Patristic distances analysis

Noemi Navarro Lleó

26/05/2022

The following code shows the approach used in the paper in order prove that the proposed norovirus sequences do not belong to the GI.3 genotype.

```
#knitr::opts_knit$set(root.dir = '~/arboles_GI_VP1')
knitr::opts_chunk$set(echo = TRUE)
library(adephylo)
```

```
## Loading required package: ade4
```

```
library(ggplot2)
library(GLDEX)
```

```
## Loading required package: cluster
```

```
library(stringr)
```

Patristic distances calculation and groups comparisons

Below, we generate a table including the patristic distances that were calculated on the tree provided in the paper.

```
tree_cluster<- ape::read.tree("./GIVP1Dispersion.nwk")
#print(tree_cluster)
d_cluster <- distTips(tree_cluster, tips = "all", method = ("patristic"), useC = TRUE)
write.csv(as.matrix(d_cluster), "tabla_noemi_patristic_completa_cluster.csv")
tabla_completa_gaps <- read.csv("tabla_noemi_patristic_completa_cluster.csv", sep=",", row.names = 1)

#list(colnames(tabla_completa_gaps))

comparacion_gi3gi3 <- tabla_completa_gaps[c("GI.3/CFIA_FV_448_1b_43479_8/QRQ46942",
"GI.3/CFIA_FV_340_1.3_43479_1234/QRM14330", "GI.3/FE29_2015/BBD74616",
"GI.3/FE47_2015/BBD74618", "GI.3/NORO_90_22_06_2015/AWR17846", "GI.3/FE8_2015/BBD74614",
"GI.3/FS118_2015/BBD74626", "GI.3/FE2_2015/BBD74612",
"GI.3/FE59_2015/BBD74620", "GI.3/FS114_2015/BBD74624",
"GI.3/FE49_2014/BBD74608","GI.3/NORO_81_10_10_2014/AWR17825",
"GI.3/NORO_83_31_10_2014/AWR17831", "GI.3/NORO_87_18_02_2015/AWR17840",
"GI.3/NORO_84_04_11_2014/AWR17834", "GI.3/NORO_85_14_11_2014/AWR17837",
"GI.3/NORO_179_03_08_2015/AWR17600", "GI.3/NORO_206_03_12_2015/AWR17651",
```

```
"GI.3/FE61_2015/BBD74622", "GI.3/CFIA_FV_491_16_43538_5/QRM14429",
"GI.3/CFIA_FV_478_13_43479_9/QRQ46993","GI.3/CFIA_FV_491_23_43538_14/QRM14432",
"GI.3/NORO_89_01_04_2015/AWR17843","GI.3/NORO_198_10_11_2015/AWR17636",
"GI.3/FS421_2014/BBD74610" ,"GI.3/CS0048/QHW16218",
"GI.3/JKPG_881/ACX33982","GI.3/JKPG_883/ACX33983"
,"GI.3/0304_19/ARI71147","GI.3/Potsdam_196/AAL32455",
"GI.3/Shimizu/KK2866/AII73765","GI.3/Westport0085/QEP29903",
"GI.3/Osaka/010314/ABQ44563","GI.3/Birmingham/CAB89093",
"GI.3/VA98115/AAK84673","GI.3/AFK75851",
"GI.3/Little_Rock/316/AAL12968","GI.3/Honolulu/219/AAL12962",
"GI.3/Stav/95/AAD37377","GI.3/Amsterdam/2/1995/AHV83751",
"GI.3/EN_3/AOO95034","GI.3/DesertShield395/AAA16285",
"GI.3/Beijing55042/ACY00647","GI.3/CFIA_FVR_002/QRD99198",
"GI.3/C9/AFN06738","GI.3/C91/AFN06737","GI.3/B8/QPJ58810",
"GI.3/18N239/QQL02664","GI.3/Nashville_0047/AZJ17766",
"GI.3/G2_12/QXD98775","GI.3/BayCity_0088/AZJ17748",
"GI.3/G5_10/QXD98781","GI.3/G7_8/QXD98784",
"GI.3/Nashville_0046/AZJ17763","GI.3/G5_9/QXD98778"),
c("GI.3.CFIA_FV_448_1b_43479_8.QRQ46942","GI.3.CFIA_FV_340_1.3_43479_1234.QRM14330",
"GI.3.FE29_2015.BBD74616", "GI.3.FE47_2015.BBD74618",
"GI.3.NORO_90_22_06_2015.AWR17846", "GI.3.FE8_2015.BBD74614",
"GI.3.FS118_2015.BBD74626", "GI.3.FE2_2015.BBD74612",
"GI.3.FE59_2015.BBD74620", "GI.3.FS114_2015.BBD74624",
"GI.3.FE49_2014.BBD74608","GI.3.NORO_81_10_10_2014.AWR17825",
"GI.3.NORO_83_31_10_2014.AWR17831", "GI.3.NORO_87_18_02_2015.AWR17840",
"GI.3.NORO_84_04_11_2014.AWR17834", "GI.3.NORO_85_14_11_2014.AWR17837",
"GI.3.NORO_179_03_08_2015.AWR17600", "GI.3.NORO_206_03_12_2015.AWR17651",
"GI.3.FE61_2015.BBD74622", "GI.3.CFIA_FV_491_16_43538_5.QRM14429",
"GI.3.CFIA_FV_478_13_43479_9.QRQ46993","GI.3.CFIA_FV_491_23_43538_14.QRM14432",
"GI.3.NORO_89_01_04_2015.AWR17843","GI.3.NORO_198_10_11_2015.AWR17636" ,
"GI.3.FS421_2014.BBD74610" ,"GI.3.CS0048.QHW16218",
"GI.3.JKPG_881.ACX33982","GI.3.JKPG_883.ACX33983",
"GI.3.0304_19.ARI71147","GI.3.Potsdam_196.AAL32455",
"GI.3.Shimizu.KK2866.AII73765","GI.3.Westport0085.QEP29903",
"GI.3.Osaka.010314.ABQ44563","GI.3.Birmingham.CAB89093",
"GI.3.VA98115.AAK84673","GI.3.AFK75851",
"GI.3.Little_Rock.316.AAL12968","GI.3.Honolulu.219.AAL12962",
"GI.3.Stav.95.AAD37377","GI.3.Amsterdam.2.1995.AHV83751",
"GI.3.EN_3.AOO95034","GI.3.DesertShield395.AAA16285",
"GI.3.Beijing55042.ACY00647","GI.3.CFIA_FVR_002.QRD99198",
"GI.3.C9.AFN06738","GI.3.C91.AFN06737",
"GI.3.B8.QPJ58810","GI.3.18N239.QQL02664",
"GI.3.Nashville_0047.AZJ17766","GI.3.G2_12.QXD98775",
"GI.3.BayCity_0088.AZJ17748","GI.3.G5_10.QXD98781",
"GI.3.G7_8.QXD98784","GI.3.Nashville_0046.AZJ17763","GI.3.G5_9.QXD98778"),]
comparacion_gi3gi3 <- unlist(comparacion_gi3gi3,use.names=FALSE)
comparacion_gi3gi3 <- fun.zero.omit(comparacion_gi3gi3)
comparacion_gi3gi3 <- unique(comparacion_gi3gi3)
media_gi3gi3 <- mean(comparacion_gi3gi3)
sd_gi3gi3 <- sd(comparacion_gi3gi3)
```

```
comparacion_gi3gina1 <- tabla_completa_gaps[c("GI.3/CFIA_FV_448_1b_43479_8/QRQ46942",
"GI.3/CFIA_FV_340_1.3_43479_1234/QRM14330", "GI.3/FE29_2015/BBD74616",
"GI.3/FE47_2015/BBD74618", "GI.3/NORO_90_22_06_2015/AWR17846", "GI.3/FE8_2015/BBD74614",
"GI.3/FS118_2015/BBD74626", "GI.3/FE2_2015/BBD74612",
"GI.3/FE59_2015/BBD74620", "GI.3/FS114_2015/BBD74624",
"GI.3/FE49_2014/BBD74608","GI.3/NORO_81_10_10_2014/AWR17825",
"GI.3/NORO_83_31_10_2014/AWR17831", "GI.3/NORO_87_18_02_2015/AWR17840",
"GI.3/NORO_84_04_11_2014/AWR17834", "GI.3/NORO_85_14_11_2014/AWR17837",
"GI.3/NORO_179_03_08_2015/AWR17600", "GI.3/NORO_206_03_12_2015/AWR17651",
"GI.3/FE61_2015/BBD74622", "GI.3/CFIA_FV_491_16_43538_5/QRM14429",
"GI.3/CFIA_FV_478_13_43479_9/QRQ46993", "GI.3/CFIA_FV_491_23_43538_14/QRM14432",
"GI.3/NORO_89_01_04_2015/AWR17843","GI.3/NORO_198_10_11_2015/AWR17636" ,
"GI.3/FS421_2014/BBD74610" ,"GI.3/CS0048/QHW16218",
"GI.3/JKPG_881/ACX33982","GI.3/JKPG_883/ACX33983",
"GI.3/0304_19/ARI71147","GI.3/Potsdam_196/AAL32455",
"GI.3/Shimizu/KK2866/AII73765","GI.3/Westport0085/QEP29903",
"GI.3/Osaka/010314/ABQ44563","GI.3/Birmingham/CAB89093",
"GI.3/VA98115/AAK84673","GI.3/AFK75851",
"GI.3/Little_Rock/316/AAL12968","GI.3/Honolulu/219/AAL12962",
"GI.3/Stav/95/AAD37377","GI.3/Amsterdam/2/1995/AHV83751",
"GI.3/EN_3/AOO95034","GI.3/DesertShield395/AAA16285",
"GI.3/Beijing55042/ACY00647","GI.3/CFIA_FVR_002/QRD99198",
"GI.3/C9/AFN06738","GI.3/C91/AFN06737","GI.3/B8/QPJ58810",
"GI.3/18N239/QQL02664","GI.3/Nashville_0047/AZJ17766",
"GI.3/G2_12/QXD98775","GI.3/BayCity_0088/AZJ17748",
"GI.3/G5_10/QXD98781","GI.3/G7_8/QXD98784",
"GI.3/Nashville_0046/AZJ17763","GI.3/G5_9/QXD98778"),
c("GI.3.20200221_ASE_08_GI.3_LPV.QUF08421","GI.3.20200226_MWE_01_GI.3_LPV.QUF08423",
"GI.3.20200408_MWE_01_GI.3_LPV.QUF08432", "GI.3.20200416_ASE_06_GI.3_LPV.QUF08433",
"GI.3.10360.AFI08231", "GI.3.Beijing54660.ACY00645",
"GI.3.E8_UG.AFN06736", "GI.3.Akabane.991130.ABQ44566", "GI.3.Beijing54114.ACY00641")]
comparacion_gi3gina1 <- unlist(comparacion_gi3gina1,use.names=FALSE)
media_gi3gina1 <- mean(comparacion_gi3gina1)
sd_gi3gina1 <- sd(comparacion_gi3gina1)

comparacion_gi3gina2 <- tabla_completa_gaps[c("GI.3/CFIA_FV_448_1b_43479_8/QRQ46942",
"GI.3/CFIA_FV_340_1.3_43479_1234/QRM14330",
"GI.3/FE29_2015/BBD74616", "GI.3/FE47_2015/BBD74618",
"GI.3/NORO_90_22_06_2015/AWR17846", "GI.3/FE8_2015/BBD74614",
"GI.3/FS118_2015/BBD74626", "GI.3/FE2_2015/BBD74612",
"GI.3/FE59_2015/BBD74620", "GI.3/FS114_2015/BBD74624",
"GI.3/FE49_2014/BBD74608","GI.3/NORO_81_10_10_2014/AWR17825",
"GI.3/NORO_83_31_10_2014/AWR17831", "GI.3/NORO_87_18_02_2015/AWR17840",
"GI.3/NORO_84_04_11_2014/AWR17834", "GI.3/NORO_85_14_11_2014/AWR17837",
"GI.3/NORO_179_03_08_2015/AWR17600", "GI.3/NORO_206_03_12_2015/AWR17651",
"GI.3/FE61_2015/BBD74622", "GI.3/CFIA_FV_491_16_43538_5/QRM14429",
"GI.3/CFIA_FV_478_13_43479_9/QRQ46993", "GI.3/CFIA_FV_491_23_43538_14/QRM14432",
"GI.3/NORO_89_01_04_2015/AWR17843","GI.3/NORO_198_10_11_2015/AWR17636" ,
"GI.3/FS421_2014/BBD74610" ,"GI.3/CS0048/QHW16218",
"GI.3/JKPG_881/ACX33982","GI.3/JKPG_883/ACX33983",
"GI.3/0304_19/ARI71147","GI.3/Potsdam_196/AAL32455",
"GI.3/Shimizu/KK2866/AII73765","GI.3/Westport0085/QEP29903",
"GI.3/Osaka/010314/ABQ44563","GI.3/Birmingham/CAB89093",
```

```
"GI.3/VA98115/AAK84673","GI.3/AFK75851",
"GI.3/Little_Rock/316/AAL12968","GI.3/Honolulu/219/AAL12962",
"GI.3/Stav/95/AAD37377","GI.3/Amsterdam/2/1995/AHV83751",
"GI.3/EN_3/AOO95034","GI.3/DesertShield395/AAA16285",
"GI.3/Beijing55042/ACY00647","GI.3/CFIA_FVR_002/QRD99198",
"GI.3/C9/AFN06738","GI.3/C91/AFN06737","GI.3/B8/QPJ58810",
"GI.3/18N239/QQL02664","GI.3/Nashville_0047/AZJ17766",
"GI.3/G2_12/QXD98775","GI.3/BayCity_0088/AZJ17748",
"GI.3/G5_10/QXD98781","GI.3/G7_8/QXD98784",
"GI.3/Nashville_0046/AZJ17763","GI.3/G5_9/QXD98778"),
c("GI.3.OS_32.LC122713", "GI.3.DS275.QPJ58855", "GI.NA.3718VLC.Spain.ON033826")]
comparacion_gi3gina2 <- unlist(comparacion_gi3gina2,use.names=FALSE)
media_gi3gina2 <- mean(comparacion_gi3gina2)
sd_gi3gina2 <- sd(comparacion_gi3gina2)


comparacion_gina1gna1 <- tabla_completa_gaps[c("GI.3/20200221_ASE_08_GI.3_LPV/QUF08421",
"GI.3/20200226_MWE_01_GI.3_LPV/QUF08423", "GI.3/20200408_MWE_01_GI.3_LPV/QUF08432",
"GI.3/20200416_ASE_06_GI.3_LPV/QUF08433", "GI.3/10360/AFI08231",
"GI.3/Beijing54660/ACY00645", "GI.3/E8_UG/AFN06736",
"GI.3/Akabane/991130/ABQ44566", "GI.3/Beijing54114/ACY00641"),
c("GI.3.20200221_ASE_08_GI.3_LPV.QUF08421","GI.3.20200226_MWE_01_GI.3_LPV.QUF08423",
"GI.3.20200408_MWE_01_GI.3_LPV.QUF08432", "GI.3.20200416_ASE_06_GI.3_LPV.QUF08433",
"GI.3.10360.AFI08231", "GI.3.Beijing54660.ACY00645",
"GI.3.E8_UG.AFN06736", "GI.3.Akabane.991130.ABQ44566", "GI.3.Beijing54114.ACY00641")]
comparacion_gina1gna1 <- unlist(comparacion_gina1gna1,use.names=FALSE)
comparacion_gina1gna1 <- fun.zero.omit(comparacion_gina1gna1)
comparacion_gina1gna1 <- unique(comparacion_gina1gna1)
media_gina1gina1 <- mean(comparacion_gina1gna1)
sd_gina1gina1 <- sd(comparacion_gina1gna1)

comparacion_gina2gna2 <- tabla_completa_gaps[c("GI.3/DS275/QPJ58855",
"GI.3/OS_32/LC122713", "GI.NA/3718VLC/Spain/ON033826"),
c("GI.3.DS275.QPJ58855", "GI.3.OS_32.LC122713", "GI.NA.3718VLC.Spain.ON033826")]
comparacion_gina2gna2 <- unlist(comparacion_gina2gna2,use.names=FALSE)
comparacion_gina2gna2 <- fun.zero.omit(comparacion_gina2gna2)
comparacion_gina2gna2 <- unique(comparacion_gina2gna2)
media_gina2gna2 <- mean(comparacion_gina2gna2)
sd_gina2gna2 <- sd(comparacion_gina2gna2)


comparacion_gina1gina2 <- tabla_completa_gaps[c("GI.3/20200221_ASE_08_GI.3_LPV/QUF08421",
"GI.3/20200226_MWE_01_GI.3_LPV/QUF08423", "GI.3/20200408_MWE_01_GI.3_LPV/QUF08432",
"GI.3/20200416_ASE_06_GI.3_LPV/QUF08433", "GI.3/10360/AFI08231",
"GI.3/Beijing54660/ACY00645", "GI.3/E8_UG/AFN06736",
"GI.3/Akabane/991130/ABQ44566", "GI.3/Beijing54114/ACY00641"),
c("GI.3.DS275.QPJ58855", "GI.3.OS_32.LC122713", "GI.NA.3718VLC.Spain.ON033826")]
comparacion_gina1gina2 <- unlist(comparacion_gina1gina2,use.names=FALSE)
media_gina1gina2 <- mean(comparacion_gina1gina2)
sd_gina1gina2 <- sd(comparacion_gina1gina2)

comparaciones <- c("GI.3-to-GI.3", "GI.3-to-GI.NA1", "GI.3-to-GI.NA2",
"GI.NA1-to-GI.NA1", "GI.NA1-to-GI.NA2","GI.NA2-to-GI.NA2")
```

```
media <- c(media_gi3gi3 ,media_gi3gina1, media_gi3gina2, media_gina1gina1, media_gina1gina2, media_gina
sd <- c(sd_gi3gi3 ,sd_gi3gina1, sd_gi3gina2, sd_gina1gina1, sd_gina1gina2, sd_gina2gna2)
sd <- as.numeric(sd)

df <- data.frame(comparaciones, media, sd)
#class(df$sd)
```

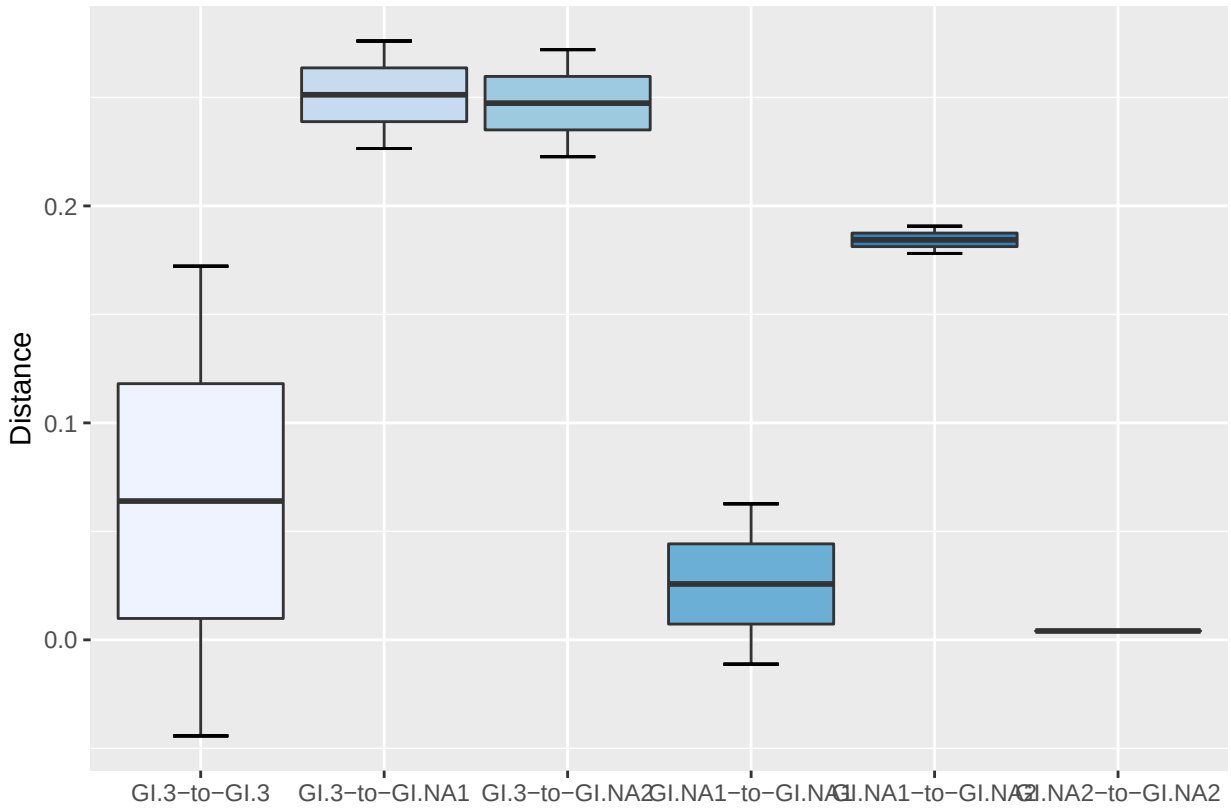
Finally, after calculating the mean and standard deviations of the patristic distances within and between groups, they were plot using a box and whisker chart type, where the bold stripe represents the mean value, box edges representing 1X SD and error bars 2X SD.

```
boxplot <- ggplot(df,aes(x=comparaciones, fill = comparaciones), show.legend=TRUE) + geom_boxplot(aes(l

r <- boxplot + scale_fill_brewer(palette="Blues") + list (
  theme(legend.position = "none"),
  ylab("Distance"),
  xlab(""))



#r + geom_hline(yintercept = 0.1722001, x=1:2, linetype="dashed", color = "red")
r <- r + geom_segment(aes(x=0.85,xend=1.15,y=0.1722001,yend=0.1722001))
r <- r + geom_segment(aes(x=0.85,xend=1.15,y=-0.04424756,yend=-0.04424756))
r <- r + geom_segment(aes(x=3.85,xend=4.15,y=0.0627412,yend=0.0627412 ))
r <- r + geom_segment(aes(x=3.85,xend=4.15,y=-0.0112166,yend=-0.0112166))
r <- r + geom_segment(aes(x=1.85,xend=2.15,y=0.2760033,yend=0.2760033))
r <- r + geom_segment(aes(x=1.85,xend=2.15,y=0.2264572,yend=0.2264572))
r <- r + geom_segment(aes(x=2.85,xend=3.15,y=0.2720274,yend=0.2720274))
r <- r + geom_segment(aes(x=2.85,xend=3.15,y=0.2226998 ,yend=0.2226998))
r <- r + geom_segment(aes(x=4.85,xend=5.15,y=0.1906248,yend=0.1906248))
r <- r + geom_segment(aes(x=4.85,xend=5.15,y=0.1781154 ,yend=0.1781154))
r
```




```
#ggsave(file="patristic.pdf", width=7, height=4.32, dpi=500)
```

To sum up, we can see that the two proposed genotypes meet the 2X SD criteria, since the mean values of the between patristic distances (with respect to GI.3) of the two proposed groups minus two times the standard deviations do not overlap with GI.3 mean within plus two times SD -0.2264572(GI.3-GI.NA1) and 0.2226998 (GI.3-GI.NA2) $>$ 0.1722001 (GI.3-GI.3)-. Likewise, between and within distances of GI3.NA1 and GI.NA2 do not overlap -0.1781154 (GI.NA1-GI.NA2) $>$ -0.0112166 (GI.NA1-GI.NA1) and 0.1781154 (GI.NA1-GI.NA2) $>$ 0.0041185 (GI.NA2-GI.NA2) -.