGAAAGCGCC CCGGCTGAGG AACCAGCTGG
     421  GGCGAGCCAC GCACCGCTGC GCCGCCCAGA CCCACGCCCC CGACGGCGGC AGTTGCCACG
     481  TGGCATGCGG TCGTACCGCT TAACGGTACG ACTTCTGGAA GCGGGCGCGG GCACCGCGAC
     541  CACCGAACCT GTGGGCGTAA GGAAATACAC GAGGAGAATA GCCGGTCAGC CACGTGGTCA
     601  GGTGTGCGGC GTTGAAAGGA AAGCTTGCGG GACGCCGTGT GCGGGACCCG GGATGCGACC
     661  CGTGTCAAGA GCAATGTCTT TGCTGACCCT AAAGAACTCC CTGGGCTTGT GTTTTAGCAA
     721  GAAGCTCGTT GCGTCATTGC ATCCCCCACC ATTTACCACG CATCCCTCCC GCCATTCAAT
     781  GCTTTGGGCG TCACTCATCG CTCGTTCTTT GCAAAAGCTT TCCCAAACCC GCCTCCCCAG
     841  CATTCGCGGC GTCCCGGCAC AAAGCTCCAG CGCAACCACT TGGAGCCCGA CCCATCGGCA
     901  CATTCGTTTT CGCCCGACCT CCAACCCCCC AGCAGCCCCT CCCTAGCGCC CCAAGCCCCC
     961  CCACGTCGTC CCAGAGCACC CACTTCTTGG GCTCGCAGCG GCGGGGGTCA GCCACCAGCA
    1021  GCGTCCTGTC GAACTGCAGG AGCGCGTCCT TGATCTGCTG CTTGCTCTGC TCGTCCACGT
    1081  ACTTCTGGTA GAAGGCAACC ACGGCCTTGG CGATGGCCTG GCGCAGAGCT GTGGGGTTAA
    1141  AAAGCAGGAA GGCATATGAA AGAGAGGTTT AGAATGCGGA TGAGATGCGA GGGATGTGTT
    1201  TCCGCGTGCT ATGGGTTAGG GTCCGGTGGC CAGGAGCCAG GCCGCCGCGG CTCCTCCAAC
    1261  GCAAGCTCCC ATGCTGCTCT GAGCTGCAGT CATTTGCGTG ACTGGGTGCC ACAAATGCTT
    1321  GCGTTCCGGG CGAGCTGCTG CAGAGCTGCT CCCGGCGCCC ACGCCCCGCA GGCGCCCCGC
    1381  CACTCGAATC GTTGCGGCAC ATCCCGATGA TATGCCGCCC TCTGCTGCAA TGATCCCTGG
    1441  TACTCGCCGC CGCCGCCCCC CCGTGGCGTC GACACGTTGC GCCACGCGGT TCCCGGTAAC
    1501  CAACGGCGGC GGCAACATGC GGCCCCGCGG CCACGTTTGC GGCTTGCCAT AATATGCTCA
    1561  CAATCCTTCG CCCGCAGAGC CTCTGTAGCT GCTGCTAGCT ACGGTACTGC GCCATGACGC
    1621  GCCATTTGCA TGGGCTAAGC TGCTGCGCAT ACGTCCGCTC CCCCGGCCCA AGCCTGCTAC
    1681  TCCCGCAGTC ACCGCGCTGC TCACTGCACT TTTCTCCAGC TCCAGCTGCT GCTCCACCGG
    1741  CCGCCAAGCC GCGCCGCCTC GCCGCTCTGA CCTCGTCGCT GCTCAGGCCC ACGCTTCGCC
    1801  AGTTTGGTGC TCAGCAGCGC TCGCACGAGG TTGCCGCGGC CGGGCGGGGG GCATCGTAAT
    1861  CGATACCCGT AGCCATGCAA TTCCCCCATA GCCATTCCGC CCGGCATGTC CCACGTCACA
    1921  CCCCCAACCC CTCCCTCCCA CTTCCTTCAC ATGTCGAACT TCCACGCACC GTAGATCTGC
```

```
    1981  GACACGTGGC CGCCGCCCTT GACGCGCAGG CGGATGTCCA GGTTCTTCAC CTTGGCGCGG
    2041  CCAATCACCA GGATGGGCTC AAACACCTTG AAGCGCAGAG CCTCGGGCTT CACCAGCTCC
    2101  AGGGGCTGGC CTGTGGTTGG AGATGCGTAA TATTGGGACG TCAAAGGCAG GTGCGCGGGC
    2161  GTCAGAGGAA GGGAGAAAGC TGGATACGGA GTGGTAAGGG CTTTGGTATG CCGCGTACAG
    2221  GCTGTGCAAC GCGCTGGGGC GCTGTGAGCC GTTGAGCCAC CAGGCTGGGC AGAGCTGGGC
    2281  TGGACTGGGA GGAGGACAAC CGTCCTGTGG GCTGAACTGT TTGGCGCTGC TGCCACCAGG
    2341  CTCGCTGCTG CGGGGCCGGG CGTCTGTGCT GACGGCCTGT AAGCGCCGCT GCACCCCGTG
    2401  CAGCGAGGGC GCAACCCCGC CCCCGCCCAG CAGTCTCCTG CGCACTGCGC GCCTCCAGCC
    2461  TCCTGCCCCC ACTCTGACGC ATGCTGCGTC CCTCCAGCAA CTCCACATAT GGCGCTGCTG
    2521  CCCGCGCGCC AGAGCATTCG CTGTGGTACA GCTCCGGAGG CTGTGTCCGC CAGTGCGGCA
    2581  CAGTGCTTCC CCGTGTTCTG CGCCATGAGC CGACGGCTGC CGCCCAGGAG CCATAACTGT
    2641  TTCGGCGTGT AACACGCCAG CAGCATGCCT GATCAGTTAT TCGGCTGCGA GCTCGCAGGA
    2701  CGCGAAGGCT CTGTCGCAGC CGATGGCCGG CCCAGCGCTC CCGAGCCAGC ACATTTGATT
    2761  CCACCCCGCT TCCTGTTTGT GTAAAGACTC CTTCAGGCCG CTCCATATAC GTCTGACTGC
    2821  TGCGCCGCAG CATCCGCCGC CCAGCCTCGC CCGCCGCCCT GGTAGAGCCC GCGCTGCTAC
    2881  TGTGATGTAC CCTGGGCTCT TGCTAGAGCC TTTGTCCTTC GGCTGCGCCA GTCCTGGCCA
    2941  TTGTTGAGCG ATCCACCGGC TTTGCACTCA CCATTCAGCT TAATGAGGCC GCTGCCCTGC
    3001  TTGCAGTAGG CCACCGCGAC GGCGGTCTTC TGCAATGGCA TAAACCAGCA CAAGGCGAGC
    3061  GTTCAGCGCT GTTCCCCGTG ACATTATGCC CGGCCGCCGC AAGGAGCCAT CGCAACCCAA
    3121  GGCCCGCCTC ACCTTGCGGC CGAAAGTCTG CACCGACTGT TGCATCCATG CCAGAGCAGC
    3181  GTAGGCACGG GGTATGGGCA GAAGGGCACA AGCCGTCAGC ATGTGTTTTG GCGGCACTGC
    3241  CCCGCCCGTC GCTGCTCCTT ATGCCTACCT GGGTCGCCAT TTTGGCAGCC TTTCTTTTCG
    3301  CAATGCGGAA GAGGAATTTT CCTGGTCGCC AGCGCGAGGT GTAGCCCCCC GCTCGCTACC
    3361  CGAGTGGAAT GGGGCCCCAT GGCTTTTCCT GCCGAGCCCT CCTTCCCGAA GGACTGATAA
    3421  TGCAAGCTAT TATATGTGCA GCTGCCTATC TATGACAAGC TTAGCCGTTT TGAAGGGTCC
    3481  CTTCGGTCCA TAGCCTCCGA GAATTGCAAA CTTTGCCCGC TGCGCAGACG CGCCGGGGTT
    3541  ACGCGTGCGG ACCAGTTCCT CCTGAAGAGT GATGTAAACT ACATGGCGTT TGCATTTCTA
    3601  GAACGGAACT GGGCGGCGGC CCCGACCCCG GCCTCTCGGC TGCCAGTGCG TTGCTCGGTC
    3661  GCGCGCGGCA CCGCGGCGGG GACTCCTGCC TCGGATGCAC CAGAGCACCT GCGACAGTTC
    3721  AGCATACCGC TGCAGAATCC TCAGAGACGA CGTCTCTCAT CCGCGGTGGC TGCAAGCGTG
    3781

//

LOCUS          11     2227 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      11
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES             CHR  ---
     CDS          join(500..530,649..803,993..1070,1365..1458,1672..1727)
                  /translation="MPLGGKLVLKGGTTLKELDGGVEKKKSKKPKKLKEGGEGAEGG
                  EGAVATADGAPADPTKPKPAVPTAINVMSGKNYEQEFDIEMKRVQEGKSKSTPWGCG
                  YRAPPAILHGYTSKVTGKTAEERLDLRSAAKSDKFCK"
ORIGIN
       1  TTCCCACCAG CGAAATGGCT ATCGTGAGCG GTTATGACGT GAACCTTAAA CATGACGAAC
      61  CGCGCTGGGG ACAGTGGAGC CATGGCCGGG TGAGGGCGTG CCCACGCTAG TTCAGCCTAT
     121  TCGCTCAGGG CTTGCACACA AGGTGACATG AAGACATTCA CATGGCAGCT GCTTTTGATG
     181  AGTGTAAATT ACAGTATACT GCCGAAATGG AAGATGAGAC GACCTGCATC GCAAACTTTG
     241  CCTGTTTCAA GGCCCTAATT GATTCTTCTT TATGCTCCGC TGACTTGACC ACGTAACACT
     301  TGTAATACTT TTAGAAGCTT GCCCGTTCAG CGAGTTGGTT ACCTGAATAA CGTCACTGCT
     361  CCTCGTTTTG CATCTAGAAT ACACATAACC TACTGCTTGC ACAGACAAAG TGTAGAAGCG
     421  ACGGGCCTGC TGCTGTAATA GACGTGCTGC TGGAAGGGAG CATTCTTTGC GTGCCTGGTC
     481  TGTACAGTGC ATCGCAGCTA TGCCACTAGG TGGCAAGCTG GTGCTGAAAG GTGTGTGTCC
     541  CGTTGGATGA GCGTGGGACA GCTCAGGTAC TTGGTCGCTG GTCACTTCAG CGCCTTGCGG
     601  TGCCGTGTCG CGTGCACCCC CACCAGCTTC CGGGCTCATG CTGCGCAGGC GGGACGACGC
     661  TCAAGGAGCT GGATGGCGGC GTGGAGAAGA AGAAGAGCAA GAAGCCCAAG AAGCTGAAGG
     721  AGGGAGGCGA GGGTGCGGAG GGAGGCGAGG GCGCTGTCGC CACGGCGGAC GGCGCACCTG
```

```
     781  CCGACCCCAC GAAACCCAAG CCGGTGCGTG GCTTGCGGGG CAGCAACGCG GGAGCTCCCT
     841  AGCCCCGCCC CGCCCAGTCA GCTGGCACTT GTCCGCTGCA CCGGCGCCCC GCTGCCATGC
     901  CACGCCCTTG GATTTCTGGA TATGGTTCGC GAGCCGTGAA CCGACACTGC CGGCTGACAA
     961  GTGTGCCCGC GCCATCGTGC TGGTGTGCGC AGGCGGTGCC CACGGCCATC AACGTCATGA
    1021  GCGGCAAGAA CTACGAGCAG GAGTTCGACA TCGAGATGAA GCGGGTGCAG GTGCGGGCGC
    1081  AGCCGCGCGG TGTGGGGGCC GGGCCGCGAG GCCAGGAGCC GGGGGCTCAC GCGGCCGCGC
    1141  TGCGAGCAGG GCCAGCTGGG GCGACCCGCG CCACGCTGCC CGGTTGCTGC AAGGCCGCAA
    1201  TGGGTGTTGG GTTGTGGAGG GGGTTGCGGC ACGACATATC GCCTCAGAGC CTGCTTACAT
    1261  CTCCTTTGCG GCGCGAGCGC ACAAGGCGGC ACGGGCGGGC CAGGCCTCCT CAACTTCCTA
    1321  CTGAGGCCAC GCCCCAACTA CGGTGCTTGC CGGTACACGC GCAGGAGGGC AAGTCCAAGA
    1381  GCACCCCGTG GGGCTGCGGC TACCGCGCGC CGCCTGCCAT CCTCCACGGC TACACAAGCA
    1441  AGGTCACTGG CAAGACAGGT GAGAGAGCTA CTAGAGCTAC TGAGCTACTG AGTGCCAAAG
    1501  CCTACACCAC CTACACACAC CCGCAACGCT TGCACTACTG CCTTGGCACT GCGTCCGGAC
    1561  CGCCTGCAGC ATTTCCTAGC TGTGCCGTGC CGTGCCCGCC GCGCCGCTTG CGAGGGTCAC
    1621  AAACTTTCTT GCCTGCCTGC TGCTGTATGC CCCTTTTCCG CCCACATTCA GCCGAAGAGC
    1681  GGCTAGACCT CAGGTCCGCT GCCAAGAGCG ACAAGTTCTG CAAGTAAAGG CCGGCGGGAC
    1741  GCGGAAAACA CGCGGCGACA CGTGCACACC GGCCGGTGCT GCTGTAGTCC TATTAATTCA
    1801  CACGCCGGCA CAATAAAGCG ATCAGCAGCT GTGCGTGGCG GCTCAGTCTG GCCTTCTATT
    1861  TGGCGGCGGG CGGCAGCAGG ATGTCCCGAA TGGCTTCGCG GAAGGCCGCG GGCTTGAAGC
    1921  CCAGGTCGGA CTGGAGGCGG GAGATGTCCA TGGAGATATC GGGGGGGCTC ACCACGCCCC
    1981  TGCGGCACCG GGGTGTGCGG CAGTGGGGTT ACATTAACGA TGAGTTAATG AATGGGTAGA
    2041  CGGTGGACAG TCGTGGGGAA CAACAATCGT TAACGGGTGG GGGCTGTGCG CGGAAGAGAT
    2101  GCGTGGCGGT GTAAGGGCTC GGTGGGTACA GGTATTAGAA AGGCGGCATG GTTACGGGAG
    2161  GAGCTGGCTG GCTGTCTGGC AAGTGGAGTT GTGACCAACA GCTTTGCTAC AGCGGCTGCG
    2221  GCATGTT

//

LOCUS          12     1852 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      12
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES             CHR  ---
     CDS             join(500..569,691..980,1224..1352)
                     /translation="MLDDLKDLPSKGPLRELAAIPANSTHDTAPPKDQVIHTDNSTA
                     ALIRIVRKGASKPQTGKGKQAAGVEAAPTTGKRTHEASAADGDERPTKKAAGAGTSG
                     AAAVAKLLLPIATFTAEQLKAKTIPQLKELLKARGLAVSGAKDELIRRIVDHQRANK
                     YAPAK"
ORIGIN
       1  GTAAATGCTA GGGAGGCCTC ACAGGGATGC TGCGATGCGC CTAGAGGTAT CACCTGTGCA
      61  GCATAAGGGG GCTGCACCCC TGGCCGCACC TGCTTGCAAT CATGGTACTG CAAGAAGAAC
     121  TCCACTGTCT GAGAGCGCCA GCGCGGGGGC CCCGCTGGCA TCTGCACTGC CTGAGCATTT
     181  GGCACCTGAC CTGTCAACTG GGGCATTCAC TTGAAGGAGC GCGGCAGGGG AGGCGGATAG
     241  CGGGGAGCAT TTAGGACGCG AGACCTGCGA GACCAACGTC GGGTGACCAG GGGCACCCAT
     301  GAGACCCACG CGTGCTCCGA TGGATCCACC GGCAGCAGCT GTCTGCTTTT GTCATGCATT
     361  GTAGCAACTG TTATATTCGT TACTCCGAAA TAATAACGTT AATCCGCACT CGAGTCCGCT
     421  GCACTCGTTA TATGCAAGGT GCTCCTCGGC ACCGTTGACA AGGGCTGGGC ATGCACCTAG
     481  AGTTACTCAA GTATAAAAAA TGCTGGATGA CTTAAAAGAC CTGCCTTCCA AGGGCCCCCT
     541  GCGCGAATTG GCAGCCATCC CGGCCAACTG TGCGCACACA TACATAGGTA AGTGGTCTGC
     601  AAAACGCGCA CTAGCGACGT GACATGATTA GCAGCAATTG ATGTGCGGTG TCCTGACATG
     661  AACTGCAATG CGGTTCCCGC CCGTTTGCAG CGACACACGA CACGGCCCCG CCGAAGGACC
     721  AAGTCATTCA CACAGACAAC AGCACCGCAG CGCTGATCCG CATAGTAAGG AAGGGCGCAT
     781  CAAAGCCACA GACAGGGAAA GGGAAACAGG CAGCCGGGGT GGAGGCGGCT CCCACCACCG
     841  GCAAGCGGAC ACACGAGGCT AGCGCTGCAG ATGGCGACGA GCGGCCTACG AAGAAGGCCG
     901  CGGGCGCGGG TACGTCGGGG GCAGCAGCCG TGGCAAAGCT GCTACTGCCG ATTGCTACCT
     961  TCACGGCCGA ACAGCTGAAG GTGGGTAGCG TGGTGCGATC TCCGTGTTTC GCACGTGTAT
    1021  GGTACCGTGT GTGGGTGCGG ATCCAGTGAG GACGGTTCTG CCAAAAGCCA CAGGGCCATA
```

```
    1081  CGCCCATACA TCTTCCCGAA TGCCTGGGCT GACGGCCCGC GAAATGATGC GAAGCGTATT
    1141  GCATTCGGTG CCCCTCTGGC ACGTTCCTCT CTCCACGCCA TCTCGTCTTT ACTCCGCTCA
    1201  CTCCACCTCC ACGACACAAA CAGGCCAAGA CCATTCCTCA GCTCAAGGAG CTATTGAAGG
    1261  CGCGCGGCCT GGCGGTGAGC GGCGCCAAGG ACGAACTCAT CCGACGGATA GTGGACCACC
    1321  AGCGTGCAAA CAAGTACGCA CCAGCCAAGT AGCGAGCCCT TAGCAGCCAT GTAAGAAGGA
    1381  CCATACTGCT ACCGTGCCGC GTACAGCGCC GTGCAGCCCG CCGCCCTTCA CGTACACGAA
    1441  ACAGCGGCGC ACCAAGTGGC ATACACCAGC ATTGCAGCTG TCCGAGTGCT GAGATCCAAA
    1501  GCTCTTGCCG CTCCGCGTTA GTTAGAGCAT GTTGCCACTC TGGCGGCTGA GCAGAGCCTC
    1561  GAACATCACT GTGGGCAGGG CAGGCAAAAG GAGCAGTGCT TATAATCGTG TCAGTCCGGT
    1621  AGCTGGGTAC TACCAACGAA TGTTTCATGC ATAGGTAATC TGGATGCATT TTCTGGATTT
    1681  GGGGTATAGA AAGACGAGCT CGAAGCCAAA GGAGCATGCG ACTCACTGTA CATTTCAGGT
    1741  CCAGCCCCGT GCATGCCGAA GATGTCCTTC GCCCGCCGGA GCGTTTCCTC GCGCGTTTGC
    1801  CGCCCAGCAT TGAGCTCCTT GATTTGGTGG AGGAACTGGC GGCACGTGCG TG

//

LOCUS          13     2827 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      13
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   C.reinhardtii
FEATURES             CHR  ---
     CDS             join(500..721,1146..1326,1784..1845,2082..2327)
                     /translation="MSKITGGKKASGVDNTARRTWDREEYRAIAEEKEKEKKEAAKA
                     RGAKDGDDDDHEETAADIRRRKRQERDPLHHGLIVERSLLKQRDYAIDLTSRLGKTQ
                     VVGFNTPLNQQAGWFCNVCNCVLRDSQSYLDHINGKWHNRALGMNMKVEKSTLEQVK
                     NKFEELKSRKSPPPDEYVPDGFDGAASAEAKEREEKREKKRRKAAEAAAAGGGGGAA
                     GGEMDPEMMQMMGFGGFGSTKR"
ORIGIN
       1  TCCAAACGTT GCACCACAGG CGCATACTGT AGAATACAGA AATACTGTAG AATACAGAAA
      61  TCACTGTGCC AAGAATCGGG CCAACTCCTG ACCAAGCTTC TGCGTTACTG CTTGACTTTA
     121  GCCCATCGTA ACATATTGAT GACAAGCGTC AGAATTCTTT CCCAGAGTTG AAGCTTAGCG
     181  AGGAGGGCCC GGAGTCTTCT GCGCTGCTTT TGTCACTGCT TGGGCTCGAT TCTTGAGAAG
     241  AGGTGAGCTC AGCAACACTT GCGACATGTC ATCTCCATGG CGCCTCACGA GGCGGAGCCT
     301  TCAGATGGCA GCGACCTTGC TCGCAGCTCC ATTTGCGTGC GCTGTTCCAT TCTTTACAAT
     361  CCCTTTGACA GAGCGCCAAC TCTAGCTGCG GAGCGAGGCG GGTGCAGTAG CTGCGACCTA
     421  GTCACCCACG ACTGGCGTGT GAACCCGCCT AGCGGCTCCC GTGCCACCGT TGCCACCGTT
     481  GCCACCAGCC GTCGGCGCCA TGAGCAAGAT AACGGGGGGC AAGAAGGCGT CGGGCGTGGA
     541  TAACACGGCC CGTCGCACGT GGGACCGTGA GGAGTACCGG GCGATAGCGG AGGAGAAGGA
     601  GAAGGAGAAG AAGGAAGCGG CCAAGGCCCG GGGGGCCAAG GACGGTGACG ATGACGACCA
     661  CGAGGAGACG GCGGCGGATA TCCGCCGCCG CAAGCGCCAG GAGCGCGACC CCCTGCATCA
     721  TGTGGGGGGC GTGGCGCATG GGGCGTGACG TGAGGGGCCT GGCGTAGGGG CATGGTTCAG
     781  CGCCAGGGTG CAGGGAGTGG GCAGGGCAGG CGGGAGGGCA GGTGAGGCGT GCGGGCTGCA
     841  GCGGCGTGTG GGGGCGGGTG GGCCGGGCTG CTGGAGCAGT CAGCCGCGGC AGCGGGTGAA
     901  GCGGCCCACC GGGACATCCA GCACGTGGAG ACTGGTTTAC GGGCGCGGCT GCAATCCCTC
     961  TACGCATGTG GCACCTGTGC CCACCTCACC ACATGCGCCC CACCTCCTGC GGGCCCCAAC
    1021  CACCTCGAGC GTCCCCACCG CATGCGGCCC CCACCAACCA CCTCCTGCAA CCCCCTTCCC
    1081  CAAATTGGTT CACAGTTAAC CCTCCCCCAA ATTGGTTCAA AGTTAACCCT CCCCCAAACC
    1141  TACAGGGCCT GATCGTGGAG CGCTCGCTGC TCAAGCAGCG CGACTATGCC ATCGACCTCA
    1201  CCTCGCGACT TGGCAAGACG CAGGTGGTGG GATTCAACAC GCCGCTCAAC CAGCAGGCGG
    1261  GCTGGTTCTG CAACGTGTGC AACTGCGTGC TGCGCGACTC GCAGAGCTAC CTCGACCACA
    1321  TCAACGGTGA GGGCGTGCGG GGGCGGTGTG TAGGGAGATA AGAAGGAAGA AAAGGGGATG
    1381  GAGGGAGGGG AATGGAGACT GTGCAGCAGG AGCAGGGTTG CGGGAGGGTG GCTGTGGGGT
    1441  GCGTTTGTAG CGCCGCCATG AGGTGAGCAC TGTGAGTCTT ATGGCACCAA GCTGCAGGGA
    1501  GGAGGCTAAG AACGCATCAC GGGGGCGTTT GCTAGCACGT GGGTGACATT GGAATGGCGG
    1561  AGTTGCAATG CGGGCAGCAT GGTGGTACAA CCGCGTGGCT GGGAGGGCAC CATCCGGCAT
    1621  GGCTCGGCTC TGGAATGGTG GGGTGGCCCG CGAGGTCACT CGTGCAAAGC CACGTGCGTT
    1681  GTGCTCATGC AGTCCTCTGC CTTCTCACCC ACCCCCACCC ACCCACCCCA CCCACCCTTA
```

```
    1741  TCATGAATGT GACCCCTTTC CTTACCATTC CAACCCCACG CAGGCAAGTG GCACAACCGC
    1801  GCTCTGGGCA TGAACATGAA AGTGGAGAAG TCCACACTGG AGCAGGTGCG TGCGGCAATG
    1861  TGATGGAGCG CGCGTGGCCT TGTGAGGTGG TGCTGTATGG AAACGTTCCA GGCTTCTGGC
    1921  AGGAGTAGCC TTTGTGCGGT GTTGAGTGGA GCAGCGCGTT ACGGCACTAG GCCACGCAAG
    1981  GCTTGACGAT TGGCATGATG TTGGCGGCAG CGCGGTATGC CTGACATGCG GGCCGAAGTA
    2041  GACTGTTCCA ATGTCCGTAC GCATCGCTTT TCTTCCTGCA GGTGAAGAAC AAGTTCGAGG
    2101  AGCTCAAATC CCGCAAGAGC CCGCCGCCGG ACGAGTACGT GCCTGACGGC TTTGACGGCG
    2161  CGGCGTCGGC GGAGGCCAAG GAGCGGGAGG AGAAGCGCGA GAAGAAGCGG CGGAAGGCGG
    2221  CGGAGGCGGC GGCAGCGGGC GGCGGTGGCG GGGCGGCAGG CGGTGAGATG GATCCCGAGA
    2281  TGATGCAGAT GATGGGTTTT GGCGGATTCG GGTCAACCAA GAGGTAGACG ACGGGCCAGA
    2341  AAGCGGGATT ATGGATAAGA ATGTCGGTGG CTATGCGACG GGAGGAGGTA GCCTTTGGAG
    2401  TGGGAAAGGT GCGCGCAAGC TATCGGATGC CTGCCCAGGG CCCGCATGCA CTGTTGACAC
    2461  GAGCATGAGG TGTGGTGTCG TTTTGGGGTA ATGGATTGGG AAGTAGACTG GCCAGGCATC
    2521  GGATGTGGCA TCGACAGCAG CAGAAAGGAG CAGGCGGCGC TTTCAAAGTC GCAAGCTGAT
    2581  CTGGGCTTGC GCATGTCACG TCTAATGAGC ATGGGTGCGG GCAGGTAGAA AGGCCGCCGT
    2641  GATGATGGTC GGGGCTTTCA AACGGTCTAA TCTGTACGCT GTGCAGCTGT GTGTAATGCA
    2701  CACCGCACAC GGTCCTGAGT CCTGACAGGG CGGCAGCCGA CAGCTGTGAC ACTGTGACGT
    2761  GTCGGGTGGG GGACTGGATG GGCCTCCTCT GGGCGCCGGC ACGCCGCATG CGTGCACCCG
    2821  GCGCGCT

//

LOCUS          1     1279 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      1
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   D.melanogaster
FEATURES             CHR  ---
     CDS          complement(join(500..686,745..779))
                  /translation="MVNVMKGVLVECDPAMKQFLLHLDEKLALGRKFIIQDLDENHL
                  FISTDIVEVLQARVDDLMDRISFPLHDKDA"
ORIGIN
       1  ACTTGAGGCA GCGCGTGTCC TGGTGCGTTT GCAGGCATCT AACGCAACAG TATCGCGTTC
      61  CGCAGGCGGT GCACGAGTAC AACGAGAAGT TGCCGCAAAC GGCGCAGAAA TGCCGCAGTG
     121  GCTTTTGTGG AGCCGGCGCT GCGGCGCTCT CGTAGTTGGG CTGCTTGTCC TTGTCCTCCT
     181  CGAGCAGTTG CTGGAAGTTC TTGCGATACT TGACGAGAAA GTATTCGGCG CCCTTGCGCT
     241  TGCCTTTCTT CTCCTTGCCG GTCTTCAGGC TGTCCTGGAA TTTGGGCAAC TTCTTGGACA
     301  TGACAAGGTC GGCGTGTGGG TCATCGTGGT AGTTGTCCTG CTCCAGAGCC TCCAAGGCTT
     361  TGCGGGCACG CCGCTGGCGG GCTGTGGAGT CCAGGACACG CTTCTTTTCA GCGTCCTTGA
     421  TGCGGTTGGA TTCGCGACCC GTCATGGTTG TTTTTAGGTT TTTACTTTTG GCTATTGGAT
     481  TTTATGGGTT GTGTAGGTTC TAAGCGTCCT TGTCGTGCAG CGGAAAGCTT ATCCGATCCA
     541  TCAAGTCGTC CACTCGCGCC TGCAAAACTT CCACAATGTC CGTGGAGATA AACAAATGGT
     601  TCTCGTCCAG ATCCTGGATA ATGAATTTGC GTCCCAGAGC CAGTTTTTCG TCCAAGTGCA
     661  GCAGGAATTG TTTCATAGCA GGATCACTGC GGGGATTGGG TTGAATATAG CGAAACCAAC
     721  GCATACCAAA CGTATACCAC TTACCACTCC ACCAGCACTC CTTTCATAAC ATTTACCATG
     781  TTGCCAGCAT TTGTTTGCCG AAATTAATAA ACAATCTCAC CAATTCTCAG AGCTGAATGA
     841  CATATTCACA GTGAAACGTA GGCTAGAGGT GGCAGTCTTT ACTATCGATA TTTTTTTGAC
     901  TTTACTTGAT AATTTAATCA ATAATTTCAG CATGGCAATA TAATATTTAC AATCAACGAA
     961  ATATAACACA TTGCTTTAAA CCAGCACCCG ATATTAACTA TATTATGTTA GTTTCACCGT
    1021  CGGAAGTTTC GTACTTTAAG AAATATAGCT ATTTTGCCAG CCCTATCAGT GCTACTCACT
    1081  ACGAAAAGTA GACAGAAGTA GCGGGGTATT TGCCAGTCAC ACGGACAAAC ACCAAGGCAA
    1141  AAGTGTGAGG AAATATTGAA TTCCAAATTA AATTCCTAAT TAAATCAGAC CATAAGCACG
    1201  CAATGGAAAG GCTGACCGGA AGGAATGTGG CACTACTGGT GCTGTGCCTG TGCGCCGGAT
    1261  ACGCCCTGGT GTTCGCCGA

//

LOCUS          2     1352 bp     DNA    linear PLN 25-FEB-2004
```

```
ACCESSION      2
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   D.melanogaster
FEATURES             CHR  ---
     CDS             complement(join(500..634,703..852))
                     /translation="MTNLQRWLFYASLFAIPYLSVVLGTVQTPLTTKYFLHIQLLPL
                     LLLVIFGIYSVWTVLYRTLTFNDCPEAAKELQDEIQEARKDLIAKGFRFRD"
ORIGIN
       1  TTTTGCAGAA GAGAAAATTA AAGTTTTGAA ACAGTATATT TTTTTCTCTT TCGCATTGTA
      61  ATTAAGAGAA TTTTAAAAAA ATTGATTTAT TGGAAATAGA TAGCCAAAGC ATGGATTTGC
     121  AATATATGGT TCGCACTTAC GAGTTTTCCA TATACCATGT TCATTAAGTA AAAATATGTT
     181  TTCCGTAAAG AATATTATTA ATCTCTTTAA ACGTATTTTA CCTCATATAC TTTTTTAGTT
     241  TTCTTATATT TTTTATTTTG AGTAATTCAC AGGGGAGCTA AGCCAATTAA ACATAAACAT
     301  ATTCCATACT TTTTCAAGCA TATCCTTTCA AAACCTCAAC ATAATTTGGC CACCAAACTT
     361  CCGACACCTG GGGAAATCAG CCCCCAACAA CGCGTCCGCC CACACAAAAT GCGGTCTAGA
     421  GTAAATTTGC GTAATTTTTG TTTTTATAAA TAATGGAATT ACAGATTAAC ATGCACAAGT
     481  CCTACGAGCT GGAAGTCTCC TAATCTCGAA ACCGAAATCC CTTGGCTATC AAATCCTTGC
     541  GAGCCTCCTG AATTTCATCC TGCAGCTCCT TGGCGGCCTC GGGACAATCG TTAAAAGTCA
     601  GAGTTCTATA TAGAACAGTC CAAACGGAAT ATATCTGTGG GTAATATATA GTAGTAGTCA
     661  ATAAGAGGTG TTATCTGCAT TGATTCATAT CCAAAAACTC ACTCCAAAAA TCACGAGGAG
     721  CAAAAGTGGT AAAAGCTGAA TGTGCAGGAA ATACTTGGTA GTTAGTGGCG TTTGCACTGT
     781  TCCCAAAACA ACGGAGAGAT AGGGAATCGC AAAGAGCGAT GCGTAAAATA GCCAGCGTTG
     841  TAGATTTGTC ATAGCTAACG AAATTTATTG GTAAACAAAA CTAAACCGAA AATAAAGTTC
     901  TACACGTAAA CATAGAGCTG GAACTTAAGC CGATGGTCAC CCGGTAATTC GATATCGATA
     961  TTTACCGGCT ACTATCGATA GGAGGTTTAC CTATTCAAAT CTAGTTAGAG GTTTTCCTAT
    1021  TCATGTCTAG TTGGAGATTT TCTAATTAGT AATTCGTAGC CTGACGGCTT CCAGTTAGAC
    1081  ATTGTTTACT TTTCTAATTG GCTAACTTAA GTGTGACCAA GCCCTAAGCA CGTGCCATAG
    1141  CACAAAGCTA TCCACAGTCA GCAAATCAAA TGCCATACTA CTCAGCTAAA CACAGTCAGC
    1201  AATTTATCGA TAGATCAAAG TTGCGCCATA ACAGCCAGCC TTCTACAGCC AGCAATTTAA
    1261  ATCAGAGTAA TCGATTGACG CAGTTTCAAA CTTGTTAAAA CTAAGTTTTA ATTAAACGTT
    1321  GGTGGGCGTG GTTTCTGCGG CGCCCACTGT AC

//

LOCUS          3     1305 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      3
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   D.melanogaster
FEATURES             CHR  ---
     CDS             complement(500..805)
                     /translation="MAGTKVLRSLLHELRQASPNGCIKDSLAARYILAQYKKFATTE
                     QQFCKARNEATFLGQTYLTYLASQRRYLELYKEYHGRGERSVRDTADLVGFKLPSDP
                     K"
ORIGIN
       1  ACCTTGGAGG GCTTCACCTA TGTGATCTTC ATCGGGACTT CTGTGATGCA GGTCTACCTG
      61  GTGTGCTATT ACGGTCAGCA AGTTCTCGAC TTGGTGGGTA ATGGCTGAGA TCCTTGGTAT
     121  GCCCACTAAC TGAGGTTCTT CGCAGAGCGG CGAGGTGGCC CACGCCGTGT ACAATCATGA
     181  TTTTCACGAT GCTTCTATAG CGTACAAGAG GTACCTGCTC ATAATCATTA TCAGGGCGCA
     241  GCAGCCCGTG GAACTTAATG CCATGGGCTA CCTGTCCATT TCGCTGGACA CCTTTAAACA
     301  GGTAAAATGC AAACACCACG TATACTTTAT ATGACATTAT GTGAATTTCG CTTTCAACTG
     361  CAGCTGATGA GCGTCTCCTA CCGGGTTATA ACCATGCTCA TGCAGATGAT TCAGTAGACA
     421  GCAAACACAA GCACGGAGTT TGCAAAATCG GGACGCTTTA ATCGCACTAC GCTAGTAACT
     481  ATCTTTCTAT ATAACTAGAT CACTTGGGAT CCGAGGGCAG CTTGAAGCCC ACCAAATCGG
     541  CGGTATCCCT TACGGATCTC TCGCCCCTTC CGTGATATTC CTTGTAGAGC TCCAAGTACC
     601  GGCGCTGGCT GGCTAGGTAG GTCAGGTAGG TTTGACCCAG GAAAGTCGCT TCGTTGCGCG
     661  CCTTGCAGAA TTGTTGTTCC GTGGTGGCGA ACTTCTTGTA TTGCGCCAAA ATGTAGCGCG
     721  CGGCCAGAGA GTCCTTGATG CAGCCATTTG GAGATGCCTG GCGCAATTCG TGCAGCAGGG
```

```
     781  AGCGCAGCAC CTTGGTTCCC GCCATTCTAA ACACAAAATA TTAATTGCAT TTGCAAGAAA
     841  CTATAAACAA TCAACGACTT ACGTGATTTT TAAGAAAATT CCAATTGCTT CTGCCGAACA
     901  GCTGTTCTCA ACAGAGTTGC CACCTAGTCA TACAATGAAT ATACAGTGAT CACTCTCACA
     961  AATTGACACT TTGAAAAATA ACACTGGATC TTATTTTTTA TAATTATCAA TAAATCTGAA
    1021  AAATCATCAA AAGTTTGACT ACGAGCACAG TAAAACTCTG TGACAAACGA AATATTTAAA
    1081  ATTTATTTAT TTTAATTATA TTATATTCTC ACTTTTGAAA TCGTTTTGCT ATAAATGTCT
    1141  AAAAGTCGAA CGGTCTGATT GGAAAACAAG AACAACCAAT AAGTAAAGTA CACAATCCAA
    1201  CTGGACTGTC AAACTAAAAA TACAAACAGT AAATCCAGTT GGAAAAAACC AATTTATTTT
    1261  AAAAGTTATT TAGCAGTACA CCGAACTCGT TGCGTGGGCT GTTTA

//

LOCUS          4     1366 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      4
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   D.melanogaster
FEATURES             CHR  ---
     CDS          join(500..614,670..866)
                  /translation="MNDIKRLKDQQREKHPGFDGYMDCMTRSLFTGLATFCLSFSGT
                  YFAQKIVQSKIRYPIKYNILISSLVATGVSYQITSTRTKACQAAWMAFEDKHSVLKE
                  KTF"
ORIGIN
       1  AAAATATCAC TAGATCGGTT TGTTCTGCTG TGTTCACCTG CTACTGGTTG TGTGGCCGCT
      61  TCATCGCGTC CTGCCACTCG ATATCCACTC CTTTTTGCGA GCCTTTAATT GTTATTTTTC
     121  AGCGTGAAGA CGCCGCAATT CCAACTTGGG TATACTAAAT TCGTAAGGGT TGCCAGGCCA
     181  CCGGACGAGC TGTTATCGGT GGCTTGACTC ATGCGATAGT ATCGAACCGA GGCGGTTGAA
     241  ATCAAATTGT TCAATATGCG CACTATCATA TTACTTATAT GAAATACTAA GTAAAGAAAG
     301  TTAATATTAT ATAAGCATTC TCTCCTTAAT CGTATTTTCT TTATTACAAA CTGTAACGTG
     361  AATAATTTTA AAACATAAGA TATTTCCCTA TCGATACTAT TTGTAATCGC TTAGCCACCT
     421  CTGGGACAAA CAGCTGATTT CGCGCACCGG CTAAAAGCCA ATGAAAATAA ACGCTTAATT
     481  AGTTCACCGG CCGTTTATTA TGAACGATAT TAAGCGATTG AAGGATCAGC AGCGCGAGAA
     541  GCATCCCGGG TTCGATGGTT ACATGGACTG CATGACCCGC TCCCTTTTCA CGGGGCTCGC
     601  CACATTTTGT TTAAGTGCGT AATTAGCTTC AGTTTTTCGA TAATTGCATA ATAATTGCTC
     661  AAACTACAGG TTTTTCCGGC ACCTATTTCG CCCAGAAGAT AGTACAATCA AAAATTCGCT
     721  ACCCAATCAA GTACAACATA CTTATCTCCT CACTGGTGGC CACCGGCGTG TCCTACCAAA
     781  TCACATCGAC ACGGACTAAA GCCTGCCAAG CGGCGTGGAT GGCCTTTGAG GACAAGCATT
     841  CGGTGCTCAA AGAGAAAACC TTTTAACACC TTCAAAATTC CATTTAGCTT TAGTCATGTC
     901  TGCTCTACGA ATATTGGCCC GATTTGTCAG CACTAGGACT CTGCGAATTC GTCCGATTCT
     961  GGAGATTCCC GCACGCTGCA AGTACACCCA ACATCAAGGA GTTAAGGGCA TTCGGAACAG
    1021  GAAGAGCTTC TTCGAAGCCC AAATTCAGGC AGGTCAACAA GGTGAGGAGG ATGCGGACTT
    1081  AGAGCGGACA GCCAATAGCG ATCTGGCGGA TATGCGGTTC TTGGATGATC GGTAGGCTTA
    1141  AAAATCAAAG AAAAAAAAAG AGATATTACT AACCTTTTAT TTAAGAGCCT ACGACGAGGT
    1201  GGCTGGAGGA GCCATGAGAA TCACACGTGA CATAGCCAGC TCTCAGCAGG TGCTCATCCT
    1261  CCAACCGTAT GTAAAATGGG CAGCCAAACG ACAGAACGCA CCCGTCGATG TTCGGCCCGA
    1321  GGATCAACTC GCGGAGGCCA CAGCGCTGAT CCACTCCCTG CCCAAT

//

LOCUS          5     1388 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      5
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   D.melanogaster
FEATURES             CHR  ---
     CDS          complement(join(500..793,847..888))
                  /translation="MFDTREETSWNSDEGTPQFITISLEEPQKVAGFSFQFQGGFSG
                  QKSELIMYSADGAQVHQEPFYPEDINSPQLFQIAESVRENPCSKLKFVFESSTDLFG
                  RIIVYDLQLFG"
```

```
ORIGIN
       1  ATAAATGAAG CATTCGAATC ATAAAGCTAT GCCTAAATTA TCTGTGCGGT TTTGGACAAA
      61  TAAACAATGG TTTTGGTGCA ATATACATAA ATGACGAAAT TTCAATTGGA ATGGGCTAAG
     121  CTGATTGATG ACGCTTCTAT ATGGGGTGAT AAGACATGGA CTAGAGTGTC ACCTTCGCTA
     181  TGAAGTTGCG TGTTTTTCGA GCCAGCTCCA TGCAAATGAT ATTCAAATTG GGGAATCGCG
     241  TGGAAATATT TTCATTGAAC CTTGATCCAC ACTCGTAAGC ACTTAAAGGG CTTACCTAGT
     301  TTGAAATATT TATATGAAAT TAGTCGCATA TCAATTTCAG CAGTGGCAGG CTGAATCAGC
     361  ACAATTATTG CTGATAACCA TGACCATCCT ATTACGGACC AATGTTTTTG ATAATACCAC
     421  AATAGATAAT GATCCGATTT TATACACTTT GATTAGAGAA GATTTTGTAC ACATTTAGTA
     481  AGATCTATTG GGTCAGAGAC TAGCCGAACA GCTGGAGATC GTAAACAATG ATTCTACCAA
     541  ATAGATCGGT GCTGGACTCG AAGACGAACT TTAGCTTTGA GCAGGGGTTT TCCCGGACTG
     601  ATTCCGCAAT CTGAAAAAGC TGTGGGGAGT TGATATCTTC TGGGTAGAAT GGCTCTTGGT
     661  GAACTTGGGC TCCGTCAGCG GAGTACATGA TCAGTTCGGA TTTTTGGCCA GAGAATCCGC
     721  CTTGGAACTG GAAACTAAAT CCAGCTACTT TCTGTGGCTC CTCCAAGCTA ATGGTAATGA
     781  ATTGGGGGGT TCCCTGAAAT GGCCGTTTGG AGTTATATTC TTTTTTTCCA AATCTTAGGA
     841  TGATACCTCA TCTGAATTCC AAGAAGTTTC TTCCCGCGTA TCAAACATGA ATTGCTTTCC
     901  GTGCTGCTTC ACATCTCTGT TTAAAACCGA GCTCACCCTG CAAACTATAC GTTAACTAAA
     961  GGGAGTTACA AAATAAAACC CAGAGTACCT GCAGCTGTAT TTTACTTTTG TCAAAATATT
    1021  CATAATTCAA TAAAAATAAG AAATAAACAA ATCGGAGACC GCGAGGTTAA AGGCATTAGA
    1081  GATGGGCAAC AATCAGAAAT CCATAGAATT TTTTTTTGTA CGGATTTTTA ACACCCGATT
    1141  TAAGAAATTG CTTTAAAAAA TATATTTTCC CTAACTTTCT AAAATATGCA TGATTCTTAT
    1201  TAGATTTTTG AAATAAATAC GTCAGATAAA GATATATCGA TATTATAGAT GCTATACCAA
    1261  GTATTGATTT TGCAGCCCTC GTCGAAAAAG CAAAACAGCT GGCAGTTTAA ACATTTTCAT
    1321  TTTTTTTGTT GACTTATCTT AGAAGTTGAG ATGATAAAAG CCATTCTCGT TTTTAACAAT
    1381  CACGGGAA

//

LOCUS          6     2736 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      6
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   D.melanogaster
FEATURES            CHR  ---
     CDS          join(500..581,1714..1817,2039..2236)
                  /translation="MPYIIIRGNLASYSHKYPWRVLVSGLKADDIEQLNKFSCGGYS
                  DESTIVYLVHPCRILSALEILGFRVVASSSTAVKQDYNEYMWTMRKEFDEPEPLEAE
                  SVVRENLSNIGREAASLGNYHKVDSPE"
ORIGIN
       1  GAAAGAATGT CCAGTAATAT CTGGTTCTAG AATGTCCCCT GAACCATATA GCCATCTAGG
      61  TATATGTTCC GCCTTGGCCA CTGGGTTGGG TCATATCGTA CATATATCGG TCGAGTGATT
     121  ATCCTGGCCA GAGTGCGTCA CTTTGTCATT TGGGTATCTG CCACCTTACG CCGCAGGCTA
     181  ATCCAATAAA AAAAATAGCT GCTAACCATT TCAGAGAGCA GCGCTGCGCG ACGCCCAAAA
     241  GCGAAATAAA AATACGAAAC CAAACCAAAC AAAATAGCCA AAGCAGTAGA GCTGAGCTAG
     301  CGCTGGGAAA ACCTCTTGGT CAACAACGTG TTACCCGATT TCGGTTTCAG TCGAGAGTCA
     361  GCCTTCAGCA GAGCGCCACA GCGAGCATTA GCTGCCTCAC GGTTTCGTAT CTCATTTGGT
     421  TTTCACTTAC GTTTCACGGC TTTCGTTGGC ACGCCCGCGC TAATTAATCC GCAGATATAT
     481  TCAGCAGTAA TCGGGCCAAA TGCCGTACAT AATCATACGG GGGAATCTAG CCTCCTACAG
     541  CCACAAATAT CCATGGAGGG TGCTCGTCTC CGGGTTGAAA GGTAAAGCCT GAACCGCGTA
     601  GATACAGTAT CTCTTTTTTT TTTGGTTTCG TTTCGCTTTT ATTTGTAGTG TGTTTACCAT
     661  TGCGACACAC ACAATTCATT TCGAGAGCTT GGCAAACAGA TACGAACACA CGATAGATTT
     721  CTGAACGCTC TTTTCGGTTT TTCCACTTTG CACCACAAAA AAAAAAAAAA AAAAAACAAA
     781  TTTTTGAATG GTTAGACAGT GGCTGTATAA ATATTATATA CATATATTAT TTTCAAACTA
     841  TTGTTTCAAT ATCTAGCATG TATTTTGCTC AAGCTTGGAA ACATGACCAA ATGCACGCCT
     901  ATTTGATGTC ATAATAATAC AATGTAATGT TTTAAAACTT TATCTAGTTA TATTCTTTAA
     961  AATACTGAAA GCATCTTGAA AATCCAATCC TATACTTATT TTTTGACATT CAAATTCTAC
    1021  TCCCATTTTG AATTGTAAAG TTACAGTAAA TCTCACTGAA ACAATAATAA AAAAAAAATA
    1081  AATAAAAAAA TCTACGGCAA GCTATTTTTA TTTGCTGGTT ACCCAATAAC CATTAGCGCT
```

```
    1141  GTGCCGTAGT TATTTTCCTC ATAAAATCAA CAGCTGCCCT TTAATTGCAC TCCATTTGCC
    1201  CCTGGAAATC CGCCTGCCCG CACCATTCCG CTTTTTTCGA CCGCCACTGT GCGAAGTTTG
    1261  GTGCCCGCAG ACGAGGTCTT CAGTTCGCCA GTTTGGGTTT GTTTTTCGCT CCGTGCAGTT
    1321  CGCCTGGCAG AGCGGCCTGT CGTGCTTATA TTTGTAGATA TCTATATATA CAGCCGTACG
    1381  TAGTGCCACA TATATGCACT CGATTCCGGT GCATATTTCT GGAAGTAGTA TCTGTGCGGC
    1441  GCAGAAGTCA ATGCACTGGG ATCGTCTAGG GACAAAGAAA ATACGATACT CCCGAGATAC
    1501  TCCCATTTTA CGTTTATTAT GAAACAGCAG TCAATATTTC TTTCGCCATT TCGCTTAATT
    1561  AGCACACTCA CATTGAAAGA CAATTGCCCT GGGCTGACGT CACTTTTCGA CACAGCCGAG
    1621  TTTATCTTTT GTTTGCACTG TTTAAACATT AGCGAAAAGG AAAGCGATTT TATCAAAAGT
    1681  AGCACCCATG ATAACCATTG CCCATTCTTC CAGCTGACGA CATCGAGCAG TTGAACAAGT
    1741  TCTCCTGCGG CGGCTACAGC GACGAGTCCA CCATCGTCTA CTTAGTGCAT CCTTGTCGGA
    1801  TTTTATCGGC GCTAGAAGTG AGTACACCTT CCAGATCCCA TACCGACCTA CTCAGAATGA
    1861  TATTTTCTTA TAGACACACA TTTCGCTATC TAATGAGACA GCAGTTATTT TATCAACTTA
    1921  AACAATTCGC CGAGCAGGCA CTGATTAATT ATTTTAGTAA AACTTTCAAC CCCACTTCCC
    1981  TTCCTATTGA ACTTTCAATT ATCCAGTGAC TCATCGATTT CTGCATTTTT CCTTCTAGAT
    2041  CCTGGGATTT CGCGTAGTGG CCAGCTCATC GACTGCCGTG AAGCAGGACT ACAACGAGTA
    2101  CATGTGGACG ATGCGCAAGG AGTTCGACGA ACCAGAACCC TTGGAAGCCG AGTCCGTGGT
    2161  GCGAGAAAAT CTATCGAATA TTGGCCGCGA GGCAGCCAGT TTAGGCAACT ATCACAAGGT
    2221  GGATTCGCCT GAATAGACCC GCATCCCATC CAGTCCAAAA CAAACAAAAA AAAAAAGAGC
    2281  AGCAATCAAC CAAGAACTCT ACTCTAGTTG TATAATTTAG TGTGTGTGTG TGTGAATTGT
    2341  GTGCGTGTTT CGGTGTGAAA AAAGAAAAAA AAAAAATTAT AGTAACCTAG GGAATTTTGT
    2401  TCCATAACAT ACACTGAGCA AAGGAGGCGA CCAAGCGAGT CCCTTCAGAC GAACGAACTA
    2461  TGAGTGAGGG CGCAGTCGGG CTGAAGGAAC CTCAGAATCT CAATAGCGAA GTTGTTTCGA
    2521  CTGTCGTGTC CGCGGCATCG GCTGCATCCT TGAACGCCGC CACCGCCAAC GTACCGGCCA
    2581  ATGTGGTGAC CATACCGGTG CCCATTCTGC AGTCGGAGGG CAACTTTGCC TACGTGACCG
    2641  TCAAGGGCTC GCTGCACGAT TACACCTGCA CCATTTTTGG CCTGAATCAA GCGGAAGTTC
    2701  AAGCCCTGTC CAAGCGCTTC GAGAGCGGTG TTAAGG

//

LOCUS          7     1587 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      7
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   D.melanogaster
FEATURES            CHR  ---
     CDS          join(500..520,572..715,776..857,921..1087)
                  /translation="MSNLRRQVLSAFKKLHRTRQYVFQGDANALAAGRLKINESFLQ
                  NRNESSEDEIQKMIKLAQDVDLELRTNVIQAQKKEDGVYELRITPETTRLDNVVFNP
                  DAIIEKPRRQRGDKNTEGCCGGAAMAALEAEVQARNK"
ORIGIN
       1  AACAGTCCTT CTTAGTAATG CGCTCCAGGT TACAAATTTT CTGAATTAAA GTCGCTGTGG
      61  CGCTGCACTG TCTGATTTCG CACGGTAATT GTGATTAATT TAGTTAACTT TTTGATGCCG
     121  CATTGCCTTT CGACCCGTTT TGCGCTGGTC ACTAATTTTC TACTAGCATT CGAGCCGATT
     181  TACGACCGAA ACGTTAAAAG ATTTCTATAA ACTGGCGTCA GATTGACCGC ATCTGCTAGA
     241  GGTGGACAAT TGACGATAGG TCGATAGCTT TCGATGTATG TGCTTACGTT TAGAAAGGAA
     301  AATATTGAAT CTCTAAAAAA AAAAAAATTG AATTGGAAAG AAAATACTGA ATTTAAAATT
     361  TTTATTGACC TCTATTATTA GAGAAAAAAA TATTATACCA ATGCAATATC TTCAAGTTAA
     421  GAAAACTGTA ACCCTAAAAA CTATCAATTG GTTCAGACAT TCTGGCAACC TTACATCTGT
     481  CAGCTGTTAC CATACAAAAA TGTCCAATTT GCGGCGCCAG GTAAATTAAC TTAAAATCAT
     541  AAATAATCGG TTAATATTGT TTGTATTCCA GGTTCTCAGT GCTTTCAAGA AGCTCCACCG
     601  TACACGACAA TACGTTTTTC AGGGGGATGC CAATGCTTTG GCGGCGGGAC GACTAAAAAT
     661  CAACGAATCT TTCCTGCAGA ATCGTAATGA AAGCAGCGAG GATGAAATTC AAAAGGTTTA
     721  ACCACTTTTT GTAAATTCTA ACTAAGTACA TATATGGCAA TTTGGCAAAT TACAGATGAT
     781  CAAGCTGGCG CAGGACGTGG ATTTGGAGCT GCGAACGAAT GTCATCCAAG CCCAGAAAAA
     841  AGAGGATGGT GTTTATGGTC AGTGGGGCAT GGCGATTACA TAAAAGAGCG CTCAAAATTA
     901  AAATTTGGCT ATATTTACAG AACTAAGGAT CACGCCAGAG ACAACGCGAC TCGACAATGT
     961  GGTCTTTAAT CCTGACGCAA TTATCGAAAA ACCGCGCAGG CAACGCGGCG ACAAAAACAC
```

```
    1021  TGAAGGTTGC TGCGGTGGCG CTGCGATGGC GGCGCTGGAA GCCGAAGTCC AGGCCCGCAA
    1081  CAAATAAAAA TATATTCAAT ATTTAGCTGT AGTCAAGATG GTTTTGTTCT GATTAATTAC
    1141  GTCGTTTATC TTTAATTTCT ATTATGTTTC TTCTTCGATT TGTGACTACT CTTTGACTTG
    1201  GACTTCTTTT GCTTGGATTT TGTACTCTTT TTGCGTTTGC TGCGATCGTC CGACGATGAG
    1261  GAGCTGCTAG AGCTGTCCTC GGAACTATCT GAAGATGTGG AGTCGCTTTC GGATTCCGAG
    1321  GAACTTTCAC TCGAATCTGA ACTGCTGCTA GCTGCTCGTT TCGATTTCTT GGCTTGCTTT
    1381  TTGCTCTTTT CCACTGAAAG TTTTCTTTTC TTCTCATGAC TCTTGTTAGG GCTTGAGTGA
    1441  GTGTCCATTC TATGGTCTTC CCGGTTCCTG GATTGTTGCC TATAAAAAGA TAGATAGATA
    1501  GTTAAAGATT TTGTATACAT CCAGATGCTT AACATTTTTA CCTCCGCTCT GGCGATCTTC
    1561  TTTGATTCCT GGAGGTTTCT CTATGCG

//

LOCUS          8     1544 bp     DNA    linear PLN 25-FEB-2004
ACCESSION      8
VERSION      GeneMark.hmm v3.0 self_training
ORGANISM   D.melanogaster
FEATURES            CHR  ---
     CDS          join(500..562,620..752,806..1044)
                  /translation="MADLVSGEASESDEELQEKNKIFAAEISGEASEEDDEDIYEQG
                  RSSNQVNSTASIYRNNLLQRKLIENNIAIWRSLTAFTRSFVLSASKQLVTTDQMLIK
                  SQHSLQSAHTSLQQAQKNAQDLQNRVADVITSNFLPHISIQKAT"
ORIGIN
       1  ACGGTGGGTG CACTAAATAT TTTACATTAT AAAGATTAAT TATCATTTGT AACCTGTGTA
      61  CTGTGTAGGT TCCCCGCGTC TTTATCGATG GCAAATTCAT TGGCGGTGGC ACTGACATCA
     121  AACGAATGTT CGAGACAGGA GCTCTGCAAA AATATTTCCA ATAAATAGTG TTCTTACTCT
     181  GTTCAGAACT CCGTTTTATG TAAAAAGTTT GTTAATCTTC AAAGTCCTTA ATAAATGCTT
     241  TTAAAATTTA TGTTTTGCGC TTATTGAGAT TACATCGTCT TCTGAGTAAT GGGCAAAGTG
     301  TTACTAATAC CATTCATTAT CCGTTAATAA TTTTTATGAT TGTAGTTCTA CGGATCTGTT
     361  AGCATAGATA TCACAAACAT AAGAATATCC CCTTCAACTG CGCCCATGTG AGCAACAGTT
     421  TGCAGTTTTA ACTGATCTGG CAGCGCCGTA ACTGCAGTAA ATAAAAAATT GTTTACAATA
     481  GTATTTAATA AACTGAATGA TGGCTGATTT GGTGAGCGGT GAAGCTTCTG AATCCGACGA
     541  GGAGCTGCAG GAGAAGAATA AGGTTGTTGA GGATATCATT TATAAATAAC AGTCCAATAA
     601  TAAAACCATT TTGTGACAGA TTTTTGCTGC TGAAATTTCG GGCGAAGCCT CCGAAGAGGA
     661  TGATGAAGAC ATTTACGAGC AAGGCCGAAG CAGCAACCAG GTCAACAGCA CGGCTTCCAT
     721  TTACCGGAAT AATCTGCTTC AGCGGAAACT AAGTATGTCA CCAATTCTGT GCCTTCAAAC
     781  ATTTCAACTT AATGCCCTTT TCCAGTCGAA AACAACATTG CCATTTGGAG ATCCCTTACC
     841  GCTTTCACTC GCAGCTTTGT GTTGAGTGCA TCAAAGCAGC TCGTAACCAC TGATCAAATG
     901  CTCATCAAGT CGCAACATAG TCTTCAGTCG GCTCACACAT CCCTGCAGCA GGCCCAGAAG
     961  AACGCCCAGG ACCTGCAAAA CAGGGTGGCT GATGTTATTA CCAGCAACTT CCTCCCGCAC
    1021  ATAAGCATCC AGAAGGCAAC CTGACTTAGA CGTATAGTGT ATTTGCTGAT AAACGATCAA
    1081  ATCTAGGCTA ATCCTCTGTA TCGGAATTGG GCGCATAGAC CTGAGTGTTG GGTGAGGGCT
    1141  CCCGCGGCAC GTAGGTCCTC TGAAAGCAAC GTGAGAAGAT GGTTCGCACC TTGTTAGGGC
    1201  GCTCAGGTGA CTGGGGTATA GTTTCGTCCT CGTCGGAGGA GCTTTTCGCA GGTTTAGGTT
    1261  GTTGAGAGTC GTCTGTGTTG ATAAAGCAAA CCTCCGTGCT GTTCGGCGGC TCTGGAGCAG
    1321  GTGCCATGCT GGCCTCCAAA GCGGCATCCG CGGCTGCCAG CATGAGAGCC TCTTGTTCCG
    1381  CCTCATCTTC AATCAGAACC ACAGTGTCCA CCTCGGCTAG GATGCGAGGG TGCTGAACAG
    1441  ATGGCGCCTC CGAGTCCCGG AACTGAAACA GTGGATTATC TAGGCTCGAT TCTAATTGGG
    1501  TCTCCTCAAC TCGCCGTTTT GGCTGCACTG CAGCCGGTTC CGAA

//
```