Supplemental code

MATLAB scripts for generating cwt heatmaps and magnitude graphs for either GC% or phastCons scores.

MATLAB SCRIPT FOR CWT ON DNA SEQUENCE

```
% dnacwt.m
% Converts an input DNA sequence into a format in which
% 1 represents a A/T and 2 represents C/G. Performs a cwt and plots the
% scalogram.  Plots the average magnitude of the cwt for the sequence.
%% get input sequence from a file and enter a name for it

tic % keep track of total processing time
% enter the desired sequence and name
inputseq = 'ACGTACGT_SEQUENCEHERE';
seqname = 'Sequence Name';

%alternatively, enter an input sequence filename here and name the seq
%seqfilename = 'seqfilename.txt';
%seqname = 'Seq Name';
%inputfile = fopen(seqfilename);
%seq = textscan(inputfile,'%s');
%inputseq = char(seq{1});

%% Convert sequence to numbers and get windowed average
% 1 for A/T, 2 for G/C, 1.5 for anything
% ambiguous. Average these values over 100nt windows.

seqlength = length(inputseq);

% preallocate an array of final size for efficiency
seqasnum = zeros(1,seqlength);

% keep track of nt composition
ATcount = 0;
GCcount = 0;
ambiguouscount = 0;

for i = 1:length(inputseq)
    if inputseq(i) == "A"
        ATcount = ATcount + 1;
        seqasnum(i) = 1;
    elseif inputseq(i) == "T"
        ATcount = ATcount + 1;
        seqasnum(i) = 1;
    elseif inputseq(i) == "C"
        GCcount = GCcount + 1;
        seqasnum(i) = 2;
    elseif inputseq(i) == "G"
        GCcount = GCcount + 1;
        seqasnum(i) = 2;
    else
        seqasnum(i) = 1.5;
        ambiguouscount = ambiguouscount + 1;
        continue
    end
```

```
end

% get nt percentages
atpct = ATcount/seqlength
gcpct = GCcount/seqlength
ambigpct = ambiguouscount/seqlength

% compute the average value
avgvalue = (ATcount + 2*GCcount + 1.5*ambiguouscount)/seqlength;

% values will be averaged over a window of the specified size for smoothing
windowsize = 200;

% for efficiency, preallocate an array of the needed size
smoothedvalues = zeros(1,seqlength - windowsize);

% put the BG-subtracted window avg in the relevant array position
for i = 1:(seqlength - windowsize)
    currentwindowavg = mean(seqasnum(i:i+windowsize-1));
    smoothedvalues(i) = currentwindowavg - avgvalue; %re-center to average
end

%% Perform cwt.  Plot averages and scalogram.
fullCWTarray = abs(cwt(smoothedvalues,'NumOctaves',14));

% get average cwt magnitude for each frequency and plot
% specify total number of frequency divisions; default is 10 per octave
for i=1:140
    % avoid edge effects: exclude 200 bp at each end of sequence from avg
    results(i)=mean(fullCWTarray(i,200:(seqlength-200)));
end

plot(1:140,results(1:140));
xlabel('frequency division');
ylabel('average magnitude');
titletext = strcat(seqname,' Magnitude Plot');
title(titletext);

% output values to file
csvtitle = strcat(seqname,' Average GC% cwt Magnitude.csv');
csvwrite(csvtitle,results);

% plot scalogram
figure;
cwt(smoothedvalues,'NumOctaves',14);
cwttitle = strcat(seqname,' GC% windowsize=100');
title(cwttitle);
toc
```

MATLAB SCRIPT FOR CWT ON PHASTCONS SCORES

```
% phastConsCwt.m
%cwt analysis on phastCons scores

tic % track processing time
```

```
%% import the phastCons scores from a file
% text file name with comma-separated phastCons scores
datatextfilename = 'Filename.txt';
shortname = 'Name for Region' %will be used in output file name
seqscores = importdata(datatextfilename,',');

% scale the score range to match G+C% excursion range
%default value is 1.0
seqscores = 1.0*(seqscores);
seqlength = length(seqscores);

% window size over which scores will be averaged for smoothing
windowsize = 200;
avgvalue = sum(seqscores)/seqlength; %average will be used for centering

% preallocate an array for efficiency
smoothedversion = zeros(1,seqlength - windowsize);
for i=1:(seqlength - windowsize)
    thiswindowavg = mean(seqscores(i:i+windowsize-1));
    % center the window average and populate the relevant array position
    smoothedversion(i) = thiswindowavg - avgvalue;
end

%% perform cwt and plot

fullCWTarray=abs(cwt(smoothedversion,'NumOctaves',14));

% get average cwt magnitudes and plot
for i=1:140 %specify total number of frequency bins; default is 10 divisions
per octave
   results(i)=mean(fullCWTarray(i,200:(seqlength-200)));
   %avoid edge effects; remove 200 bp at either end of sequence
end

plot(1:140,results(1:140))

% send results to an appropriately named .csv file
csvtitle = strcat(shortname,' phastCons cwt Magnitude window 100.csv');
csvwrite(csvtitle,results);

%plot the cwt scalogram
figure;
cwt(smoothedversion,'NumOctaves',14);
title(shortname,' phastCons scores 14 octaves')
toc
```