Figure 4. Assembly-based core-SNP phylogenetic tree and clade distribution of nontypeable *H. influenzae*

Anna Carrera-Salinas, Aida González-Díaz, Laura Calatayud, Julieta Mercado-Maza, Carmen Puig, Dàmaris Berbel, Jordi Càmara, Fe Tubau, Imma Grau, M Ángeles Domínguez, Carmen Ardanuy, Sara Martí

2021-07-22

```r
library(ggplot2)
library(ggtree)
```

```
## ggtree v2.2.4  For help: https://yulab-smu.github.io/treedata-book/
##
## If you use ggtree in published research, please cite the most appropriate paper(s):
##
## Guangchuang Yu. Using ggtree to visualize data on tree-like structures. Current Protocols
in Bioinformatics, 2020, 69:e96. doi:10.1002/cpbi.96
## Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods for mapping and visu
alizing associated data on phylogeny using ggtree. Molecular Biology and Evolution 2018, 35(1
2):3041-3043. doi:10.1093/molbev/msy194
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam. ggtree: an R packag
e for visualization and annotation of phylogenetic trees with their covariates and other asso
ciated data. Methods in Ecology and Evolution 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
```

```r
library(extrafontdb)
library(extrafont)
```

```
## Registering fonts with R
```

```r
loadfonts(device = "win", quiet = TRUE)
```

1.Load data

```r
##Define the working directory
setwd("path/to/file")

##Load the tree file obtained using Parsnp from the Harvest suite (.nwk file)
tree<-read.tree("treefile.nwk")
```

2.Tree and metadata

```
##Visualize the tree
p <- ggtree(tree,size=0.2,right=TRUE,branch.length = "none") +
  geom_text(aes(label=node),check_overlap=TRUE)+
  theme(text=element_text(family="Arial Nova Light"))

#Rotate the tree branches
p2<- flip(p,214,226)%>% flip(413,295)%>% flip(144,145)%>% flip(88,302)%>% flip(296,298)
p2$layers[[3]] <- NULL
p3 <- rotate(p2, 294) %>% rotate(320)%>% rotate(348)%>% rotate(335)

#Add metadata
metadata<-read.csv("metadata.csv",
                   header=TRUE, sep = ";")
heatmapData=read.csv("metadata.csv",sep = ";", row.names=1)
rn <- rownames(heatmapData)
heatmapData <- as.data.frame(sapply(heatmapData, as.character))
rownames(heatmapData) <- rn
```

3.Construction of the phylogenetic tree, including metadata and bootstrap values

```
#Create the tree with metadata
tree_metadata<-gheatmap(p3, heatmapData, offset = 2,
                        width=5,
                        colnames_position="top",
                        colnames_angle=90,
                        hjust=0,
                        color="lightgrey",
                        font.size=3.5,
                        family="Arial Nova Light") +
  geom_treescale(x=0.05, y=230, offset=2, fontsize = 4,family="Arial Nova Light")+
  geom_tiplab(align=TRUE,linetype='dotted',linesize=.25,color="#525252",size=1)+
  geom_nodelab(aes(label=label,x=branch),size=1,vjust=-.5)+
  geom_point2(aes(subset=node==226), color="#66c2a5", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==303), color="#9e0142", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==336), color="#e6f598", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==359), color="#f46d43", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==348), color="#fee08b", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==145), color="#f46d43", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==144), color="#fee08b", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==300), color="#5e4fa2", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==317), color="#5e4fa2", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==380), color="#5e4fa2", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==296), color="#5e4fa2", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==321), color="#5e4fa2", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==413), color="#5e4fa2", alpha=0.5, size=5)+
  geom_point2(aes(subset=node==214), color="#5e4fa2", alpha=0.5, size=5)+
  scale_y_continuous(expand=c(0, 20))+
  scale_fill_manual(name="",values = c("#BA6261", "#65B0CB", "#A1D6DD","#9e0142", "#f46d43",
"#fee08b","#e6f598","#66c2a5", "#5e4fa2","red","grey", "white"),limits=c("Carrera-Salinas",
"De Chiara", "Pinto","I","II","III","IV","V","VI","Unclassified","Presence","Abscence"))+
  theme(legend.position="none")
```

Figure 4.Assembly-based core-SNP phylogenetic tree and clade distribution of nontypeable *H. influenzae*