

## Visualisation of mapped sequencing reads

The following packages were used for this analysis: 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

```
p1 <- ggplot(reads, aes(x = pos, y = ..density..)) +
  geom_histogram(bins = 155, color = "green4", fill = "green2") +
  geom_density(color = "darkgreen", lwd=1.3, bw = 70000)+
  # geom_histogram(reads)
  labs(x = "Genome Position in Mb", y = "Density of reads mapped",
    caption = "Fig. 1: Density of the reads mapped to the E.coli genome.") +
  theme_classic()

p2 <- ggplot(reads) +
  geom_histogram(data = reads[reads$strand == "s",],aes(x = pos/10^6, y = ..count..),
    bins = 155,color="lightblue",lwd = 0.2,fill="#0040ff") +
  geom_histogram(data = reads[reads$strand == "a",],aes(x = pos/10^6, y = -..count..),
    bins = 155,color="orange",lwd = 0.2,fill="#ff4000") +
  scale_y_continuous(labels = abs) +
  labs(x = "Genome Position in Mb", y = "reads mapped per 30 Kb segment",
```

```

    caption = "Fig 2: Orientation of mapped Reads") +
    theme_classic()

p3 <- ggplot(pie1, aes(x="", y=amount, fill=Group)) +
  geom_bar(stat="identity", width=1, color = "black") +
  coord_polar("y", start=0)+
  theme_void()+
  ## theme(legend.position="bottom", legend.key.size = unit(1,'cm'),
  ## legend.text = element_text(size = 10)) +
  ## theme(legend.title= element_blank()) +
  geom_text(aes(x = 1, y = ypos, label = amount), color = "white", size=5)+
  scale_fill_manual(values= c("#ff4000", "#0040ff")) +
  labs(caption = "Fig 3")

p2p3 <- ggpubr::ggarrange(p2,p3,nrow = 1, widths = c(1,0.5), align = "h",
  common.legend = TRUE)

p4 <- ggplot(pie2, aes(x="", y=value, fill=Group)) +
  geom_bar(stat="identity", width=1, color = "black") +
  coord_polar("y", start=0)+
  theme_void()+
  # theme(legend.position="bottom") +
  # theme(legend.title= element_blank()) +
  geom_label_repel(aes(y=ypos, label = value), colour = "white",
    segment.colour = "black", size = 7, nudge_x = 0.9, show.legend = FALSE) +
  # geom_text(aes(x = 1.15, y = ypos, label = value), color = "white", size=5)+
  scale_fill_manual(values= c("#0040ff", "#ff4000", "orange", "grey"))#+
  # labs(caption = "Fig 4")

p5 <- ggplot(MapNum, aes(x = factor(Mappings), y=Hits, fill=factor(Mappings))) +
  geom_bar(stat="identity", width=1, color = "black") +
  # coord_polar("y", start=0)+
  theme_classic()+
  scale_fill_brewer(name="Frequency of read", palette="Greens", guide = "none")+
  geom_label_repel(data = MapNum,
    aes(y = pos, label = paste0(Hits, " (", Fraction, "%)")),
    size = 5, nudge_y = 3, show.legend = FALSE) +
  labs(x = "Number of hits per read", caption = "Fig 5")

```

## 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

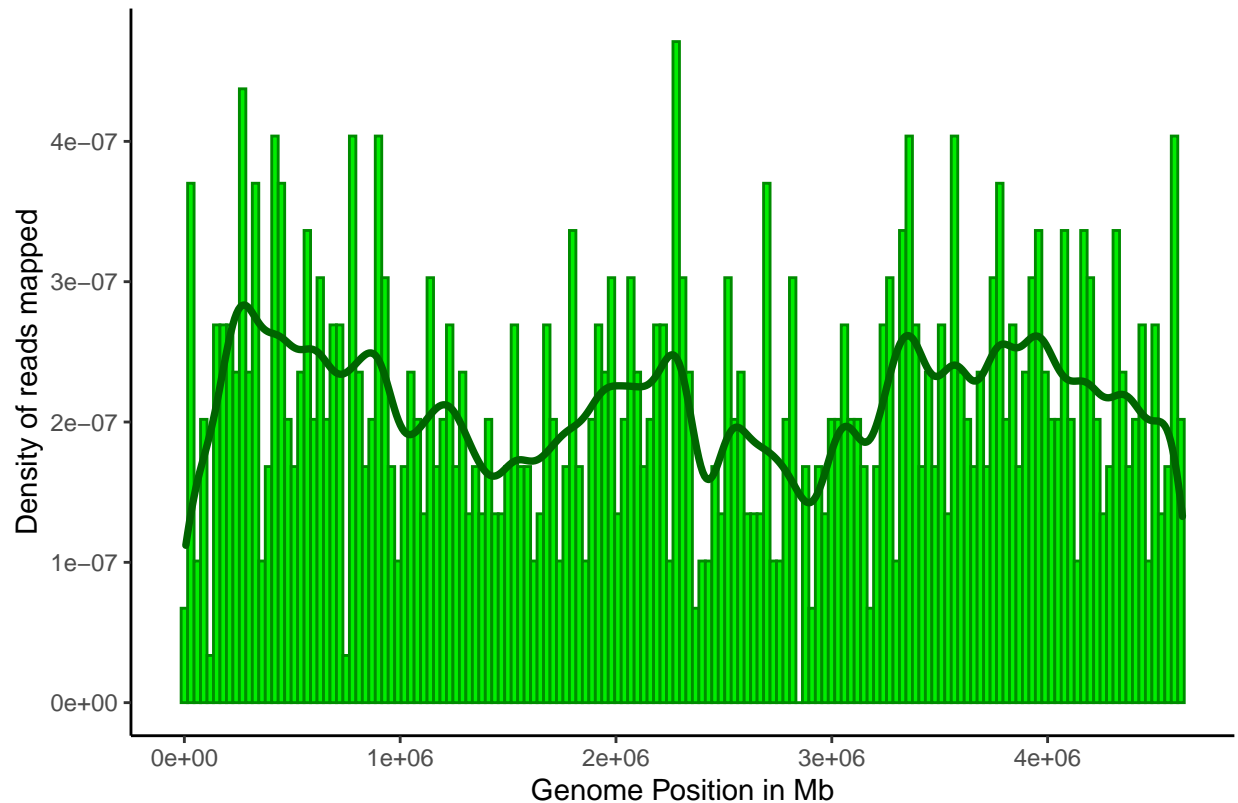


Fig. 1: Density of the reads mapped to the E.coli genome.

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

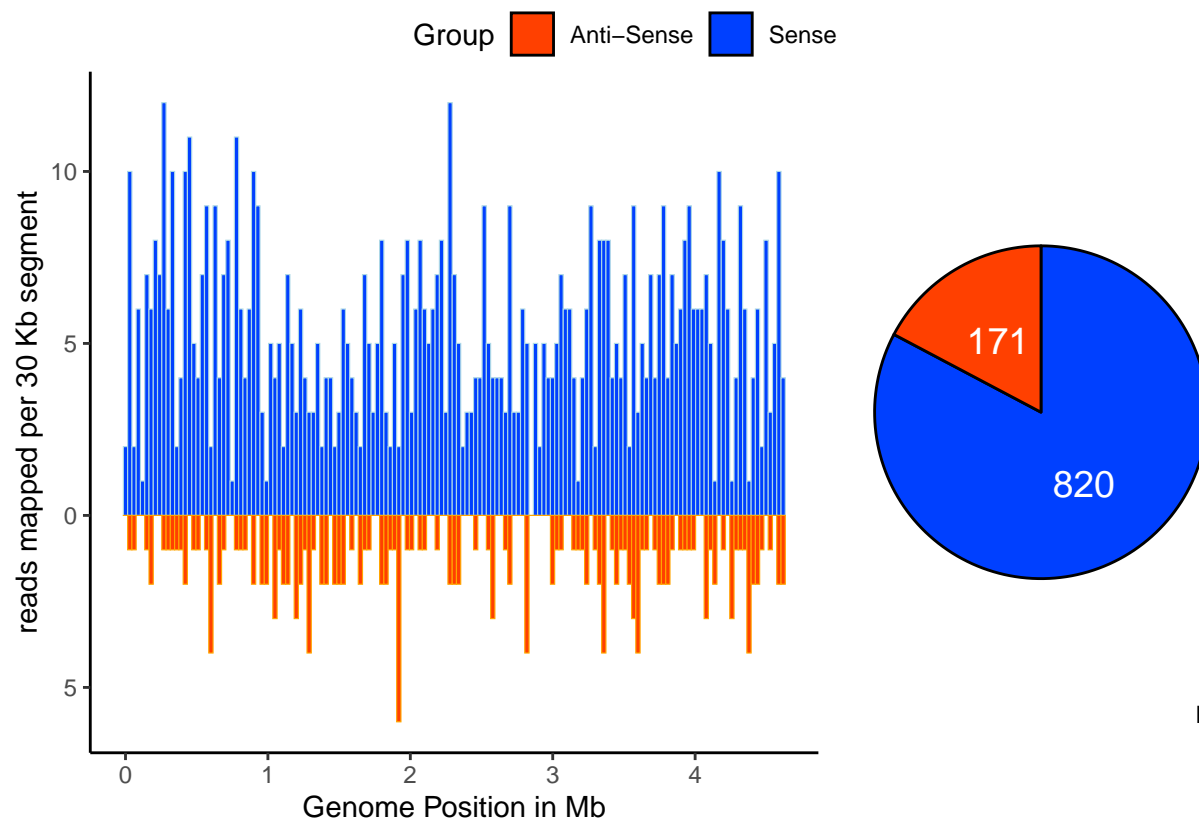
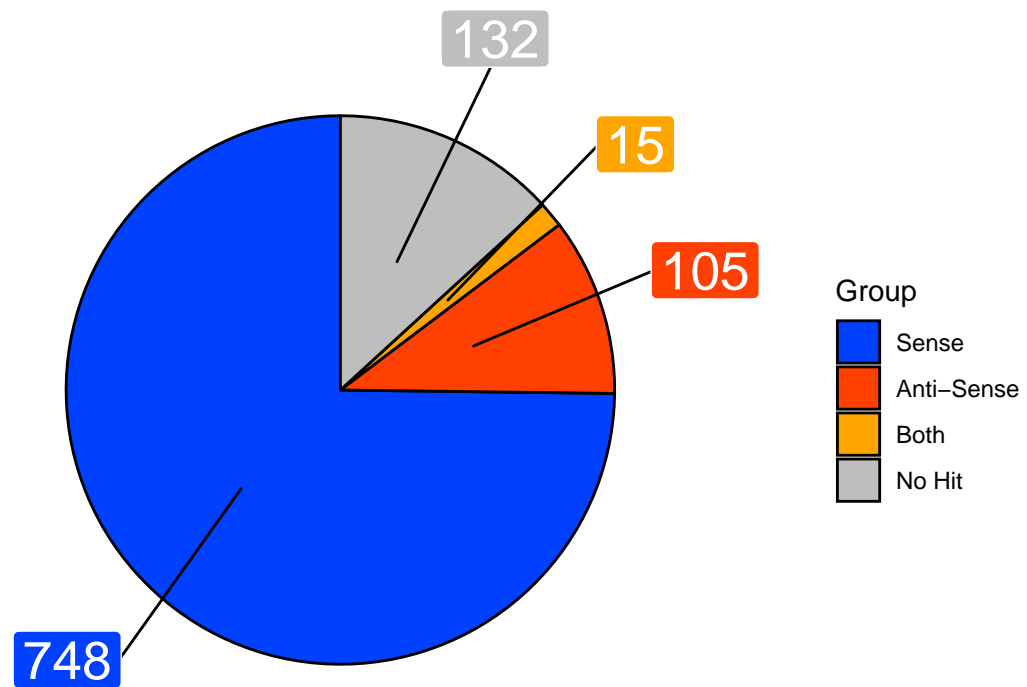


Fig 3

**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

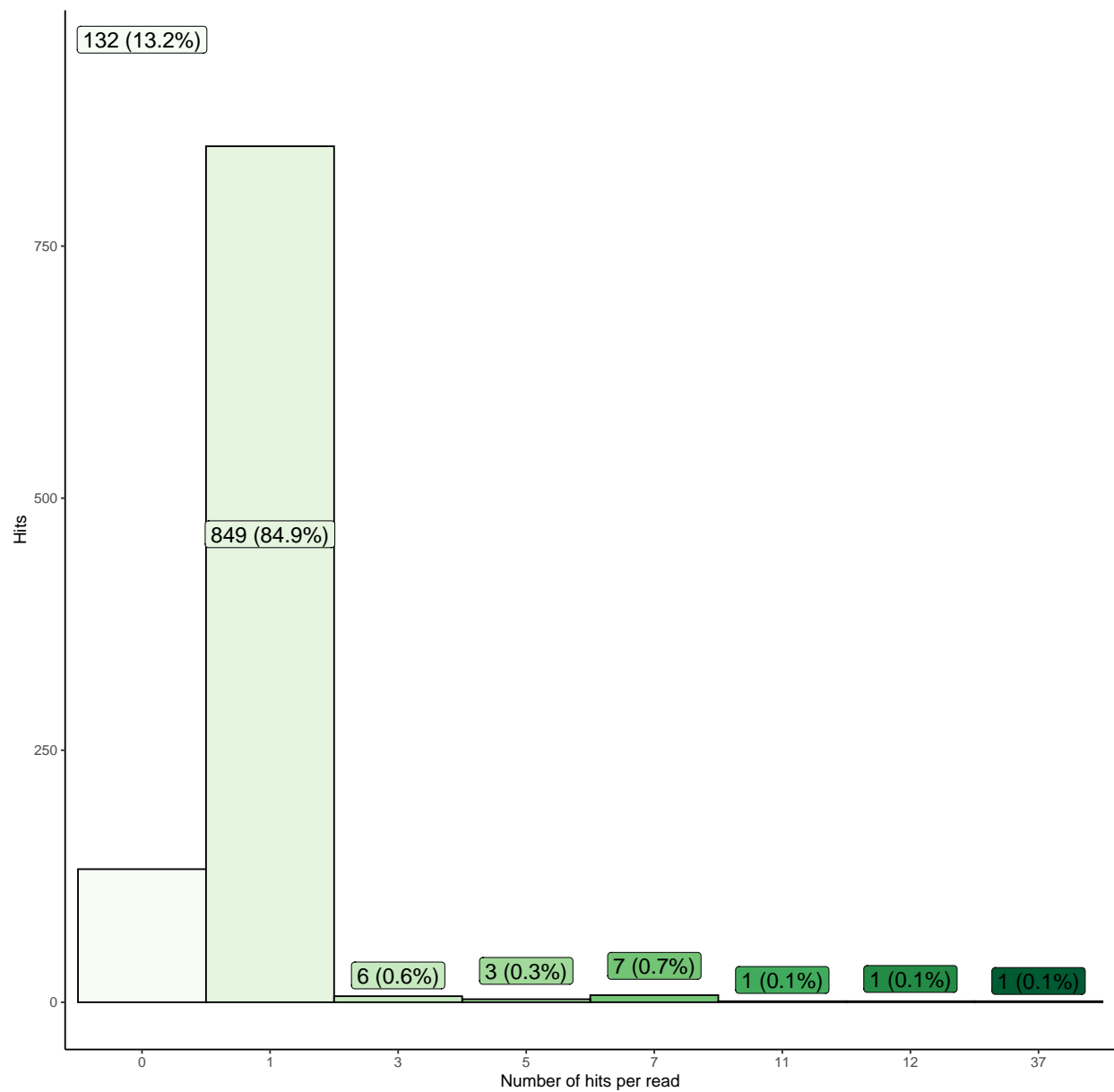


Fig 5

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