Read mapping project - 01/2023 - Group 13

Niklas Horner, Abdullah Cetinkaya, Philip Wolper

Link to GitHub Repository

Link to GitHub repository: https://github.com/pwolper/read-mapping

Visualisation of mapped sequencing reads

The following packages were used for this anlysis: ggplot2, dplyr, purrr, here, ape, ggrepel, ggpubr.

Define in- and output directories in case of running script separately:

```
## [1] "/home/philip/GitHub/read-mapping/data/Ecoli_genome.fasta"
```

```
## [1] "2023-01-27_NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome.csv"
```

```
genome <- read.dna(inputGenome,format="fasta")
sequence <- as.character(genome)
genome_length <- length(sequence)
```

Importing data from /output/*.csv

```
mapped <- read.csv(pathMap,sep = ";")

# mapped = mapped[50:60,]

mapped = mapped %>%
  mutate_at(vars(Sense, Antisense), ~map(strsplit(. ,split=","), ~map_int(.x, as.integer)))
# mapped %>% str()

sense = unlist(mapped$Sense)
anti = unlist(mapped$Antisense)

sense %>% length()
```

```
## [1] 820
```

```
anti %>% length()
```

```
## [1] 171
```

```
reads <- data.frame(pos = c(sense, anti),
                    strand = as.factor(rep(c("s","a"),
                    times = c(length(sense),length(anti)))))
```

Plotting mapped reads

Saving figures to file

```
ggsave("F1_Genome_pos_total.png",p1, device = "png", dpi = 300, path = output)
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("F2_Genome_pos_orient.png",p2, device = "png", dpi = 300, path = output)
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("F3_Hits_sense_vs_antisense.png",p3 ,device = "png", dpi = 300, path = output)
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("F4_Read_distribution.png",p4, device = "png", dpi = 300, path = output)
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("F5_Reads_by_number_of_hits.png",p5 ,device = "png", height = 10, width = 10,
       dpi = 300, path = output)
ggsave("F2_and_F3_orient.png", p2p3, device = "png", dpi = 300, path = here("doc/"))
```

```
## Saving 6.5 x 4.5 in image
```

Figures 1 - 5

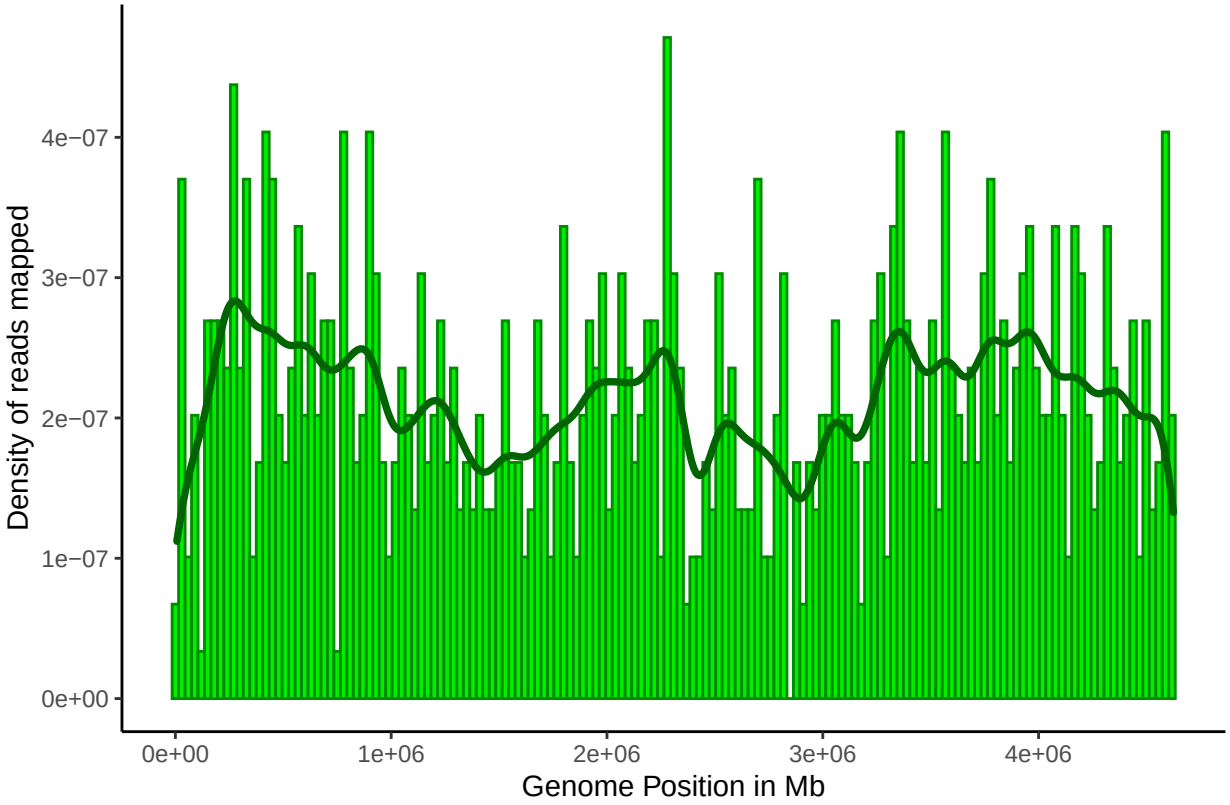


Fig. 1: Density of the reads mapped to the E.coli genome.

Description Fig 1: Density of reads mapped to the E.coli genome. Each bin size represents a 30 kb long segment along the genome.

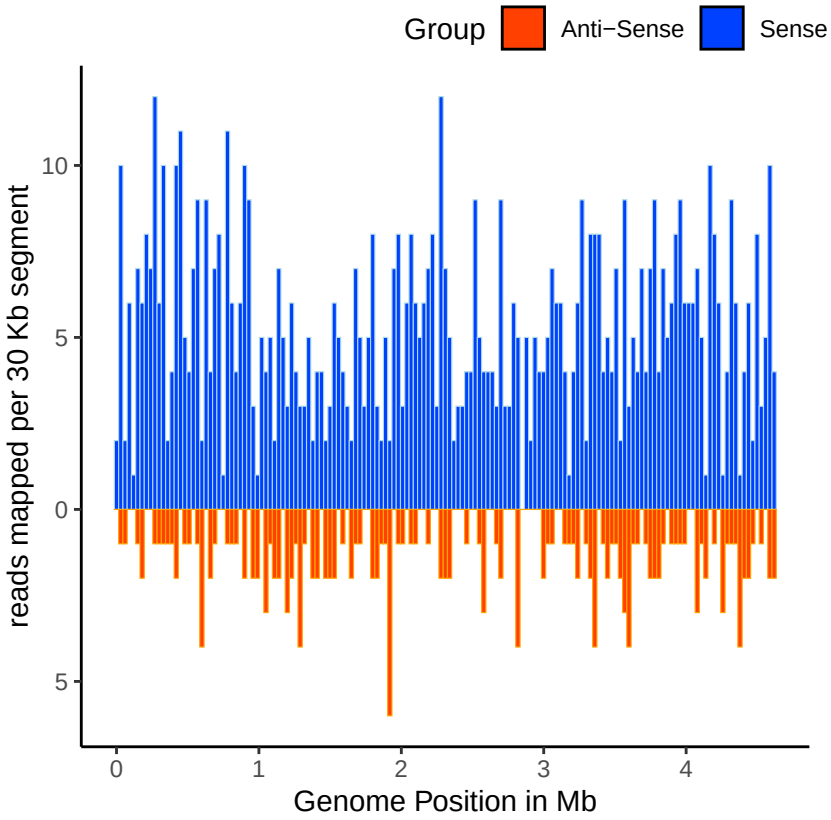


Fig 2: Orientation of mapped Reads

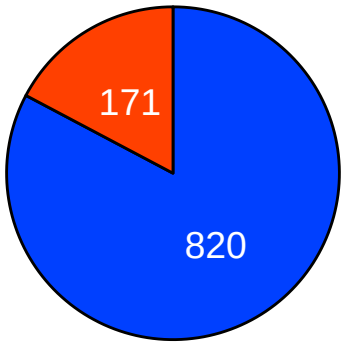
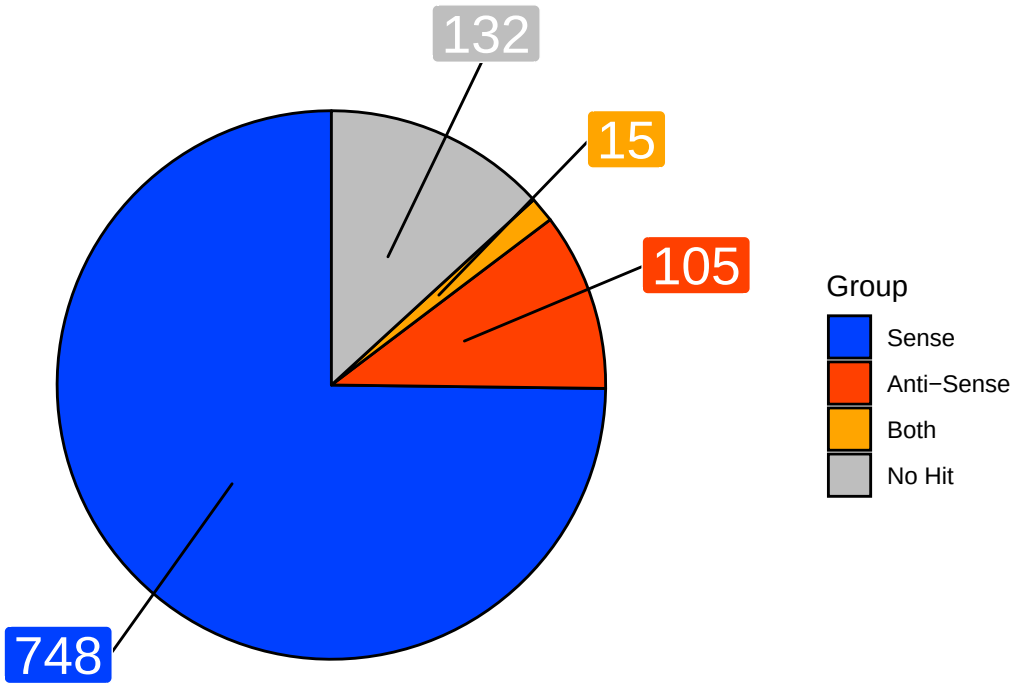
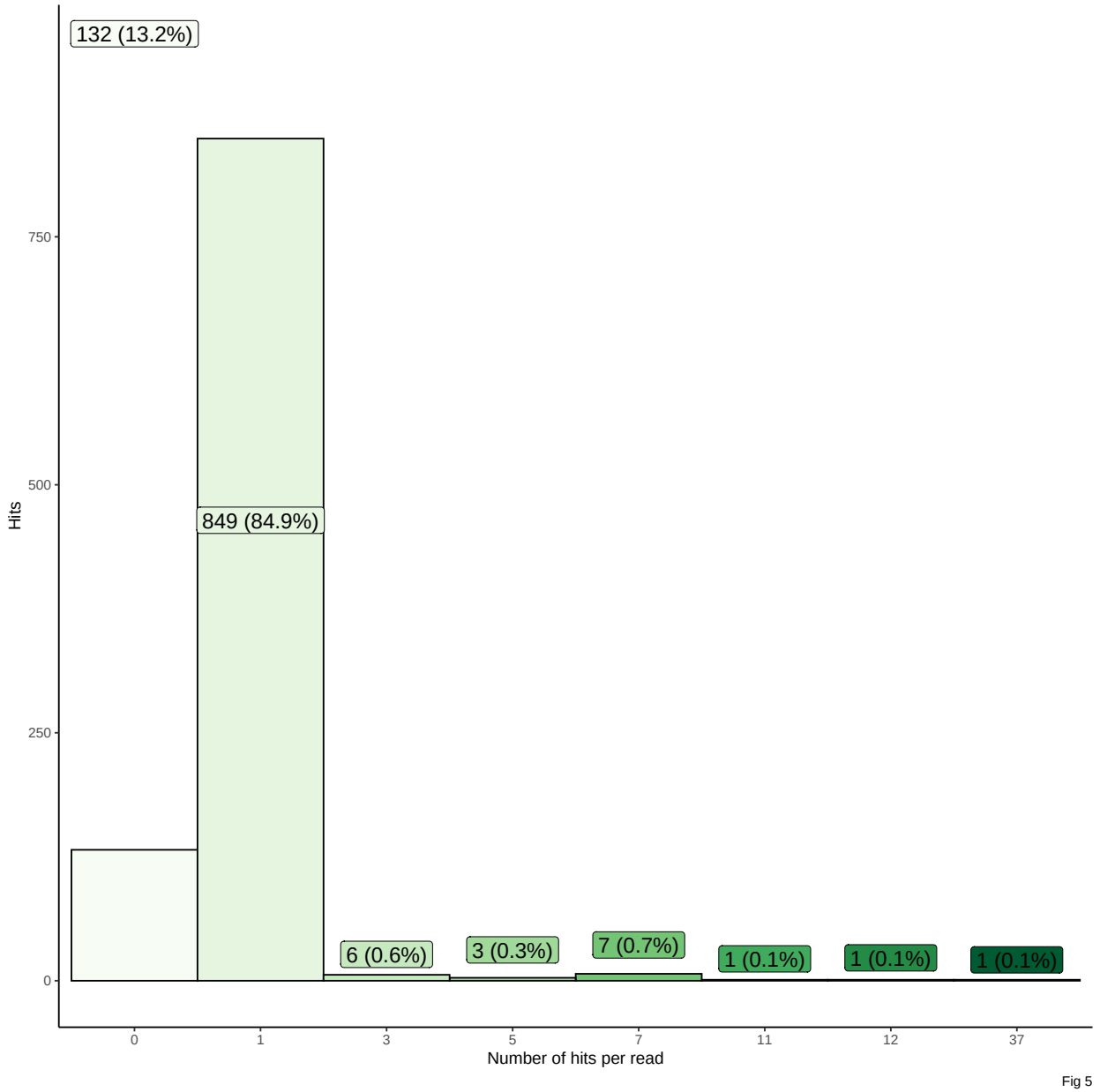


Fig 3

Description Fig 2: Sense vs. Antisense density of mapped reads. The y-axis shows a number of counts per 30 Kb segment.
Description Fig 3: Number of successful hits to sense or anti-sense strand.



Description Fig 4: Number of reads mapped to sense strand, anti-sense strand, both strands or none in Genome



Description Fig 5: Frequency of hits per read. Most reads occur once, few at higher frequencies in the genome.