

```
#!/usr/bin/perl -w
# CodonAnalyzer_v1.pl    (C) Ulrich Krauss (2011)
#
# DESCRIPTION
# Reads in a DNA Sequence file (only DNA sequence no header) containing one or several
# open reading frames pasted as one piece of DNA without intergenic regions
# and generates a "out-file.txt" containing a list with the occurrence (in %) of (user
# defined) rare codons.
#
# NOTE1: Currently, the tool checks only for rarely used E.coli codons
#        In order to analyze for other rare codons, the script has to be modified
#
# NOTE2: The input DNA sequence file (e.g. INPUT.txt) has to be located in the same
#        directory as the script

use strict;
use warnings;

#Declare variables and Input Data

$/ = '\777'; # entire input to be read in one slurp
my $ORFs = <>; # read input, assigning to single string

my $length = length($ORFs);
my $codoncount = $length/3;

#Print Input Data
print " \n\n";
print "Input data: \n\n";
print "sequence length: $length \n";
print "number codons: $codoncount \n";
print "query sequence: $ORFs \n\n";

#Split Sequence in Codons
my @sequence = split( ' ', $ORFs );
my @codons = ( $ORFs =~ m/.../g );

#Initialize Rare Codon Counts
my $count_of_g = 0;
my $count_of_c = 0;
my $count_of_gga = 0;
my $count_of_ggt = 0;
my $count_of_cct = 0;
my $count_of_gca = 0;
my $count_of_gta = 0;
my $count_of_tta = 0;
my $count_of_cta = 0;
my $count_of_ata = 0;
my $count_of_att = 0;
my $count_of_tgt = 0;
my $count_of_tac = 0;
my $count_of_aga = 0;
my $count_of_agt = 0;
```

```
my $count_of_agc = 0;
my $count_of_tca = 0;
my $count_of_tct = 0;
my $count_of_acg = 0;
my $count_of_aca = 0;
my $count_of_act = 0;

#Count Codons

foreach my $nuc (@sequence) {

    if ($nuc eq 'g') { ++$count_of_g;}
    if ($nuc eq 'c') { ++$count_of_c;}
}

foreach my $triplet (@codons) {

    if ($triplet eq 'gga') { ++$count_of_gga;}
    if ($triplet eq 'ggt') { ++$count_of_ggt;}
    if ($triplet eq 'cct') { ++$count_of_cct;}
    if ($triplet eq 'gca') { ++$count_of_gca;}
    if ($triplet eq 'gta') { ++$count_of_gta;}
    if ($triplet eq 'tta') { ++$count_of_tta;}
    if ($triplet eq 'cta') { ++$count_of_cta;}
    if ($triplet eq 'ata') { ++$count_of_ata;}
    if ($triplet eq 'att') { ++$count_of_att;}
    if ($triplet eq 'tgt') { ++$count_of_tgt;}
    if ($triplet eq 'tac') { ++$count_of_tac;}
    if ($triplet eq 'aga') { ++$count_of_aga;}
    if ($triplet eq 'agt') { ++$count_of_agt;}
    if ($triplet eq 'agc') { ++$count_of_agc;}
    if ($triplet eq 'tca') { ++$count_of_tca;}
    if ($triplet eq 'tct') { ++$count_of_tct;}
    if ($triplet eq 'acg') { ++$count_of_acg;}
    if ($triplet eq 'aca') { ++$count_of_aca;}
    if ($triplet eq 'act') { ++$count_of_act;}

    #else { print "\n";}
}

#Compute percentage of rare codons
my $percent_gga = (($count_of_gga/$codoncount)*100);
my $percent_gga = sprintf("%.2f", $percent_gga);

my $percent_ggt = (($count_of_ggt/$codoncount)*100);
my $percent_ggt = sprintf("%.2f", $percent_ggt);

my $percent_cct = (($count_of_cct/$codoncount)*100);
my $percent_cct = sprintf("%.2f", $percent_cct);

my $percent_gca = (($count_of_gca/$codoncount)*100);
my $percent_gca = sprintf("%.2f", $percent_gca);

my $percent_gta = (($count_of_gta/$codoncount)*100);
```

```
my $percent_gta = sprintf ( "%.2f" , $percent_gta);

my $percent_tta = (($count_of_tta/$codoncount)*100);
my $percent_tta = sprintf ( "%.2f" , $percent_tta);

my $percent_cta = (($count_of_cta/$codoncount)*100);
my $percent_cta = sprintf ( "%.2f" , $percent_cta);

my $percent_ata = (($count_of_ata/$codoncount)*100);
my $percent_ata = sprintf ( "%.2f" , $percent_ata);

my $percent_att = (($count_of_att/$codoncount)*100);
my $percent_att = sprintf ( "%.2f" , $percent_att);

my $percent_tgt = (($count_of_tgt/$codoncount)*100);
my $percent_tgt = sprintf ( "%.2f" , $percent_tgt);

my $percent_tac = (($count_of_tac/$codoncount)*100);
my $percent_tac = sprintf ( "%.2f" , $percent_tac);

my $percent_aga = (($count_of_aga/$codoncount)*100);
my $percent_aga = sprintf ( "%.2f" , $percent_aga);

my $percent_agt = (($count_of_agt/$codoncount)*100);
my $percent_agt = sprintf ( "%.2f" , $percent_agt);

my $percent_agc = (($count_of_agc/$codoncount)*100);
my $percent_agc = sprintf ( "%.2f" , $percent_agc);

my $percent_tca = (($count_of_tca/$codoncount)*100);
my $percent_tca = sprintf ( "%.2f" , $percent_tca);

my $percent_tct = (($count_of_tct/$codoncount)*100);
my $percent_tct = sprintf ( "%.2f" , $percent_tct);

my $percent_acg = (($count_of_acg/$codoncount)*100);
my $percent_acg = sprintf ( "%.2f" , $percent_acg);

my $percent_aca = (($count_of_aca/$codoncount)*100);
my $percent_aca = sprintf ( "%.2f" , $percent_aca);

my $percent_act = (($count_of_act/$codoncount)*100);
my $percent_act = sprintf ( "%.2f" , $percent_act);

my $gc_content = ((($count_of_g + $count_of_c)/$length)*100);
my $gc_content = sprintf ( "%.2f" , $gc_content);
```

```
#Print Results
```

```
open (MYFILE, '>>out-file.txt');
```

```
print MYFILE " \n\n";
print MYFILE "Input data: \n\n";
print MYFILE "sequence length: $length \n";
print MYFILE "number codons: $codoncount \n";
```

```
print MYFILE "query sequence: $ORFs \n\n";

print MYFILE "Results: \n\n";
print MYFILE "GC content: $gc_content \n";
print MYFILE "GGA = $count_of_gga codon(s)... equals to $percent_gga % of all codons \n";
print MYFILE "GGT = $count_of_ggt codon(s)... equals to $percent_ggt % of all codons \n";
print MYFILE "CCT = $count_of_cct codon(s)... equals to $percent_cct % of all codons \n";
print MYFILE "GCA = $count_of_gca codon(s)... equals to $percent_gca % of all codons \n";
print MYFILE "GTA = $count_of_gta codon(s)... equals to $percent_gta % of all codons \n";
print MYFILE "TTA = $count_of_tta codon(s)... equals to $percent_tta % of all codons \n";
print MYFILE "CTA = $count_of_cta codon(s)... equals to $percent_cta % of all codons \n";
print MYFILE "ATA = $count_of_ata codon(s)... equals to $percent_ata % of all codons \n";
print MYFILE "ATT = $count_of_att codon(s)... equals to $percent_att % of all codons \n";
print MYFILE "TGT = $count_of_tgt codon(s)... equals to $percent_tgt % of all codons \n";
print MYFILE "TAC = $count_of_tac codon(s)... equals to $percent_tac % of all codons \n";
print MYFILE "AGA = $count_of_aga codon(s)... equals to $percent_aga % of all codons \n";
print MYFILE "AGT = $count_of_agt codon(s)... equals to $percent_agt % of all codons \n";
print MYFILE "AGC = $count_of_agc codon(s)... equals to $percent_agc % of all codons \n";
print MYFILE "TCA = $count_of_tca codon(s)... equals to $percent_tca % of all codons \n";
print MYFILE "TCT = $count_of_tct codon(s)... equals to $percent_tct % of all codons \n";
print MYFILE "ACG = $count_of_acg codon(s)... equals to $percent_acg % of all codons \n";
print MYFILE "ACA = $count_of_aca codon(s)... equals to $percent_aca % of all codons \n";
print MYFILE "ACT = $count_of_act codon(s)... equals to $percent_act % of all codons \n";
close (MYFILE);
```

```
exit;
```