

Plot Mauve Alignements

IM

June 7, 2018

Synteny analysis of strains

Genome organisation can have strong impacts in gene expression and could tell us about the evolution of chromosomes. Synteny analyses, can also tell us about large differences between the strains measured. Notably, Genomic Islands (GI) can be detected by looking the synteny of the homolog region on two different strains.

A common way to measure the synteny between homolog regions, is to align the strains with the software MAUVE <http://darlinglab.org/mauve/mauve.html> The advantage of Mauve compared to other genome assemblers, is that Mauve do a Multiple genome alignment. meaning that in a group of strains to analyse, it will align all the strain to each other. This is in opposition to single genome alignments, that compared all strains to a Reference strain.

Mauve multiple sequence alignment

The software Mauve can be downloaded from <http://darlinglab.org/mauve/mauve.html>. It can be opened by opening the on the shell at its good location: `./Mauve`. Mauve came with a GUI that allow an easy use.

```
knitr::include_graphics("/media/imateus/USB_/UPBLO/RMarkdown/Screensho_Mauve_synteny.png")
```

To make a multiple alignment, you need to compare at least 3 sequences/regions/strains/species.

And then use the progressiveMauve algorithm to align the sequences. You need to find the sequences to compare and then specify the name of the outputs generated by Mauve.

After the analysis, Mauve, will show an interactive plot were you can observe the synteny between the sequences (Figure 1), zoom in/out and identify the coordinates of possible breakpoints.

Import Mauve data

```
# add libraries
library("genoPlotR")

## Loading required package: ade4
## Loading required package: grid
library(RColorBrewer)
# Set working directory
setwd("~/Documents/EPFL/A_baumannii/Reference_sequences/ALL_Abaum_vf/ALL_Analysis_V2/SyntenyATCC17978St)

align <- read_mauve_backbone("Synteny_all_ATCC17978.backbone")
```

Show summary of Mauve input

```
summary(align)
```

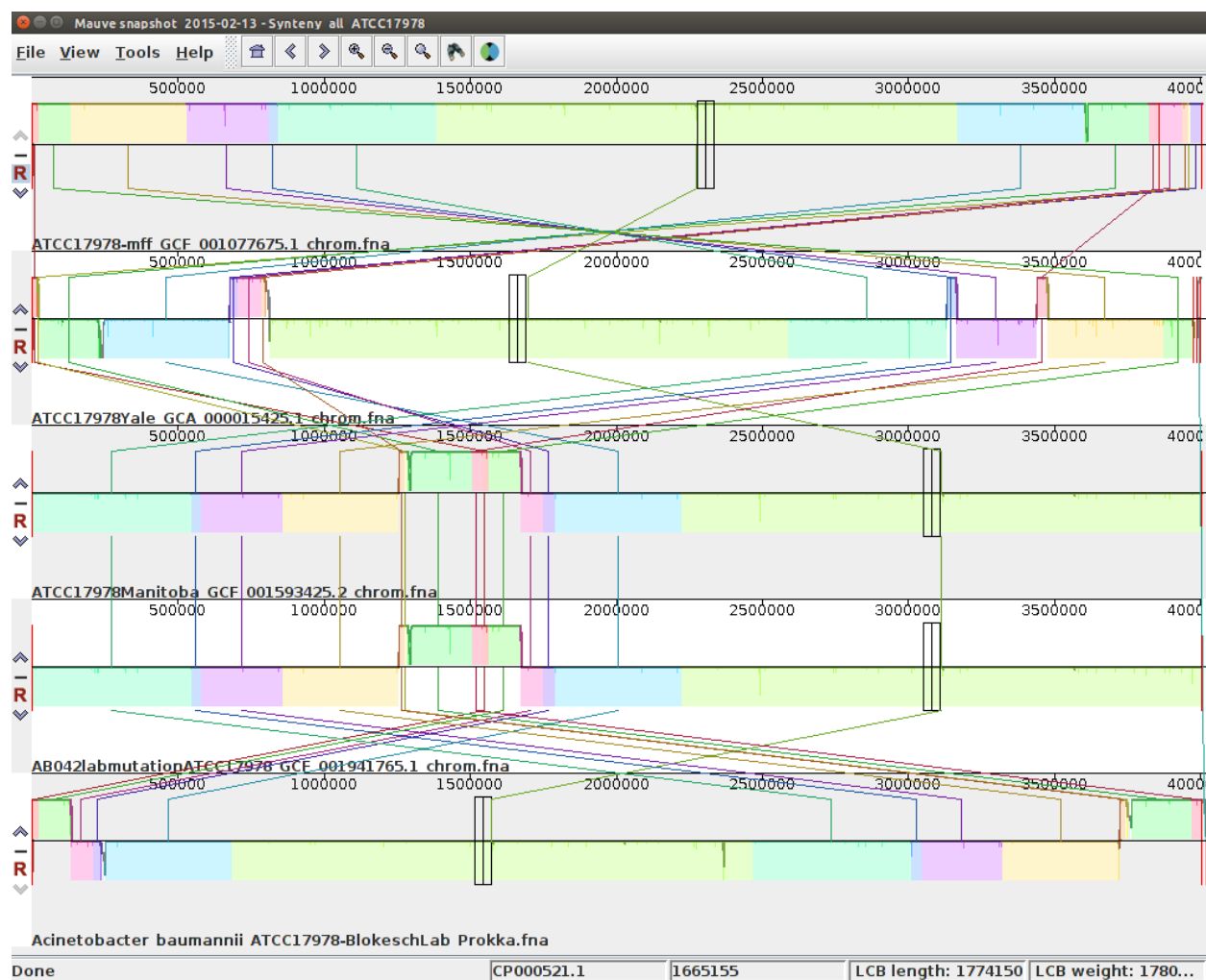


Figure 1: Screenshot of Mauve alignment. Comparison of ATCC17978 genome assemblies

```
##          Length Class Mode
## dna_segs    5    -none- list
## comparisons 4    -none- list
```

Change color parameters

```
colors <- rainbow(length(align[[1]][[1]]$col))

for (i in align[[1]][[1]]$col) {
  for (ca in 1:length(align[[1]])) {
    align[[1]][[ca]]$col[align[[1]][[ca]]$col == i] <- colors[match(i,
      align[[1]][[ca]]$col)]
  }
}
```

Change color of syntenic matches

```
for (i in 1:length(aligned[[2]])) {  
  aligned[[2]][[i]]$col <- rep("gray59", length(aligned[[2]][[i]]$col))  
}
```

Add Names

look at file order in bash # cat MAUVESYNTENY_OUTPUT | grep "#Sequence*.File" | cut -d'/' -f10 | cut -d'_' -f3 | cut -d'-' -f1

```
names <- c("ATCC17978-mff", "ATCC17978-Yale", "ATCC17978-Manitoba",  
  "AB042labmutation-ATCC17978", "ATCC17978-BlokeschLab")
```

Plot

```
plot_gene_map(aligned$dna_segs, aligned$comparisons, main = "ATCC17978 strains",  
  dna_seg_labels = names)
```

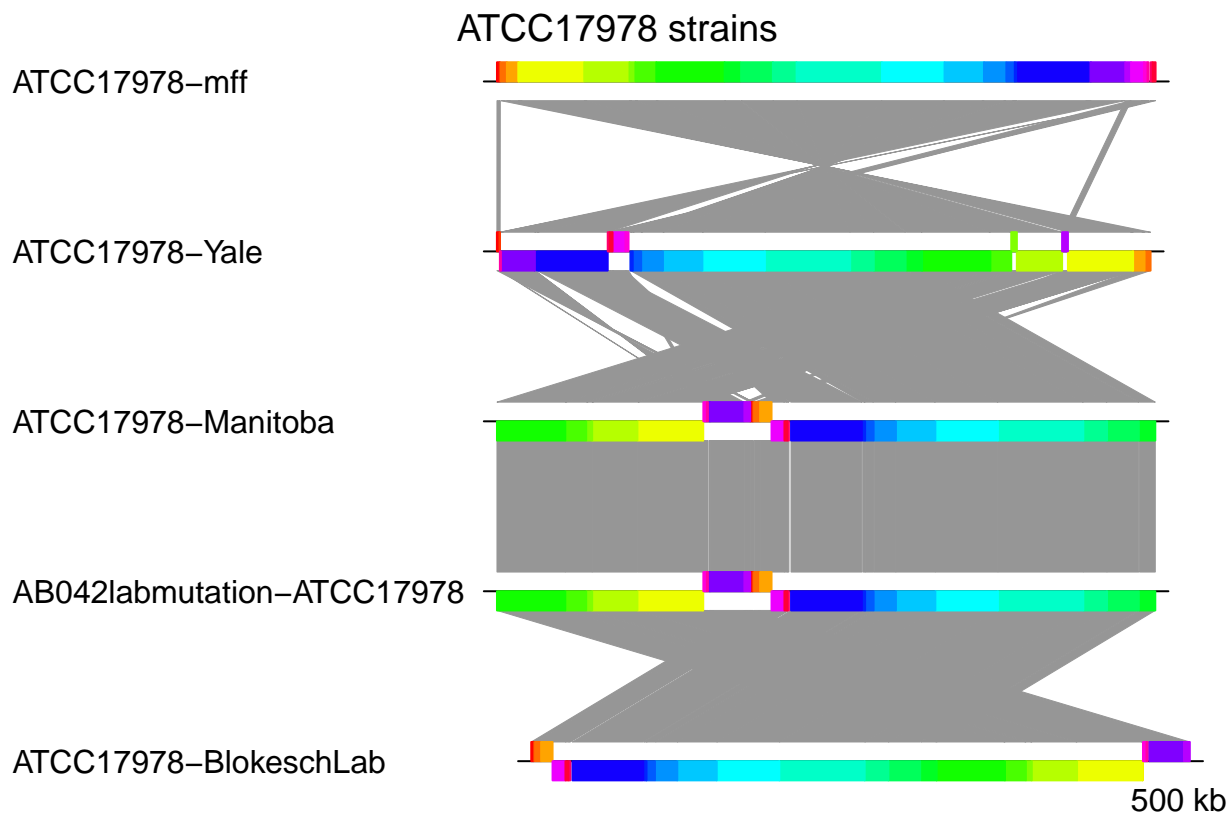


Figure 2: Aligement of ATCC17978 strains

Note that wide genomic rearrangements are observed between the different genome assemblies of the ATCC17978 strain (Figure 2). Could we discard a methodological bias? Is this the result of biological processes occurring in consequence of the culturing regime, or these are true variations that evidence genome plasticity?