Table of Contents

Lab 8 - Tyler Bradley

```
clc;clear;
```

Lab 8.1.1

```
hbb = genbankread("hbb_region_chr11.gb");

% get the coding and non coding regions of the 5 coding region
region_five = get_coding(hbb.Sequence, hbb.CDS(5).indices);


coding_conv = convert_seq(region_five.coding);
noncoding_conv = convert_seq(region_five.non_coding);
```

Lab 8.2.1

```
% 2
ar_coding = lpc(coding_conv, 100);
ar_noncoding = lpc(noncoding_conv, 100);
```

Lab 8.3.1

```
1.

est_coding = filter([0 -ar_coding(2:length(ar_coding))], 1,
 coding_conv);
est_noncoding = filter([0 -ar_noncoding(2:length(ar_coding))], 1,
 noncoding_conv);

% 2.
x = 1:length(coding_conv(201:300));
subplot(2, 1, 1);
plot(x, coding_conv(201:300));
hold on;
plot(x, est_coding(201:300));
title("Coding Region")
hold off

subplot(2, 1, 2);
plot(x, noncoding_conv(201:300));
hold on;
```

```matlab
plot(x, est_coding(201:300));
title("Non-coding region");
hold off;


% 3. MSE
n_coding = length(coding_conv);
mse_coding = (1/n_coding)*sum((coding_conv-est_coding).^2);
% mse_coding =
%
%    0.8581

n_noncoding = length(noncoding_conv);
mse_noncoding = (1/n_noncoding)*sum((noncoding_conv-
est_noncoding).^2);
% mse_noncoding =
%
%    1.4164

% 4. This method is able to better predict the numerical
 representation of
% the sequence when it is located within a coding region of the
 genome. As
% a result, this method could be used to calcualte whether a region is
 in a
% coding region by converting the sequence to its numerical
 representation
% and then taking the lpc of the sequence. This could be used to find
 the
% est of the sequence. Once you have the estimate, the MSE or the ME
 can be
% calculated for the sequence. If the value is high than it is not
 likely
% in a coding region. To be thorough, it would be desired to find what
 the
% distribution of MSE is across coding regions of known genomes so
 that we
% cna establish some kind of statistical significance test to
 determine if
% it is within the normal range of coding regions or not.
```

Coding Region
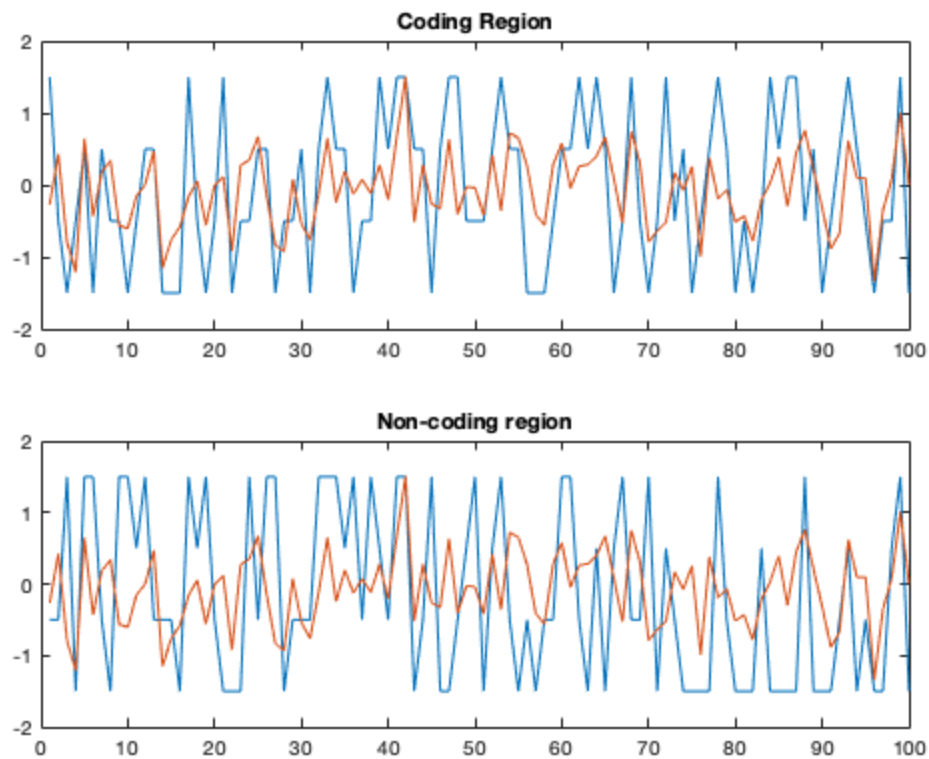
Non-coding region

Functions for lab

Function from lab 2 to get coding and non coding regions

```
function output = get_coding(seq, indices)
num_indices = length(indices);
coding = repelem("a", num_indices/2);
non_coding = repelem("a", (num_indices/2)-1);
n_code = 1;
n_non_code = 1;
for i = 1:num_indices-1
    if mod(i,2) == 1
        coding(n_code) = seq(indices(i):indices(i+1));
        n_code = n_code + 1;
    else
        non_coding(n_non_code) = seq(indices(i)+1:indices(i+1)-1);
        n_non_code = n_non_code+1;
    end
end

output.coding = char(strjoin(coding, ''));
output.non_coding = char(strjoin(non_coding, ''));
end


function output = convert_seq(seq)
```

```matlab
    len_seq = length(seq);
    output = zeros(len_seq, 1);
    for i = 1:len_seq
      if seq(i) == "a"
          output(i) = 1.5;
      elseif seq(i) == "c"
          output(i) = 0.5;
      elseif seq(i) == "g"
          output(i) = -0.5;
      else
          output(i) = -1.5;
      end
    end
end
```

Published with MATLAB® R2018b