miniproject

Ebony Michelle Argaez (PID: A59026556)

Loading sequence

```
CTCF_seq <- "MEGDAVEAIVEESETFIKGKERKTYQRRREGGQEEDACHLPQNQTDGGEVVQDVNSSVQMVMMEQLDPTLLQMKTEV
```

```
count_RK_freq <- function(sequence) {
  # Convert the sequence to uppercase
  sequence <- toupper(sequence)

  # Set the chunk size
  chunk_size <- 5

  # Initialize a dataframe to store the results
  results <- data.frame(site = integer(),
                        R_freq = numeric(),
                        K_freq = numeric(),
                        RK_freq = numeric())

  # Scan the sequence
  for (i in 1:(nchar(sequence) - chunk_size + 1)) {
    chunk <- substr(sequence, i, i + chunk_size - 1)

    # Calculate frequencies
    r_freq <- sum(strsplit(chunk, "")[[1]] == "R") / chunk_size
    k_freq <- sum(strsplit(chunk, "")[[1]] == "K") / chunk_size
    rk_freq <- r_freq + k_freq

    # Add to dataframe
    results <- rbind(results, data.frame(site = i, R_freq = r_freq, K_freq = k_freq, RK_fr
  }
```
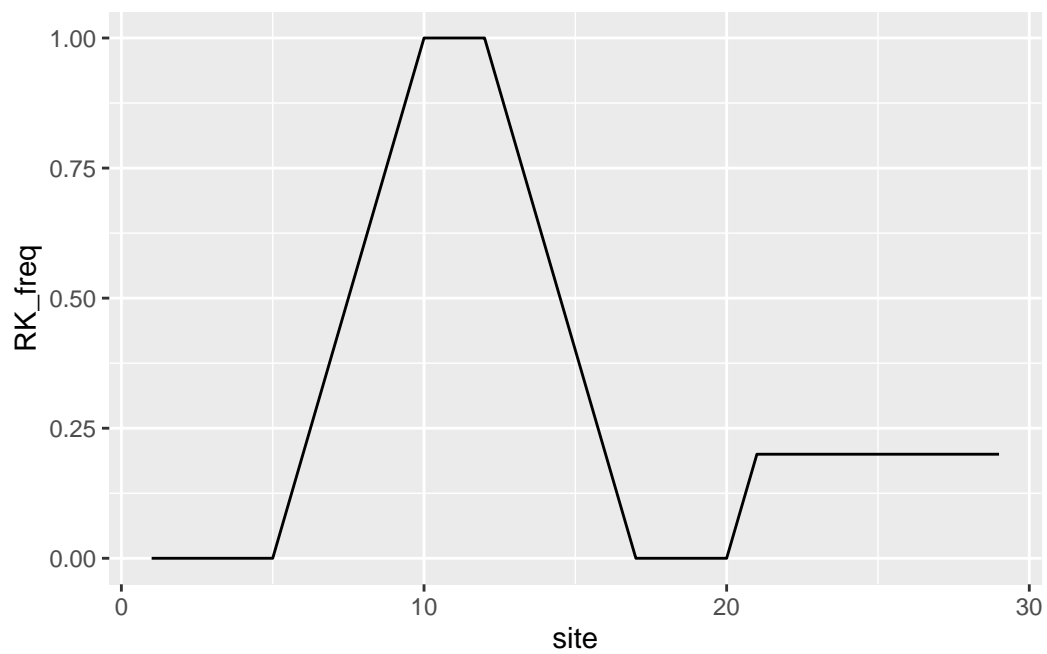
```
    return(results)
}

test_seq <- "mmmmmmmmmmRRrKKKKnnnnnnnnRnnnnKnnn"

test_seq <- count_RK_freq(test_seq)

library(ggplot2)

ggplot(test_seq) + aes(x = site, y = RK_freq) + geom_line()
```
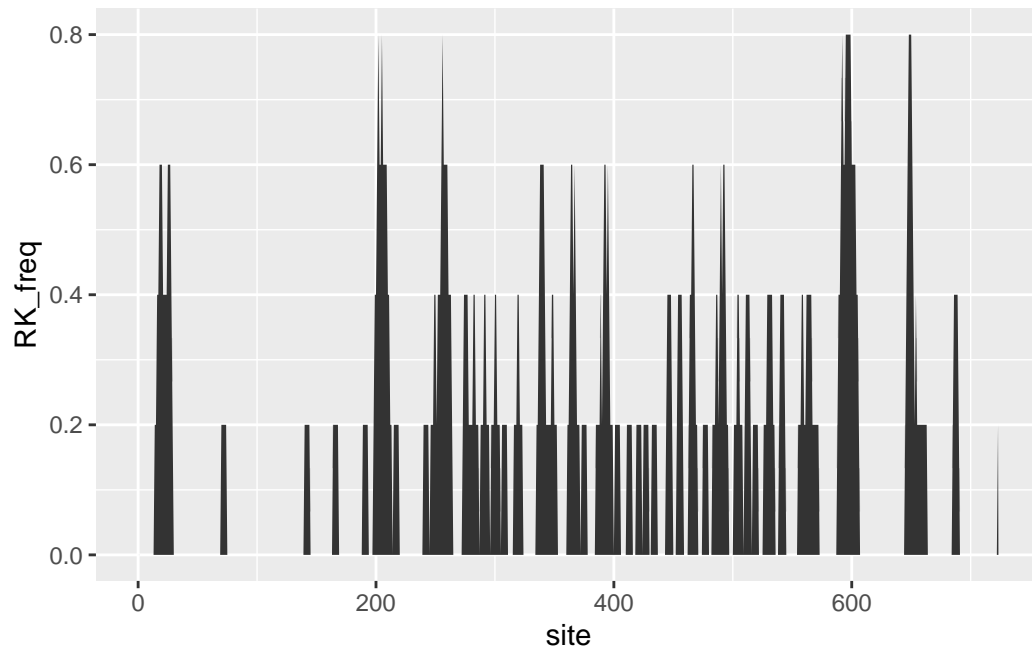


```
# Calculate the frequencies of R, K, and RK in CTCF

CTCF_KR_freq <- count_RK_freq(CTCF_seq)

# Plot KR_freq vs site by geom_area
ggplot(CTCF_KR_freq) +
  aes(x = site, y = RK_freq) +
  geom_area()
```

```
# Plot K_freq, R_freq, and KR_freq vs site by geom_area filled with different colors

ggplot(CTCF_KR_freq) +
  geom_area(aes(x = site, y = RK_freq, fill = "RK"), alpha = 0.7) +
  geom_area(aes(x = site, y = K_freq, fill = "K"), alpha = 0.9) +
  geom_area(aes(x = site, y = R_freq, fill = "R"), alpha = 0.9) +
  # Add a legend
  scale_fill_manual(name = "AA", values = c("RK" = "black", "K" = "lightpink", "R" = "ligh
  # Add a title and labels
  labs(title = "Arginine and Lysine Frequency in CTCF", x = "Site", y = "Frequency")
```

Arginine and Lysine Frequency in CTCF
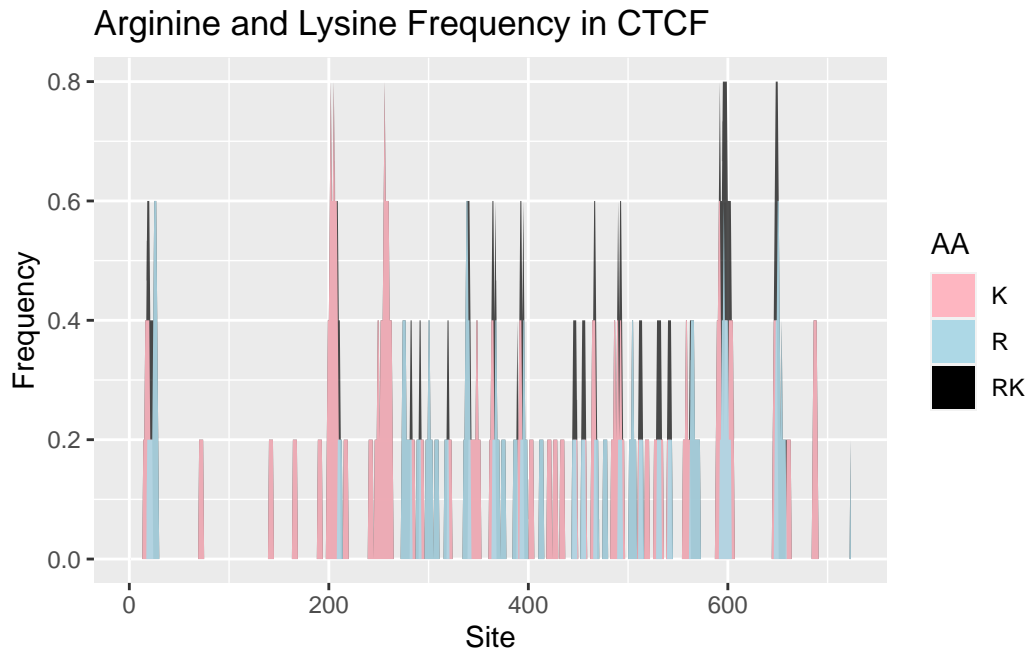

```r
ggplot(CTCF_KR_freq) +
  geom_area(aes(x = site, y = RK_freq, fill = "RK"), alpha = 0.7) +
  geom_area(aes(x = site, y = K_freq, fill = "K"), alpha = 0.9) +
  geom_area(aes(x = site, y = R_freq, fill = "R"), alpha = 0.9) +
  # Add a legend
  scale_fill_manual(name = "AA", values = c("RK" = "black", "K" = "lightpink", "R" = "ligh
  # Add a title and labels
  labs(title = "Arginine and Lysine Frequency in CTCF", x = "Site", y = "Frequency") +
  geom_rect(xmin = 266, xmax = 577, ymin = 0, ymax = 0.02, fill = "lightyellow", alpha = 0
```

Arginine and Lysine Frequency in CTCF