Genome Functions to Use Later

Brief description of each function

Libraries used

```
# library(stringr)
# library(stringi)
source("genome-functions.R")
```

Example sequences:

```
l_seq <- "gtatgggaatcagccgggtctcactatgtgcaaa"
s_seq <- "gtatgggaat"

long_sequence <- toupper(l_seq)
long_sequence
```

```
[1] "GTATGGGAATCAGCCGGGTCTCACTATGTGCAAA"
```

```
short_sequence <- toupper(s_seq)
short_sequence
```

```
[1] "GTATGGGAAT"
```

Counts per Base

```
bases_count(long_sequence)
```

```
 A  T  C  G
 9  8  7 10
```

```
bases_count(short_sequence)
```

```
A T C G
3 3 0 4
```

Percentage per Base

```
bases_percentage(long_sequence)
```

```
        A         T         C         G
0.2647059 0.2352941 0.2058824 0.2941176
```

```
        bases_percentage(short_sequence)
```

```
  A   T   C   G
0.3 0.3 0.0 0.4
```

GC Percentage

```
        gc_percentage(long_sequence)
```

```
GC%
0.5
```

```
        gc_percentage(short_sequence)
```

```
GC%
0.4
```

Base Highlight

```
        highlight_base(long_sequence, "a")
```

```
[1] "gtAtgggAAtcAgccgggtctcActAtgtgcAAA"
```

```
        highlight_base(short_sequence, "a")
```

```
[1] "gtAtgggAAt"
```

Reverse Complementary

```
        long_sequence
```

```
[1] "GTATGGGAATCAGCCGGGTCTCACTATGTGCAAA"
```

```
        rev_complement(long_sequence)
```

```
[1] "TTTGCACATAGTGAGACCCGGCTGATTCCCATAC"
```

```
        short_sequence
```

```
[1] "GTATGGGAAT"
```

```
        rev_complement(short_sequence)
```

```
[1] "ATTCCCATAC"
```

Kmer Combinations

```
        combi_kmers()
```

```
 [1] "AA" "TA" "CA" "GA" "AT" "TT" "CT" "GT" "AC" "TC" "CC" "GC" "AG" "TG" "CG"
[16] "GG"
```

```
        combi_kmers(k = 3)
```

```
 [1] "AAA" "TAA" "CAA" "GAA" "ATA" "TTA" "CTA" "GTA" "ACA" "TCA" "CCA" "GCA"
[13] "AGA" "TGA" "CGA" "GGA" "AAT" "TAT" "CAT" "GAT" "ATT" "TTT" "CTT" "GTT"
[25] "ACT" "TCT" "CCT" "GCT" "AGT" "TGT" "CGT" "GGT" "AAC" "TAC" "CAC" "GAC"
[37] "ATC" "TTC" "CTC" "GTC" "ACC" "TCC" "CCC" "GCC" "AGC" "TGC" "CGC" "GGC"
[49] "AAG" "TAG" "CAG" "GAG" "ATG" "TTG" "CTG" "GTG" "ACG" "TCG" "CCG" "GCG"
[61] "AGG" "TGG" "CGG" "GGG"
```

Counts per Kmer

```
        count_kmers(short_sequence)
```

```
AA TA CA GA AT TT CT GT AC TC CC GC AG TG CG GG
 1  1  0  1  2  0  0  1  0  0  0  0  0  1  0  2
```

```
        count_kmers(short_sequence, combi_kmers(k = 3))
```

```
AAA TAA CAA GAA ATA TTA CTA GTA ACA TCA CCA GCA AGA TGA CGA GGA AAT TAT CAT GAT
  0   0   0   1   0   0   0   1   0   0   0   0   0   0   0   1   1   1   0   0
ATT TTT CTT GTT ACT TCT CCT GCT AGT TGT CGT GGT AAC TAC CAC GAC ATC TTC CTC GTC
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
ACC TCC CCC GCC AGC TGC CGC GGC AAG TAG CAG GAG ATG TTG CTG GTG ACG TCG CCG GCG
  0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0
AGG TGG CGG GGG
  0   1   0   1
```

```
        count_kmers(short_sequence, c("CG"))
```

```
CG
 0
```

```
        another_short_sequence <- "GCGCGCGCATTCGC"
        count_kmers(another_short_sequence, c("CG"))
```

```
CG
 4
```

Kmer Windows

```
        short_sequence
```

[1] "GTATGGGAAT"

```
        kmer_windws(short_sequence)
```

[1] "GT" "TA" "AT" "TG" "GG" "GG" "GA" "AA" "AT"

```
        kmer_windws(short_sequence, k = 3)
```

[1] "GTA" "TAT" "ATG" "TGG" "GGG" "GGA" "GAA" "AAT"

TM Calculation

```
        tm_calc(short_sequence)
```

```
TM in Celcius
         28
```

```
        tm_calc(long_sequence)
```

```
TM in Celcius
    65.62353
```

TM Calculation (Sequence Length less than 14 bp)

```
        tm_len_lt14(short_sequence)
```

```
TM in Celcius
         28
```

TM Calculation (Sequence Length more than 13 bp)

```
        tm_len_mt13(long_sequence)
```

```
TM in Celcius
    65.62353
```