

Genomes in BSstring packages - Caenorhabditis elegans

Caenorhabditis elegans reference genome structure

BSgenome.Celegans.UCSC.ce11 contains 7 Biostrings.

```
library(BSgenome.Celegans.UCSC.ce11)
eleg <- BSgenome.Celegans.UCSC.ce11
```

Sequences have different length, with median 15279421

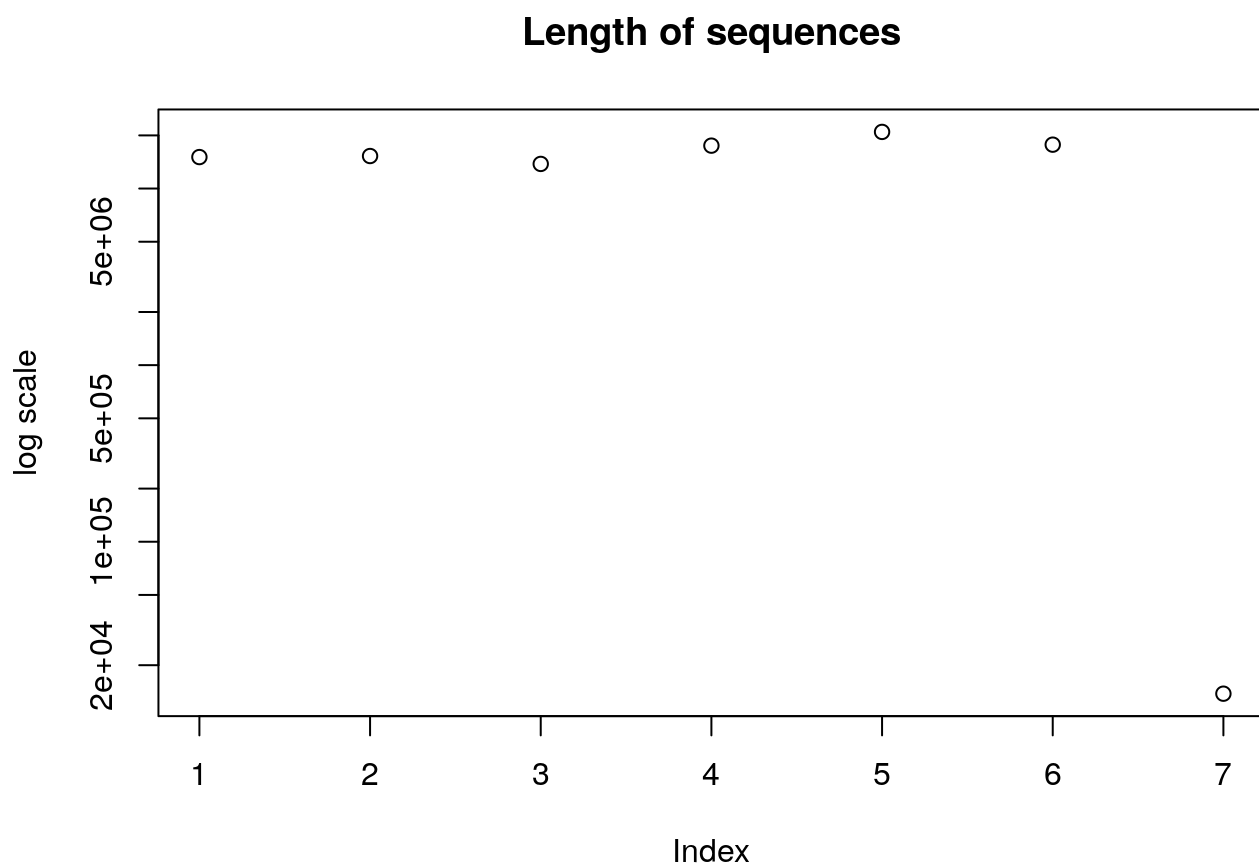
```
seqlengths(eleg)
```

```
##      chrI      chrII      chrIII      chrIV      chrV      chrX      chrM
## 15072434 15279421 13783801 17493829 20924180 17718942 13794
```

```
median(seqlengths(eleg))
```

```
## [1] 15279421
```

After visualization is visible that Mitochondrial genome is significantly smaller.



Let's look into first chromosome:

```
eleg$chrI
```

```
## 15072434-letter "DNAString" instance
## seq: GCCTAAGCCTAAGCCTAAGCCTAAGCCTAAGCCTAA...TTAGGCTTAGGCTTAGGCTTAGGTTTAGGCTTAGGC
```

To check what is the % of AT or GC in chr2L there is a function:

```
## percentage of AT: 0.642523231483382
```

```
## percentage of GC: 0.357476768516618
```

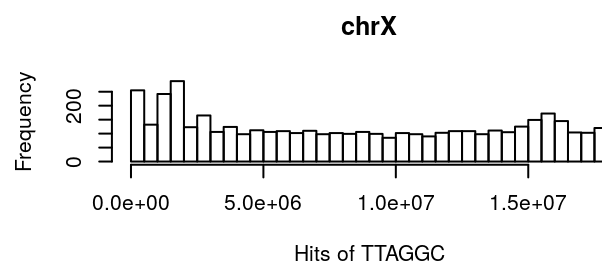
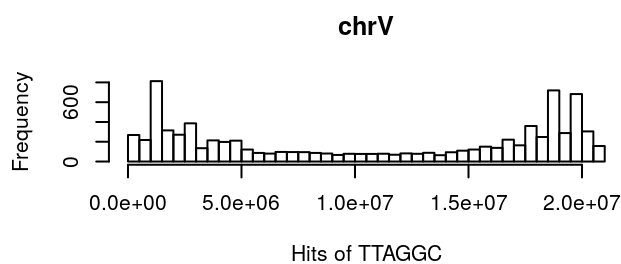
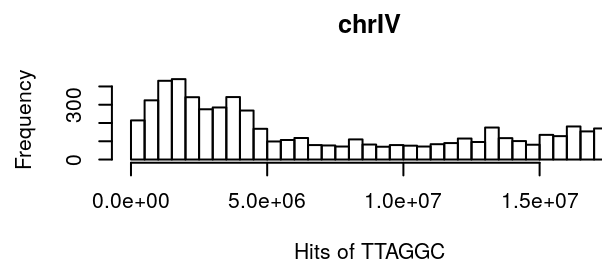
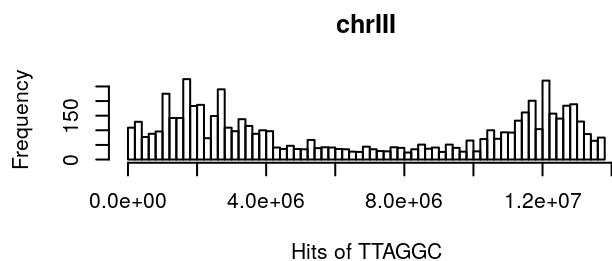
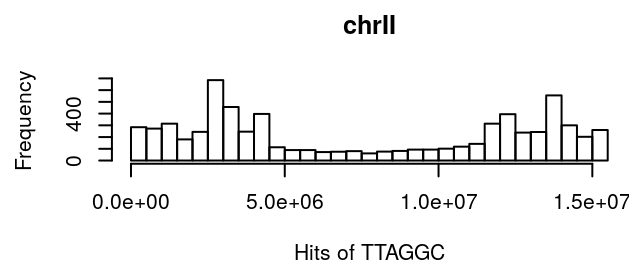
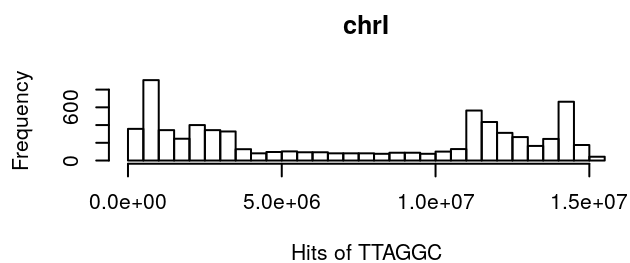
Lets check how many "TTAGGC" pattern is in *C.elegans* How many "TTAGGG" sequences are in the genome in reverse and forward?

```
dna_seq = DNAString("TTAGGC")
tel <- vmatchPattern(dna_seq, eleg)
length(tel)
```

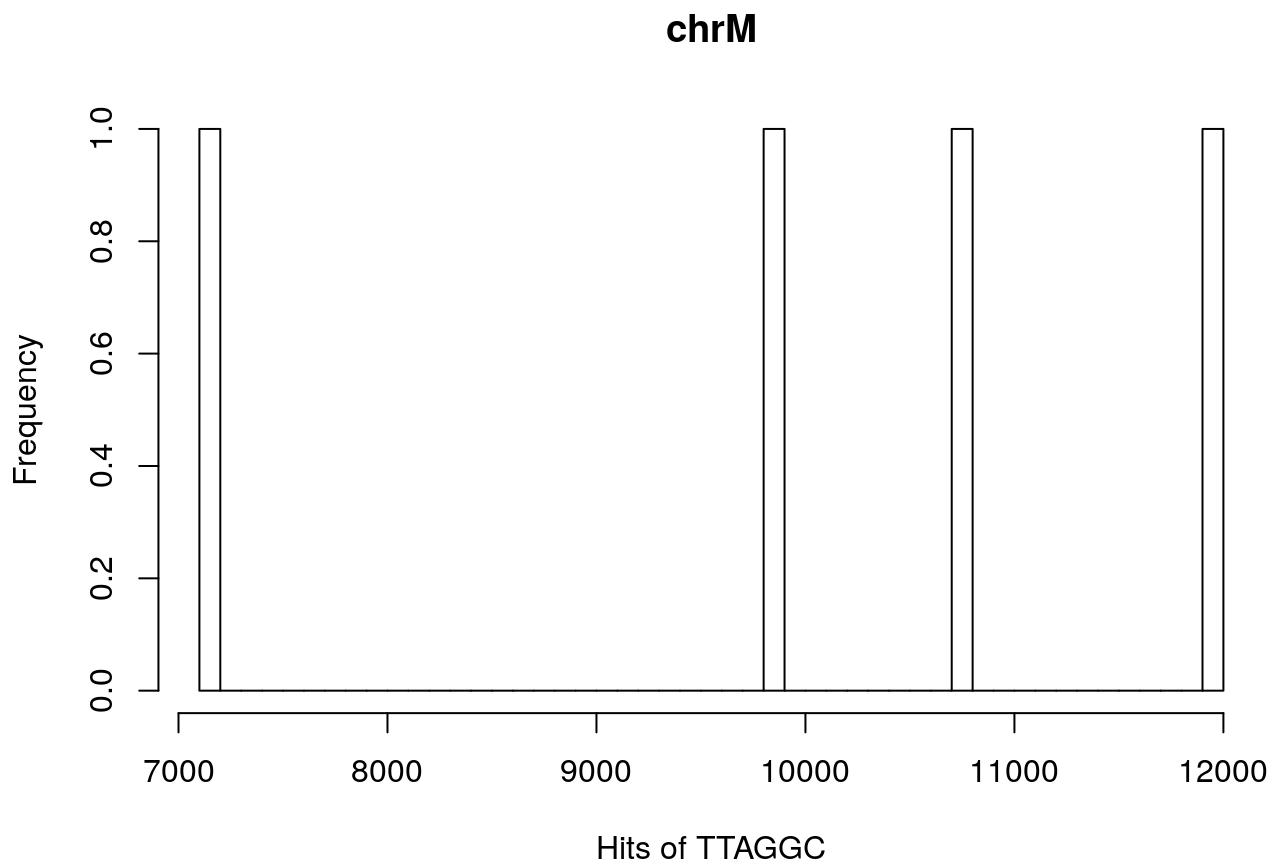
```
## [1] 38935
```

What is the distribution of this pattern depending on the chromosome? According to literature repeated

sequence in *C. elegans* is TTAGGC (http://telomerase.asu.edu/sequences_telomere.html
(http://telomerase.asu.edu/sequences_telomere.html))



Mitochondrial DNA does not contain telomere pattern at the ends. Why? Does mitochondrial DNA contain ends of chromosome? No, because it is circular.



And how about TTAGGC repeated 10 times in order?

```
dna_seq = DNASTring("TTAGGC")
tel <- vmatchPattern(rep(dna_seq, 10), eleg)
length(tel)
```

```
## [1] 310
```

What is the distribution of this pattern depending on the chromosome? According to literature repeated sequence in *C. elegans* is TTAGGC

