```
clc;
close all;
clear;
```

Retrieve sequence information from GenBank database

```
getgenbank('nm_000520','toFile', 'nm_000520.txt');
% load in Matlab
s = genbankread('nm_000520.txt');
% Load Sequence only
seq = getgenbank('nm_000520','SequenceOnly',true)
```

seq =

*TCACATCACAACGACTTGTGGTTTTAATCCTCCGTTTTTCTGCTTCTGAAGTTACTTCAGCCTGGCAAGTCCTTTACCTCCCCGTAGGCCTGGCGAGCTG GATTCGCCGATAAGTCACGGGGGCCCCCTCACCTGACCAGGGTCTCACGTGGCCAGCCCCCTCCGAGAGGGGAGACCAGCCGGCCATGACAAGCTCCAGGCTTT A CAACTTTCAATTCCAGTACGATGTCAGCTCGGCCGCGCAGCCCGGCTGCTCAGTCCTCGACGACGGCCTTCCAGCGCTATCGTGACCTGCTTTTCGGTTCCGGGT $\tt CTTGGCCCCGTCCTTACCTCACAGGGAAACGGCATACACTGGAGAAGAATGTGTTTGGTTGTCTCTGTAGTCACACCTGGATGTAACCAGCTTCCTACTTTTGGAGT$ ${\tt CAGTGGAGAATTATACCCTGACCATAAATGATGACCAGTGTTTACTCCTCTCTGAGACTGTCTGGGGAGCTCTCCGAGGTCTGGAGACTTTTAGCCAGCTTGTTT}$ GGAAATCTGCTGAGGGCACATTCTTTATCAACAAGACTGAGATTGAGGACTTTCCCCGCTTTCCTCACCGGGGCTTGCTGTTGGATACATCTCGCCATTACCTGC $\tt CTTTTCCAGAGCTCATGAGAAAGGGGTCCTACAACCCTGTCACCCACATCTACACAGGACGGTGTGAAGGAGGTCATTGAATACGCACGGCTCCGGGGTATCC$ $\tt GTGTGCTTGCAGAGTTTGACACTCCTGGCCACACTTTGTCCTGGGGACCAGGTATCCCTGGATTACTGACTCCTTGCTACTCTGGCTCTGAGCCCTCTGGCACCT$ ATGGCCCTGACTGGAAGGATTTCTACATAGTGGAACCCCTGGCATTTGAAGGTACCCCTGAGCAGAAGGCTCTGGTGATTGGTGGAGAGGCTTGTATGTGGGGAA AATATGTGGACAACACAAACCTGGTCCCCAGGCTCTGGCCCAGAGCAGGGGCTGTTGCCGAAAGGCTGTGGACAACAAGTTGACATCTGACCTTGACATTTGCCTATGAACGTTTGTCACACTTCCGCTGTGAATTGCTGAGGCGAGGTGTCCAGGCCCAACCCCTCAATGTAGGCTTCTGTGAGCAGGAGTTTGAACAGACCTGAGCCC $\tt CGTGCCCTGTGCTTGGAGAGAAGGGGCCGGTGCTGGCTCTCGCATTCAATAAAGAGTAATGTGGCATTTTTCTATAATAAACATGGATTACCTGTGTTTAAAA$ AAAAAAGTGTGAATGGCGTTAGGGTAAGGGCACAGCCAGGCTGGAGTCAGTGTCTGCCCCTGAGGTCTTTTAAGTTGAGGGCTGGGAATGAAACCTATAGCCTTT $\tt GTGCTGTTCTGCCTTGCCTGTGAGCTATGTCACTCCCCTCCCACTCCTGACCATATTCCAGACACCCTGCCCTAATCCTCAGCCTGCTCACTTCACTTCTGCATTA$ ${\tt CATGCATACTCATCCACCTCCCCTAGAGCTATTCTCCTTTGGGTTTCTTGCTGCTTCAATTTTATACAACCATTATTTAAATATTATTAAACACACATATTGT}$ TCTCTAGGAAAAAAAAAAAAAAAAA

Format long sequence output for easy viewing

```
seqdisp(s.Sequence)
% Count nucleotides in sequence
[seq_counts] = basecount(s.Sequence)% Plot density of nucleotides along sequence
figure(1)
seq_density = ntdensity(s.Sequence)
% Count dimers in nucleotide sequence
figure(2)
[Dimers, Percent] = dimercount(s.Sequence, 'chart', 'pie')
% Count 3-mer in nucleotide sequence
trimer = nmercount(s.Sequence, 3)
```

```
ans = 46×71 char array
```

- ' 1 TCACATCACA ACGACTTGTG GTTTTAATCC TCCGTTTTTC TGCTTCTGAA GTTACTTCAG'
- ' 61 CCTGGCAAGT CCTTTACCTC CCCGTAGGCC TGGCGAGCTG CATCACAACA TTCAAGATTC'

```
' 121 ACCCTAGAGC CATCTGGGAA ACTTTCTTCT CCAGGTCGCC CTGCGTCCTC GCCTCCCCAC'
' 181 CCCGTTCTTC TCGAGTCGGG TGAGCTGTCT AGTTCCATCA CGGCCGGCAC GGCCGCAGGG'
      GTGGCCGGTT ATTTACTGCT CTACTGGGCC CGTGAACAGT CTGGCGAGCC GAGCAGTTGC
      CGACGCCCGG CACAATCCGC TGCACGTAGC AGGAGCCTCA GGTCCAGGCC GGAAGTGAAA'
      GGGCAGGGTG TGGGTCCTCC TGGGGTCGCA GGCGCAGAGC CGCCTCTGGT CACGTGATTC
      GCCGATAAGT CACGGGGGCG CCGCTCACCT GACCAGGGTC TCACGTGGCC AGCCCCCTCC
      GAGAGGGGAG ACCAGCGGGC CATGACAAGC TCCAGGCTTT GGTTTTCGCT GCTGCTGGCG'
      GCAGCGTTCG CAGGACGGCC GACGGCCCTC TGGCCCTGGC CTCAGAACTT CCAAACCTCC'
      GACCAGCGCT ACGTCCTTTA CCCGAACAAC TTTCAATTCC AGTACGATGT CAGCTCGGCC'
      GCGCAGCCCG GCTGCTCAGT CCTCGACGAG GCCTTCCAGC GCTATCGTGA CCTGCTTTTC'
721
      GGTTCCGGGT CTTGGCCCCG TCCTTACCTC ACAGGGAAAC GGCATACACT GGAGAAGAAT'
781
      GTGTTGGTTG TCTCTGTAGT CACACCTGGA TGTAACCAGC TTCCTACTTT GGAGTCAGTG'
      GAGAATTATA CCCTGACCAT AAATGATGAC CAGTGTTTAC TCCTCTGA GACTGTCTGG'
' 901 GGAGCTCTCC GAGGTCTGGA GACTTTTAGC CAGCTTGTTT GGAAATCTGC TGAGGGCACA'
' 961 TTCTTTATCA ACAAGACTGA GATTGAGGAC TTTCCCCGCT TTCCTCACCG GGGCTTGCTG'
'1021 TTGGATACAT CTCGCCATTA CCTGCCACTC TCTAGCATCC TGGACACTCT GGATGTCATG'
'1081 GCGTACAATA AATTGAACGT GTTCCACTGG CATCTGGTAG ATGATCCTTC CTTCCCATAT'
'1141 GAGAGCTTCA CTTTTCCAGA GCTCATGAGA AAGGGGTCCT ACAACCCTGT CACCCACATC'
'1201 TACACAGCAC AGGATGTGAA GGAGGTCATT GAATACGCAC GGCTCCGGGG TATCCGTGTG'
'1261 CTTGCAGAGT TTGACACTCC TGGCCACACT TTGTCCTGGG GACCAGGTAT CCCTGGATTA'
'1321 CTGACTCCTT GCTACTCTGG GTCTGAGCCC TCTGGCACCT TTGGACCAGT GAATCCCAGT'
'1381 CTCAATAATA CCTATGAGTT CATGAGCACA TTCTTCTTAG AAGTCAGCTC TGTCTTCCCA'
'1441 GATTTTATC TTCATCTTGG AGGAGATGAG GTTGATTTCA CCTGCTGGAA GTCCAACCCA'
'1501 GAGATCCAGG ACTTTATGAG GAAGAAAGGC TTCGGTGAGG ACTTCAAGCA GCTGGAGTCC'
'1621 CAGGAGGTGT TTGATAATAA AGTAAAGATT CAGCCAGACA CAATCATACA GGTGTGGCGA'
'1681 GAGGATATTC CAGTGAACTA TATGAAGGAG CTGGAACTGG TCACCAAGGC CGGCTTCCGG'
'1741 GCCCTTCTCT CTGCCCCCTG GTACCTGAAC CGTATATCCT ATGGCCCTGA CTGGAAGGAT'
'1801 TTCTACATAG TGGAACCCCT GGCATTTGAA GGTACCCCTG AGCAGAAGGC TCTGGTGATT'
'1861 GGTGGAGAGG CTTGTATGTG GGGAGAATAT GTGGACAACA CAAACCTGGT CCCCAGGCTC'
'1921 TGGCCCAGAG CAGGGGCTGT TGCCGAAAGG CTGTGGAGCA ACAAGTTGAC ATCTGACCTG'
'1981 ACATTTGCCT ATGAACGTTT GTCACACTTC CGCTGTGAAT TGCTGAGGCG AGGTGTCCAG'
'2041 GCCCAACCC TCAATGTAGG CTTCTGTGAG CAGGAGTTTG AACAGACCTG AGCCCCAGGC'
'2101 ACCGAGGAGG GTGCTGGCTG TAGGTGAATG GTAGTGGAGC CAGGCTTCCA CTGCATCCTG'
'2161 GCCAGGGGAC GGAGCCCCTT GCCTTCGTGC CCCTTGCCTG CGTGCCCCTG TGCTTGGAGA'
'2221 GAAAGGGCC GGTGCTGGCG CTCGCATTCA ATAAAGAGTA ATGTGGCATT TTTCTATAAT'
'2281 AAACATGGAT TACCTGTGTT TAAAAAAAAA AGTGTGAATG GCGTTAGGGT AAGGGCACAG'
'2341 CCAGGCTGGA GTCAGTGTCT GCCCCTGAGG TCTTTTAAGT TGAGGGCTGG GAATGAAACC'
'2401 TATAGCCTTT GTGCTGTTCT GCCTTGCCTG TGAGCTATGT CACTCCCTC CCACTCCTGA'
'2461 CCATATTCCA GACACCTGCC CTAATCCTCA GCCTGCTCAC TTCACTTCTG CATTATATCT'
'2521 CCAAGGCGTT GGTATATGGA AAAAGATGTA GGGGCTTGGA GGTGTTCTGG ACAGTGGGGA'
'2581 GGGCTCCAGA CCCAACCTGG TCACAGAAGA GCCTCTCCCC CATGCATACT CATCCACCTC'
'2641 CCTCCCCTAG AGCTATTCTC CTTTGGGTTT CTTGCTGCTT CAATTTTATA CAACCATTAT'
C: 750
```

seq_counts =

struct with fields:

A: 593

G: 716

T: 692

seq density =

struct with fields:

A: [1×2751 double]

C: [1×2751 double]

G: [1×2751 double]

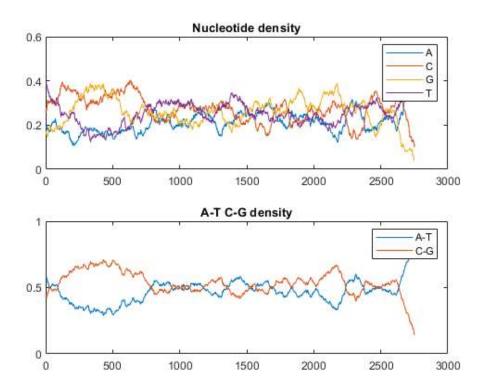
T: [1×2751 double]

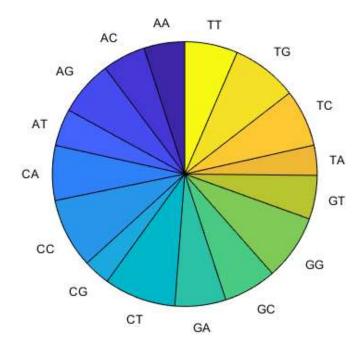
Dimers =

struct with fields:

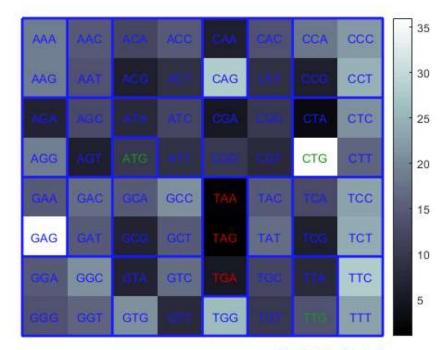
```
AA: 137
    AC: 145
    AG: 185
    AT: 125
    CA: 184
    CC: 235
    CG: 90
    CT: 241
    GA: 171
    GC: 178
    GG: 220
    GT: 147
    TA: 101
    TC: 192
    TG: 221
    TT: 178
Percent =
                                     0.0455
    0.0498
               0.0527
                          0.0673
                          0.0327
                                     0.0876
    0.0669
               0.0855
               0.0647
    0.0622
                          0.0800
                                     0.0535
                                     0.0647
    0.0367
               0.0698
                          0.0804
trimer =
  64×2 cell array
    {'ctg'}
                {[95]}
    { 'cct'}
                {[81]}
    { 'tgg' }
                {[77]}
    { 'gag ' }
                {[68]}
    { 'ggc'}
                {[68]}
    { 'ccc'}
                {[66]}
    { 'tcc' }
                {[66]}
    { 'agg'}
                {[65]}
    { 'cag' }
                {[65]}
    { 'tct'}
                {[63]}
    { 'gct ' }
                {[62]}
    {'ctc'}
                {[61]}
    { 'ctt'}
                {[61]}
    {'gcc'}
                {[59]}
    { 'ttc' }
                {[58]}
                {[55]}
    {'gga'}
    {'tga'}
                { [55] }
    { 'cca' }
                {[54]}
    { 'ggg'}
                { [53] }
    {'gtg'}
                {[51]}
    { 'ttt'}
                {[51]}
    { 'cac'}
                {[50]}
    {'tgt'}
                {[48]}
    { 'aaa ' }
                {[46]}
    { 'aca' }
                {[46]}
    { 'tca' }
                {[46]}
    {'agc'}
                {[45]}
    { 'acc'}
                {[44]}
    {'ggt'}
                {[44]}
    {'gtc'}
                {[43]}
    { 'aga'}
                {[42]}
    { 'gaa ' }
                {[42]}
    {'ttg'}
                {[42]}
    {'tgc'}
                {[41]}
    { 'cat'}
                {[39]}
    { 'gca' }
                {[38]}
    {'gac'}
                {[37]}
    { 'act'}
                {[35]}
    { 'tat'}
                {[35]}
    { 'aag ' }
                {[34]}
```

```
{ 'att'}
             {[34]}
             {[34]}
{ 'ccg'}
             {[33]}
{ 'agt' }
{ 'atg'}
             {[33]}
{ 'gtt' }
             {[32]}
{ 'caa' }
             {[30]}
{ 'aac'}
             {[29]}
{ 'ata' }
             {[29]}
{ 'atc' }
             {[29]}
{ 'tac'}
             {[29]}
{ 'aat'}
             {[27]}
{ 'tta' }
             {[27]}
{ 'cgg'}
             {[25]}
{ 'cgc'}
             {[24]}
{ 'cta' }
             {[24]}
{ 'gat'}
             {[24]}
{ 'cgt'}
             {[22]}
{ 'gta' }
             {[21]}
{ 'acg' }
             {[20]}
             {[19]}
{ 'cga' }
{ 'gcg' }
             {[19]}
{ 'taa'}
             {[19]}
{ 'tag' }
             {[18]}
{'tcg'}
             {[17]}
```

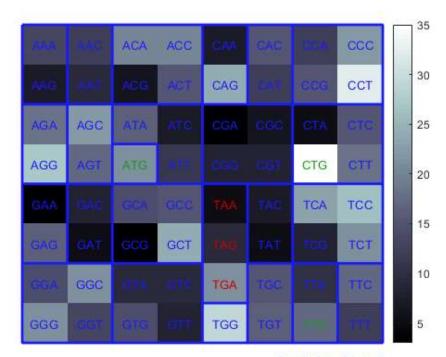




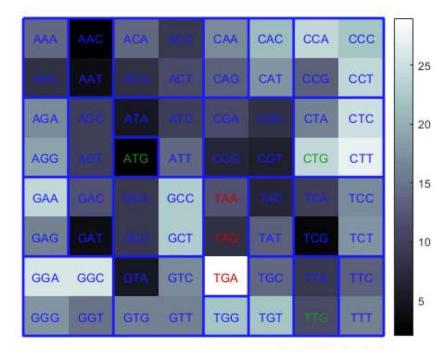
```
%lab1.3.1
% returns codon counts for the first reading frame
% and plot the results in a heat map
%First 3 frames
figure(3)
rlcodons = codoncount(s.Sequence,'frame',1,'figure',true);
figure(4)
r2codons = codoncount(s.Sequence,'frame',2,'figure',true);
figure(5)
r3codons = codoncount(s.Sequence,'frame',3,'figure',true);
figure(6)
%Reversing the sequence to get last 3 frames
r4codons = codoncount(s.Sequence,'Reverse','true','frame',1,'figure',true);
figure(7)
r5codons = codoncount(s.Sequence, 'Reverse', 'true', 'frame', 2, 'figure', true);
figure(8)
r6codons = codoncount(s.Sequence,'Reverse','true','frame',3,'figure',true);
```



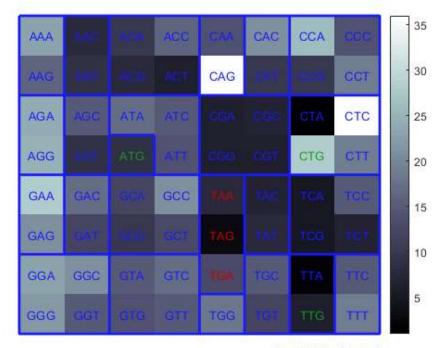
Genetic Code: Standard



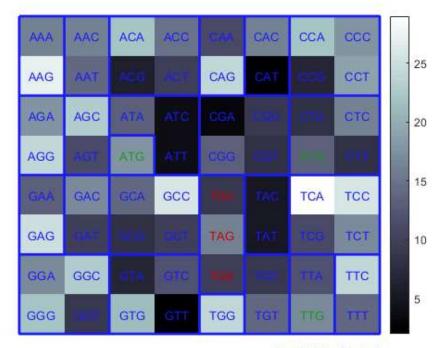
Genetic Code: Standard



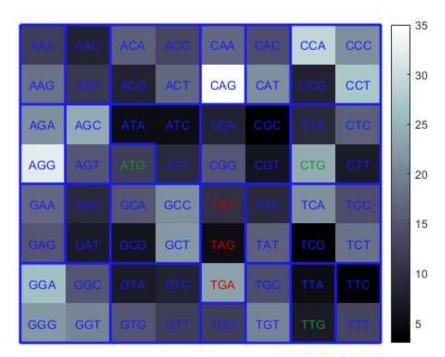
Genetic Code: Standard



Genetic Code: Standard



Genetic Code: Standard

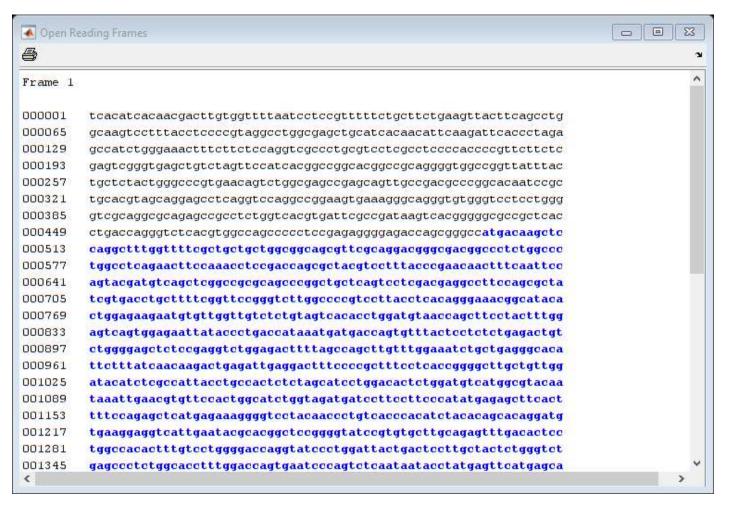


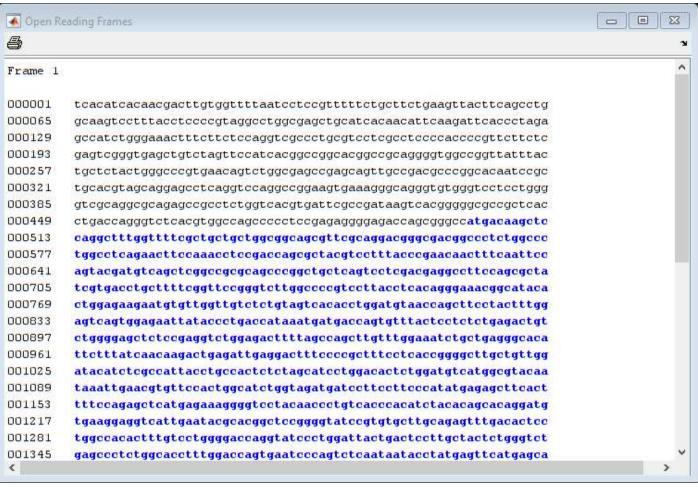
Genetic Code: Standard

```
%Lab1.3.2

%ORFs of minimum length 50 in Frame 1
orfminlength_50=seqshoworfs(s.Sequence,'Frames',1,'MinimumLength',50);

%ORFs of minimum length 500 in Frame 1
orfminlength_500=seqshoworfs(s.Sequence,'Frames',1,'MinimumLength',500);
```





```
% Find the total of stop codons in each frame
frame1_Pstop = r1codons.TAA + r1codons.TAG + r1codons.TGA
frame2_Pstop = r2codons.TAA + r2codons.TAG + r2codons.TGA
frame3_Pstop = r3codons.TAA + r3codons.TAG + r3codons.TGA
frame4_Pstop = r4codons.TAA + r4codons.TAG + r4codons.TGA
frame5_Pstop = r5codons.TAA + r5codons.TAG + r5codons.TGA
frame6_Pstop = r6codons.TAA + r6codons.TAG + r6codons.TGA
%sum of all frames above
Frametotal = frame1_Pstop + frame2_Pstop + frame3_Pstop + frame4_Pstop + frame5_Pstop + frame6_Pstop
%Total number of codons in sequence
% calculate the value in each codon
c1=struct2cell(r1codons);
c2=struct2cel1(r2codons);
c3=struct2cel1(r3codons);
c4=struct2cell(r4codons);
c5=struct2cell(r5codons);
c6=struct2cell(r6codons);
%total codons in sequence
\label{eq:total} \texttt{Total} = \text{sum}([\texttt{c1}\{:\}]) + \text{sum}([\texttt{c2}\{:\}]) + \text{sum}([\texttt{c3}\{:\}]) + \text{sum}([\texttt{c4}\{:\}]) + \text{sum}([\texttt{c5}\{:\}]) + \text{sum}([\texttt{c6}\{:\}])
\probability of stop codons
Pstop = (Frametotal./Total)
frame1_Pstop =
     8
frame2 Pstop =
    35
frame3 Pstop =
    49
frame4_Pstop =
    24
frame5_Pstop =
    35
frame6_Pstop =
    38
Frametotal =
   189
Total =
         5498
Pstop =
    0.0344
```

```
%probability of k non stop
%Given confident level alpha
alpha= 0.05 ;
```

k is the length of the sequence

```
k = log(0.05) / log(1-(Pstop));
%optimal length
knew = k + 1 + 1
```

knew =

87.6391

Published with MATLAB® R2018b