

Borrelia burgdorferi

```
Sequences <- read.csv("./A5_MAHRTSMITH_AMELIE_Sequences.csv")

#first, create some empty vectors for output
SequenceID <- NULL
A <- NULL
T <- NULL
C <- NULL
G <- NULL
GCcontent <- NULL

for (i in 1:nrow(Sequences)){
  SequenceID <- c(SequenceID, gsub(">(\w+\\.\\.\\d).*", "\\1", Sequences$Name[i]))
  cat(SequenceID[i], "has the sequence:\n", Sequences$Sequence[i], "\n") #print each sequence
  A <- c(A, nchar(gsub("[TCG]", "", Sequences$Sequence[i]))) #number of A's in each sequence
  T <- c(T, nchar(gsub("[ACG]", "", Sequences$Sequence[i]))) #number of T's in each sequence
  C <- c(C, nchar(gsub("[ATG]", "", Sequences$Sequence[i]))) #number of C's in each sequence
  G <- c(G, nchar(gsub("[ATC]", "", Sequences$Sequence[i]))) #number of G's in each sequence
  GCcontent <- c(GCcontent, paste(round(((C[i] + G[i])/(A[i] + T[i] + C[i] + G[i]))*100, 2), "%"
)) #percentage of sequence that is G or C
}
```

```
## HQ433692.1 has the sequence:
## AGCATGCAAGTCAAACGAGATGTAGCAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTAT
TAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTAT
TAGTTAGTTGGTAGGGTAAATGCCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAAGTGAAGACACGGTCCAGACT
CCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTC
TTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTA
ATACG
## HQ433694.1 has the sequence:
## AGCATGCAAGTCAAACGGGATGTAGCAATACATTCAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTAT
TAGAAATAGTAGCTAATACCGAATAAGGTCAGTTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTAT
TAGCTAGTTGGTAGGGTAAATGCCTACCAAGGCAATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAAGTGAAGATACGGTCCAGACT
CCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTC
TTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACAAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCAGCGGTA
ATACG
## HQ433691.1 has the sequence:
## AGCATGCAAGTCAAACGAGATGTAGTAATACATCTAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTATGAGATGGGGATAACTAT
TAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTTGTTAATTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTAT
TAGTTAGTTGGTAGGGTAAATGCCTACCAAGGCGATGATAAGTAACCGGCCTGAGAGGGTGAACGGTCACACTGGAAGTGAAGACACGGTCCAGACT
CCTACGGGAGGCAGCAGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACACTGCGTGAATGAAGAAGGTCGAAAGATTGTAAAATTC
TTTTATAAATGAGGAATAAGCTTTGTAGGAAATGACGAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCCGCGGTA
ATACG
```

Table 1. The number of nucleotides in each sequence.

| Sequence ID | A | T | C | G |
|-------------|---|---|---|---|
|-------------|---|---|---|---|

| Sequence ID | A | T | C | G |
|-------------|-----|-----|----|-----|
| HQ433692.1 | 154 | 114 | 82 | 131 |
| HQ433694.1 | 155 | 114 | 81 | 131 |
| HQ433691.1 | 154 | 115 | 81 | 131 |

Table 2. The percentage of each sequence that contains G or C.

| Sequence ID | GC Content |
|-------------|------------|
| HQ433692.1 | 44.28 % |
| HQ433694.1 | 44.07 % |
| HQ433691.1 | 44.07 % |



Image: The spirochete bacterium *Borrelia burgdorferi* that causes Lyme disease.

Source: Bay Area Lyme Foundation (<https://www.bayarealyme.org/about-lyme/what-causes-lyme-disease/borrelia-burgdorferi/>)

To learn more about *B. burgdorferi*, visit the Wikipedia page (https://en.wikipedia.org/wiki/Borrelia_burgdorferi).