Analysis.Rmd

Braeden Grenier

2023-02-14

# 4. Importing Sequences

mydata = read.csv("Sequences")  
print(mydata)

## X  
## 1 1  
## 2 2  
## 3 3  
## 4 4  
## 5 5  
## 6 6  
## 7 7  
## 8 8  
## 9 9  
## 10 10  
## 11 11  
## 12 12  
## 13 13  
## 14 14  
## 15 15  
## 16 16  
## 17 17  
## 18 18  
## 19 19  
## 20 20  
## 21 21  
## 22 22  
## 23 23  
## 24 24  
## 25 25  
## 26 26  
## 27 27  
## Name  
## 1 >HQ433692.1 Borrelia burgdorferi strain QLZP1 16S ribosomal RNA gene, partial sequence  
## 2 AGCATGCAAGTCAAACGAGATGTAGCAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA  
## 3 CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGA  
## 4 TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATG  
## 5 CCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCA  
## 6 GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG  
## 7 AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAG  
## 8 TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG  
## 9   
## 10 >HQ433694.1 Borrelia burgdorferi strain CS4 16S ribosomal RNA gene, partial sequence  
## 11 AGCATGCAAGTCAAACGGGATGTAGCAATACATTCAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA  
## 12 CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAGTTAATTTGTTAATTGA  
## 13 TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGCTAGTTGGTAGGGTAAATG  
## 14 CCTACCAAGGCAATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGATACGGTCCA  
## 15 GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG  
## 16 AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACAAAG  
## 17 TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCAGCGGTAATACG  
## 18   
## 19 >HQ433691.1 Borrelia burgdorferi strain GL18 16S ribosomal RNA gene, partial sequence  
## 20 AGCATGCAAGTCAAACGAGATGTAGTAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA  
## 21 CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGA  
## 22 TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATG  
## 23 CCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCA  
## 24 GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG  
## 25 AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAG  
## 26 TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG  
## 27   
## sequence  
## 1 >HQ433692.1 Borrelia burgdorferi strain QLZP1 16S ribosomal RNA gene, partial sequence  
## 2 AGCATGCAAGTCAAACGAGATGTAGCAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA  
## 3 CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGA  
## 4 TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATG  
## 5 CCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCA  
## 6 GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG  
## 7 AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAG  
## 8 TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG  
## 9   
## 10 >HQ433694.1 Borrelia burgdorferi strain CS4 16S ribosomal RNA gene, partial sequence  
## 11 AGCATGCAAGTCAAACGGGATGTAGCAATACATTCAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA  
## 12 CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAGTTAATTTGTTAATTGA  
## 13 TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGCTAGTTGGTAGGGTAAATG  
## 14 CCTACCAAGGCAATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGATACGGTCCA  
## 15 GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG  
## 16 AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACAAAG  
## 17 TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCAGCGGTAATACG  
## 18   
## 19 >HQ433691.1 Borrelia burgdorferi strain GL18 16S ribosomal RNA gene, partial sequence  
## 20 AGCATGCAAGTCAAACGAGATGTAGTAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA  
## 21 CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGA  
## 22 TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATG  
## 23 CCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCA  
## 24 GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG  
## 25 AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAG  
## 26 TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG  
## 27

?bind\_rows

## starting httpd help server ... done

seq1 = c(mydata$Name[2:8])  
seq1

## [1] "AGCATGCAAGTCAAACGAGATGTAGCAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA"  
## [2] "CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGA"  
## [3] "TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATG"  
## [4] "CCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCA"  
## [5] "GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG"  
## [6] "AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAG"  
## [7] "TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG"

seq2 = c(mydata$Name[11:17])  
seq2

## [1] "AGCATGCAAGTCAAACGGGATGTAGCAATACATTCAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA"  
## [2] "CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAGTTAATTTGTTAATTGA"  
## [3] "TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGCTAGTTGGTAGGGTAAATG"  
## [4] "CCTACCAAGGCAATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGATACGGTCCA"  
## [5] "GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG"  
## [6] "AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACAAAG"  
## [7] "TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCAGCGGTAATACG"

seq3 = c(mydata$Name[20:26])  
seq3

## [1] "AGCATGCAAGTCAAACGAGATGTAGTAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTA"  
## [2] "CCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGA"  
## [3] "TGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATG"  
## [4] "CCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCA"  
## [5] "GACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTG"  
## [6] "AATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAG"  
## [7] "TGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG"

Can’t get the regex expressions or anything to work

sequence1 = "AGCATGCAAGTCAAACGAGATGTAGCAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATGCCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG"  
  
sequence2 = "AGCATGCAAGTCAAACGGGATGTAGCAATACATTCAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAGTTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGCTAGTTGGTAGGGTAAATGCCTACCAAGGCAATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGATACGGTCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACAAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCAGCGGTAATACG"  
  
sequence3 = "AGCATGCAAGTCAAACGAGATGTAGTAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATGCCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG"  
  
seqdata = c(sequence1, sequence2, sequence3)  
sequencenumber = c(1,2,3)  
  
newdata = data.frame(sequencenumber, seqdata)  
print(newdata)

## sequencenumber  
## 1 1  
## 2 2  
## 3 3  
## seqdata  
## 1 AGCATGCAAGTCAAACGAGATGTAGCAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATGCCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG  
## 2 AGCATGCAAGTCAAACGGGATGTAGCAATACATTCAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAGTTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGCTAGTTGGTAGGGTAAATGCCTACCAAGGCAATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGATACGGTCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACAAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCAGCGGTAATACG  
## 3 AGCATGCAAGTCAAACGAGATGTAGTAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTATTAGTTAGTTGGTAGGGTAAATGCCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTAATACG

grep("A", sequence1)

## [1] 1

a1 = str\_count(sequence1, "A")  
c1 = str\_count(sequence1, "C")  
t1 = str\_count(sequence1, "T")  
g1 =str\_count(sequence1, "G")  
frame1 = data.frame(Base = c("A", "C", "T", "G"), Occurences = c(a1,c1,t1,g1))  
  
a2 = str\_count(sequence2, "A")  
c2 = str\_count(sequence2, "C")  
t2 = str\_count(sequence2, "T")  
g2 = str\_count(sequence2, "G")  
frame2 = data.frame(Base = c("A", "C", "T", "G"), Occurences = c(a2,c2,t2,g2))  
  
a3 = str\_count(sequence3, "A")  
c3 = str\_count(sequence3, "C")  
t3 = str\_count(sequence3, "T")  
g3 = str\_count(sequence3, "G")  
frame3 = data.frame(Base = c("A", "C", "T", "G"), Occurences = c(a3,c3,t3,g3))

print(frame1)

## Base Occurences  
## 1 A 154  
## 2 C 82  
## 3 T 114  
## 4 G 131

print(frame2)

## Base Occurences  
## 1 A 155  
## 2 C 81  
## 3 T 114  
## 4 G 131

print(frame3)

## Base Occurences  
## 1 A 154  
## 2 C 81  
## 3 T 115  
## 4 G 131

<https://en.wikipedia.org/wiki/Borrelia_burgdorferi>

knitr::include\_graphics("https://en.wikipedia.org/wiki/Borrelia\_burgdorferi#/media/File:Borrelia\_burgdorferi\_(CDC-PHIL\_-6631)\_lores.jpg")

![](data:text/html; charset=UTF-8;base64,)

GC content

seq1GC = ((c1 + g1)/(a1+c1+g1+t1))\*100  
print(seq1GC)

## [1] 44.28274

seq2GC = ((c2 + g2)/(a2+c2+g2+t2))\*100  
print(seq2GC)

## [1] 44.07484

seq3GC = ((c3 + g3)/(a3+c3+g3+t3))\*100  
print(seq1GC)

## [1] 44.28274

SequenceID = c("HQ433692.1", "HQ433694.1", "HQ433691.1")  
GC\_Content = c(seq1GC,seq2GC,seq2GC)  
print(data.frame(SequenceID, GC\_Content))

## SequenceID GC\_Content  
## 1 HQ433692.1 44.28274  
## 2 HQ433694.1 44.07484  
## 3 HQ433691.1 44.07484