# Report

Frederik Lutz

2023-05-08

Introduction

Phylogenetic trees are tools to assess the evolutionary relationships between organisms. This report compares the modeled trees of the bionformatic tools dRep and PhyloPhlan3 based on the genome of Treponema Succinifaciens as database.

Methods

The data was prepared by running dRrep (version: 3.4.3)(Olm et al. 2017) on it, which identifies and removes duplicate genomes and also generates the primary cluster dendogram based on Average Nucleotide Identity. This leads to better downstream processing due to less computational time needed. The resulting table contains all represented gene clusters, allowing one to use PhyloPhlAn3 (version: 3.0.67)(Segata et al. 2013). Due to on an additional list of marker genes for Treponemas Succinifaciens the program builds the best suited phylogenetic tree for the compilation of the previously reconstructed-/ SGB- and reference genomes. Additionally a R script can be used to increase the accuracy of the classification by creating genetic distances using different models as well as "Neighbour-joining" and creating another tree based on preselected high abundant marker genes (ammonium transporter, ABC transporter related protein.
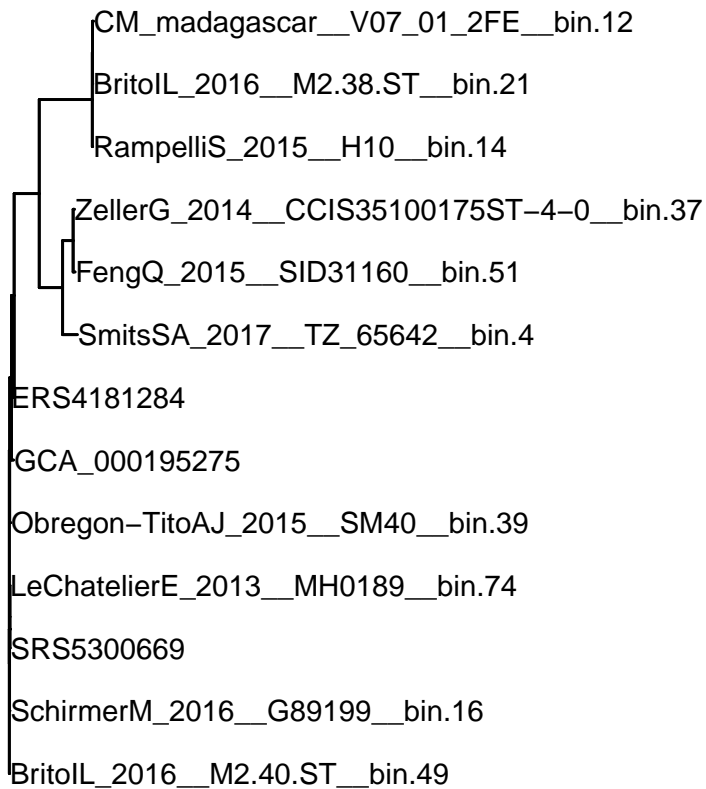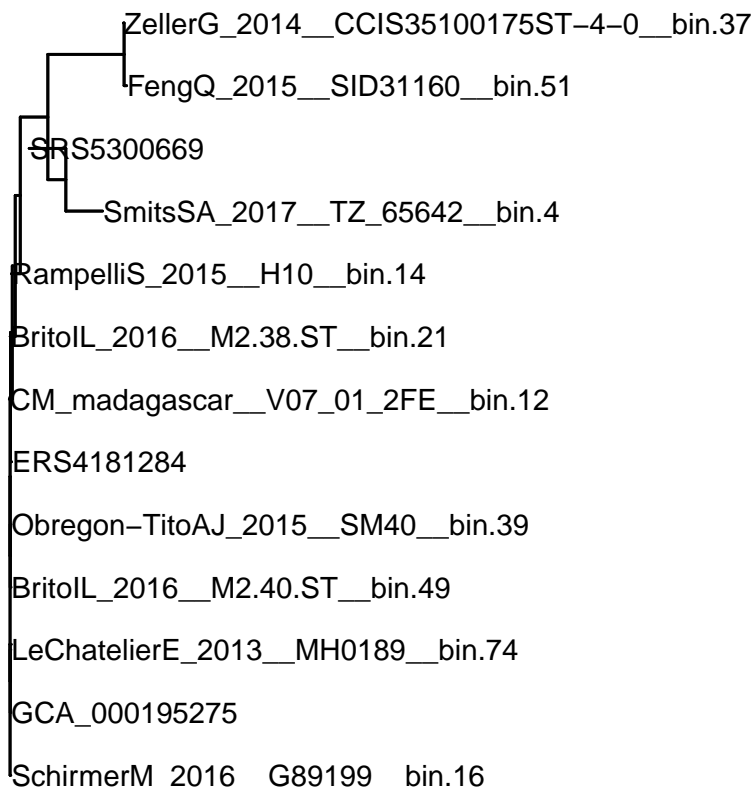
Results

MASH clustering
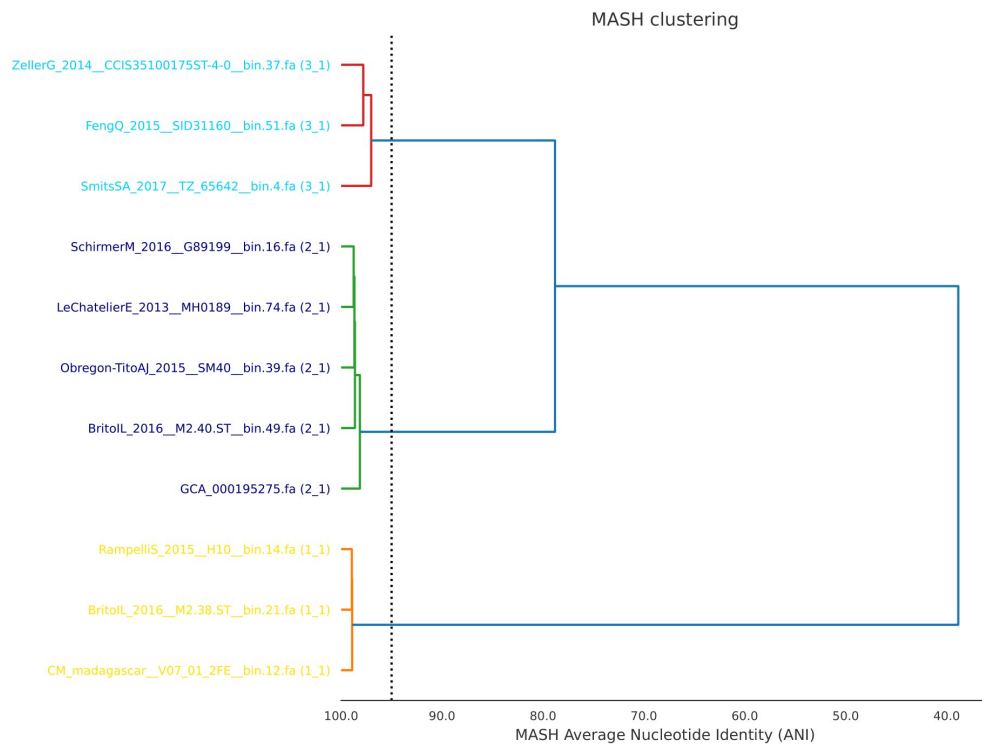
The dendrogram by dRep clusters the samples into three major groups.

```
##
## Phylogenetic tree with 13 tips and 11 internal nodes.
##
## Tip labels:
##   BritoIL_2016__M2.38.ST__bin.21, ERS4181284, FengQ_2015__SID31160__bin.51, BritoIL_2016__M2.40.ST__l
##
## Unrooted; includes branch lengths.

## List of 4
##  $ edge       : int [1:23, 1:2] 14 14 16 17 18 19 20 21 21 20 ...
##  $ edge.length: num [1:23] 0.00121 0.00129 0.00654 0.00973 0.09109 ...
##  $ tip.label  : chr [1:13] "BritoIL_2016__M2.38.ST__bin.21" "ERS4181284" "FengQ_2015__SID31160__bin.5
##  $ Nnode      : int 11
##  - attr(*, "class")= chr "phylo"
##  - attr(*, "order")= chr "cladewise"
```

CM_madagascar__V07_01_2FE__bin.12

BritoIL_2016__M2.38.ST__bin.21

RampelliS_2015__H10__bin.14

ZellerG_2014__CCIS35100175ST−4−0__bin.37

FengQ_2015__SID31160__bin.51

SmitsSA_2017__TZ_65642__bin.4

ERS4181284

GCA_000195275

Obregon−TitoAJ_2015__SM40__bin.39

LeChatelierE_2013__MH0189__bin.74

SRS5300669

SchirmerM_2016__G89199__bin.16

BritoIL_2016__M2.40.ST__bin.49

The tree using the R script distancing based on the ammonium transporter gene largely distinguishes betweeen two clusters of which one is further subdivided in two distinct groups including three taxa.

ZellerG_2014__CCIS35100175ST−4−0__bin.37

FengQ_2015__SID31160__bin.51

SRS5300669

SmitsSA_2017__TZ_65642__bin.4

RampelliS_2015__H10__bin.14

BritoIL_2016__M2.38.ST__bin.21

CM_madagascar__V07_01_2FE__bin.12

ERS4181284

Obregon−TitoAJ_2015__SM40__bin.39

BritoIL_2016__M2.40.ST__bin.49

LeChatelierE_2013__MH0189__bin.74

GCA_000195275

SchirmerM_2016__G89199__bin.16

The tree using the R script distancing based on the ABC transporter related protein divides the tree in two

larger clusters of which one is subdivided into two groups containing two taxa.

Discussion

All three trees cluster the Schirmer, LeChatelier, Obregon, GCA, BritoIL_2016___M2.40.ST___bin.49 samples together. ZellerG-, Smits- and FenQ- samples match the classification in the MASH and first R- script generated tree, as well as CM_madagaskar-, -Britol and Rampelli samples. The results show that by determining the genetic distance by modelling with R, the phylogenetic tree get mor precise.

References

Olm, Matthew R., Christopher T. Brown, Brandon Brooks, and Jillian F. Banfield. 2017. "dRep: A Tool for Fast and Accurate Genomic Comparisons That Enables Improved Genome Recovery from Metagenomes Through de-Replication." *The ISME Journal* 11 (12): 2864–8. https://doi.org/10.1038/ismej.2017.126.

Segata, Nicola, Daniela Börnigen, Xochitl C. Morgan, and Curtis Huttenhower. 2013. "PhyloPhlAn Is a New Method for Improved Phylogenetic and Taxonomic Placement of Microbes." *Nature Communications* 4 (1): 2304. https://doi.org/10.1038/ncomms3304.