

Stage_M2_NB_1_ACP

Nicolas Bettembourg

2024-04-19

Chargement des packages R

Lien vers les scripts bash et commandes utilisées issus de la publication de Wragg et al., 2022 : <https://github.com/avignal5/SeqApiPop/tree/v1.5>

Analyse en Composantes Principales (ACP)

ACP - BeeMuSe - 748 échantillons

```
# Chargement des données
eigenvec <- read.table("BeeMuse_filtered.eigenvec", header = F)
#plink2.eigenvec = non filtrées / BeeMuse_filtered.eigenvec : filtre liste marqueurs
eigenval <- read.table("BeeMuse_filtered.eigenval", header = F)
#plink2.eigenval = non filtrées / BeeMuse_filtered.eigenval : filtre liste marqueurs
colnames(eigenvec)[colnames(eigenvec) == "V1"] <- "Sample"
colnames(eigenvec)[colnames(eigenvec) == "V2"] <- "Filename"
```

```
Corres_ID_E756 <- read.table("/home/nbettembourg/Documents/Stage_NB/data/BeeMuSe_Wageningen_E756_17_03_2022.csv", header = F)
input_pedigree_BeeMuSe <- read.table("/home/nbettembourg/Documents/Stage_NB/data/BeeMuSe_Wageningen_E756_17_03_2022.csv", header = F)
```

```
#extraire 'Pool' et '-100' obtenir 'Pool-100'
eigenvec$name <- paste(sub("Beemuse_", "", eigenvec$Sample),
                      sub("_(.*)\\.\\.\\.\\1", "\\1", eigenvec$Filename),
                      sep = "-")
eigenvec$name <- str_extract(eigenvec$name, "[A-Za-z0-9]+-[0-9]+")

colnames(Corres_ID_E756)[colnames(Corres_ID_E756) == "V1"] <- "name"
colnames(Corres_ID_E756)[colnames(Corres_ID_E756) == "V2"] <- "ID_1a"

Corres_ID_E756$ID_1a <- gsub("o", "_", Corres_ID_E756$ID_1a)

# Remplacement de tous les "o" par "_" sauf quand présent dans un mot
Corres_ID_E756$ID_1a <- gsub("Pers_", "Perso", Corres_ID_E756$ID_1a)
Corres_ID_E756$ID_1a <- gsub("L_c", "Loc", Corres_ID_E756$ID_1a)

Corres_ID_E756_eigenvec <- merge(eigenvec, Corres_ID_E756, by = 'name')
merged_3 <- merge(Corres_ID_E756_eigenvec, input_pedigree_BeeMuSe, by = 'ID_1a')
colnames(merged_3)[colnames(merged_3) == "V3.x"] <- "V3"
```

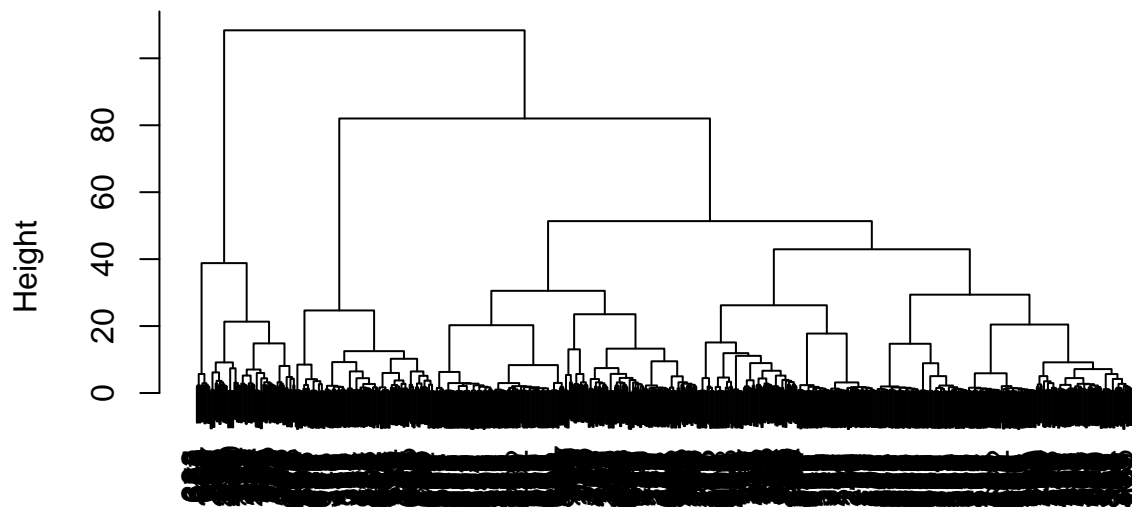
```
ind2pop_ID_2a = subset(merged_3, select = c(name, ID_2a))
```

```
id_counts <- as.data.frame(table(merged_3$ID_2a))
names(id_counts) <- c("ID_2a", "Occurrence")
```

Importation des données de matrice d'apparement dans R et visualisation par clustering hiérarchique

```
matrice_app <- read.table("plink2.rel", header = FALSE)
dist_matrice <- dist(matrice_app)
hc <- hclust(dist_matrice, method = "ward.D2")
plot(hc)
```

Cluster Dendrogram



```
dist_matrice
hclust (*, "ward.D2")
```

#heatmap(as.matrix(dist(matrice_app,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow', dua

Proportion de la variance expliquée

```
eigen_percent <- round((eigenval / (sum(eigenval) )*100),2)
```

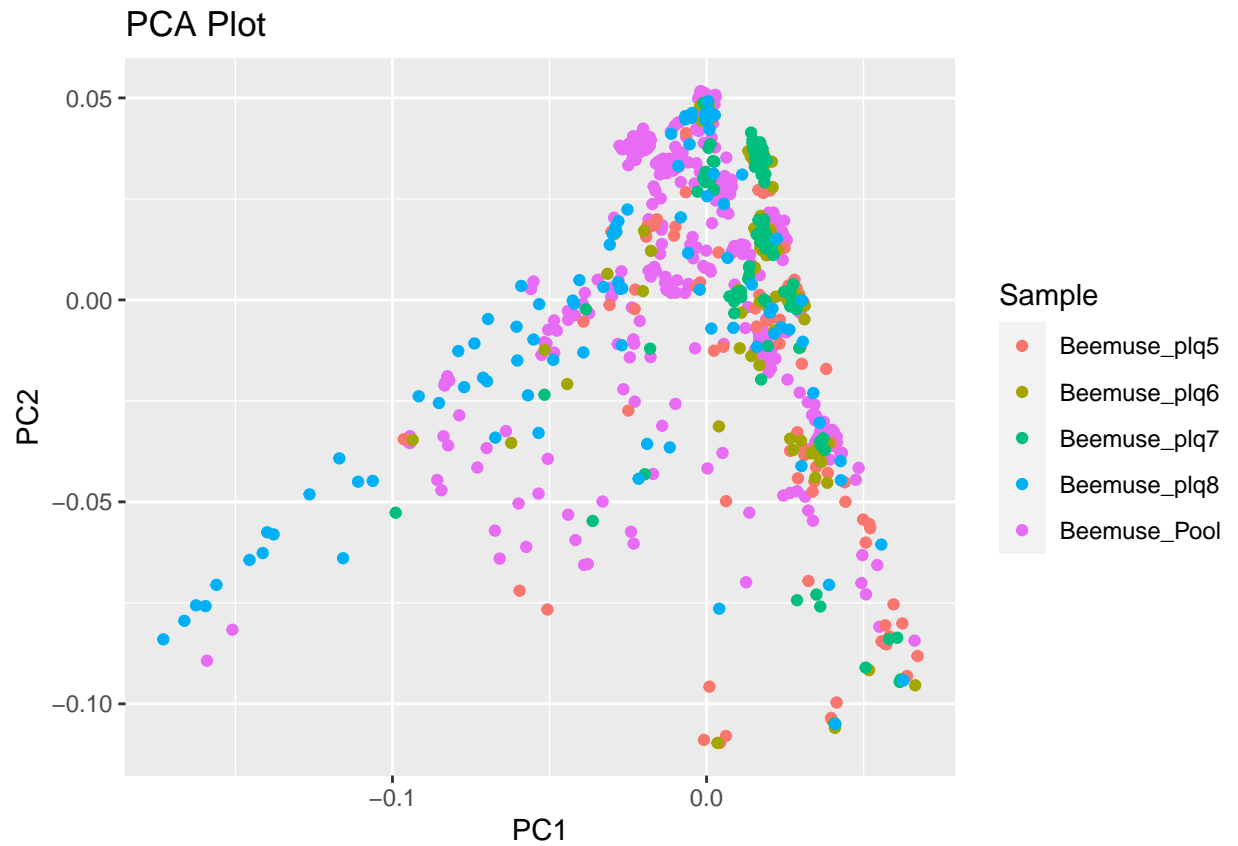
```
lambda <- eigenval$V1
```

```
variance_proportion <- lambda / sum(lambda)
```

```
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)
```

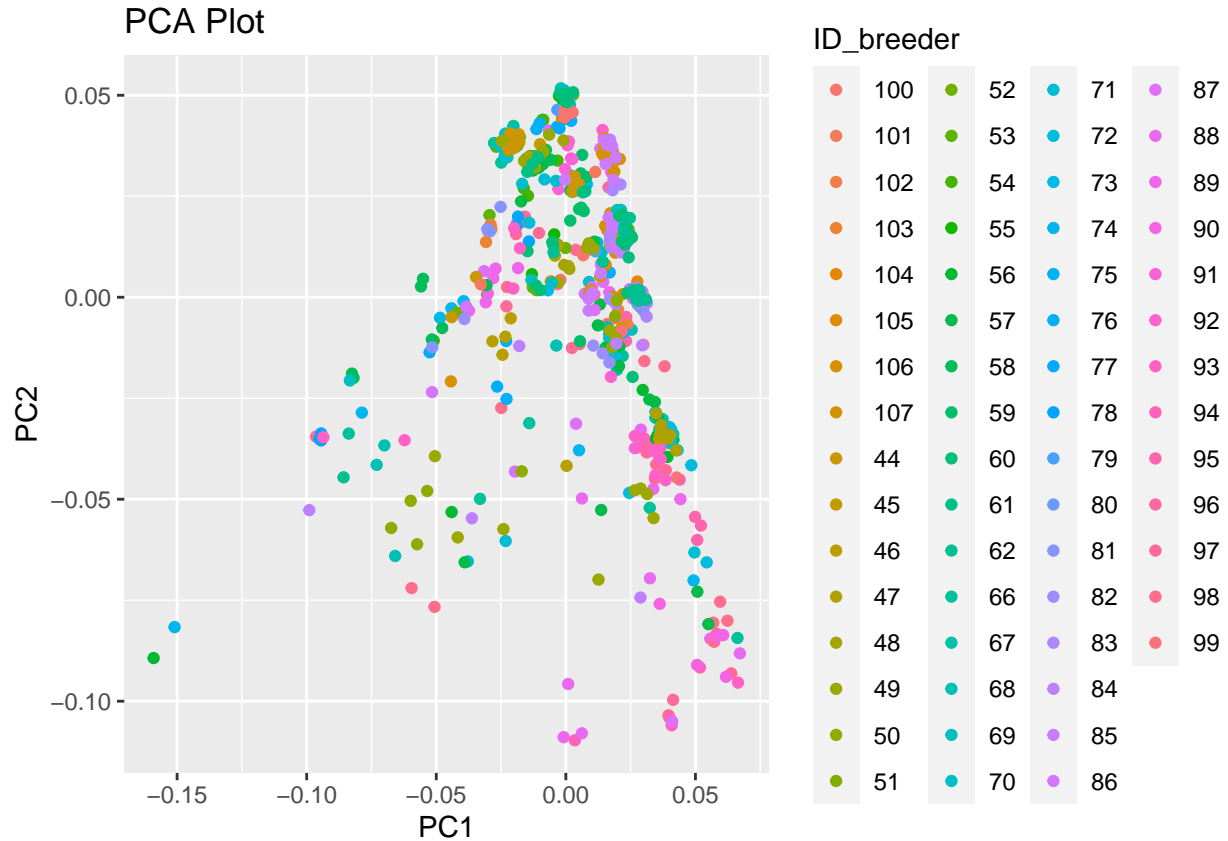
```
ggplot(eigenvec, aes(x = V3, y = V4, label = Sample, color = Sample)) +
  geom_point() +
```

```
labs(title = "PCA Plot", x = "PC1", y = "PC2") +  
theme(legend.position = "right")
```

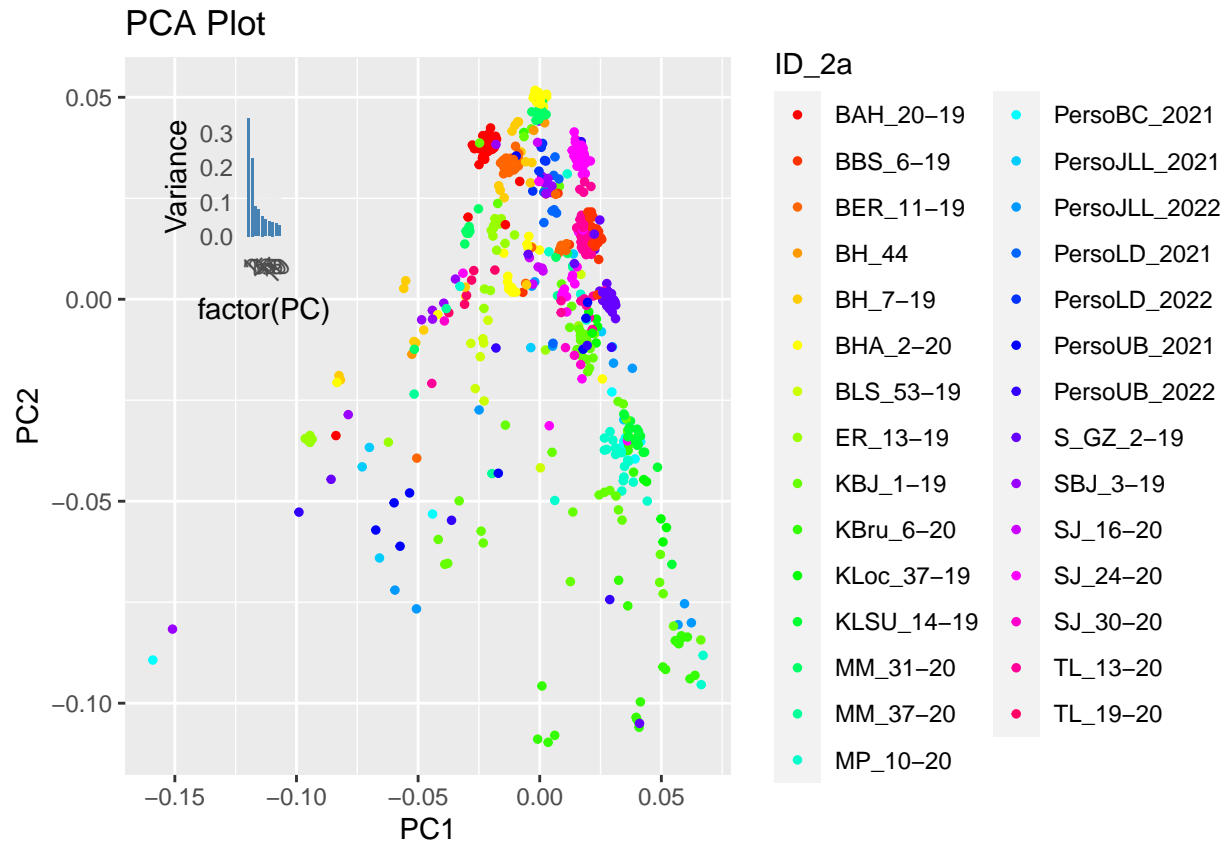


PC1/PC2

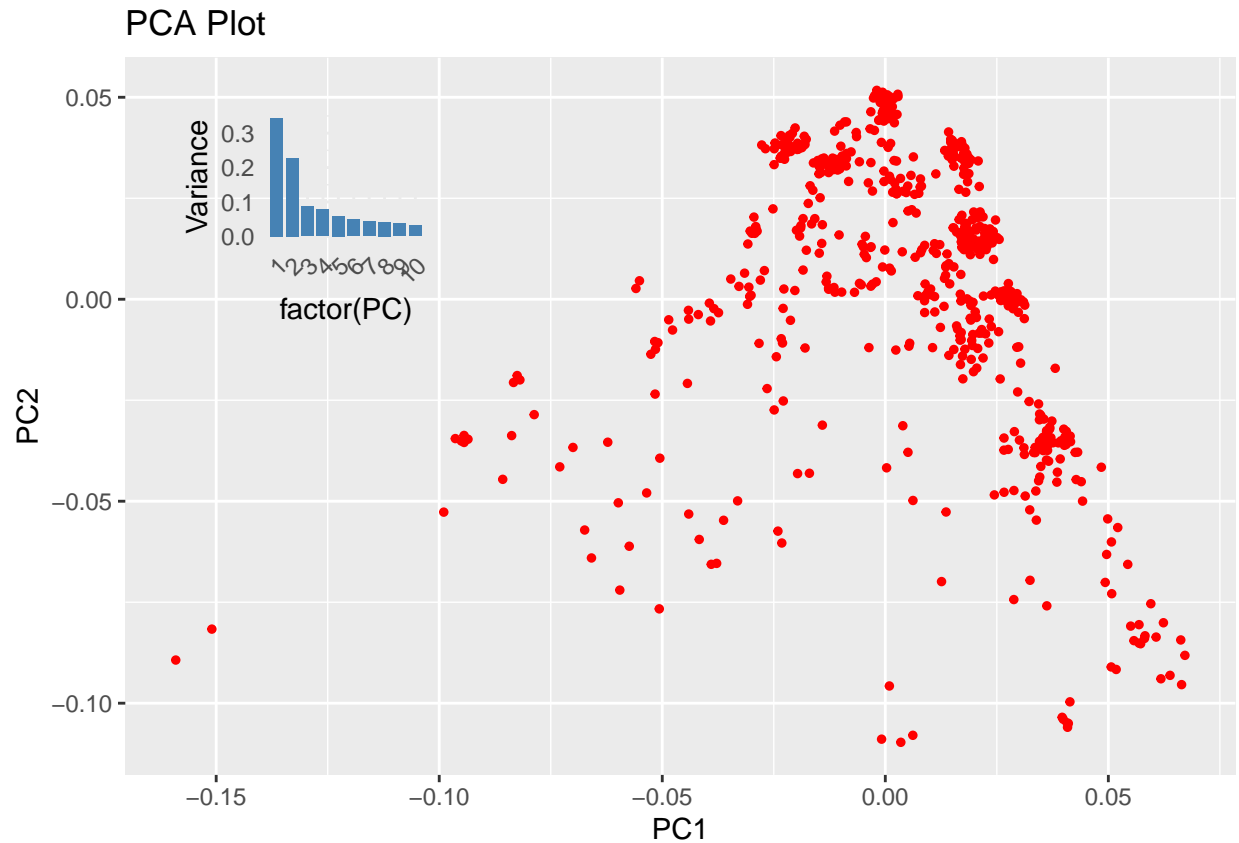
```
# ACP ID_breeder : identification de l'apiculteur de chez qui proviennent les abeilles  
ggplot(merged_3, aes(x = V3, y = V4, label = ID_breeder, color = ID_breeder)) +  
geom_point() +  
labs(title = "PCA Plot", x = "PC1", y = "PC2") +  
theme(legend.position = "right")
```



```
#plot ID_2a : reines mères des reines génotypées
ggplot(merged_3, aes(x = V3, y = V4, label = ID_2a, color = ID_2a)) +
  geom_point(size = 1) +
  scale_color_manual(values = c("#FF0000FF", "#FF3300FF", "#FF6600FF", "#FF9900FF", "#FFCC00FF",
                                "#FFFF00FF", "#CCFF00FF", "#99FF00FF", "#66FF00FF", "#33FF00FF",
                                "#00FF00FF", "#00FF33FF", "#00FF66FF", "#00FF99FF", "#00FFCCFF",
                                "#00FFFFFF", "#00CCFFFF", "#0099FFFF", "#0066FFFF", "#0033FFFF",
                                "#0000FFFF", "#3300FFFF", "#6600FFFF", "#9900FFFF", "#CC00FFFF",
                                "#FF00FFFF", "#FF00CCFF", "#FF0099FF", "#FF0066FF")) +
  labs(title = "PCA Plot", x = "PC1", y = "PC2") +
  theme(legend.position = "right") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = -0.1, xmax = -0.16, ymin = 0.05, ymax = -0.01)
```



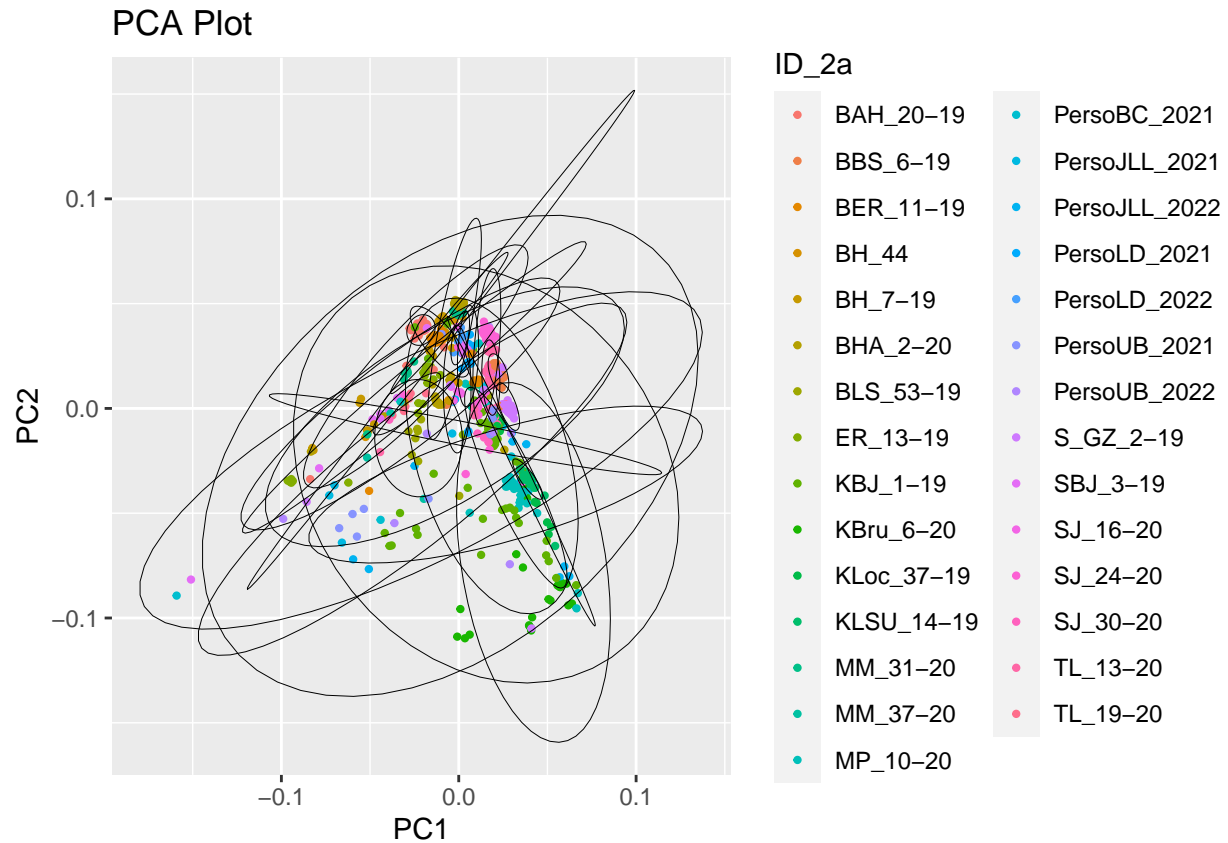
```
ggplot(merged_3, aes(x = V3, y = V4, label = ID_2a)) +
  geom_point(size = 1, color = "red") +
  labs(title = "PCA Plot", x = "PC1", y = "PC2") +
  theme(legend.position = "right") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = -0.1, xmax = -0.16, ymin = 0.05, ymax = -0.01)
```



```
# Visualisaiton ACP PC1 PC2 avec ellipses de seuil de confiance de 0.97
ggplot(merged_3, aes(x = V3, y = V4, label = ID_2a, color = ID_2a)) +
  geom_point(size = 0.8) +
  stat_ellipse(aes(group = ID_2a), geom = "polygon", level = 0.97, alpha = 0, size = 0.1, color = "black") +
  labs(title = "PCA Plot", x = "PC1", y = "PC2") +
  theme(legend.position = "right")
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

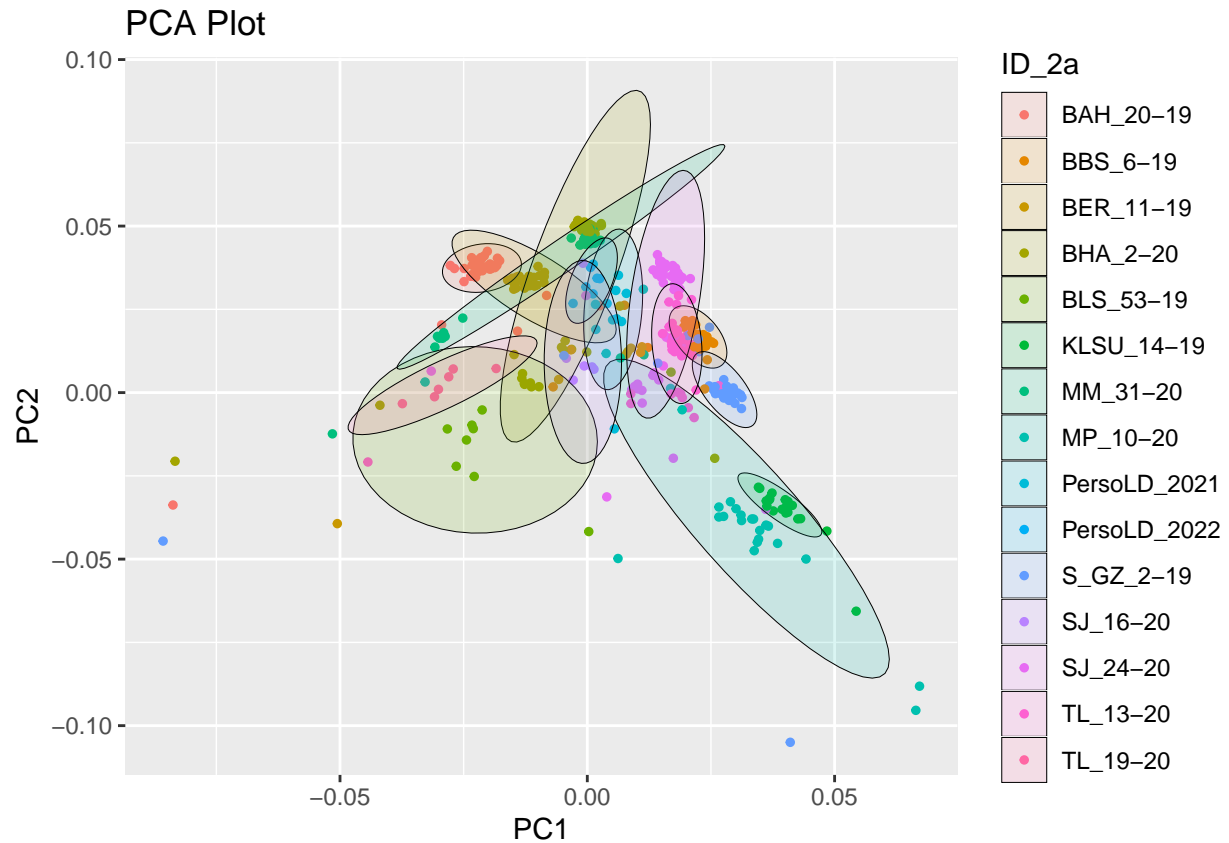
```
## Warning in MASS::cov.trob(data[, vars]): Probable convergence failure
```



#15 familles bien regroupés entre elles

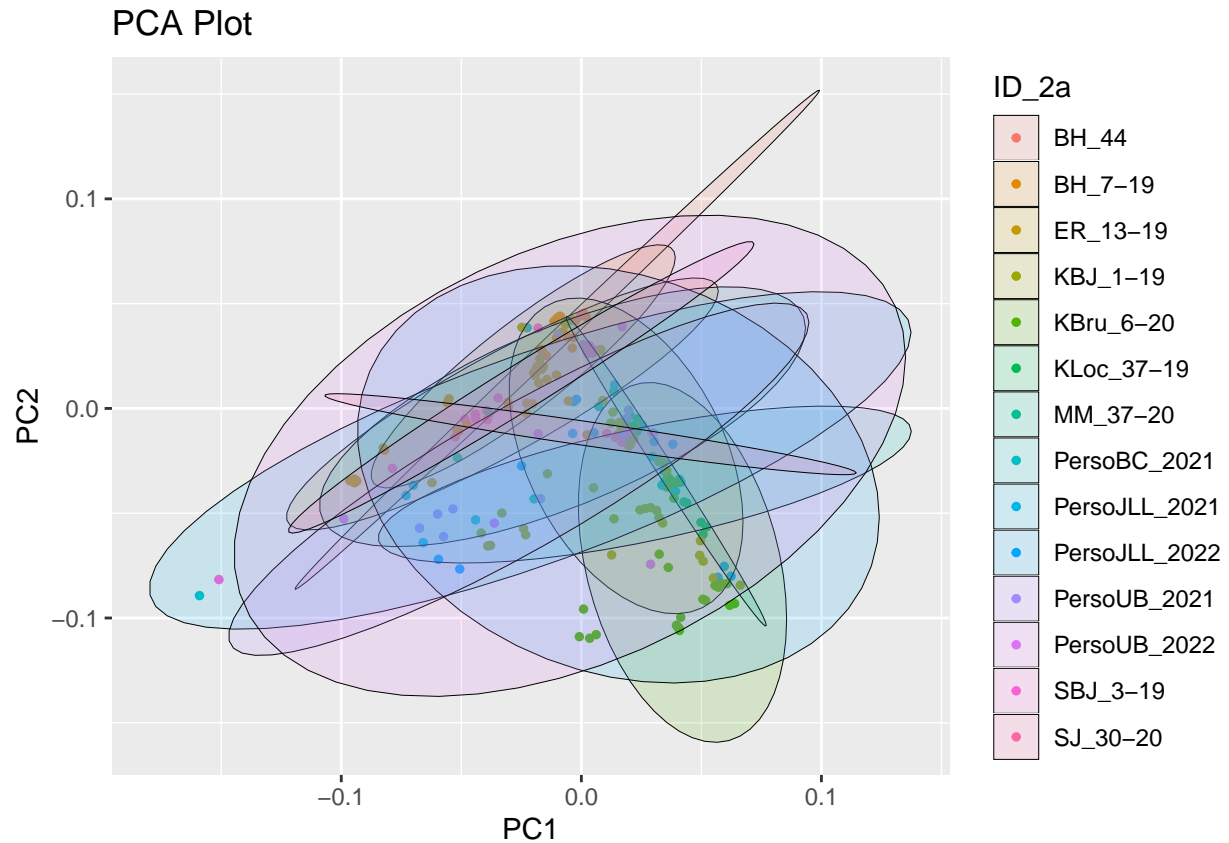
```
ggplot(subset(merged_3, ID_2a %in% c("BAH_20-19", "BBS_6-19", "BER_11-19", "BLS_53-19", "BHA_2-20", "KLSU_14-19", "MM_31-20", "MM_37-20", "MP_10-20", "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021", "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20", "TL_19-20"))) +
  geom_point(size = 1) +
  stat_ellipse(aes(group = ID_2a, fill = ID_2a), geom = "polygon", level = 0.97, alpha = 0.15, size = 0.1) +
  #stat_ellipse(aes(group = ID_2a), geom = "polygon", level = 0.97, alpha = 0, size = 0.1, color = "black") +
  labs(title = "PCA Plot", x = "PC1", y = "PC2") +
  theme(legend.position = "right")
```

Warning in MASS::cov.trob(data[, vars]): Probable convergence failure

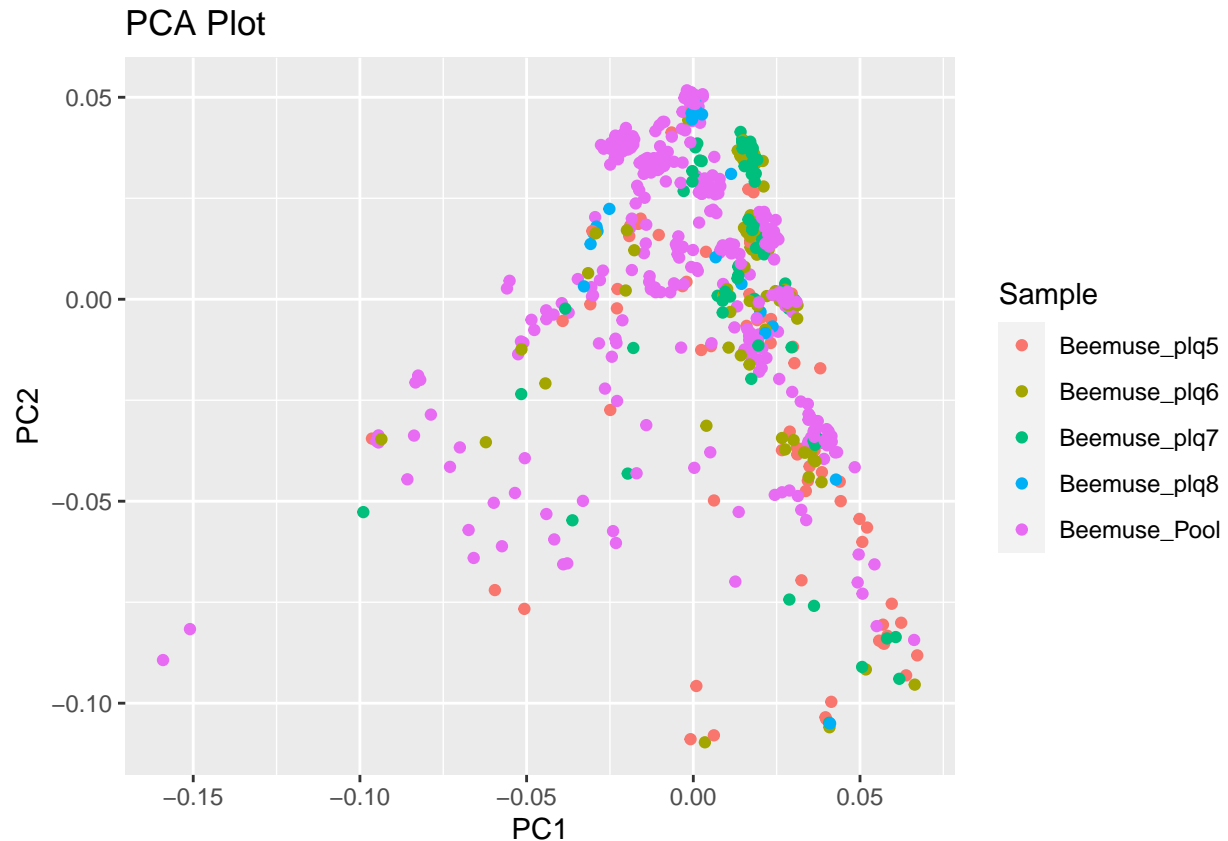


#14 familles moins bien regroupées entre elles

```
ggplot(subset(merged_3, ID_2a %in% c( "BH_44", "BH_7-19", "ER_13-19", "MM_37-20", "KBJ_1-19", "KLoc_37-19"
  geom_point(size = 1) +
  stat_ellipse(aes(group = ID_2a, fill = ID_2a), geom = "polygon", level = 0.97, alpha = 0.15, size = 0.1)
  #stat_ellipse(aes(group = ID_2a), geom = "polygon", level = 0.97, alpha = 0, size = 0.1, color = "black")
  labs(title = "PCA Plot", x = "PC1", y = "PC2") +
  theme(legend.position = "right")
```

```
# ACP selon variable Sample (plaques)
ggplot(merged_3, aes(x = V3, y = V4, label = Sample, color = Sample)) +
  geom_point() +
  # stat_ellipse(aes(group = Sample), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
  # stat_ellipse(aes(group = Sample, fill = Sample), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.2)
  labs(title = "PCA Plot", x = "PC1", y = "PC2") +
  theme(legend.position = "right")
```

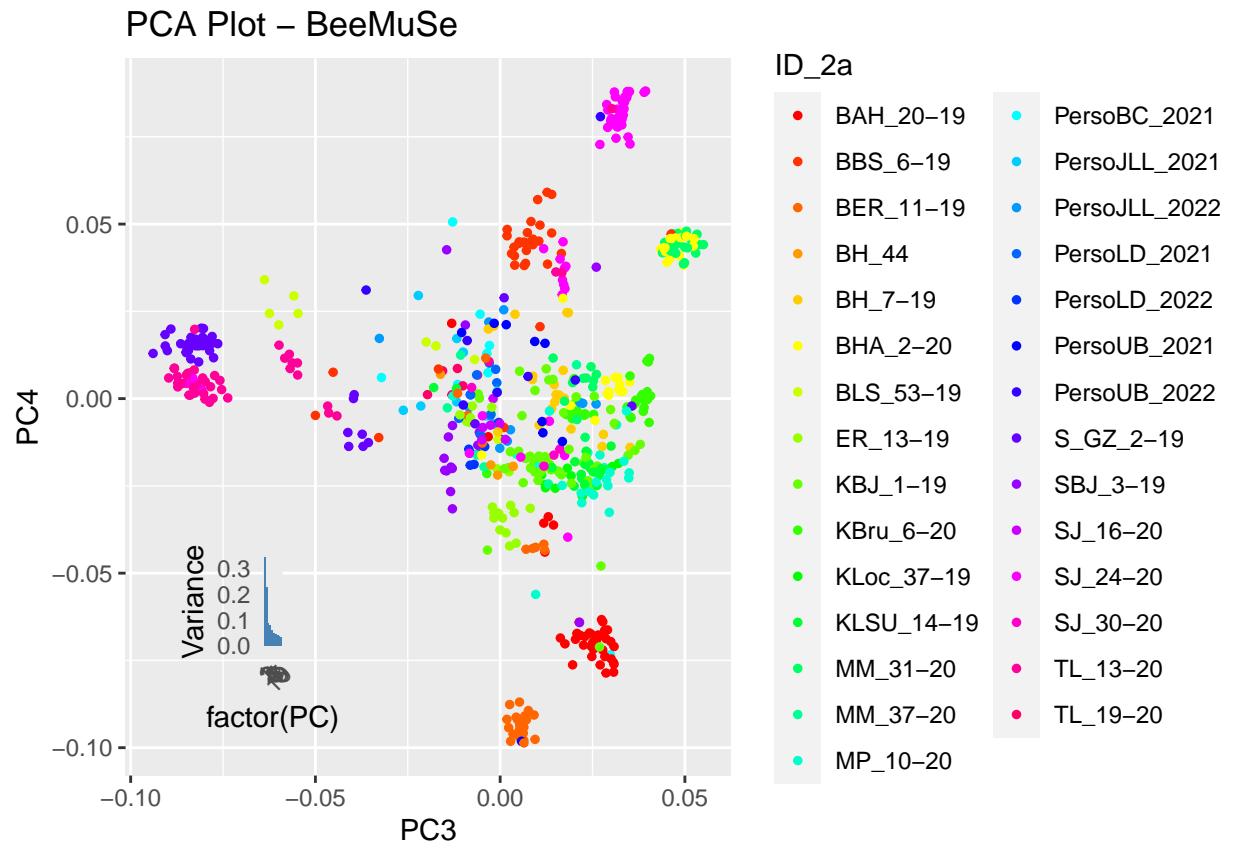


```
# 748 -> 681 -> 612
values_in_A_not_in_B <- eigenvec$name[!(eigenvec$name %in% merged_3$name)]
length(values_in_A_not_in_B)
```

```
## [1] 136
```

```
# ACP PC3 PC4 - ID_2a : reines mères des reines génotypées
ggplot(merged_3, aes(x = V5, y = V6, label = ID_2a, color = ID_2a)) +
  geom_point(size = 1) +
  scale_color_manual(values = c("#FF0000FF", "#FF3300FF", "#FF6600FF", "#FF9900FF", "#FFCC00FF",
                                "#FFFF00FF", "#CCFF00FF", "#99FF00FF", "#66FF00FF", "#33FF00FF",
                                "#00FF00FF", "#00FF33FF", "#00FF66FF", "#00FF99FF", "#00FFCCFF",
                                "#00FFFFFF", "#00CCFFFF", "#0099FFFF", "#0066FFFF", "#0033FFFF",
                                "#0000FFFF", "#3300FFFF", "#6600FFFF", "#9900FFFF", "#CC00FFFF",
                                "#FF00FFFF", "#FF00CCFF", "#FF0099FF", "#FF0066FF")) +
  labs(title = "PCA Plot - BeeMuSe", x = "PC3", y = "PC4") +
  theme(legend.position = "right") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
```

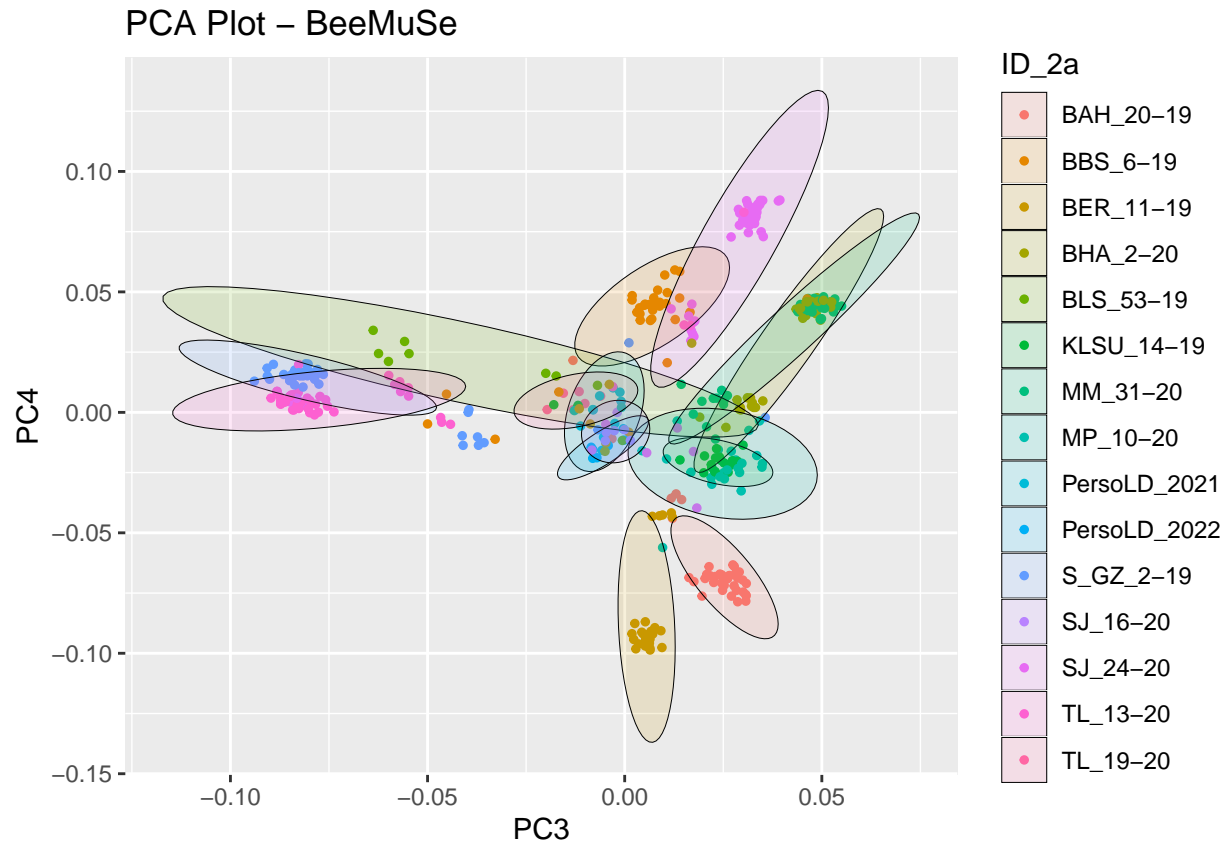
```
theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
), xmin = -0.055, xmax = -0.09, ymin = -0.04, ymax = -0.1)
```



PC3/PC4

```
#15 bien regroupés
ggplot(subset(merged_3, ID_2a %in% c("BAH_20-19", "BBS_6-19", "BER_11-19", "BLS_53-19", "BHA_2-20", "KLSU_14-19", "MM_31-20", "MM_37-20", "MP_10-20", "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021", "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20", "TL_19-20"))) +
  geom_point(size = 1) +
  stat_ellipse(aes(group = ID_2a, fill = ID_2a), geom = "polygon", level = 0.97, alpha = 0.15, size = 0.5)
#stat_ellipse(aes(group = ID_2a), geom = "polygon", level = 0.97, alpha = 0, size = 0.1, color = "black")
labs(title = "PCA Plot - BeeMuSe", x = "PC3", y = "PC4") +
theme(legend.position = "right")
```

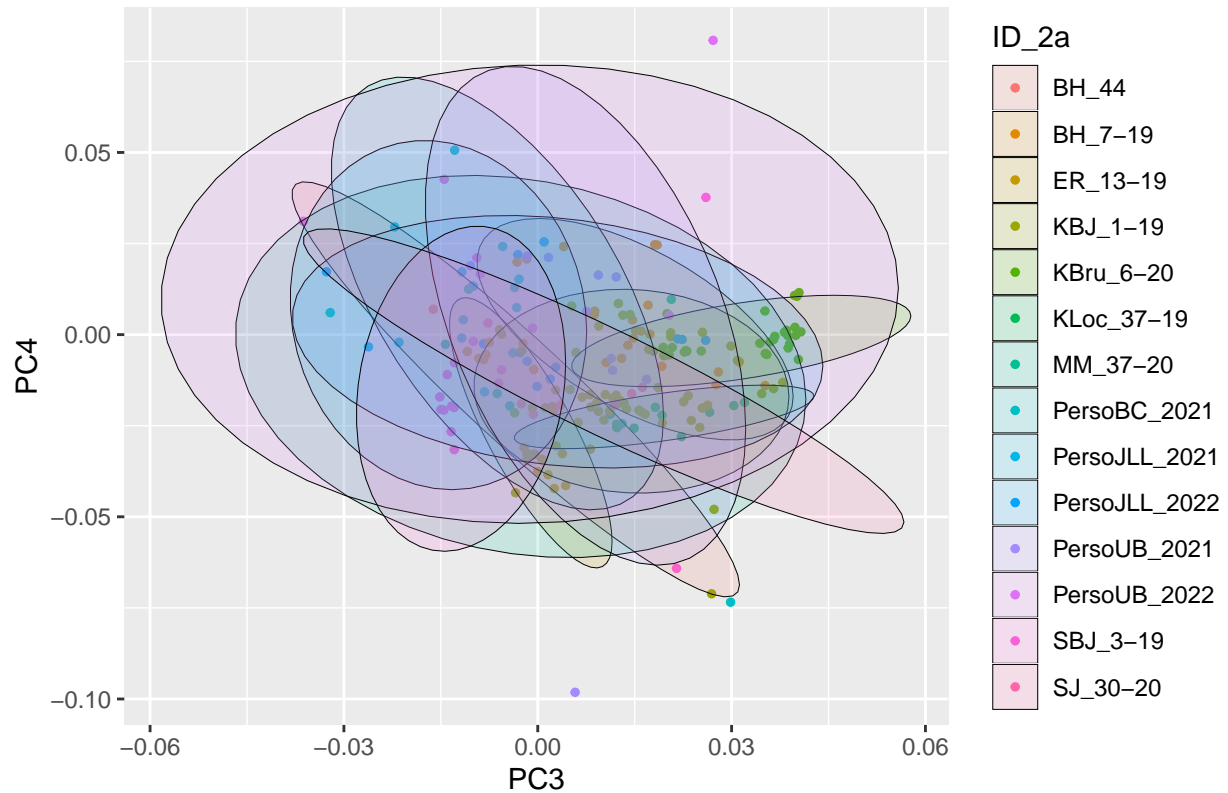
```
## Warning in MASS::cov.trob(data[, vars]): Probable convergence failure
```



```
#14 autres
ggplot(subset(merged_3, ID_2a %in% c( "BH_44", "BH_7-19", "ER_13-19", "MM_37-20", "KBJ_1-19", "KLoc_37-19" )
  geom_point(size = 1) +
  stat_ellipse(aes(group = ID_2a, fill = ID_2a), geom = "polygon", level = 0.97, alpha = 0.15, size = 0.1)
  #stat_ellipse(aes(group = ID_2a), geom = "polygon", level = 0.97, alpha = 0, size = 0.1, color = "black")
  labs(title = "PCA Plot - BeeMuSe", x = "PC3", y = "PC4") +
  theme(legend.position = "right")
```

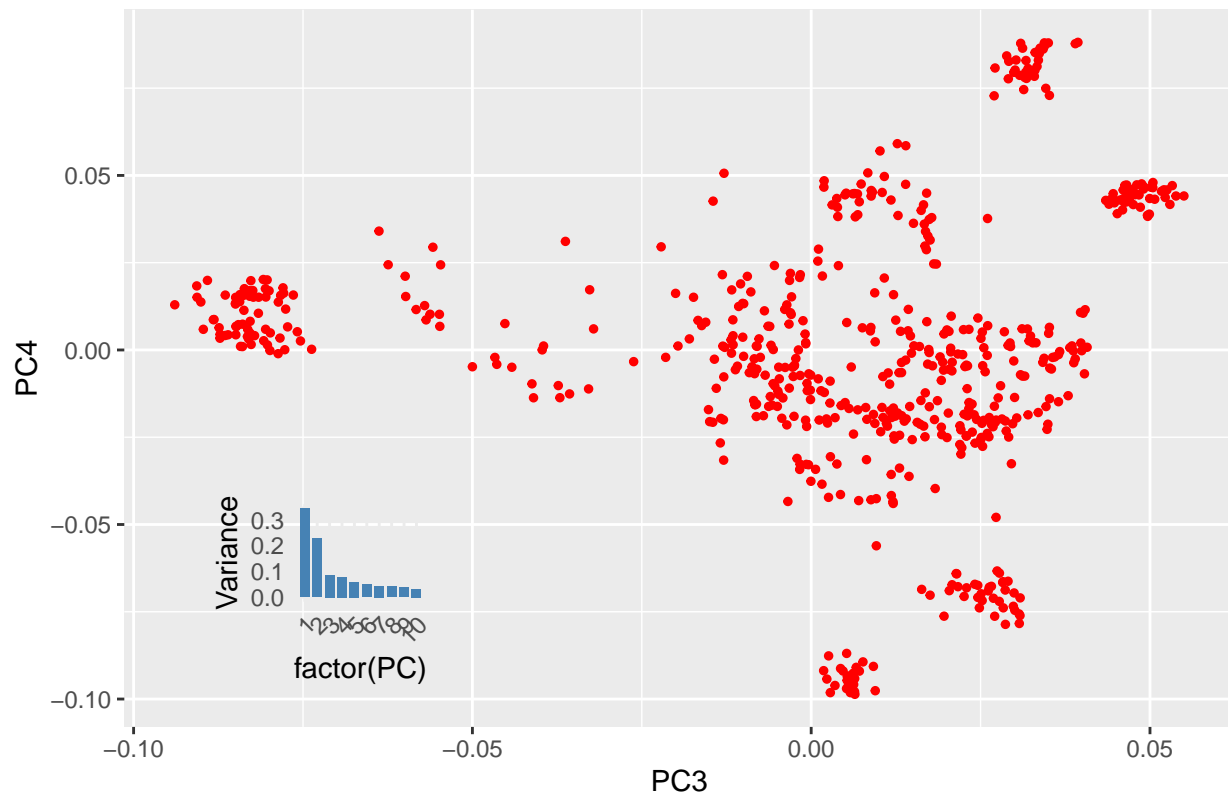
```
## Warning in MASS::cov.trob(data[, vars]): Probable convergence failure
```

PCA Plot – BeeMuSe



```
ggplot(merged_3, aes(x = V5, y = V6, label = ID_2a)) +
  geom_point(size = 1, color = "red") +
  labs(title = "PCA Plot - BeeMuSe", x = "PC3", y = "PC4") +
  theme(legend.position = "right") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = -0.055, xmax = -0.09, ymin = -0.04, ymax = -0.1)
```

PCA Plot – BeeMuSe



```
# BEEMUSE - ID_2a
beemuse_samples_ID_2a <- merged_3[, c("Sample", "Filename", "ID_2a")]
nom_fichier <- "beemuse_samples_ID_2a.txt"
write.table(beemuse_samples_ID_2a, file = nom_fichier, sep = "\t", row.names = FALSE, quote=FALSE)
```

```
# FAM MERGED SEQAPIPOP BEEMUSE 561
merged_fam <- read.table("merged_BeeMuSe_SeqApiPop_561_filtered_maf001_LD_default.fam", sep = " ", stringsAsFactors = FALSE)

merged_fam <- merged_fam[, 1:2]
colnames(merged_fam)[colnames(merged_fam) == "V1"] <- "Sample"
colnames(merged_fam)[colnames(merged_fam) == "V2"] <- "Filename"
```

```
# Créer une nouvelle colonne dans le fichier 2 pour stocker les valeurs de la troisième colonne
merged_fam$ID_2a <- NA

# Parcourir les lignes du fichier 2
for (i in 1:nrow(merged_fam)) {
```

```

# Vérifier si les valeurs des colonnes Sample et Filename du fichier 2 sont présentes dans le fichier 1
correspondance <- beemuse_samples_ID_2a$ID_2a[beemuse_samples_ID_2a$Sample == merged_fam$Sample[i]]

# Si une correspondance exacte est trouvée, ajouter la valeur correspondante de la colonne "ID_2a"
if (length(correspondance) > 0) {
  merged_fam$ID_2a[i] <- as.character(correspondance[1]) # Utilisez la correspondance trouvée
} else {
  # Sinon, ajouter "Beemuse" dans la troisième colonne du fichier 2
  merged_fam$ID_2a[i] <- "Beemuse"
}
}

# Créer fichier de correspondance entre les échantillons de BeeMuSe ID_2a avec le fichier .fam
correspondance <- merged_fam[, c("Sample", "Filename", "ID_2a")]
nom_fichier <- "correspondance_samples.txt"
write.table(correspondance, file = nom_fichier, sep = "\t", row.names = FALSE, quote=FALSE)

```

Correspondance ID_2a - fichier .fam - BeeMuSe

ACP - Merged Data - BeeMuSe - SeqApiPop

```

setwd("~/Documents/Stage_NB/data/merged_data_1055_561_not_supervised")

eigenvec_merged_maf001_LD03 <- read.table("merged_BeeMuSe_SeqApiPop_561_filtered_maf001_LD_default_acp.csv")
eigenval_merged_maf001_LD03 <- read.table("merged_BeeMuSe_SeqApiPop_561_filtered_maf001_LD_default_acp.csv")

colnames(eigenvec_merged_maf001_LD03)[colnames(eigenvec_merged_maf001_LD03) == "V1"] <- "Sample"
colnames(eigenvec_merged_maf001_LD03)[colnames(eigenvec_merged_maf001_LD03) == "V2"] <- "name"

Corres_ID_E756 <- read.table("/home/nbettebourg/Documents/Stage_NB/data/BeeMuSe_Wageningen_E756_17_03_2019.csv")
input_pedigree_BeeMuSe <- read.table("/home/nbettebourg/Documents/Stage_NB/data/BeeMuSe_Wageningen_E756_17_03_2019.csv")

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

eigenvec_merged_maf001_LD03_seq_api_labels <- merge(eigenvec_merged_maf001_LD03, seq_api_labels, by = "name")

eigen_percent <- round((eigenval_merged_maf001_LD03 / (sum(eigenval_merged_maf001_LD03) ) * 100), 2)

```

```

lambda <- eigenval_merged_maf001_LD03$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

```

```

# filter 629 -> 301 RefPop Samples
eigenvec_merged_maf001_LD03_seq_api_labels <- eigenvec_merged_maf001_LD03_seq_api_labels[eigenvec_merged_maf001_LD03_seq_api_labels$Sample %in% RefPop_Samples, ]

```

```

eigenvec_merged_maf001_LD03_seq_api_
eigenvec_merged_maf001_LD03_seq_api_
eigenvec_merged_maf001_LD03_seq_api_

# Filtrer les lignes où les facteurs des colonnes Sample et name sont différents
eigenvec_BeeMuSe_748_Samples <- eigenvec_merged_maf001_LD03[as.character(eigenvec_merged_maf001_LD03$Sample) == eigenvec_merged_maf001_LD03$name, ]

#extraire 'Pool' et '-100' obtenir 'Pool-100'
eigenvec_merged_maf001_LD03$name <- paste(sub("Beemuse_", "", eigenvec_merged_maf001_LD03$Sample),
      sub("_(.*)\\..*", "\\1", eigenvec_merged_maf001_LD03$name),
      sep = "-")
eigenvec_merged_maf001_LD03$name <- str_extract(eigenvec_merged_maf001_LD03$name, "[A-Za-z0-9]+-[0-9]+")

colnames(Corres_ID_E756)[colnames(Corres_ID_E756) == "V1"] <- "name"
colnames(Corres_ID_E756)[colnames(Corres_ID_E756) == "V2"] <- "ID_1a"

Corres_ID_E756$ID_1a <- gsub("o", "_", Corres_ID_E756$ID_1a)

# Remplacement de "o" par "_" sauf quand présent dans un mot
Corres_ID_E756$ID_1a <- gsub("Pers_", "Perso", Corres_ID_E756$ID_1a)
Corres_ID_E756$ID_1a <- gsub("L_c", "Loc", Corres_ID_E756$ID_1a)

Corres_ID_E756_eigenvec <- merge(eigenvec_merged_maf001_LD03, Corres_ID_E756, by = 'name')
eigenvec_merged_Corres_ID_E756_pedigree <- merge(Corres_ID_E756_eigenvec, input_pedigree_BeeMuSe, by = 'name')
colnames(eigenvec_merged_Corres_ID_E756_pedigree)[colnames(eigenvec_merged_Corres_ID_E756_pedigree) == "ID_1a"] <- "ID"

```

ACP - maf001 - LD03 - SeqApiPop 561 échantillons - 1055 SNPs

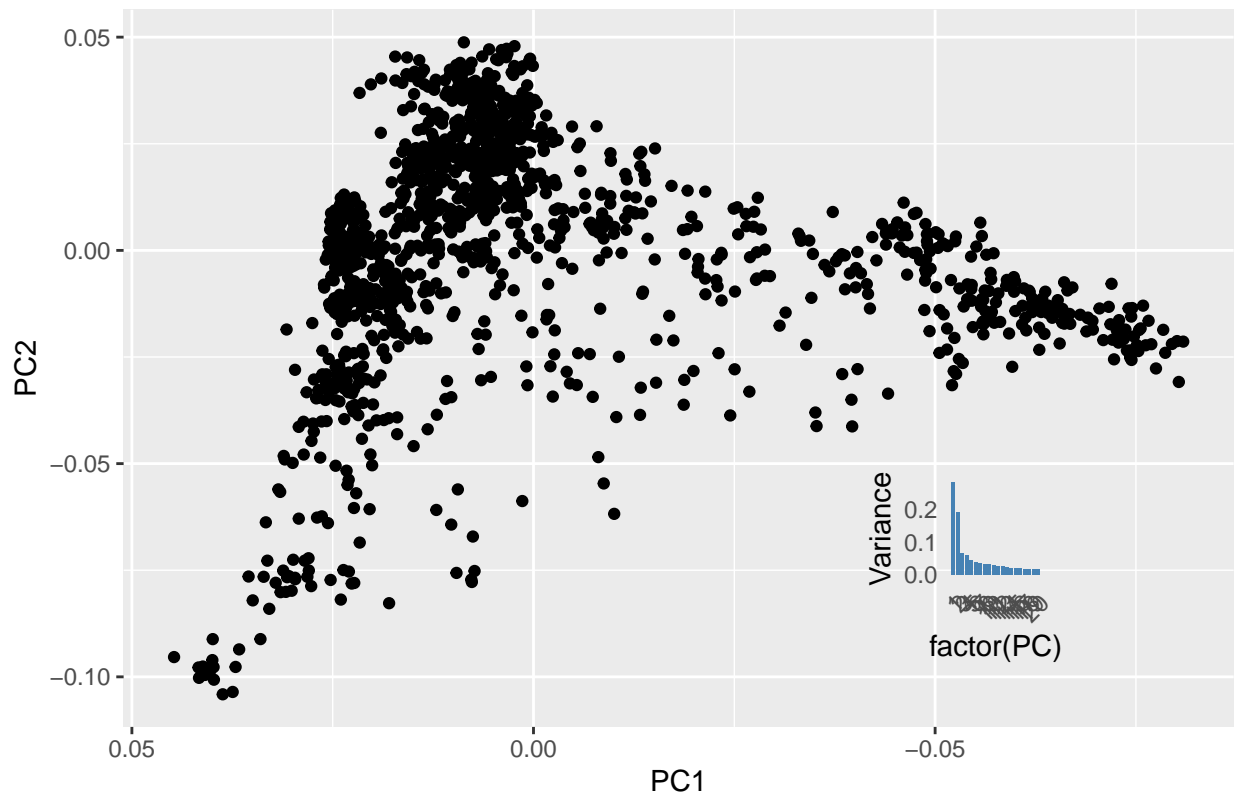
```

# ACP - merged data
lambda <- eigenval_merged_maf001_LD03$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

ggplot(eigenvec_merged_maf001_LD03, aes(x = V3, y = V4)) +
  geom_point() +
  scale_x_reverse() +
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.065, xmax = 0.04, ymin = -0.1, ymax = -0.05)

```


PCA Plot – BeeMuSe & SeqApiPop



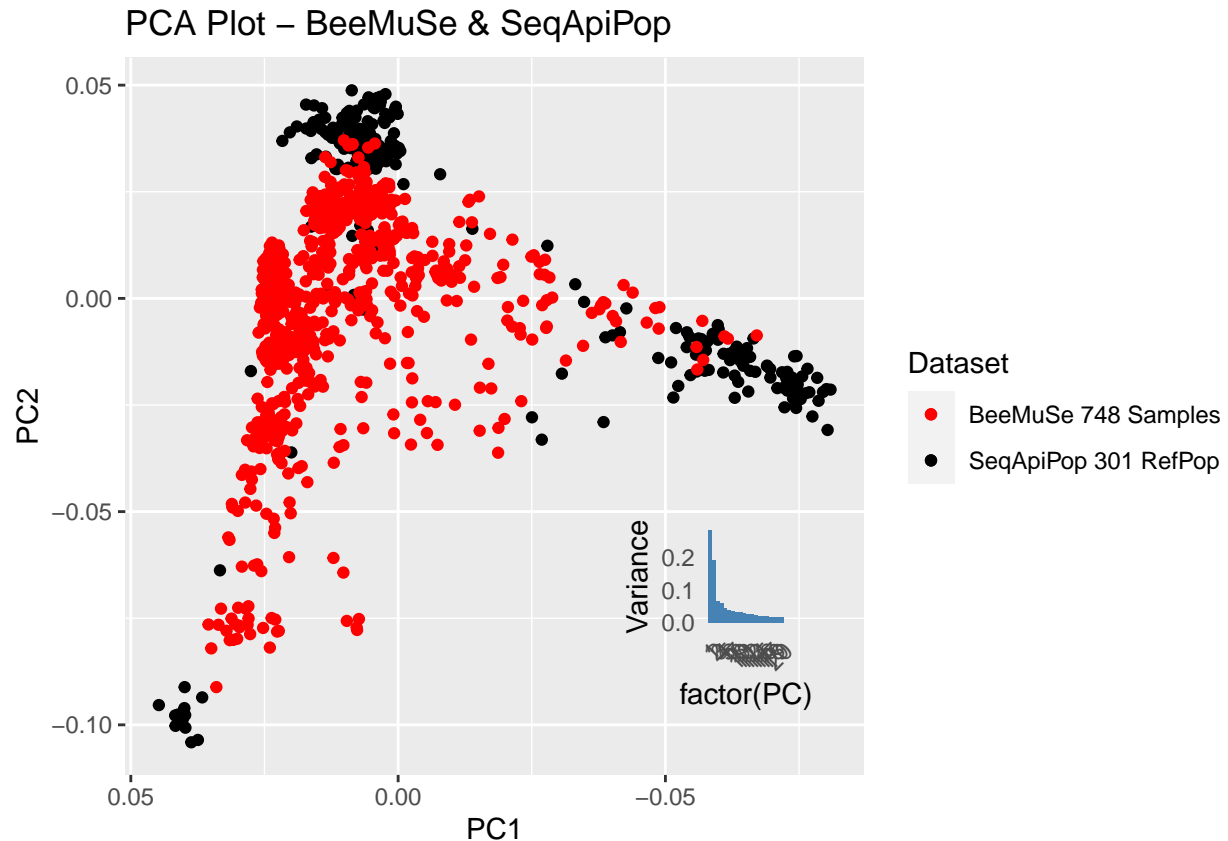
PC1/PC2

```
# ACP plot - BeeMuSe 748 & SeqApiPop 301 RefPop
# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V3, y = V4, color = "SeqApiPop")) +
  geom_point() +
  scale_x_reverse()

# Ajout ACP pour BeeMuSe
plot2 <- plot1 + geom_point(data = eigenvec_BeeMuSe_748_Samples, aes(x = V3, y = V4, color = "BeeMuSe"))
labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2")

# Ajout de la légende
plot2 + scale_color_manual(name = "Dataset",
  values = c(SeqApiPop = "black", BeeMuSe = "red"),
  labels = c(SeqApiPop = "SeqApiPop 301 RefPop", BeeMuSe = "BeeMuSe 748 Samples"))

annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
      geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
      theme_minimal() +
      theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.075, xmax = 0.04, ymin = -0.1, ymax = -0.05)
```

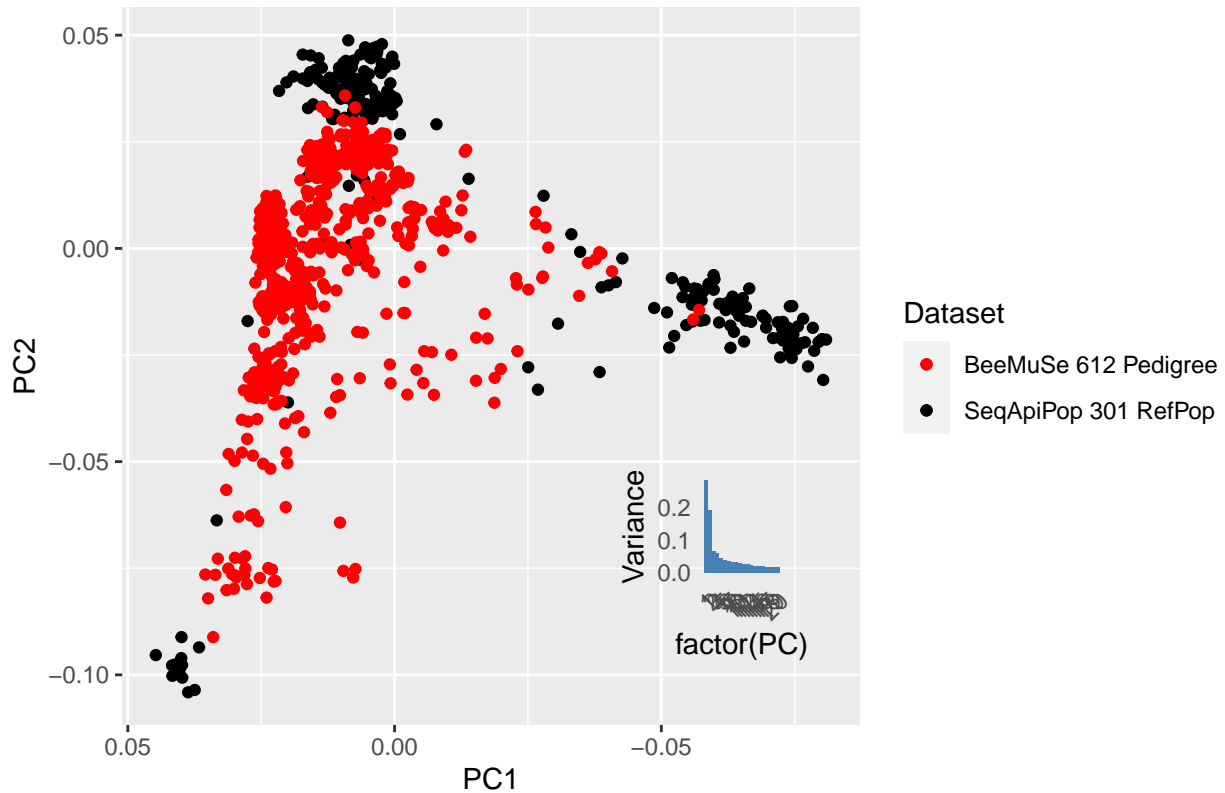


```
# ACP plot - BeeMuSe 612 Pedigree & SeqApiPop 301 RefPop
# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V3, y = V4, color = "SeqApiPop")) +
  geom_point() +
  scale_x_reverse()

# Ajout ACP pour BeeMuSe
plot2 <- plot1 + geom_point(data = eigenvec_merged_Corres_ID_E756_pedigree, aes(x = V3, y = V4, color =
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2")

# Ajout de la légende
plot2 + scale_color_manual(name = "Dataset",
  values = c(SeqApiPop = "black", BeeMuSe = "red"),
  labels = c(SeqApiPop = "SeqApiPop 301 RefPop", BeeMuSe = "BeeMuSe 612 Pedigree")) +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.075, xmax = 0.04, ymin = -0.1, ymax = -0.05)
```

PCA Plot – BeeMuSe & SeqApiPop



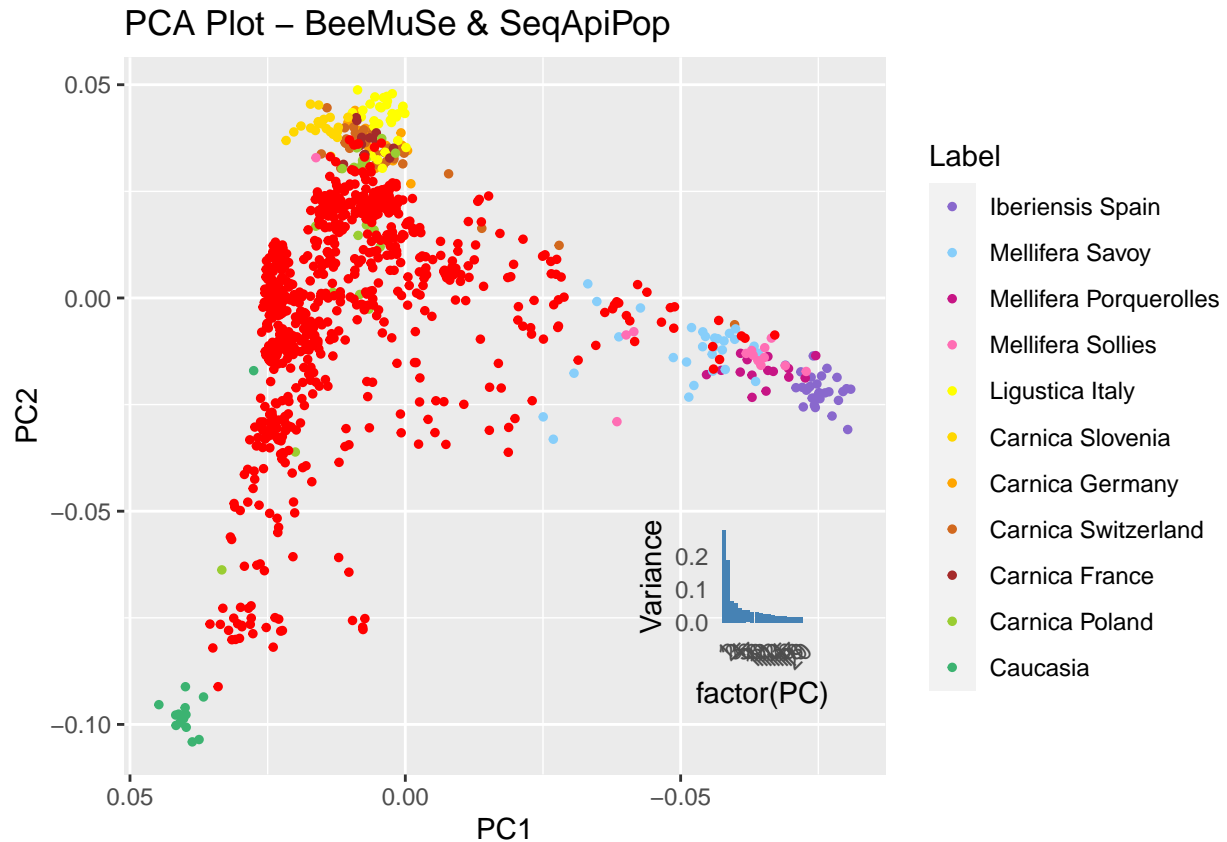
```
# Couleurs 3 lignées - Label
custom_colors_2 <- c( "black", "black", "black", "black", "orange", "orange", "orange", "orange", "orange")

# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point(size = 1) +
  scale_x_reverse() +
  #scale_y_reverse() +
  scale_color_manual(values = custom_colors_2,
    breaks = c("Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solilies Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solilies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

# Ajout ACP pour BeeMuSe
plot1 + geom_point(data = eigenvec_BeeMuSe_748_Samples, aes(x = V3, y = V4), color = "red", size = 1) +
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal())
```



```
ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
  geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
), xmin = 0.075, xmax = 0.04, ymin = -0.1, ymax = -0.05)
```



```
# SeqApiPop - 301 Samples - BeeMuSe 612 Samples - ID_2a - colors
```

```
custom_colors <- c("black", "black", "black",
  "black", "black", "black", "black", "black",
  "black",
  "#FF0000FF", "#FF3300FF", "#FF6600FF", "#FF9900FF", "#FFCC00FF",
  "#FFFF00FF", "#CCFF00FF", "#99FF00FF", "#66FF00FF", "#33FF00FF",
  "#00FF00FF", "#00FF33FF", "#00FF66FF", "#00FF99FF", "#00FFCCFF",
  "#00FFFFFF", "#00CCFFFF", "#0099FFFF", "#0066FFFF", "#0033FFFF",
  "#0000FFFF", "#3300FFFF", "#6600FFFF", "#9900FFFF", "#CC00FFFF",
  "#FF00FFFF", "#FF00CCFF", "#FF0099FF", "#FF0066FF")
```

```
# ACP SeqApiPop
```

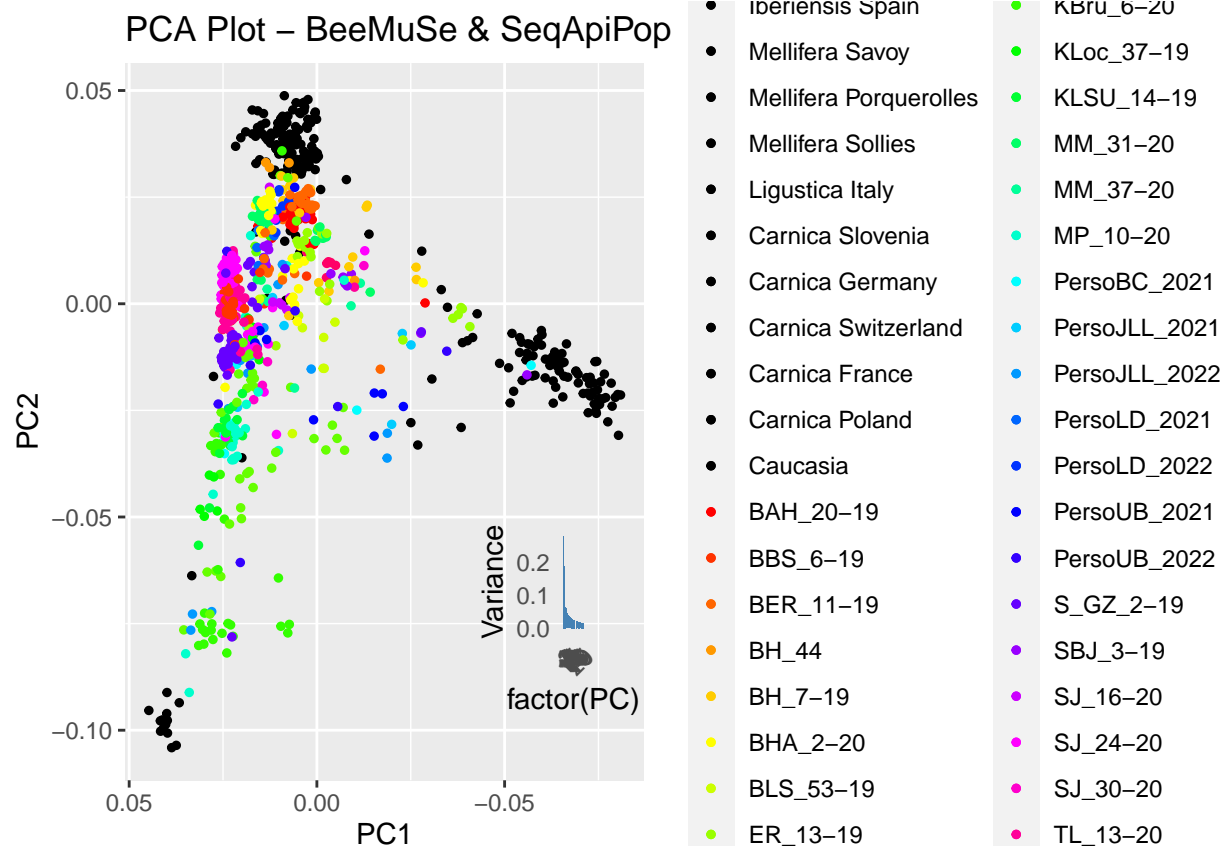
```
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point(size=1) +
  scale_x_reverse() +
  scale_color_manual(values = custom_colors,
    breaks = c("Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
```

```

"Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
"PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
"PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
"TL_19-20"),
  labels = c( "Iberiensis Spain",
              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia", "
  "KBJ_1-19", "KBru_6-20", "KLoc_37-19", "KLSU_14-19", "MM_31-20", "MM_37-20", "MP_10-20",
  "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
  "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
  "TL_19-20")) +
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2") +
  theme(legend.position = "right")

# Ajout ACP BeeMuSe
plot1 + geom_point(data = eigenvec_merged_Corres_ID_E756_pedigree, aes(x = V3, y = V4, color = ID_2a), s
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.075, xmax = 0.04, ymin = -0.1, ymax = -0.05)

```



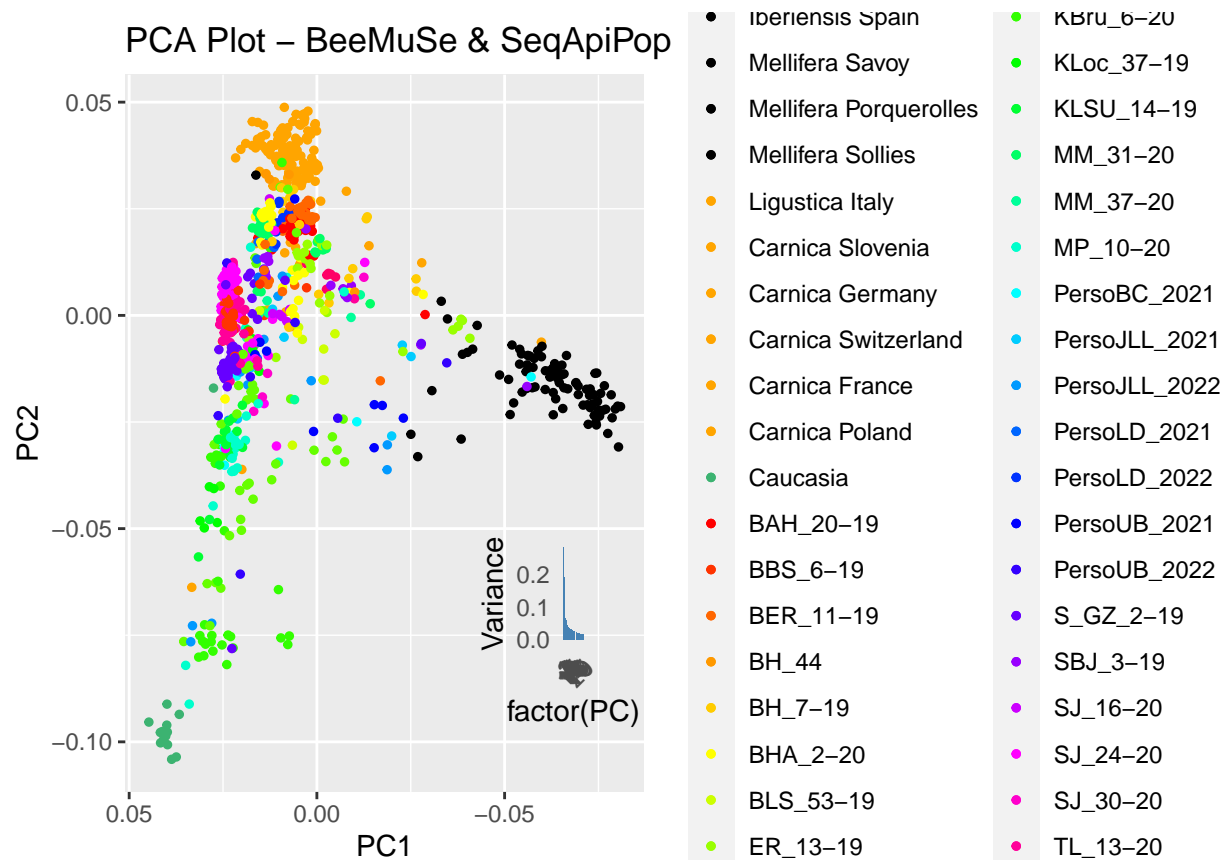
```

# SeqApiPop - 301 Samples - Label - colors - 3 lignées
custom_colors <- c("black", "black", "black",
                  "black", "orange", "orange", "orange", "orange", "orange", "orange",
                  "mediumseagreen",
                  "#FF0000FF", "#FF3300FF", "#FF6600FF", "#FF9900FF", "#FFCC00FF",
                  "#FFFF00FF", "#CCFF00FF", "#99FF00FF", "#66FF00FF", "#33FF00FF",
                  "#00FF00FF", "#00FF33FF", "#00FF66FF", "#00FF99FF", "#00FFCCFF",
                  "#00FFFFFF", "#00CCFFFF", "#0099FFFF", "#0066FFFF", "#0033FFFF",
                  "#0000FFFF", "#3300FFFF", "#6600FFFF", "#9900FFFF", "#CC00FFFF",
                  "#FF00FFFF", "#FF00CCFF", "#FF0099FF", "#FF0066FF")

# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point(size=1) +
  scale_x_reverse() +
  scale_color_manual(values = custom_colors,
                    breaks = c("Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
                              "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
                              "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
                              "TL_19-20"),
                    labels = c("Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia", "France",
                              "KBJ_1-19", "KBru_6-20", "KLoc_37-19", "KLSU_14-19", "MM_31-20", "MM_37-20", "MP_10-20",
                              "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
                              "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
                              "TL_19-20"))) +
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2") +
  theme(legend.position = "right")

# Ajout ACP BeeMuSe
plot1 + geom_point(data = eigenvec_merged_Corres_ID_E756_pedigree, aes(x = V3, y = V4, color = ID_2a), size=1) +
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.075, xmax = 0.04, ymin = -0.1, ymax = -0.05)

```



```
# SeqApiPop - 301 - Label - couleurs - 11 populations de référence
custom_colors <- c("mediumpurple3", "lightskyblue", "mediumvioletred",
  "hotpink1", "yellow", "gold", "orange", "chocolate", "brown", "olivedrab3",
  "mediumseagreen",
  "#FF0000FF", "#FF3300FF", "#FF6600FF", "#FF9900FF", "#FFCC00FF",
  "#FFFF00FF", "#CCFF00FF", "#99FF00FF", "#66FF00FF", "#33FF00FF",
  "#00FF00FF", "#00FF33FF", "#00FF66FF", "#00FF99FF", "#00FFCCFF",
  "#00FFFFFF", "#00CCFFFF", "#0099FFFF", "#0066FFFF", "#0033FFFF",
  "#0000FFFF", "#3300FFFF", "#6600FFFF", "#9900FFFF", "#CC00FFFF",
  "#FF00FFFF", "#FF00CCFF", "#FF0099FF", "#FF0066FF")

# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point(size = 1) +
  scale_x_reverse() +
  scale_color_manual(values = custom_colors,
    breaks = c("Iberiensis Spain", "Savoy Conservatory", "Porquerolles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
      "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
      "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
      "TL_19-20"),
    labels = c("Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia", "Ligustica France",
      "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
      "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20", "TL_19-20"))
```

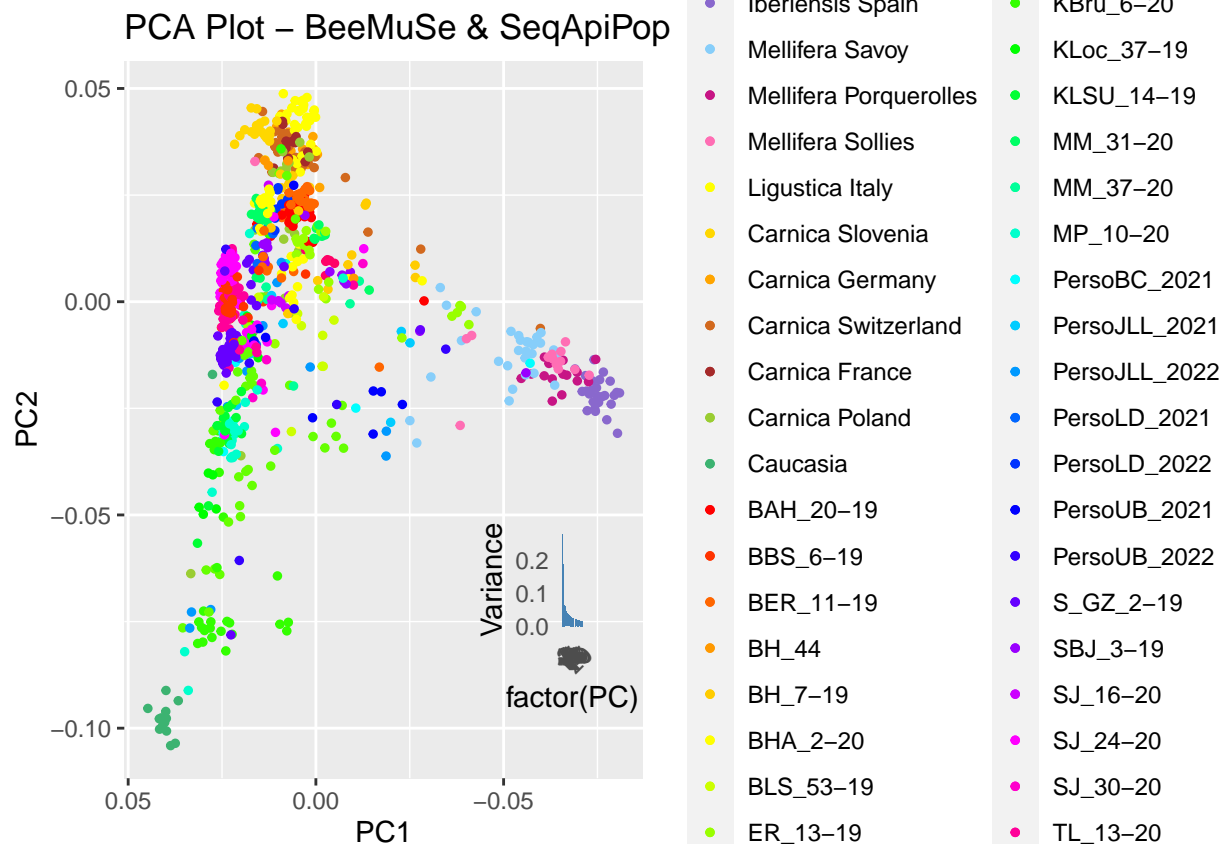


```

"KBJ_1-19", "KBru_6-20", "KLoc_37-19", "KLSU_14-19", "MM_31-20", "MM_37-20", "MP_10-20",
"PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
"PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
"TL_19-20")) +
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2") +
  theme(legend.position = "right")

# Ajout ACP BeeMuSe
plot1 + geom_point(data = eigenvec_merged_Corres_ID_E756_pedigree, aes(x = V3, y = V4, color = ID_2a),
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC1", y = "PC2") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.075, xmax = 0.04, ymin = -0.1, ymax = -0.05)

```



```

# ACP - PC3 / PC4

# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V5, y = V6, color = Label)) +

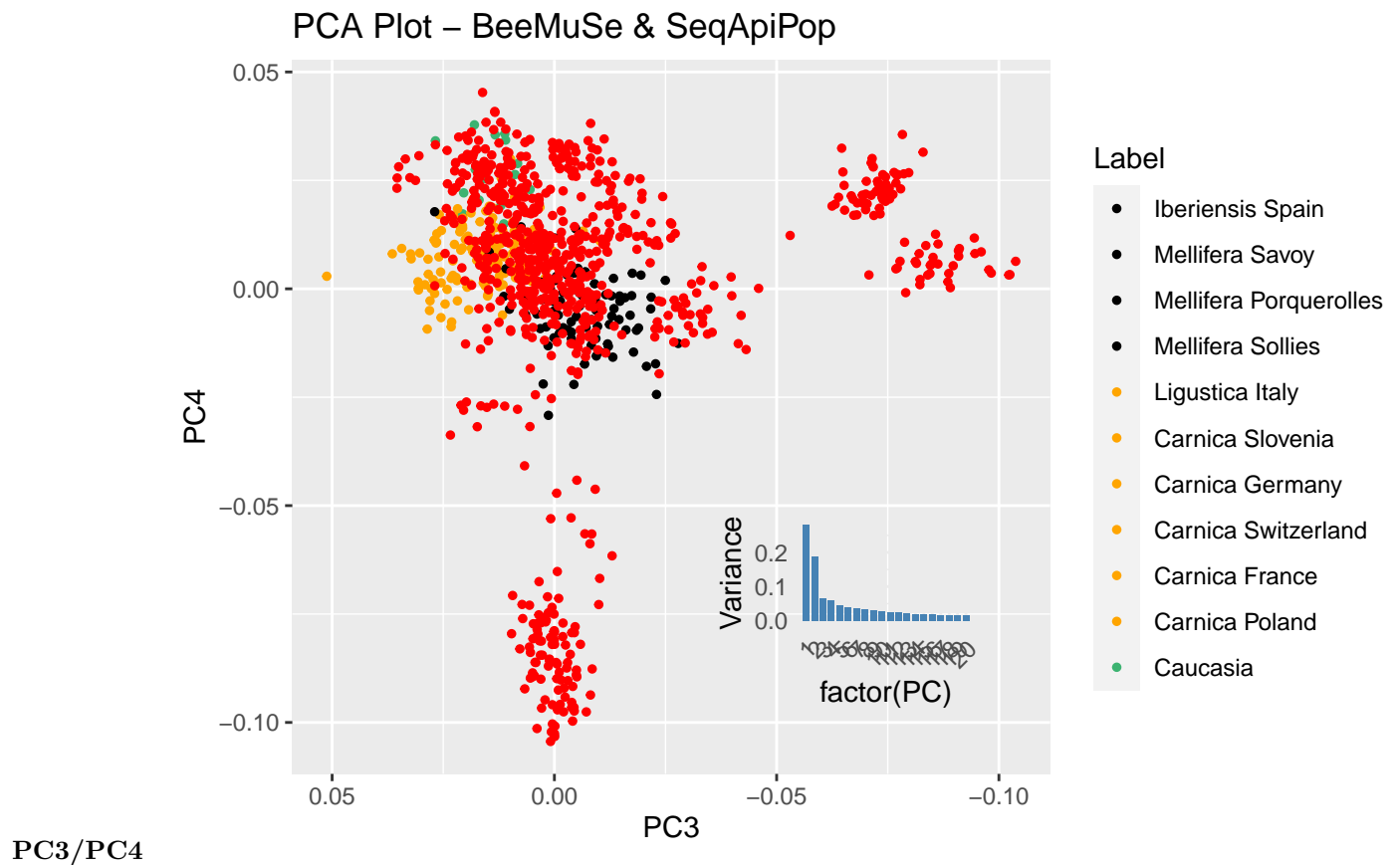
```

```

geom_point(size = 1) +
scale_x_reverse() +
scale_color_manual(values = custom_colors_2,
  breaks = c("Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
  labels = c("Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

# Ajout ACP BeeMuSe
plot1 + geom_point(data = eigenvec_BeeMuSe_748_Samples, aes(x = V5, y = V6), color = "red", size = 1) +
labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC3", y = "PC4") +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.097, xmax = 0.034, ymin = -0.1, ymax = -0.05)

```



```

# Couleurs 11 populations de référence
custom_colors <- c("mediumpurple3", "lightskyblue", "mediumvioletred", "hotpink1", "yellow", "gold", "black")

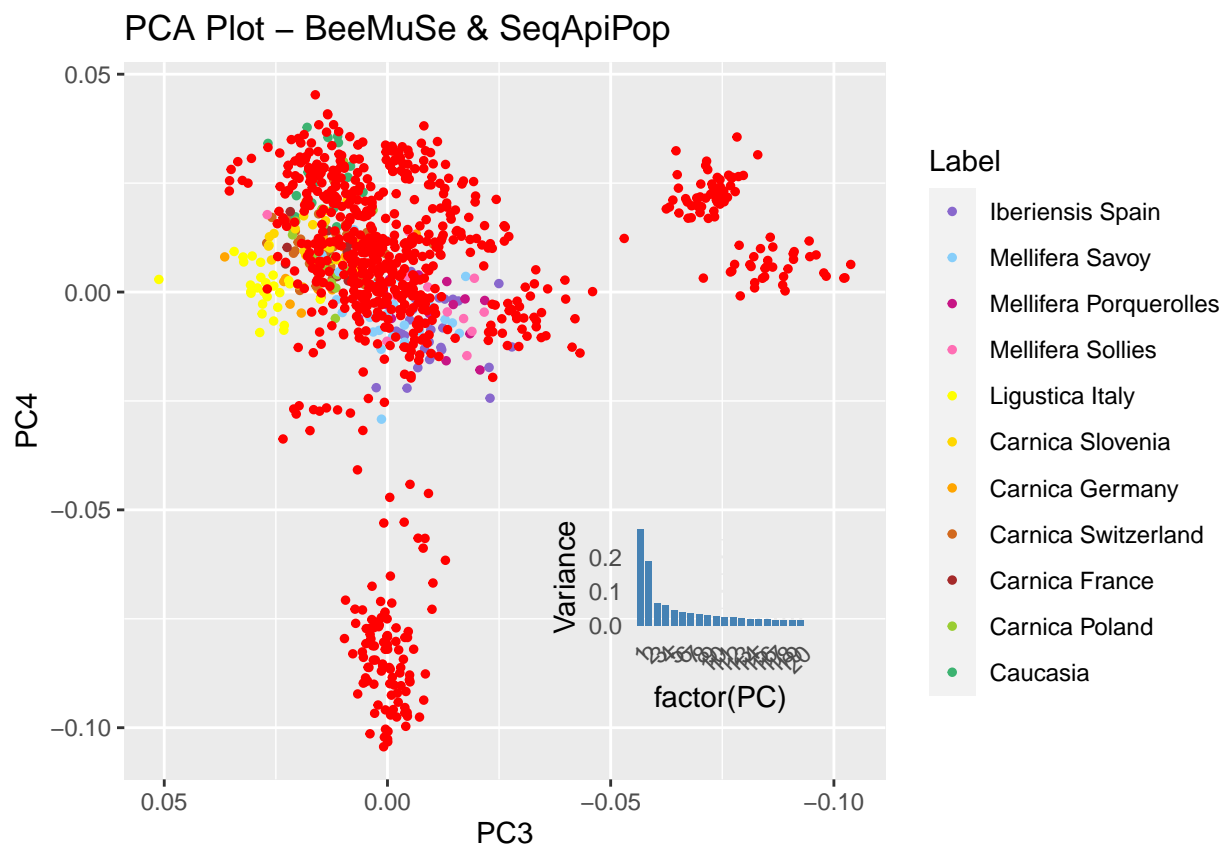
```

```

# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point(size = 1) +
  scale_x_reverse() +
  scale_color_manual(values = custom_colors,
                     breaks = c("Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                     labels = c("Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

# Ajout ACP BeeMuSe
plot1 + geom_point(data = eigenvec_BeeMuSe_748_Samples, aes(x = V5, y = V6), color = "red", size = 1) +
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC3", y = "PC4") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.097, xmax = 0.034, ymin = -0.1, ymax = -0.05)

```



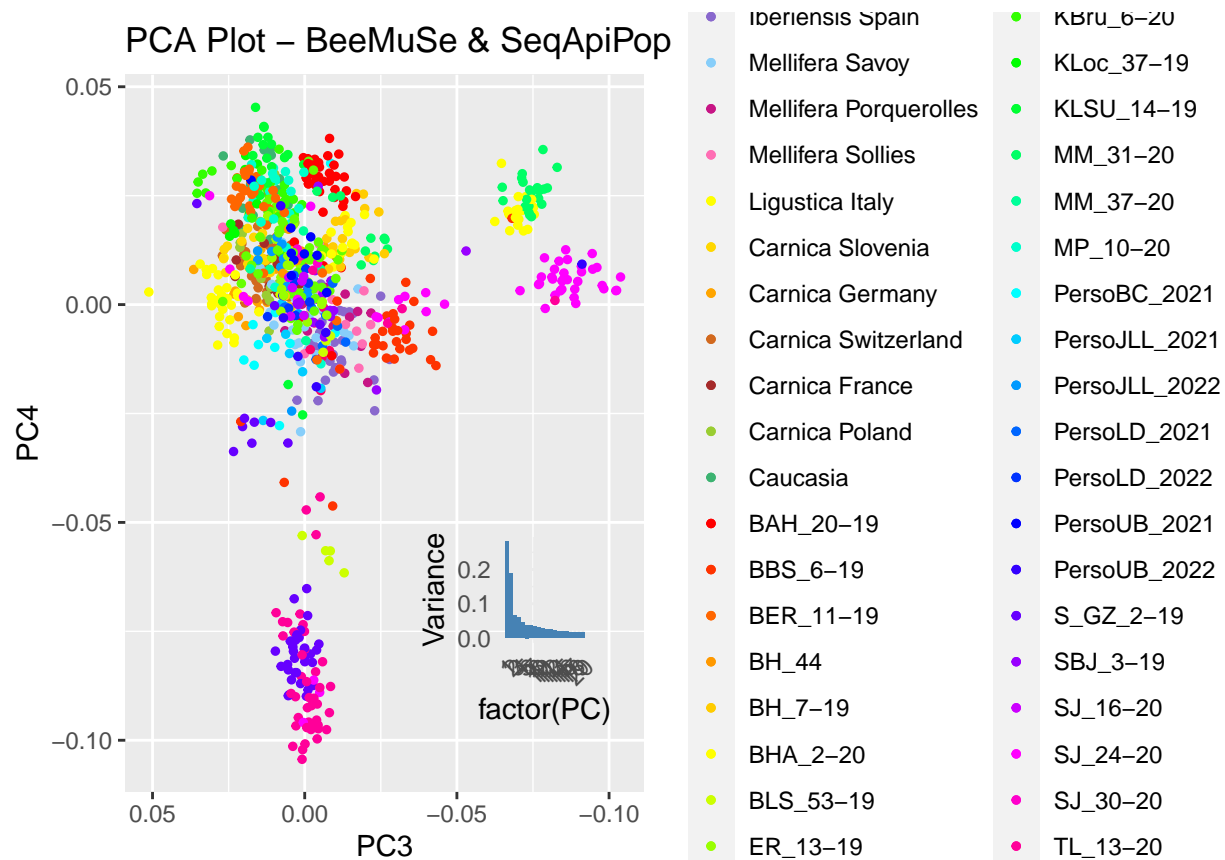
```

# Couleurs des 11 populations de référence SeqApiPop + 29 familles ID_2a
custom_colors <- c("mediumpurple3", "lightskyblue", "mediumvioletred",
  "hotpink1", "yellow", "gold", "orange", "chocolate", "brown", "olivedrab3",
  "mediumseagreen",
  "#FF0000FF", "#FF3300FF", "#FF6600FF", "#FF9900FF", "#FFCC00FF",
  "#FFFF00FF", "#CCFF00FF", "#99FF00FF", "#66FF00FF", "#33FF00FF",
  "#00FF00FF", "#00FF33FF", "#00FF66FF", "#00FF99FF", "#00FFCCFF",
  "#00FFFFFF", "#00CCFFFF", "#0099FFFF", "#0066FFFF", "#0033FFFF",
  "#0000FFFF", "#3300FFFF", "#6600FFFF", "#9900FFFF", "#CC00FFFF",
  "#FF00FFFF", "#FF00CCFF", "#FF0099FF", "#FF0066FF")

# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point(size = 1) +
  scale_x_reverse() +
  scale_color_manual(values = custom_colors,
    breaks = c("Iberiensis Spain", "Savoy Conservatory", "Porquerolles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr",
      "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
      "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
      "TL_19-20"),
    labels = c("Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia",
      "KBJ_1-19", "KBru_6-20", "KLoc_37-19", "KLSU_14-19", "MM_31-20", "MM_37-20", "MP_10-20",
      "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
      "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
      "TL_19-20")) +
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC3", y = "PC4") +
  theme(legend.position = "right")

# Ajouter ACP BeeMuSe
plot1 + geom_point(data = eigenvec_merged_Corres_ID_E756_pedigree, aes(x = V5, y = V6, color = ID_2a),
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC3", y = "PC4") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.097, xmax = 0.034, ymin = -0.1, ymax = -0.05)

```



```
# Couleurs 3 lignées SeqApiPop + 29 familles ID_2a
custom_colors <- c( "black", "black", "black",
  "black", "orange", "orange", "orange", "orange", "orange", "orange",
  "mediumseagreen",
  "#FF0000FF", "#FF3300FF", "#FF6600FF", "#FF9900FF", "#FFCC00FF",
  "#FFFF00FF", "#CCFF00FF", "#99FF00FF", "#66FF00FF", "#33FF00FF",
  "#00FF00FF", "#00FF33FF", "#00FF66FF", "#00FF99FF", "#00FFCCFF",
  "#00FFFFFF", "#00CCFFFF", "#0099FFFF", "#0066FFFF", "#0033FFFF",
  "#0000FFFF", "#3300FFFF", "#6600FFFF", "#9900FFFF", "#CC00FFFF",
  "#FF00FFFF", "#FF00CCFF", "#FF0099FF", "#FF0066FF")

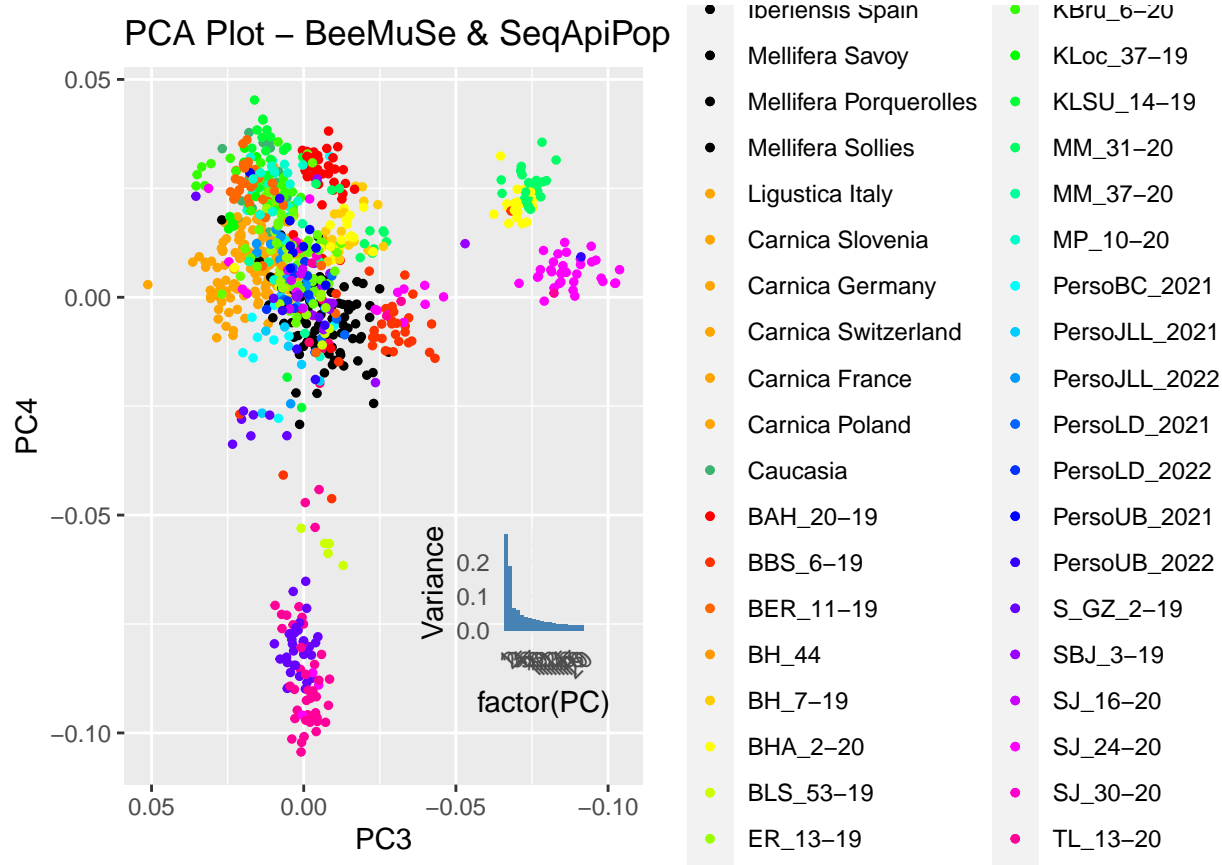
# ACP SeqApiPop
plot1 <- ggplot(eigenvec_merged_maf001_LD03_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point(size = 1) +
  scale_x_reverse() +
  scale_color_manual(values = custom_colors,
    breaks = c( "Iberiensis Spain", "Savoy Conservatory", "Porquerolles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
      "PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
      "PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
      "TL_19-20"),
    labels = c( "Iberiensis Spain", "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany", "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia", "Iberiensis Spain",
      "KBJ_1-19", "KBru_6-20", "KLoc_37-19", "KLSU_14-19", "MM_31-20", "MM_37-20", "MP_10-20",
```

```

"PersoBC_2021", "PersoJLL_2021", "PersoJLL_2022", "PersoLD_2021", "PersoLD_2022", "PersoUB_2021",
"PersoUB_2022", "S_GZ_2-19", "SBJ_3-19", "SJ_16-20", "SJ_24-20", "SJ_30-20", "TL_13-20",
"TL_19-20"))

# Ajout ACP BeeMuSe
plot1 + geom_point(data = eigenvec_merged_Corres_ID_E756_pedigree, aes(x = V5, y = V6, color = ID_2a),
  labs(title = "PCA Plot - BeeMuSe & SeqApiPop", x = "PC3", y = "PC4") +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.097, xmax = 0.034, ymin = -0.1, ymax = -0.05)

```



ACP - SeqApiPop - 629 échantillons - MAF > 0.01

```

setwd("~/Documents/Stage_NB/data/maf001_LD03")

# fichiers pour SeqApiPop 629 échantillons et filtre maf001
#eigenvec_refpop <- read.table("SeqApiPop_629_maf001_acp.eigenvec", header = F)
#eigenval_refpop <- read.table("SeqApiPop_629_maf001_acp.eigenval", header = F)

```

```

# fichiers pour SeqApiPop 629 échantillons et filtre maf001 LD pruning = 0.3 (fenêtre de 1749 SNPs et p
eigenvec_refpop <- read.table("SeqApiPop_629_maf001_LD03_acp.eigenvec", header = F)
eigenval_refpop <- read.table("SeqApiPop_629_maf001_LD03_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_refpop)[colnames(eigenvec_refpop) == "V2"] <- "name"
eigenvec_refpop_seq_api_labels <- merge(eigenvec_refpop, seq_api_labels, by = "name")

eigen_percent_refpop <- round((eigenval_refpop / (sum(eigenval_refpop) )*100),2)

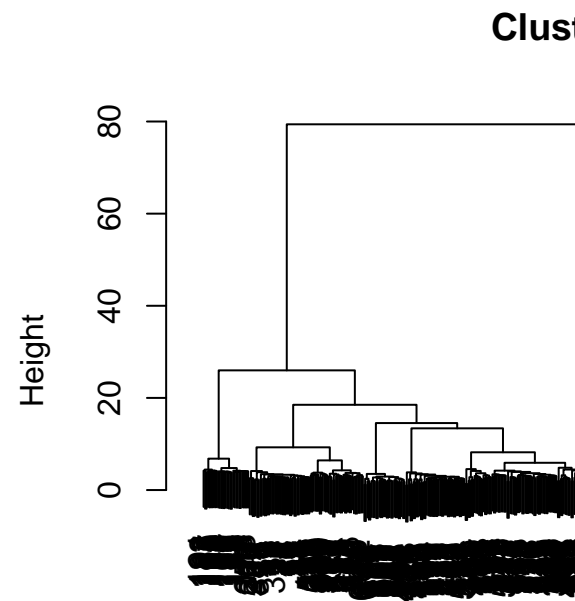
```

```

#Clustering hiérarchique
#Tree
setwd("~/Documents/Stage_NB/data/maf001_LD03")
matrice_app_refpop <- read.table("SeqApiPop_629_maf001_LD03_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)

```



ACP - LD pruning = 0.3 (fenêtre de 1749 SNPS et pas de 175 bp)

```

#heatmap
#heatplot(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')

```

```

#ACP
#filter 629 -> 301 RefPop
eigenvec_refpop_seq_api_labels <- eigenvec_refpop_seq_api_labels[eigenvec_refpop_seq_api_labels$Genetic
                                eigenvec_refpop_seq_api_labels$Label
                                eigenvec_refpop_seq_api_labels$Label
                                eigenvec_refpop_seq_api_labels$Unique
                                eigenvec_refpop_seq_api_labels$Unique
                                eigenvec_refpop_seq_api_labels$Genet.

custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred",

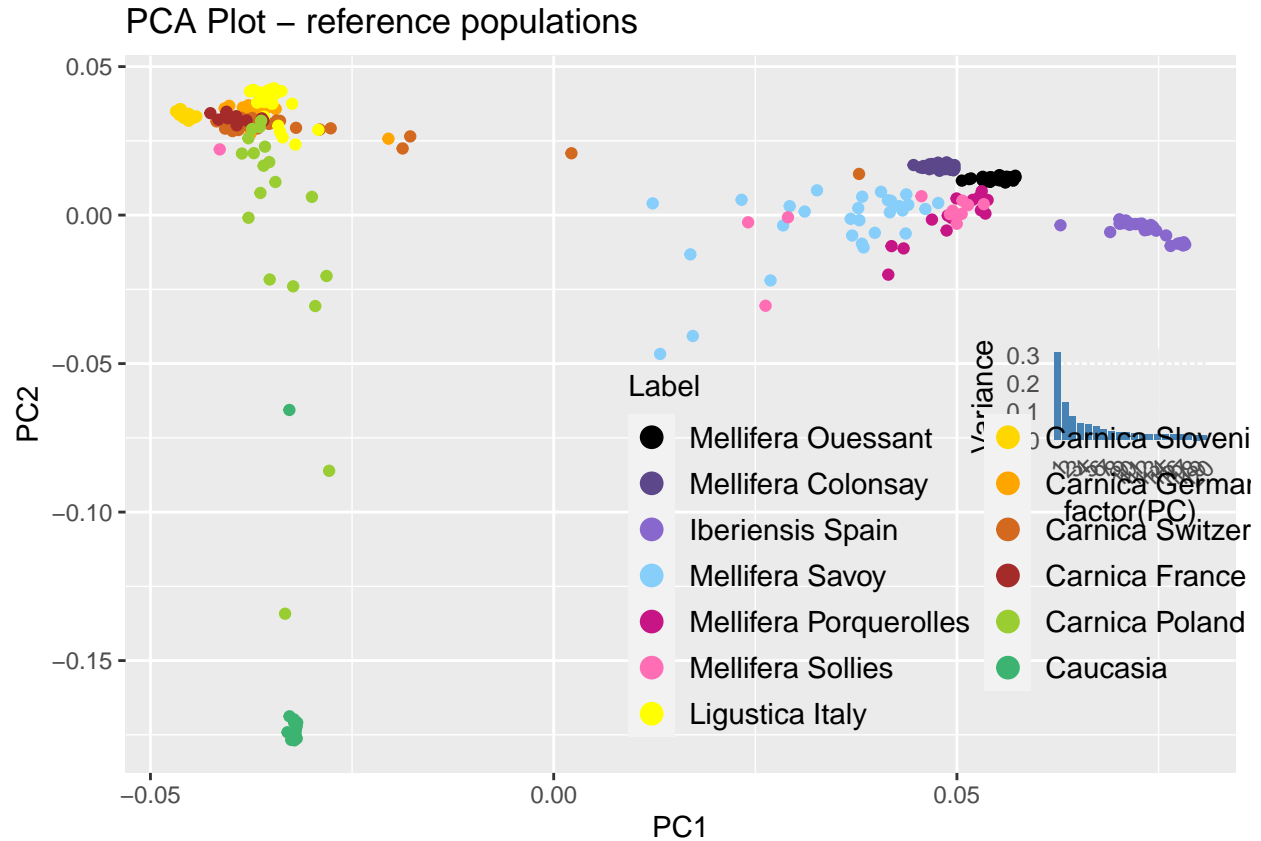
lambda <- eigenval_refpop$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

```

```

#PC1/PC2
# ACP avec variance expliquée
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                     breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                     labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.44, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
      geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
      theme_minimal() +
      theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.083, xmax = 0.05, ymin = -0.11, ymax = -0.04)

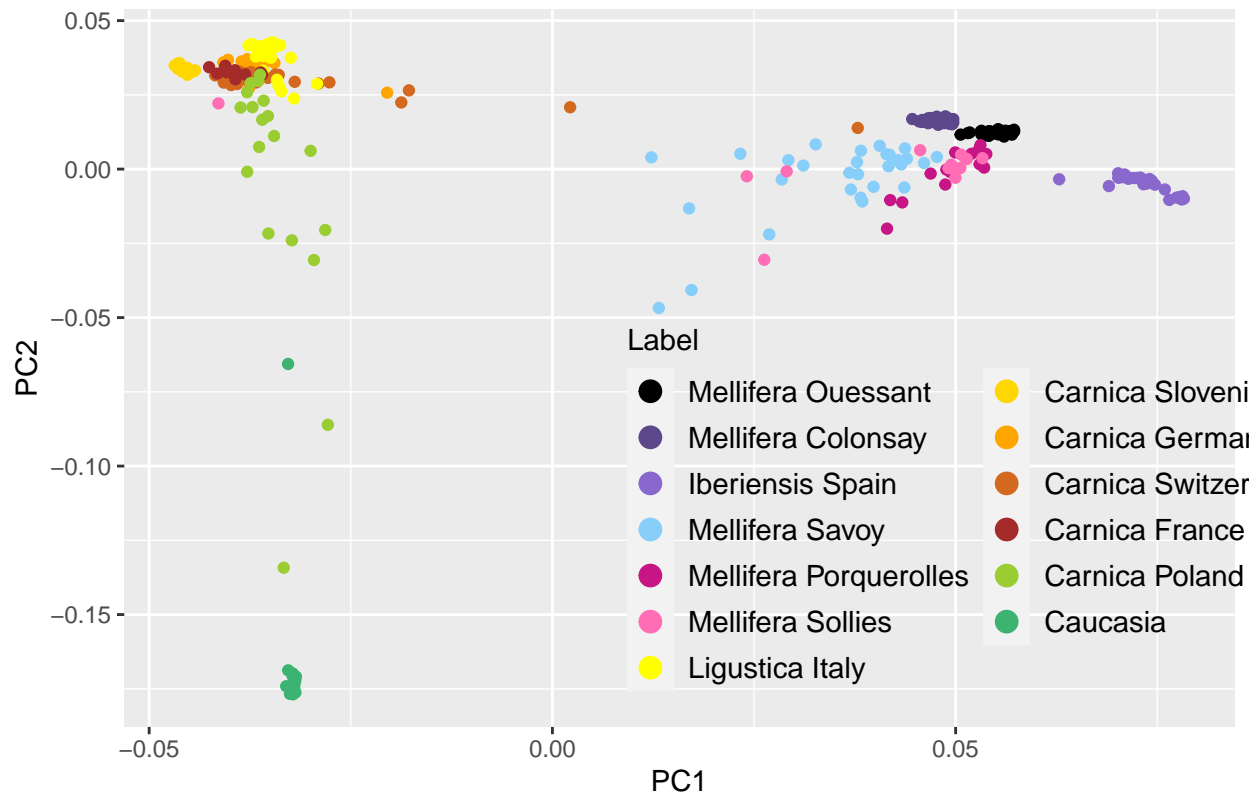
```

PC1/PC2

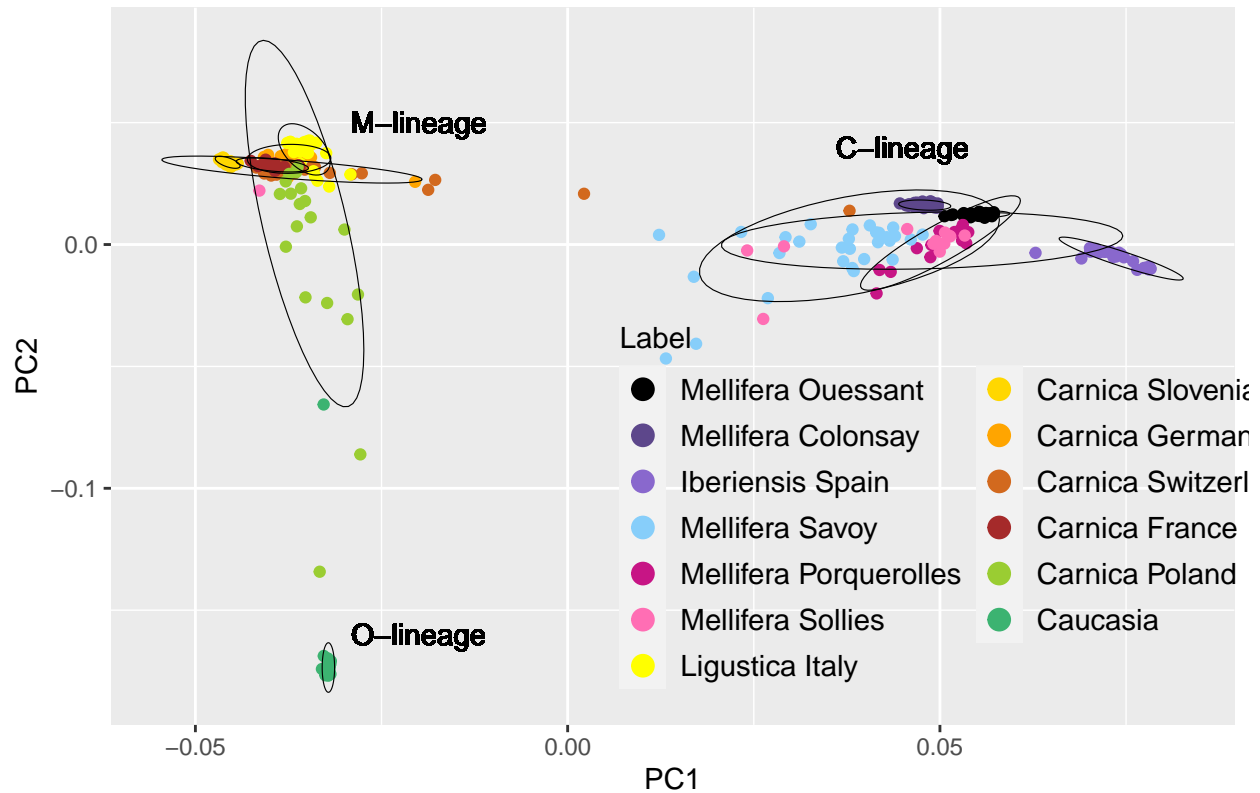
```
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensus Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensus Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.44, 0.03), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),
    legend.key.size = unit(1.2, "lines"),
    legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))
```

PCA Plot – reference populations



```
#ellipses autour des points selon Label
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black") +
  geom_text(aes(x = 0.045, y = 0.04, label = "C-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.02, y = 0.05, label = "M-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.02, y = -0.16, label = "O-lineage"), size = 4, color = "black") +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.44, 0.03), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),
    legend.key.size = unit(1.2, "lines"),
    legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))
```

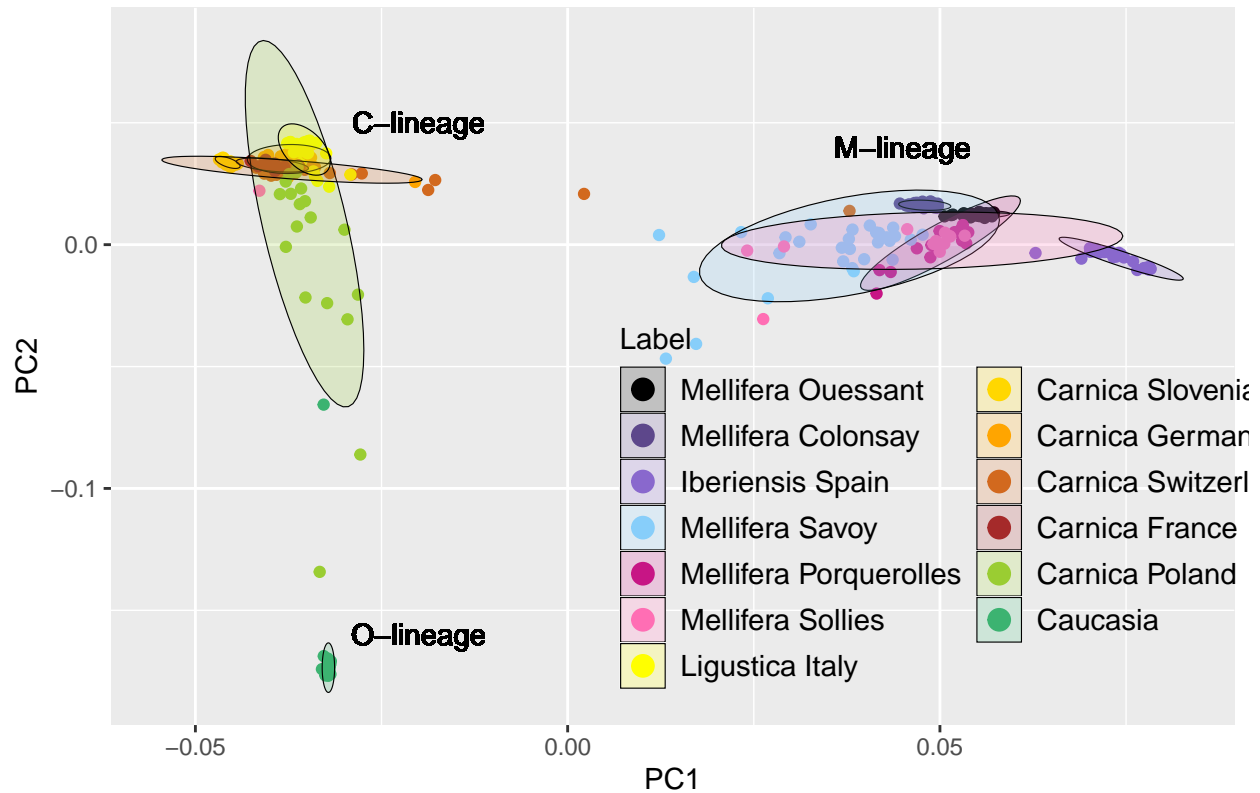
PCA Plot – reference populations



```
# ellipses avec couleur
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  geom_text(aes(x = 0.045, y = 0.04, label = "M-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.02, y = 0.05, label = "C-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.02, y = -0.16, label = "O-lineage"), size = 4, color = "black") +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
  scale_fill_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
```

```
theme(legend.position = c(0.44, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))
```

PCA Plot – reference populations



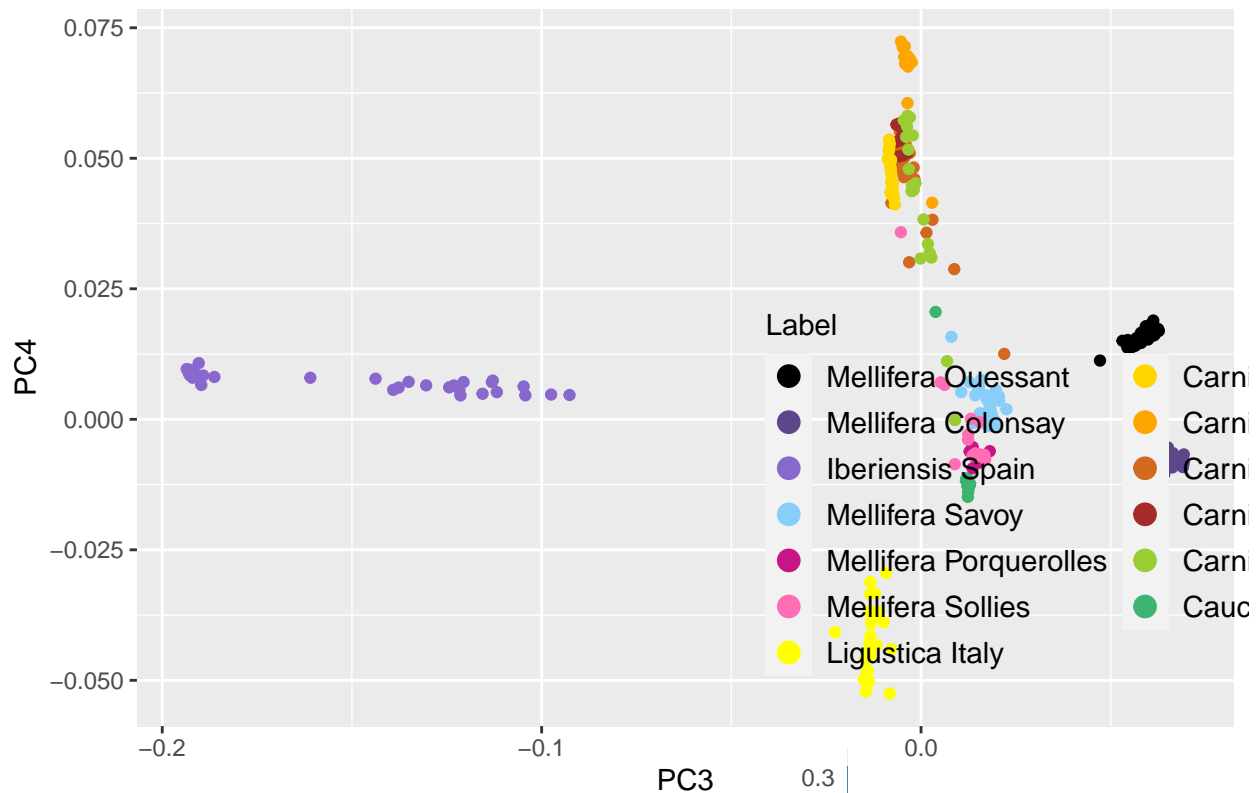
```
#ACP - plot variance - PC3 - PC4
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
                     breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                     labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.56, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
```

```

legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = -0.04, xmax = -0.02, ymin = -0.16, ymax = -0.06)

```

PCA Plot – reference populations



PC3/PC4

```

#LD02
setwd("~/Documents/Stage_NB/data/maf001_LD02")

eigenvec_LD02 <- read.table("SeqApiPop_629_maf001_LD02_acp.eigenvec", header = F)
eigenval_LD02 <- read.table("SeqApiPop_629_maf001_LD02_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_LD02)[colnames(eigenvec_LD02) == "V2"] <- "name"
eigenvec_LD02_seq_api_labels <- merge(eigenvec_LD02, seq_api_labels, by = "name")

eigen_percent_LD02 <- round((eigenval_LD02 / (sum(eigenval_LD02) ) * 100), 2)
eigen_percent_LD02

```

ACP - LD pruning = 0.2

```
##      V1
## 1  27.79
## 2  10.17
## 3   8.92
## 4   5.95
## 5   5.90
## 6   5.40
## 7   3.98
## 8   3.26
## 9   3.06
## 10  2.67
## 11  2.62
## 12  2.59
## 13  2.40
## 14  2.28
## 15  2.26
## 16  2.21
## 17  2.17
## 18  2.16
## 19  2.11
## 20  2.09
```

```
#Clustering hiérarchique
```

```
#Tree
```

```
matrice_app_refpop <- read.table("SeqApiPop_629_maf001_LD02_acp.rel", header = FALSE)
```

```
dist_matrice_refpop <- dist(matrice_app_refpop)
```

```
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
```

```
plot(hc_refpop)
```

Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

```
#heatmap
#heatplot(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')
```

```
# filter 629 -> 301 RefPop
```

```
eigenvec_LD02_seq_api_labels <- eigenvec_LD02_seq_api_labels[eigenvec_LD02_seq_api_labels$GeneticOrigin != "Quessant Conservatory", ]
eigenvec_LD02_seq_api_labels <- eigenvec_LD02_seq_api_labels[eigenvec_LD02_seq_api_labels$Label != "Colonsay Conservatory", ]
eigenvec_LD02_seq_api_labels <- eigenvec_LD02_seq_api_labels[eigenvec_LD02_seq_api_labels$UniqueID != "Iberiensis Spain", ]
eigenvec_LD02_seq_api_labels <- eigenvec_LD02_seq_api_labels[eigenvec_LD02_seq_api_labels$GeneticOrigin != "Quessant Conservatory", ]
```

```
custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred", "yellowgreen")
```

```
lambda <- eigenval_LD02$V1
```

```
variance_proportion <- lambda / sum(lambda)
```

```
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)
```

```
ggplot(data = eigenvec_LD02_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Quessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain", "Quessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain", "Quessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain"))
```

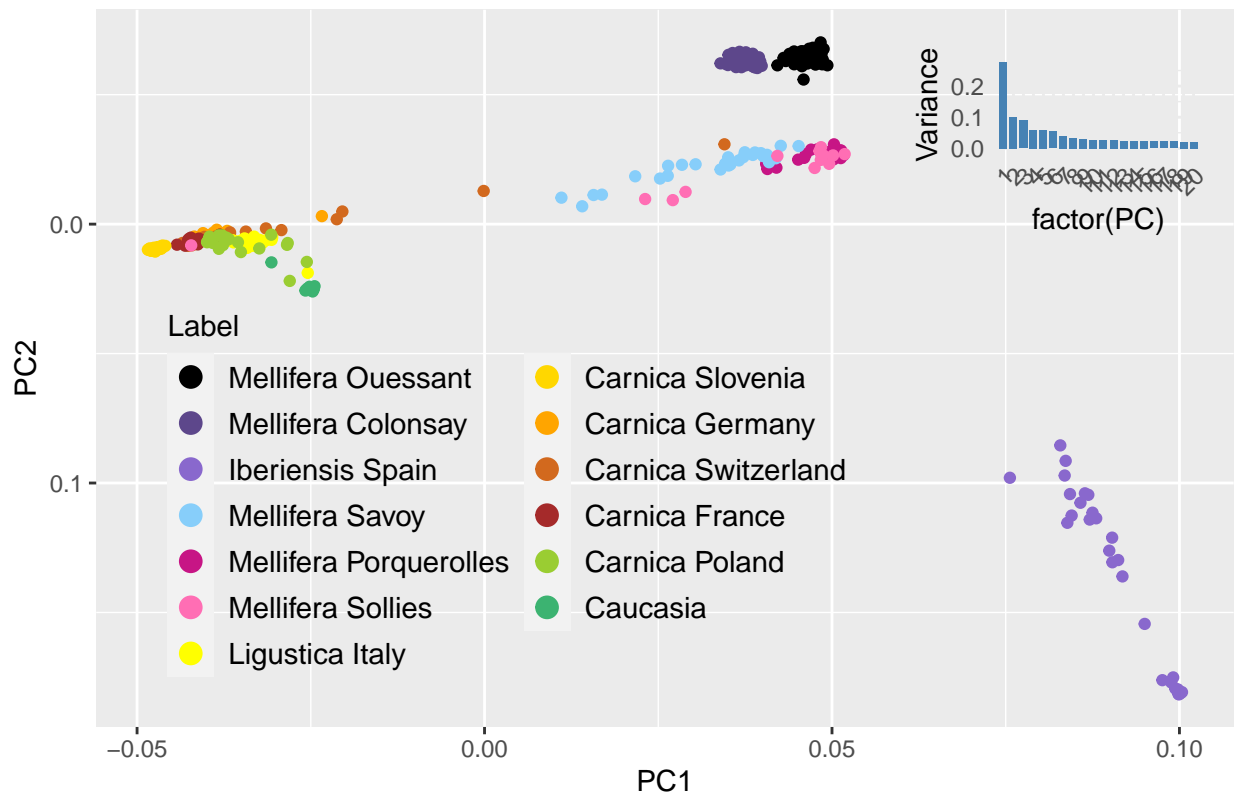
```

      "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
      labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() +
theme(legend.position = c(0.05, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.105, xmax = 0.06, ymin = -0.01, ymax = 0.07)

```

PCA Plot – reference populations



PC1/PC2

```

ggplot(data = eigenvec_LD02_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,

```



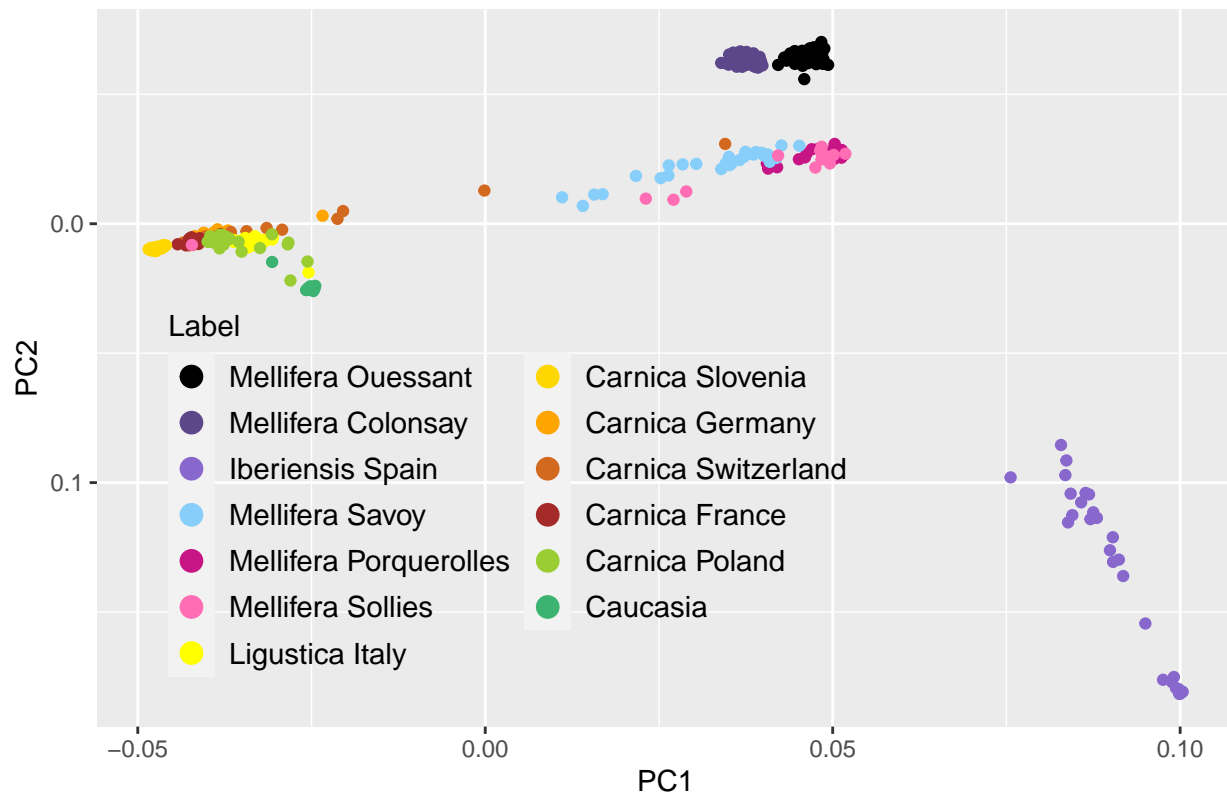
```

breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
            "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
            "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
            "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France")
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
            "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
            "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
            "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia")

scale_y_reverse() +
theme(legend.position = c(0.05, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses autour des points selon Label
ggplot(data = eigenvec_LD02_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))

```

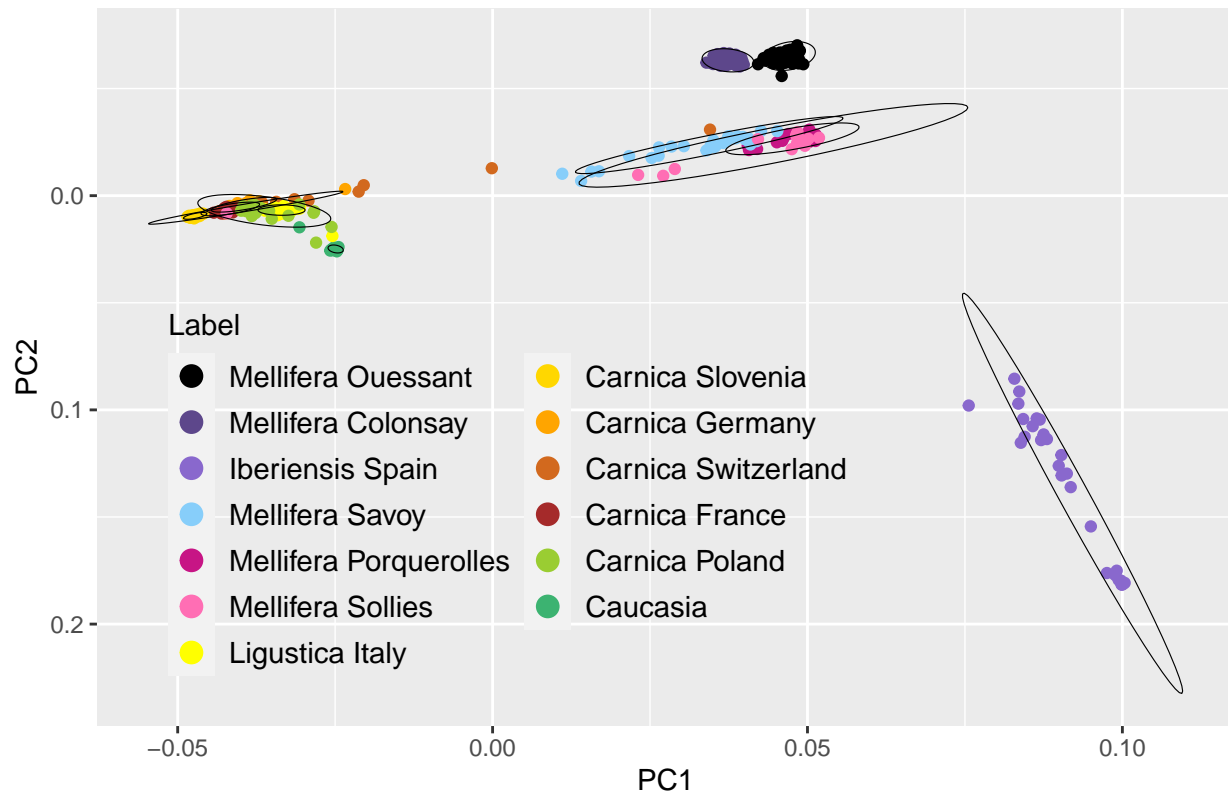
```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() +
theme(legend.position = c(0.05, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses avec couleur
ggplot(data = eigenvec_LD02_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5)
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
                   breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                   labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

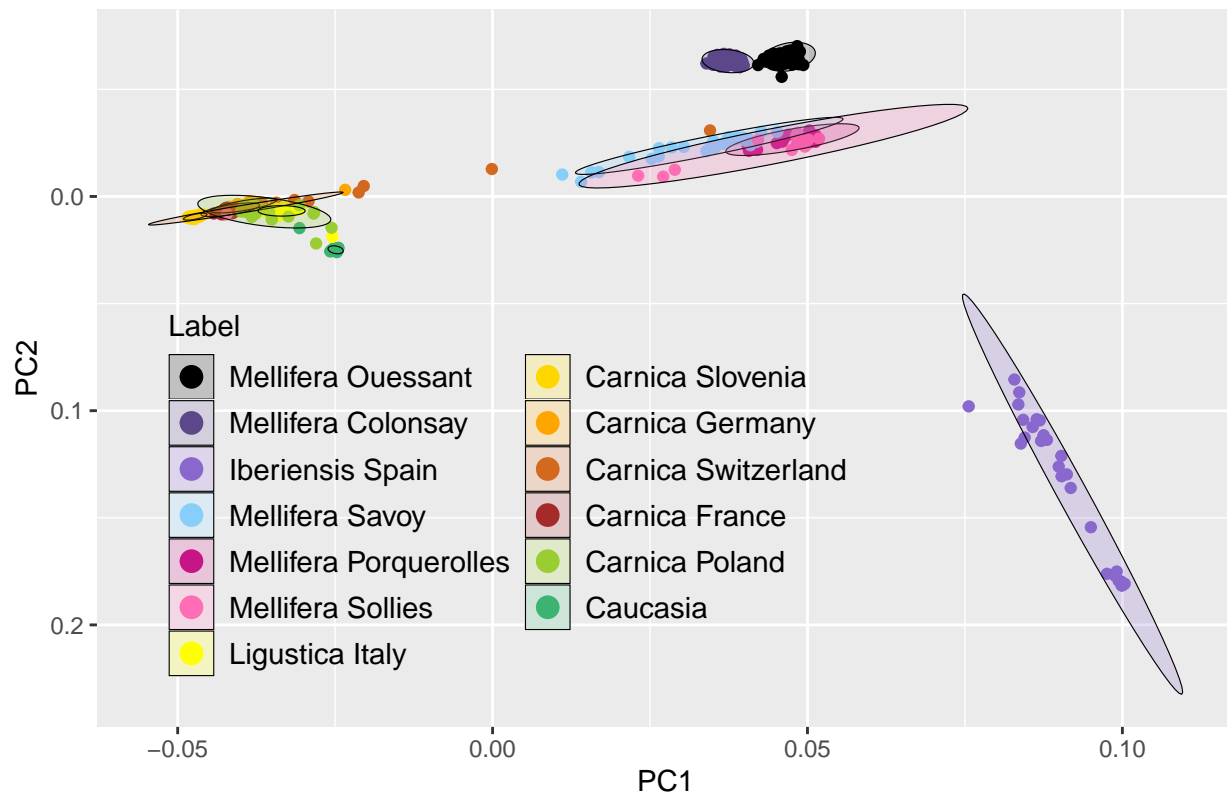
```

```

scale_fill_manual(values = custom_colors_label2,
  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
  labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia")) +
scale_y_reverse() +
theme(legend.position = c(0.05, 0.05), legend.justification = c(0, 0),
  legend.background = element_rect(fill = "transparent"),
  legend.key.size = unit(1.2, "lines"),
  legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

ggplot(data = eigenvec_LD02_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))

```

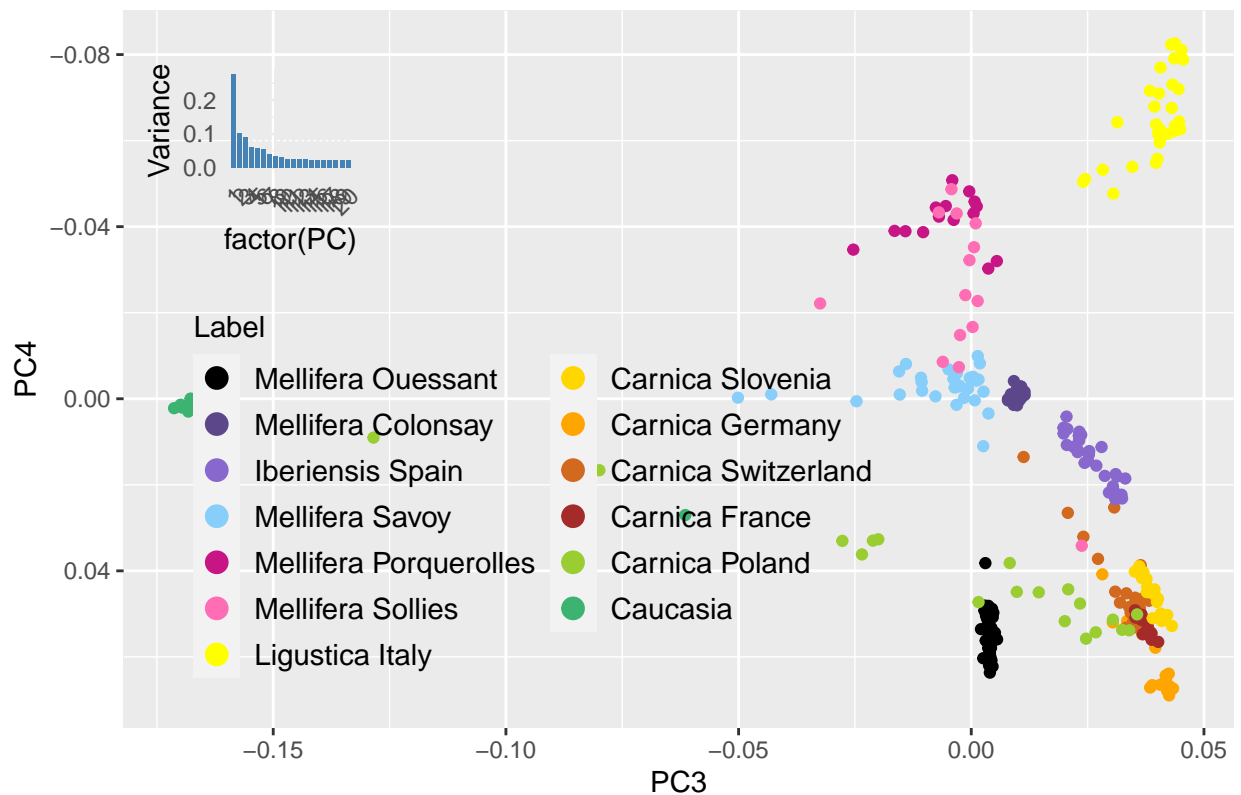
```

    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.05, 0.05), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),
    legend.key.size = unit(1.2, "lines"),
    legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
    ggplotGrob(
        ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = -0.13, xmax = -0.18, ymin = 0.03, ymax = 0.08)

```

PCA Plot – reference populations



PC3/PC4

ACP - LD pruning = 0.1

```

#LD01
setwd("~/Documents/Stage_NB/data/maf001_LD01")

eigenvec_LD01 <- read.table("SeqApiPop_629_maf001_LD01_acp.eigenvec", header = F)
eigenval_LD01 <- read.table("SeqApiPop_629_maf001_LD01_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_LD01)[colnames(eigenvec_LD01) == "V2"] <- "name"
eigenvec_LD01_seq_api_labels <- merge(eigenvec_LD01, seq_api_labels, by = "name")

eigen_percent_LD01 <- round((eigenval_LD01 / (sum(eigenval_LD01) )*100),2)
eigen_percent_LD01

```

PC1/PC2

```

##      V1
## 1  24.21
## 2  10.02
## 3   6.20
## 4   5.78
## 5   5.18
## 6   4.50
## 7   3.93
## 8   3.75
## 9   3.46
## 10  3.18
## 11  3.10
## 12  3.07
## 13  3.02
## 14  3.00
## 15  2.96
## 16  2.95
## 17  2.93
## 18  2.92
## 19  2.92
## 20  2.90

```

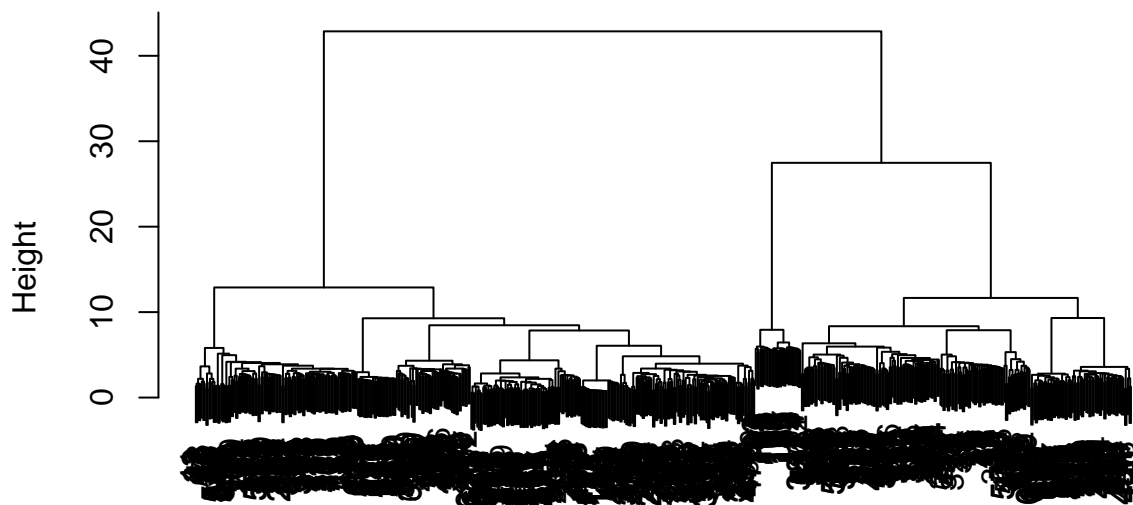
```

#Clustering hiérarchique
#Tree
matrice_app_refpop <- read.table("SeqApiPop_629_maf001_LD01_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)

```

Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

```
#heatmap
#heatmap(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')

#ACP
#filter 629 -> 301 RefPop
eigenvec_LD01_seq_api_labels <- eigenvec_LD01_seq_api_labels[eigenvec_LD01_seq_api_labels$GeneticOrigin != "A", ]
eigenvec_LD01_seq_api_labels <- eigenvec_LD01_seq_api_labels[eigenvec_LD01_seq_api_labels$Label != "B", ]
eigenvec_LD01_seq_api_labels <- eigenvec_LD01_seq_api_labels[eigenvec_LD01_seq_api_labels$UniqueInHive == "Yes", ]
eigenvec_LD01_seq_api_labels <- eigenvec_LD01_seq_api_labels[eigenvec_LD01_seq_api_labels$UniqueInHive == "Yes", ]
eigenvec_LD01_seq_api_labels <- eigenvec_LD01_seq_api_labels[eigenvec_LD01_seq_api_labels$GeneticOrigin != "A", ]

custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred", "darkgreen")

lambda <- eigenval_LD01$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

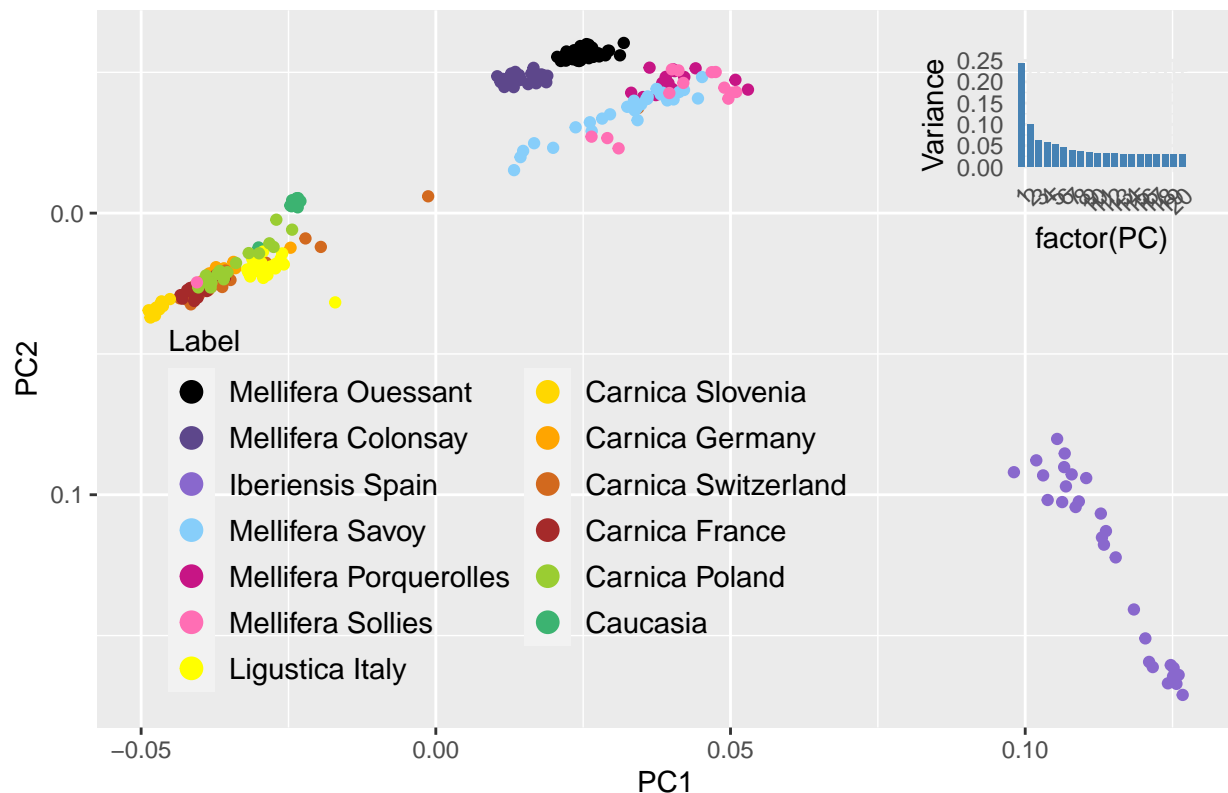
ggplot(data = eigenvec_LD01_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Quessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solies Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
```

```

    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
"Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
"Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
"Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.05, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.13, xmax = 0.08, ymin = -0.02, ymax = 0.06)

```

PCA Plot – reference populations



```

ggplot(data = eigenvec_LD01_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
geom_point() +
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory")

```

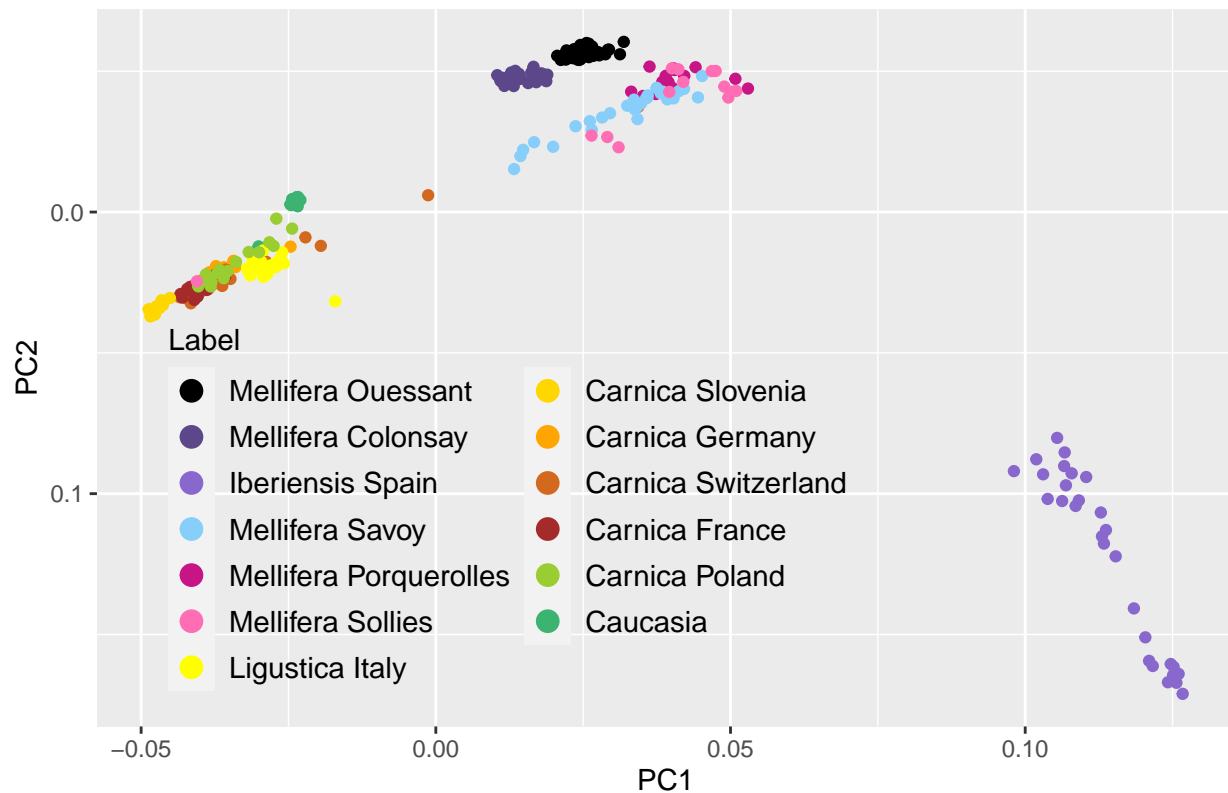
```

    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.05, 0.03), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),
    legend.key.size = unit(1.2, "lines"),
    legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#ellipses autour des points selon Label
ggplot(data = eigenvec_LD01_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",

```

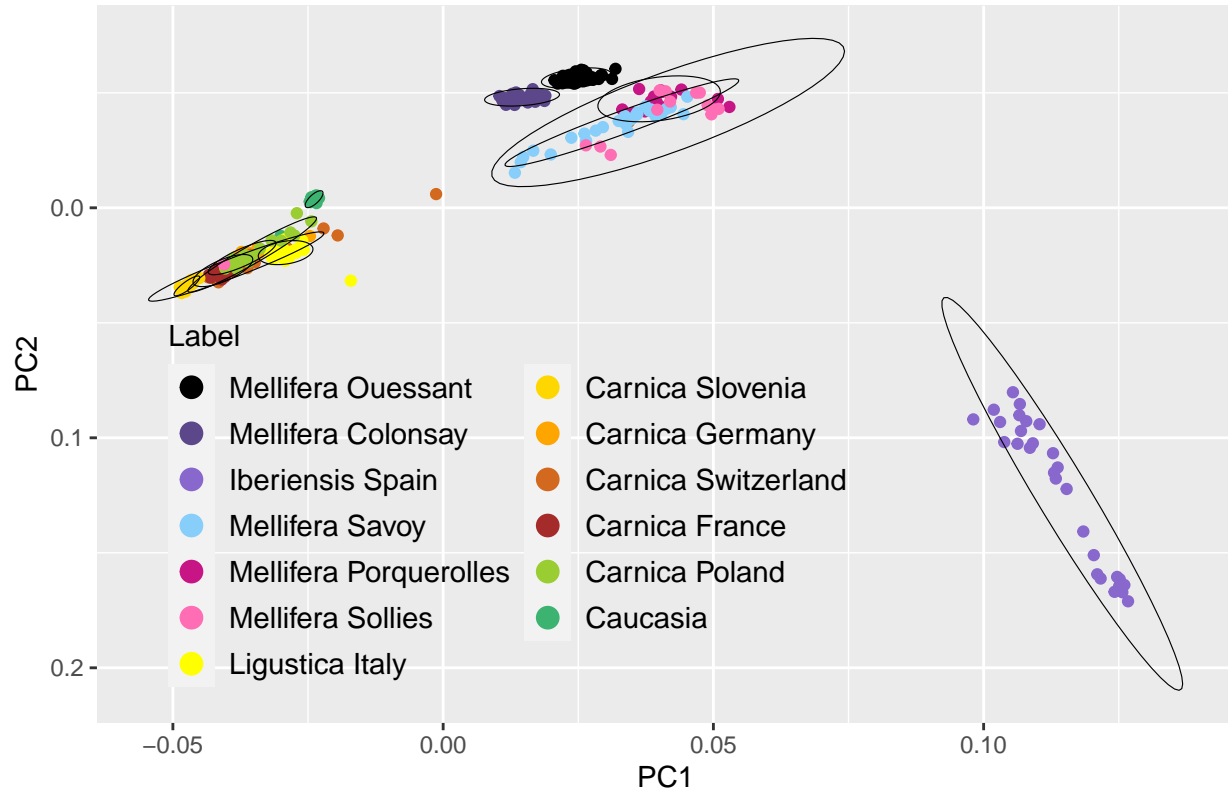


```

    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.05, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses avec couleur
ggplot(data = eigenvec_LD01_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5)
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
                   breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                   labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_fill_manual(values = custom_colors_label2,
                  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                             "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                             "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                             "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

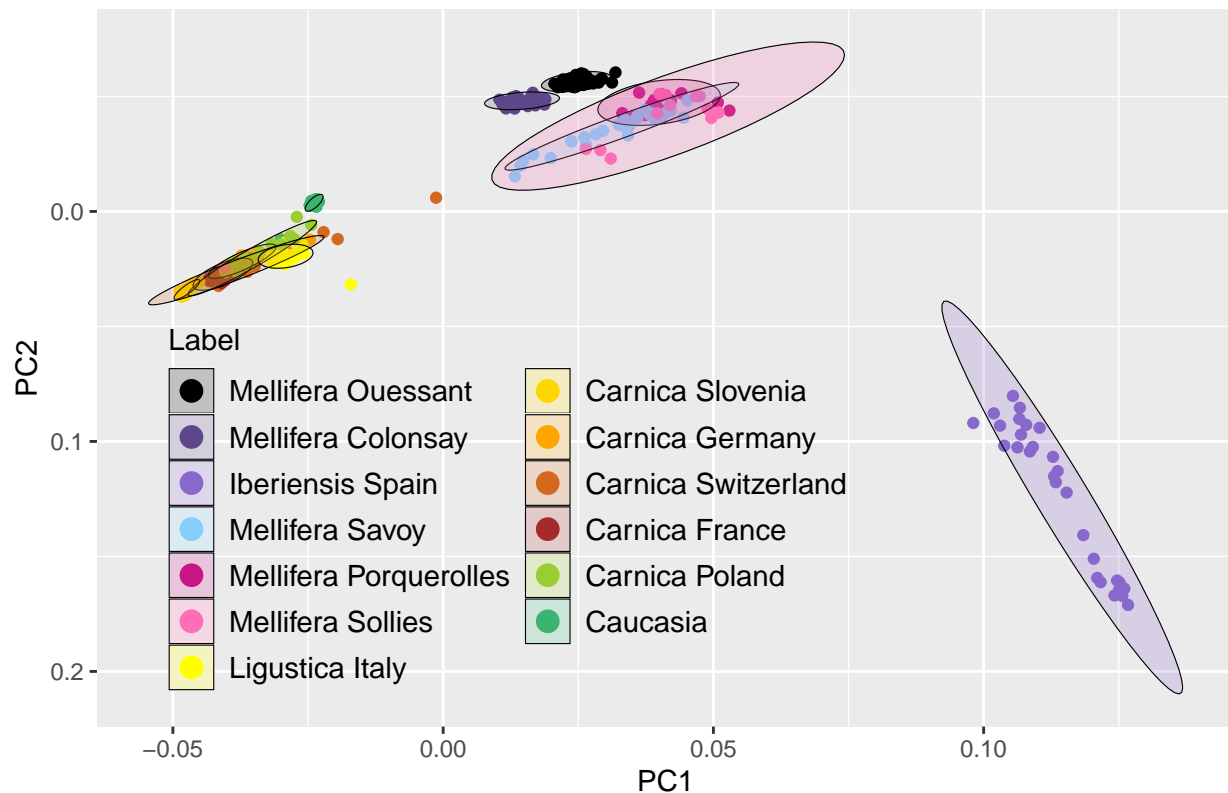
```

```

    "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.05, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#ACP - plot variance - PC3 - PC4
ggplot(data = eigenvec_LD01_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",

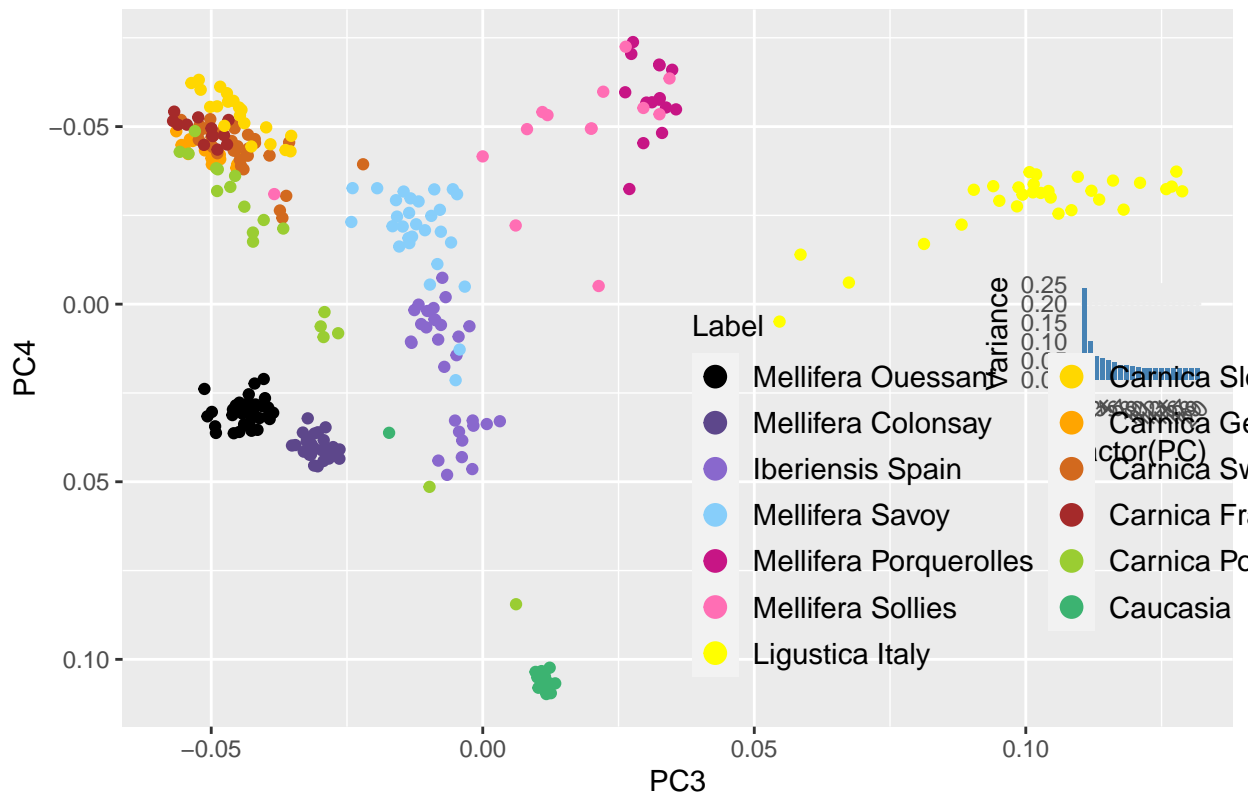
```

```

    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
"Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
"Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
"Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.5, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.09, xmax = 0.135, ymin = -0.05, ymax = 0.01)

```

PCA Plot – reference populations



PC3/PC4

ACP - LD pruning = 0.05

```

#LD005
setwd("~/Documents/Stage_NB/data/maf001_LD005")

eigenvec_LD005 <- read.table("SeqApiPop_629_maf001_LD005_acp.eigenvec", header = F)
eigenval_LD005 <- read.table("SeqApiPop_629_maf001_LD005_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_LD005)[colnames(eigenvec_LD005) == "V2"] <- "name"
eigenvec_LD005_seq_api_labels <- merge(eigenvec_LD005, seq_api_labels, by = "name")

eigen_percent_LD005 <- round((eigenval_LD005 / (sum(eigenval_LD005) )*100),2)
eigen_percent_LD005

```

PC1/PC2

```

##      V1
## 1  31.83
## 2   7.54
## 3   6.72
## 4   6.26
## 5   4.69
## 6   3.84
## 7   3.29
## 8   3.15
## 9   3.06
## 10  2.96
## 11  2.84
## 12  2.79
## 13  2.74
## 14  2.70
## 15  2.66
## 16  2.63
## 17  2.62
## 18  2.58
## 19  2.56
## 20  2.55

```

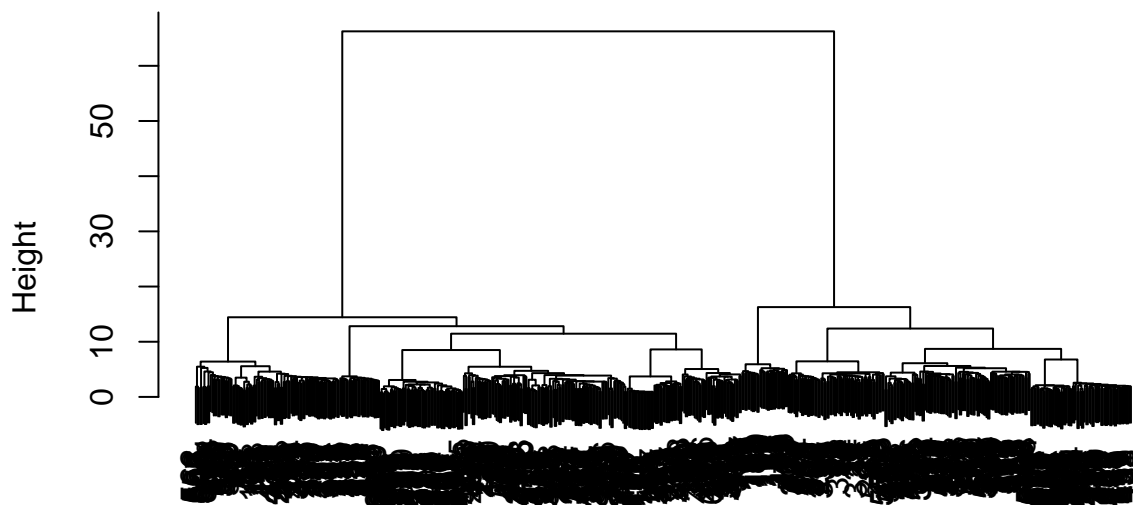
```

#Clustering hiérarchique
#Tree
matrice_app_refpop <- read.table("SeqApiPop_629_maf001_LD005_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)

```

Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

```
#heatmap
#heatmap(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')

#ACP
#filter 629 -> 301 RefPop
eigenvec_LD005_seq_api_labels <- eigenvec_LD005_seq_api_labels[eigenvec_LD005_seq_api_labels$GeneticOrigin == "RefPop",]
eigenvec_LD005_seq_api_labels$Label != 'RefPop'
eigenvec_LD005_seq_api_labels$Label != 'RefPop'
eigenvec_LD005_seq_api_labels$UniqueInHi
eigenvec_LD005_seq_api_labels$UniqueInHi
eigenvec_LD005_seq_api_labels$GeneticOrigin

custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred",

lambda <- eigenval_LD005$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

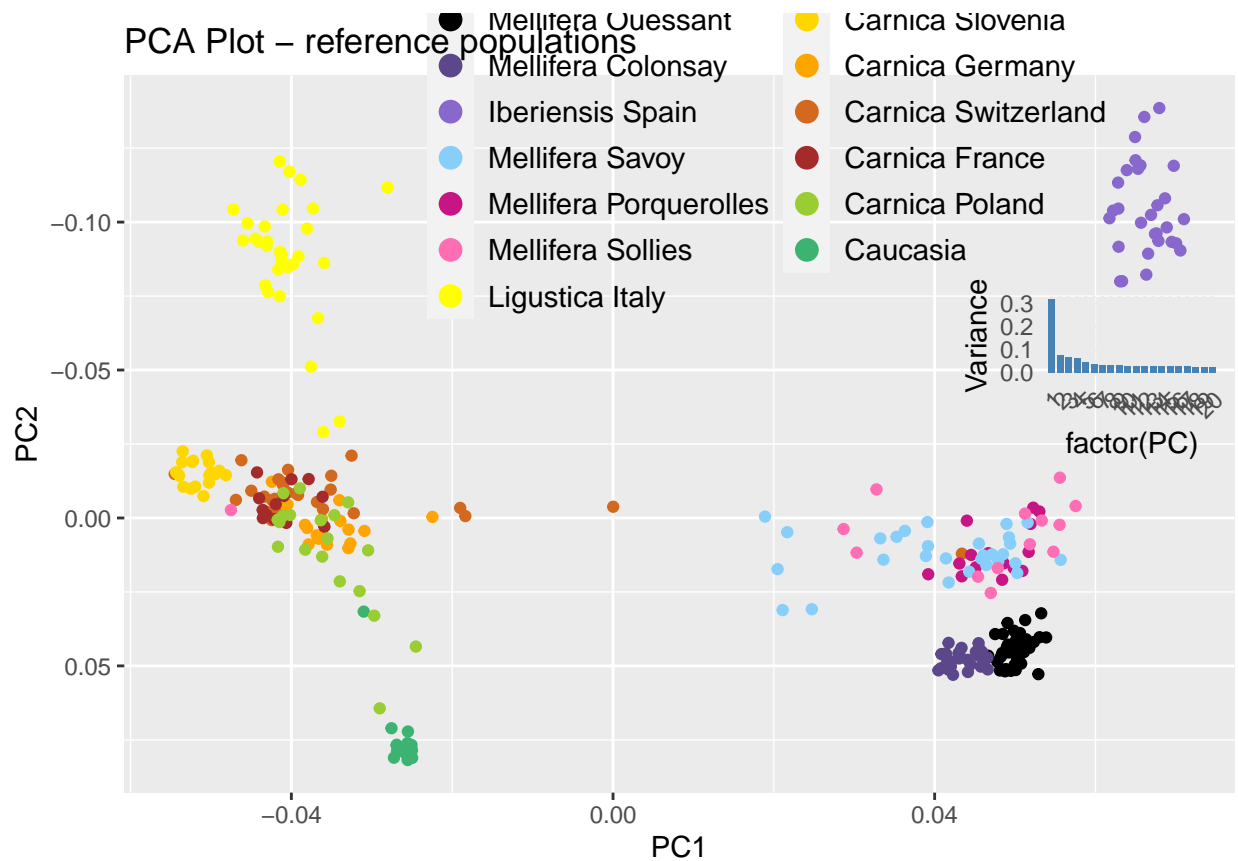
ggplot(data = eigenvec_LD005_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Quessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solies Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
```

```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.26, 0.64), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.077, xmax = 0.042, ymin = 0.015, ymax = 0.08)

```



```

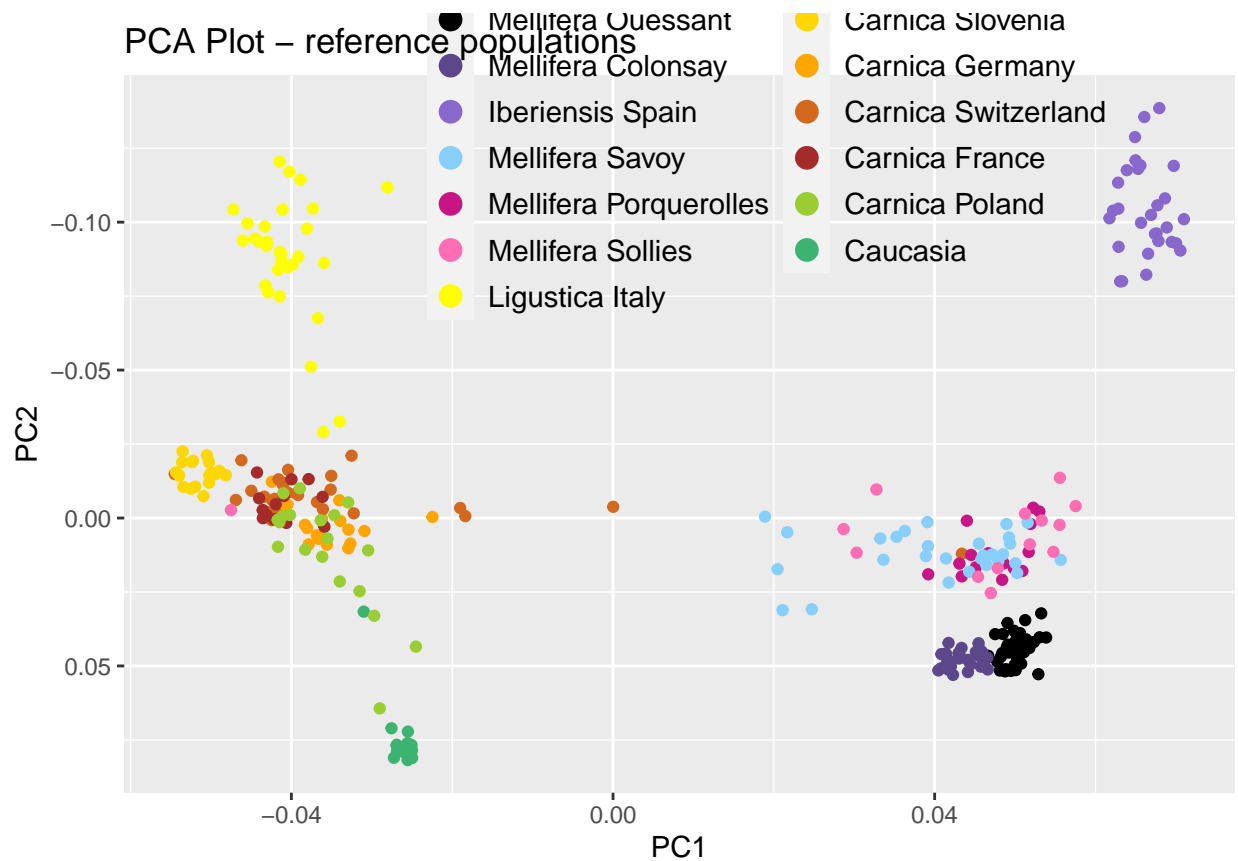
ggplot(data = eigenvec_LD005_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
geom_point() +
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
                   breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",

```

```

      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
      labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                  "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                  "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                  "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.26, 0.64), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```



```

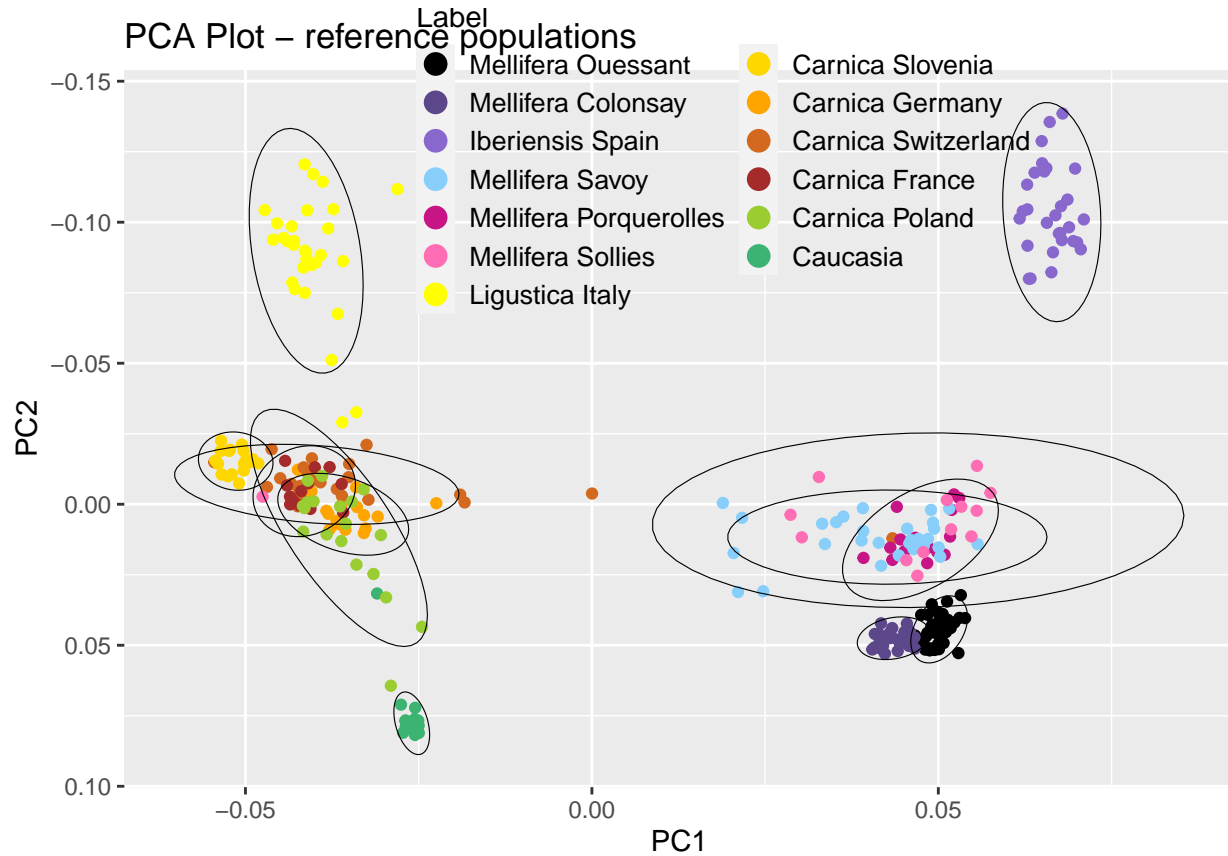
#ellipses autour des points selon Label
ggplot(data = eigenvec_LD005_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

```

```

    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.25, 0.64), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1, "lines"),
      legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```



```

# ellipses avec couleur
ggplot(data = eigenvec_LD005_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia")) +
  scale_fill_manual(values = custom_colors_label2,
                   breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

```

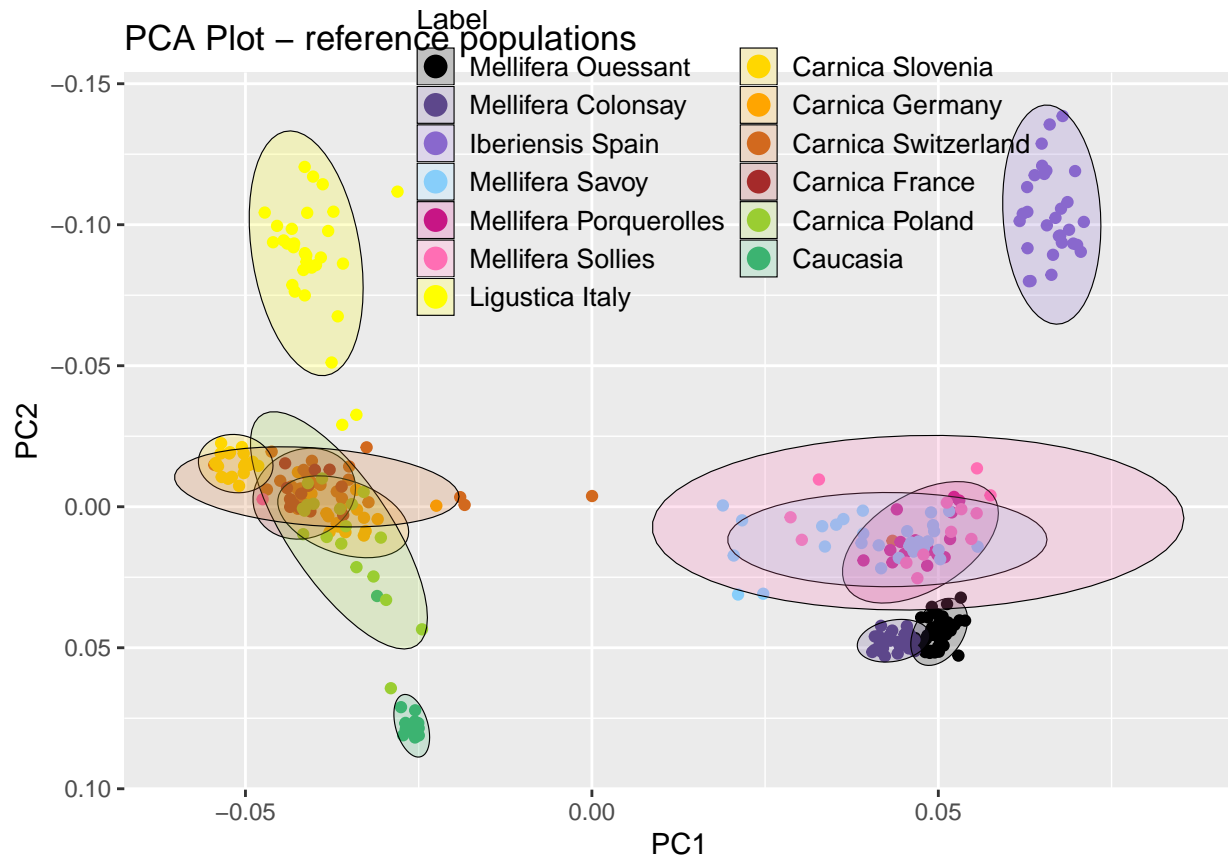


```

    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.25, 0.64), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1, "lines"),
      legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```



```

#ACP - plot variance - PC3 - PC4
ggplot(data = eigenvec_LD005_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                    "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))

```

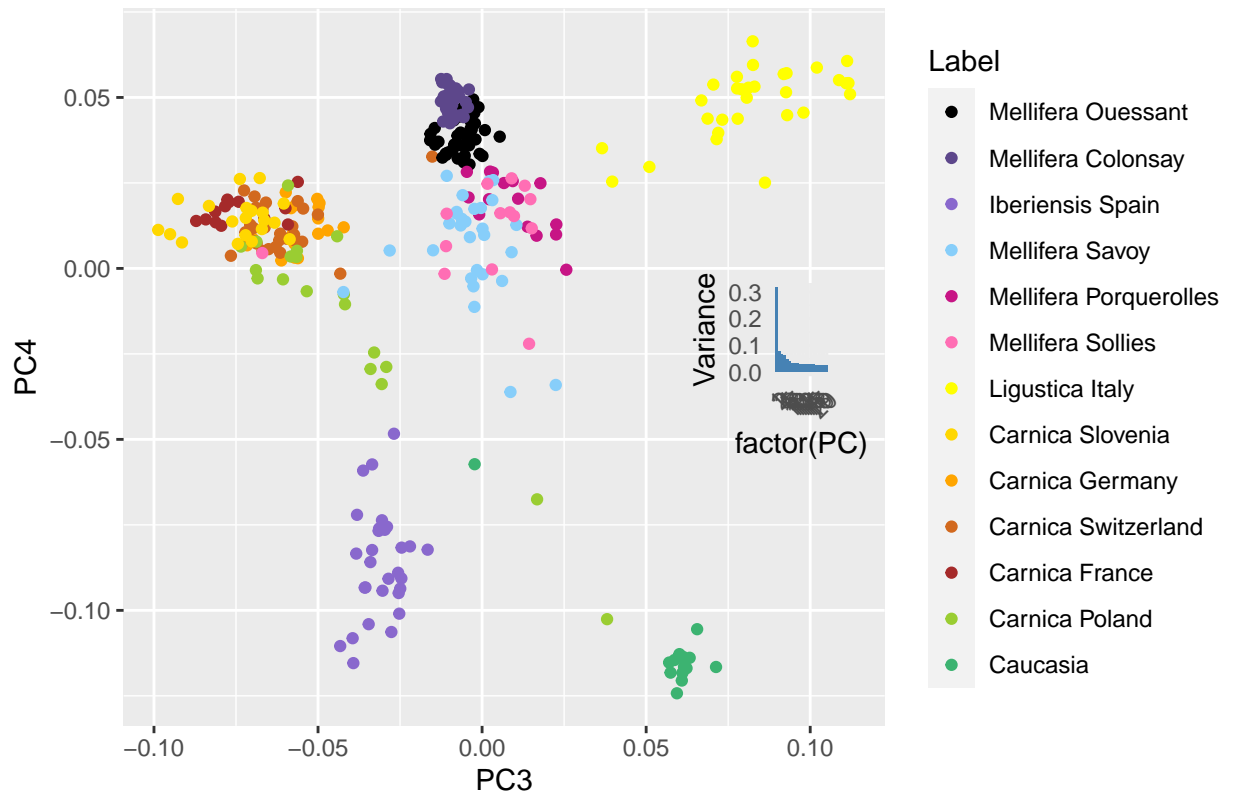
```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
      geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
      theme_minimal() +
      theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.06, xmax = 0.11, ymin = -0.06, ymax = 0)

```

PCA Plot – reference populations



PC3/PC4

ACP - LD pruning = 0.04

```

#LD004
setwd("~/Documents/Stage_NB/data/maf001_LD004")

eigenvec_LD004 <- read.table("SeqApiPop_629_maf001_LD004_acp.eigenvec", header = F)
eigenval_LD004 <- read.table("SeqApiPop_629_maf001_LD004_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

```

```

colnames(eigenvec_LD004)[colnames(eigenvec_LD004) == "V2"] <- "name"
eigenvec_LD004_seq_api_labels <- merge(eigenvec_LD004, seq_api_labels, by = "name")

eigen_percent_LD004 <- round((eigenval_LD004 / (sum(eigenval_LD004) )*100),2)
eigen_percent_LD004

```

PC1/PC2

```

##      V1
## 1  44.08
## 2   6.94
## 3   5.59
## 4   4.90
## 5   3.90
## 6   3.10
## 7   2.76
## 8   2.64
## 9   2.48
## 10  2.36
## 11  2.29
## 12  2.24
## 13  2.22
## 14  2.12
## 15  2.12
## 16  2.09
## 17  2.07
## 18  2.05
## 19  2.04
## 20  2.02

```

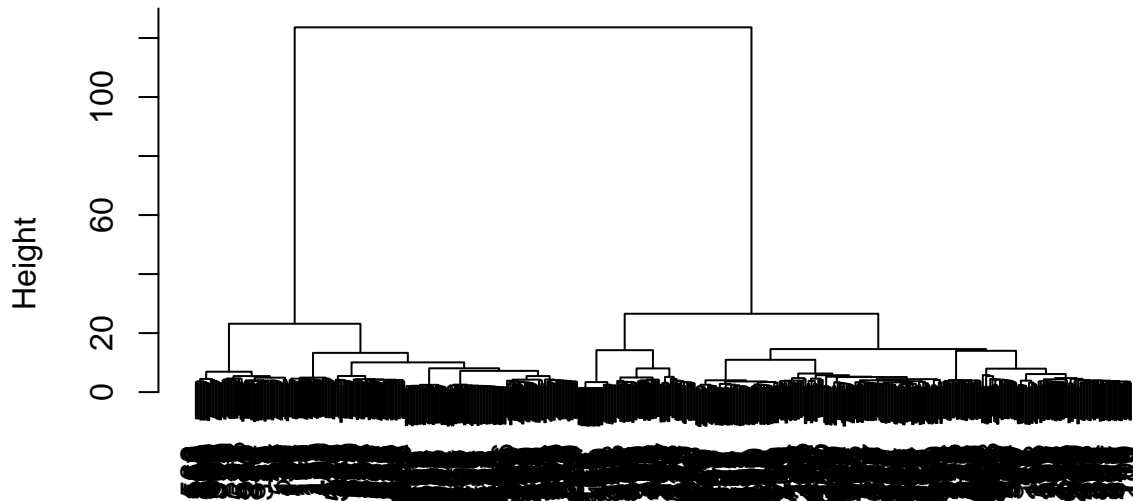
```

#Clustering hiérarchique
#Tree
matrice_app_refpop <- read.table("SeqApiPop_629_maf001_LD004_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)

```

Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

```
#heatmap
#heatmap(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')

#ACP
#filter 629 -> 301 RefPop
eigenvec_LD004_seq_api_labels <- eigenvec_LD004_seq_api_labels[eigenvec_LD004_seq_api_labels$GeneticOrigin == "RefPop", ]
eigenvec_LD004_seq_api_labels$Label != "RefPop"
eigenvec_LD004_seq_api_labels$Label != "RefPop"
eigenvec_LD004_seq_api_labels$UniqueIndividualID != "RefPop"
eigenvec_LD004_seq_api_labels$UniqueIndividualID != "RefPop"
eigenvec_LD004_seq_api_labels$GeneticOrigin != "RefPop"

custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred", "darkgreen")

lambda <- eigenval_LD004$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

ggplot(data = eigenvec_LD004_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
```

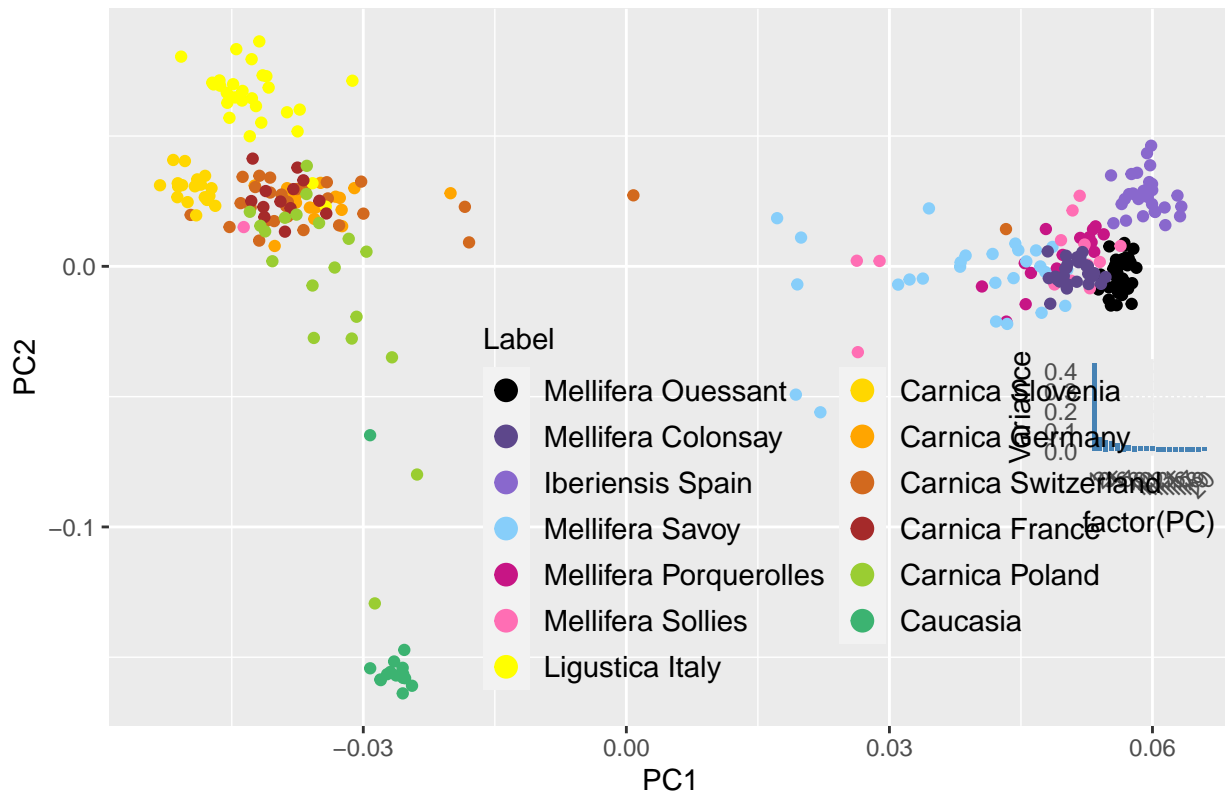
```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

theme(legend.position = c(0.32, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.068, xmax = 0.042, ymin = -0.03, ymax = -0.11)

```

PCA Plot – reference populations



```

ggplot(data = eigenvec_LD004_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))

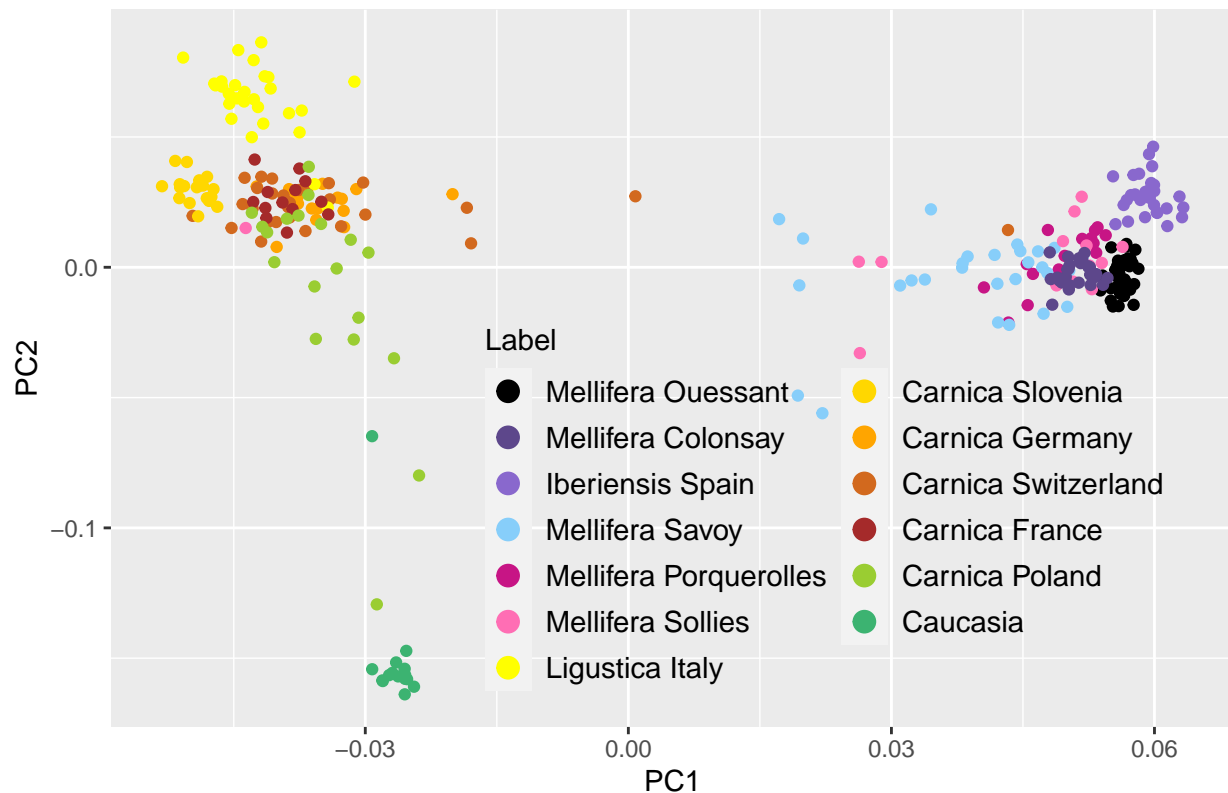
```

```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.32, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

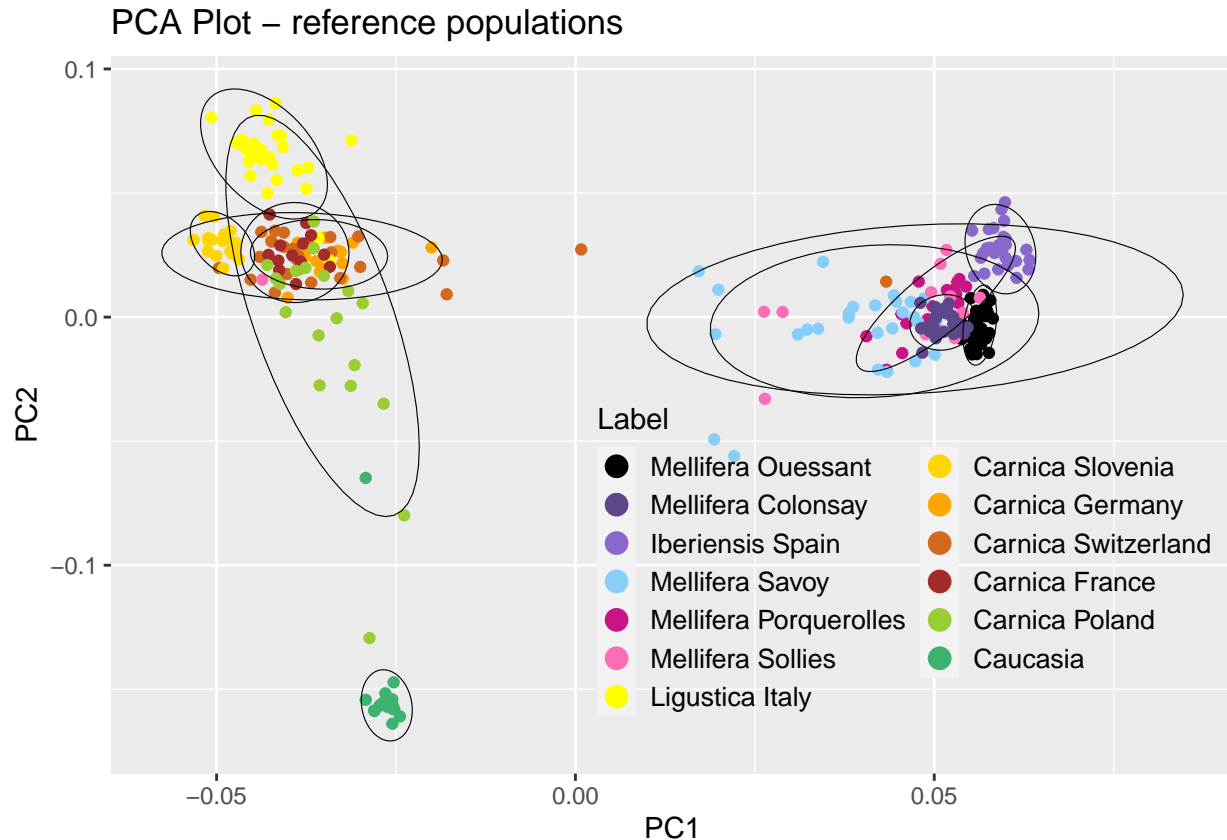
#ellipses autour des points selon Label
ggplot(data = eigenvec_LD004_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
                   breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                   labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.42, 0.06), legend.justification = c(0, 0),

```

```

legend.background = element_rect(fill = "transparent"),
legend.key.size = unit(1, "lines"),
legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```



```

# ellipses avec couleur
ggplot(data = eigenvec_LD004_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France")) +
  scale_fill_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",

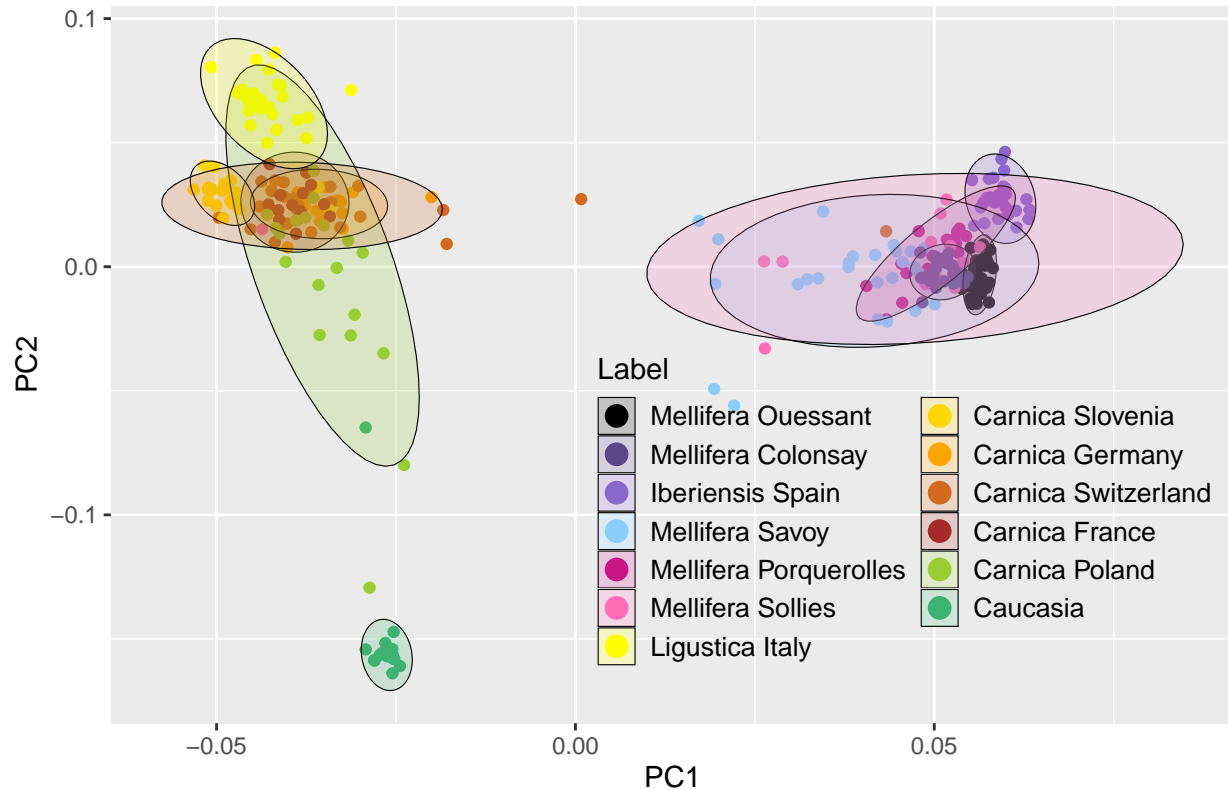
```

```

      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia")) +
theme(legend.position = c(0.42, 0.06), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1, "lines"),
      legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#ACP - plot variance - PC3 - PC4
ggplot(data = eigenvec_LD004_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
                     breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                     labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

```

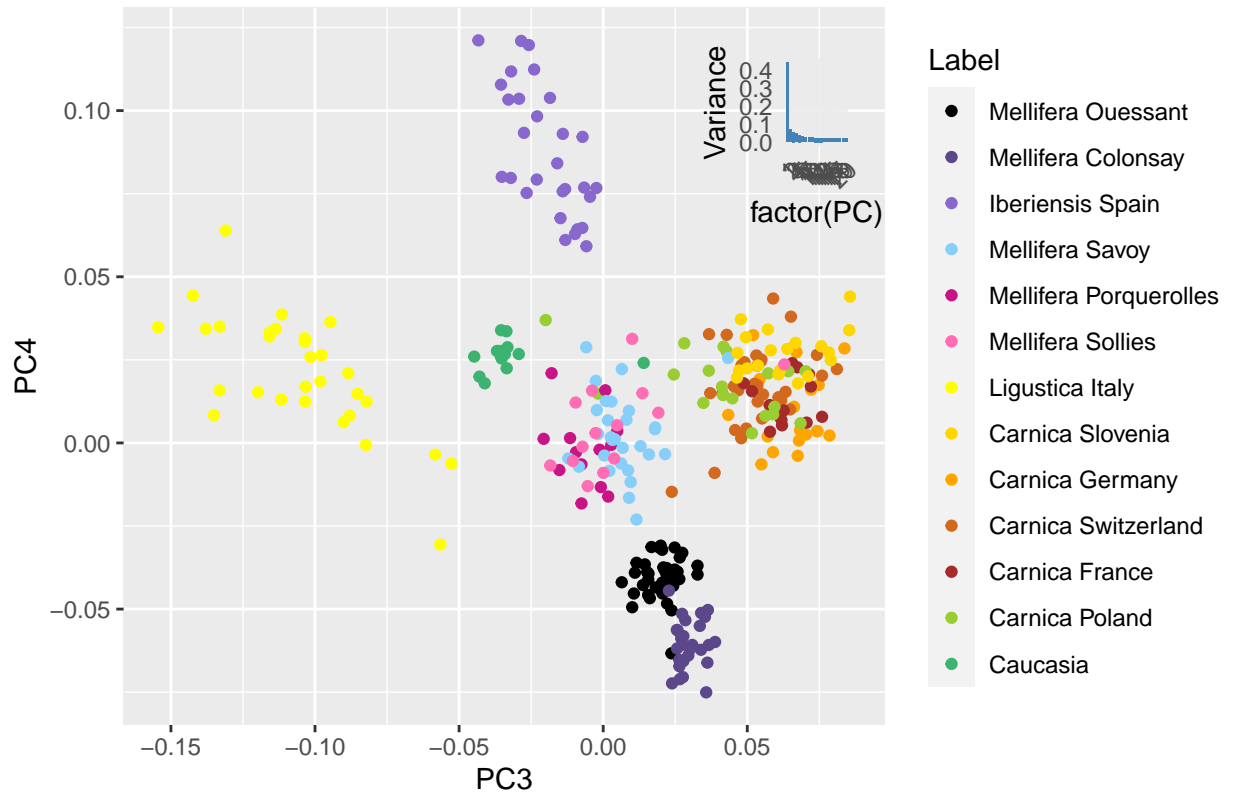


```

annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
      geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
      theme_minimal() +
      theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.03, xmax = 0.09, ymin = 0.06, ymax = 0.12)

```

PCA Plot – reference populations



PC3/PC4

ACP - LD pruning = 0.03

```

#LD003
setwd("~/Documents/Stage_NB/data/maf001_LD003")

eigenvec_LD003 <- read.table("SeqApiPop_629_maf001_LD003_acp.eigenvec", header = F)
eigenval_LD003 <- read.table("SeqApiPop_629_maf001_LD003_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_LD003)[colnames(eigenvec_LD003) == "V2"] <- "name"
eigenvec_LD003_seq_api_labels <- merge(eigenvec_LD003, seq_api_labels, by = "name")

```

```
eigen_percent_LD003 <- round((eigenval_LD003 / (sum(eigenval_LD003) )*100),2)
eigen_percent_LD003
```

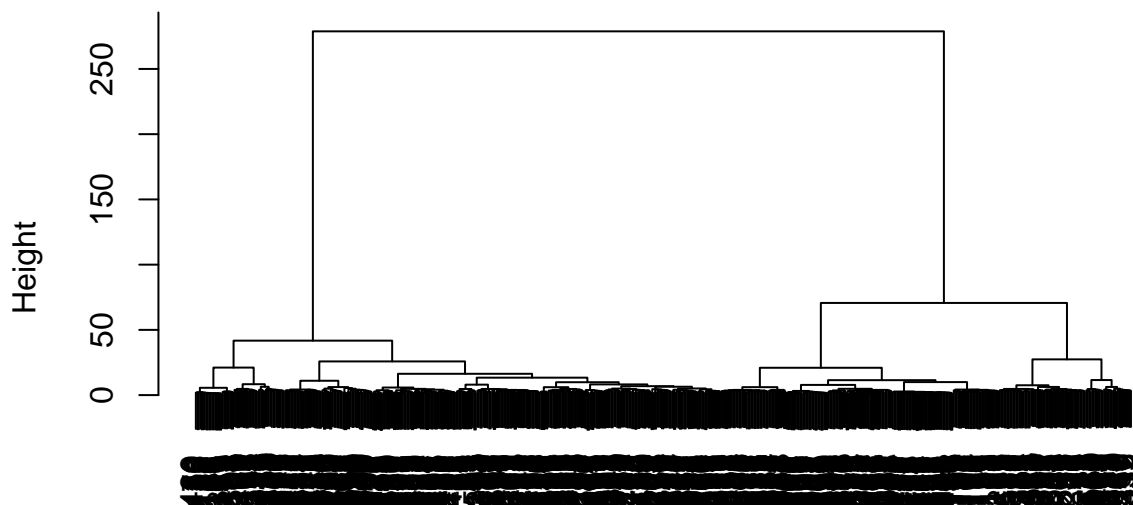
PC1/PC2

```
##      V1
## 1  60.11
## 2   6.49
## 3   3.80
## 4   3.17
## 5   2.42
## 6   2.10
## 7   1.99
## 8   1.81
## 9   1.70
## 10  1.64
## 11  1.58
## 12  1.57
## 13  1.53
## 14  1.51
## 15  1.47
## 16  1.45
## 17  1.43
## 18  1.42
## 19  1.41
## 20  1.40
```

```
#Clustering hiérarchique
#Tree
matrice_app_refpop <- read.table("SeqApiPop_629_maf001_LD003_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)
```

Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

[illegible]

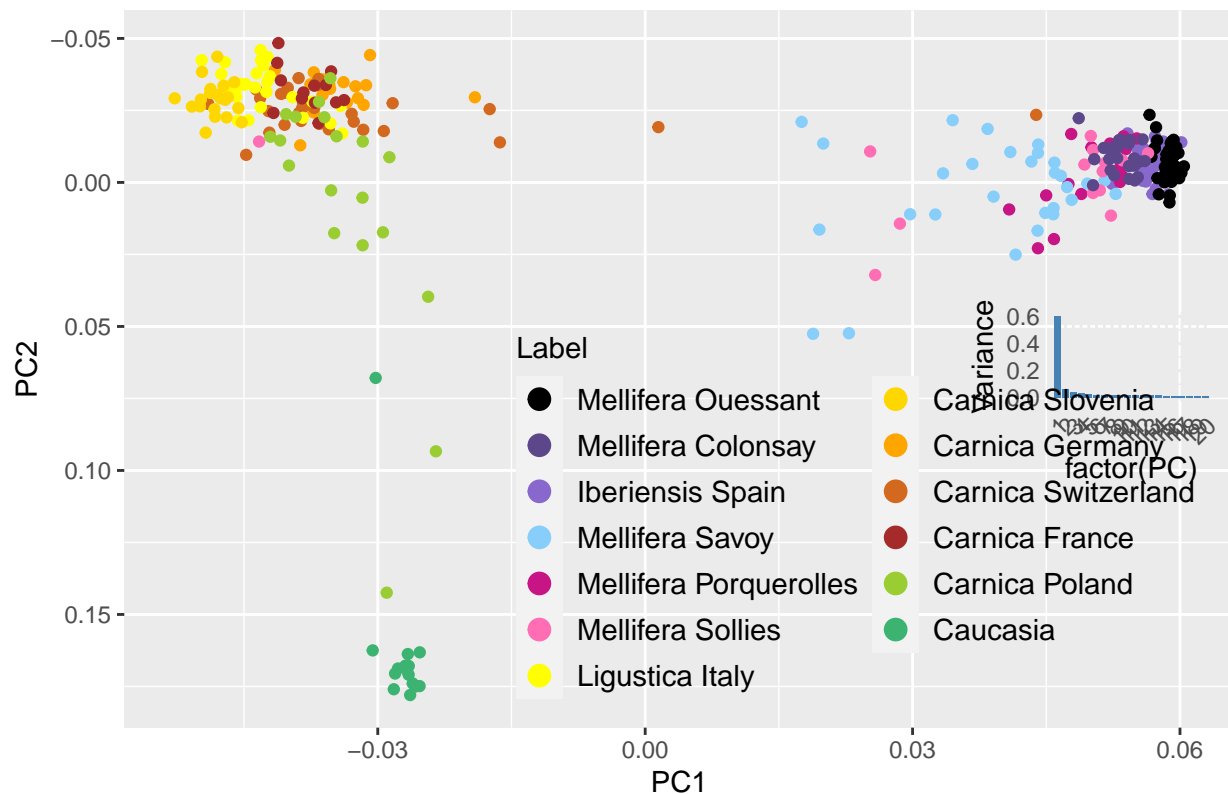
```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.035, xmax = 0.065, ymin = -0.04, ymax = -0.11)

```

PCA Plot – reference populations



```

ggplot(data = eigenvec_LD003_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",

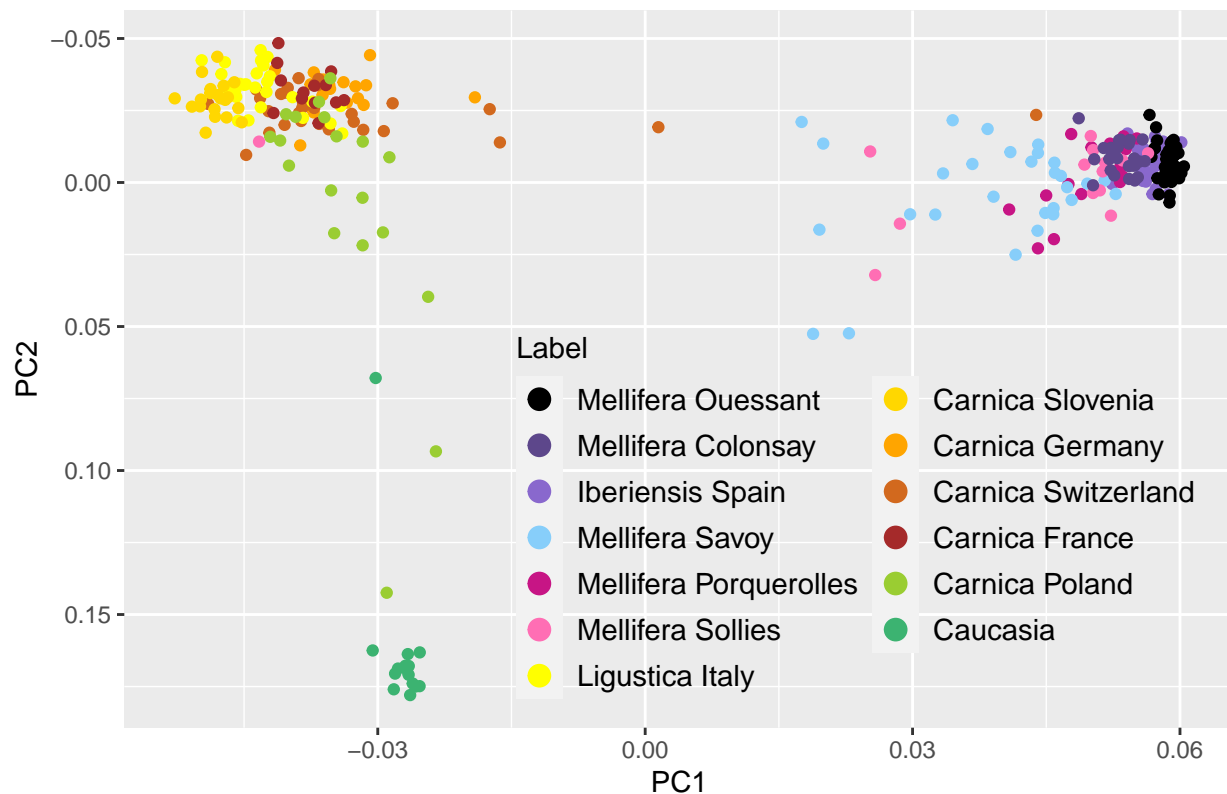
```

```

    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
"Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
"Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
"Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
legend.background = element_rect(fill = "transparent"),
legend.key.size = unit(1.2, "lines"),
legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#ellipses autour des points selon Label
ggplot(data = eigenvec_LD003_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
"Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
"Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
"Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
"Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
"Ligustica Italy", "Carnica Slovenia", "Carnica Germany",

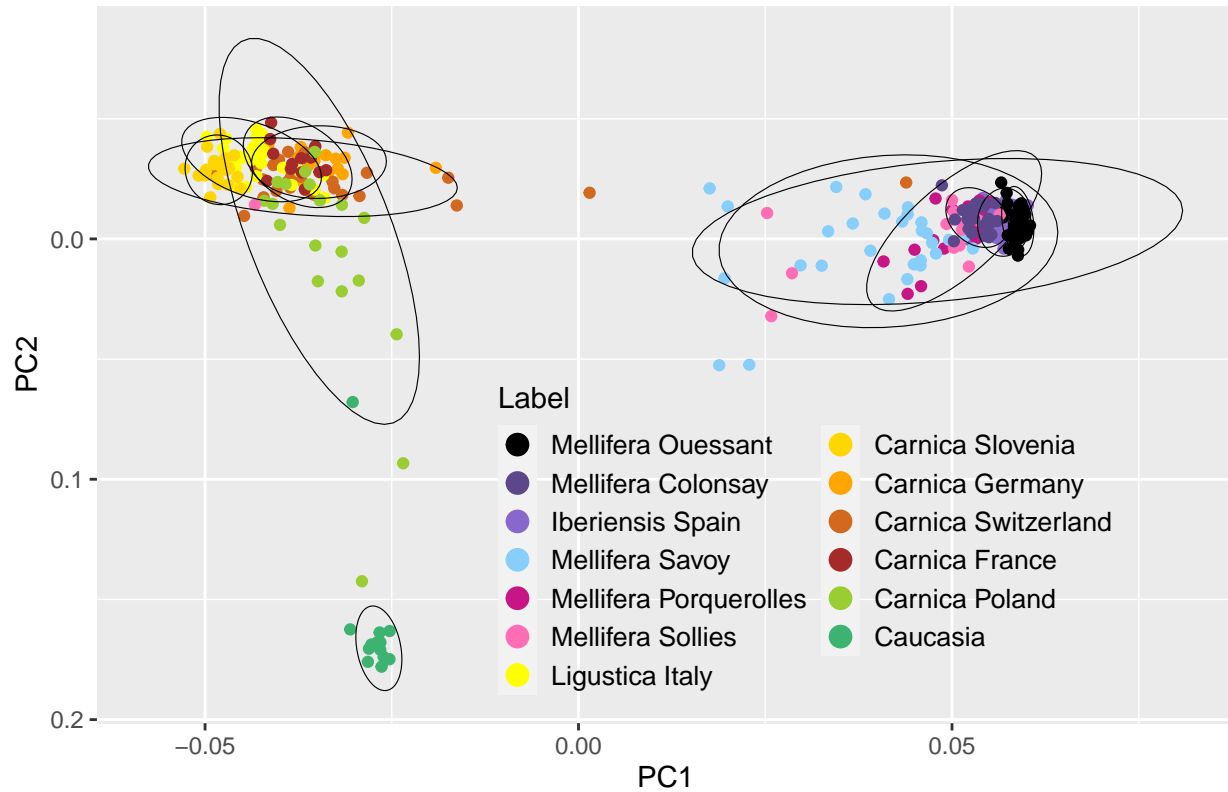
```

```

    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1, "lines"),
      legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses avec couleur
ggplot(data = eigenvec_LD003_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5)
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
             "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
             "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
             "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
  labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
             "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
             "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
             "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_fill_manual(values = custom_colors_label2,
  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
             "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
             "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
             "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

```

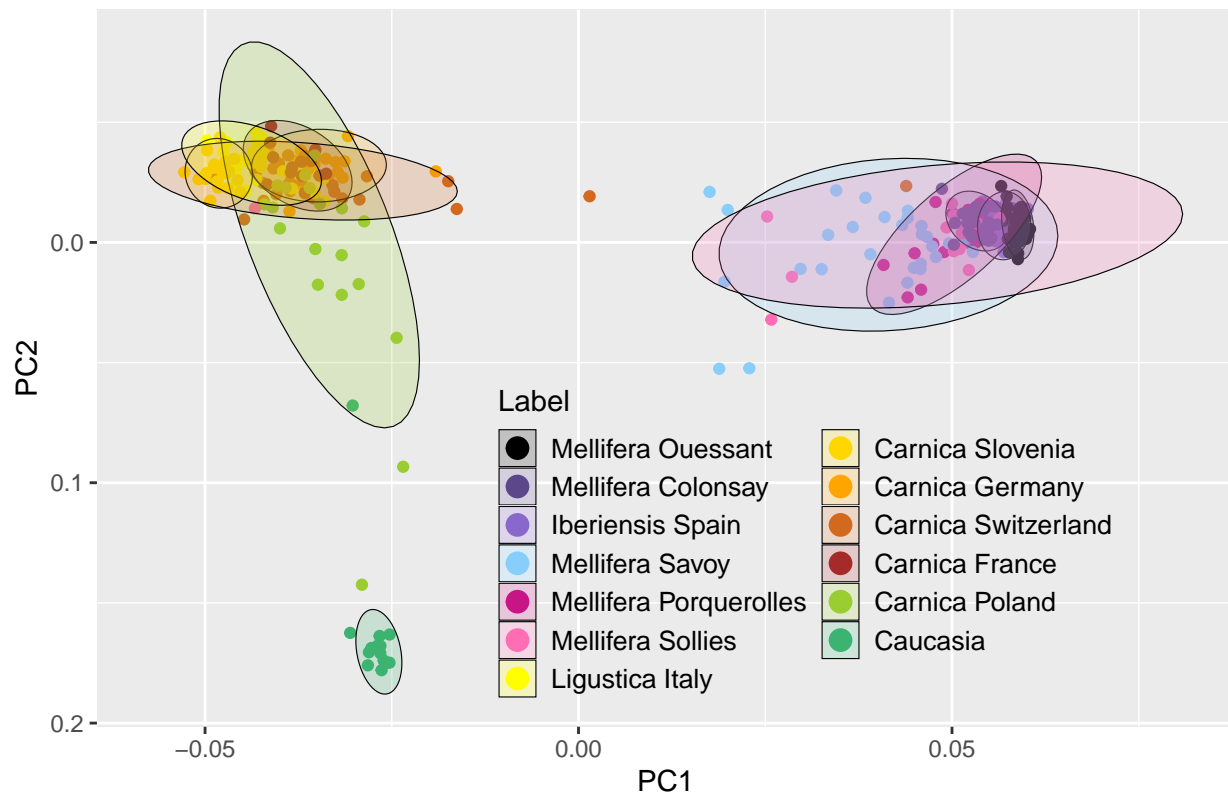
```

    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1, "lines"),
      legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#ACP - plot variance - PC3 - PC4
ggplot(data = eigenvec_LD003_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))

```

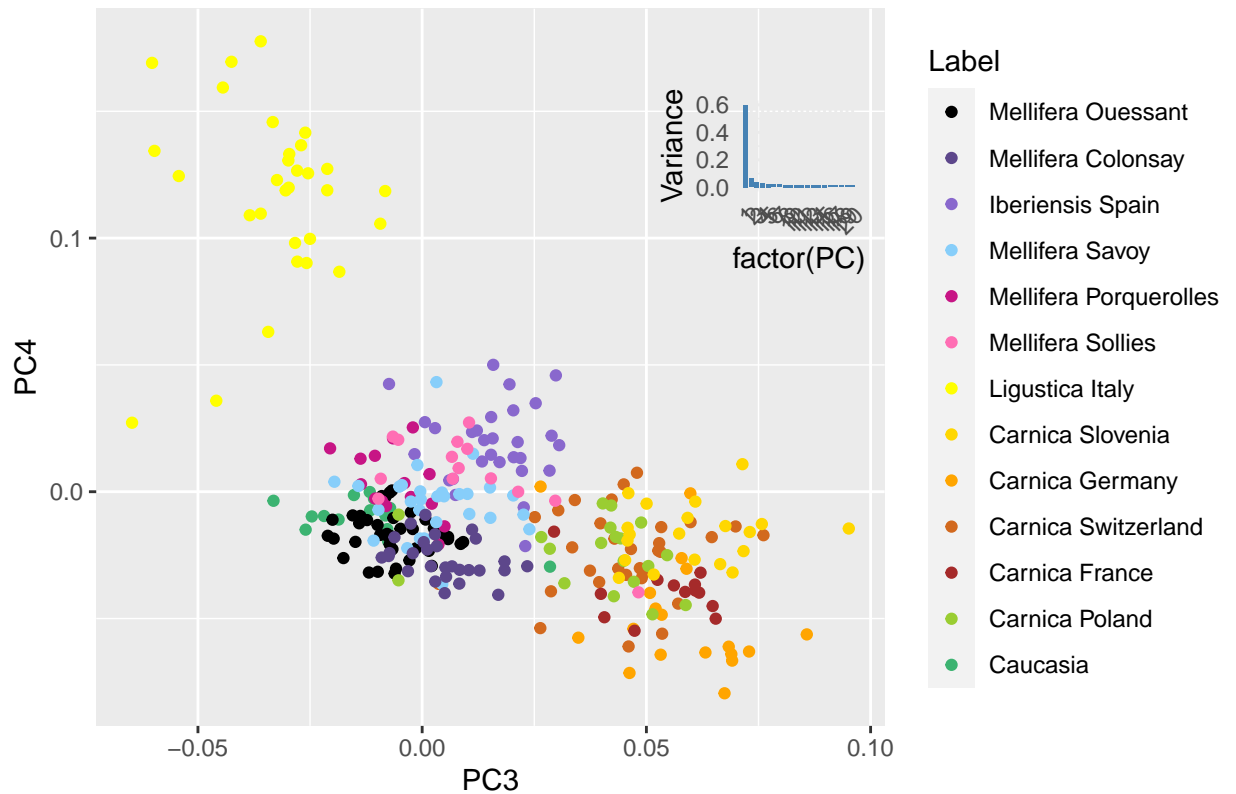
```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
      geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
      theme_minimal() +
      theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.05, xmax = 0.1, ymin = 0.08, ymax = 0.16)

```

PCA Plot – reference populations



PC3/PC4

ACP - LD pruning = 0.01

```

#LD001
setwd("~/Documents/Stage_NB/data/maf001_LD001")

eigenvec_LD001 <- read.table("SeqApiPop_629_maf001_LD001_acp.eigenvec", header = F)
eigenval_LD001 <- read.table("SeqApiPop_629_maf001_LD001_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

```



```
colnames(eigenvec_LD001)[colnames(eigenvec_LD001) == "V2"] <- "name"
eigenvec_LD001_seq_api_labels <- merge(eigenvec_LD001, seq_api_labels, by = "name")

eigen_percent_LD001 <- round((eigenval_LD001 / (sum(eigenval_LD001) )*100),2)
eigen_percent_LD001
```

PC1/PC2

```
##      V1
## 1  74.56
## 2   5.35
## 3   2.27
## 4   1.50
## 5   1.28
## 6   1.21
## 7   1.17
## 8   1.14
## 9   1.11
## 10  1.05
## 11  1.02
## 12  1.01
## 13  0.96
## 14  0.96
## 15  0.93
## 16  0.92
## 17  0.91
## 18  0.89
## 19  0.88
## 20  0.86
```

```
#Clustering hiérarchique
#Tree
matrice_app_refpop <- read.table("SeqApiPop_629_maf001_LD001_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)
```

Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

```
#heatmap
#heatplot(as.matrix(dist(matrice_app_refpop,dia=7)), cols.default = F, lowcol = 'blue', highcol='yellow')

#ACP
#filter 629 -> 301 RefPop
eigenvec_LD001_seq_api_labels <- eigenvec_LD001_seq_api_labels[eigenvec_LD001_seq_api_labels$GeneticOrigin == "Reference Populations",]
eigenvec_LD001_seq_api_labels$Label != NA #remove NAs
eigenvec_LD001_seq_api_labels$Label != "" #remove empty labels
eigenvec_LD001_seq_api_labels$UniqueID[is.na(eigenvec_LD001_seq_api_labels$UniqueID)] = eigenvec_LD001_seq_api_labels$GeneticOrigin

custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred", "darkgreen")

lambda <- eigenval_LD001$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

ggplot(data = eigenvec_LD001_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Quessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solliers Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
```

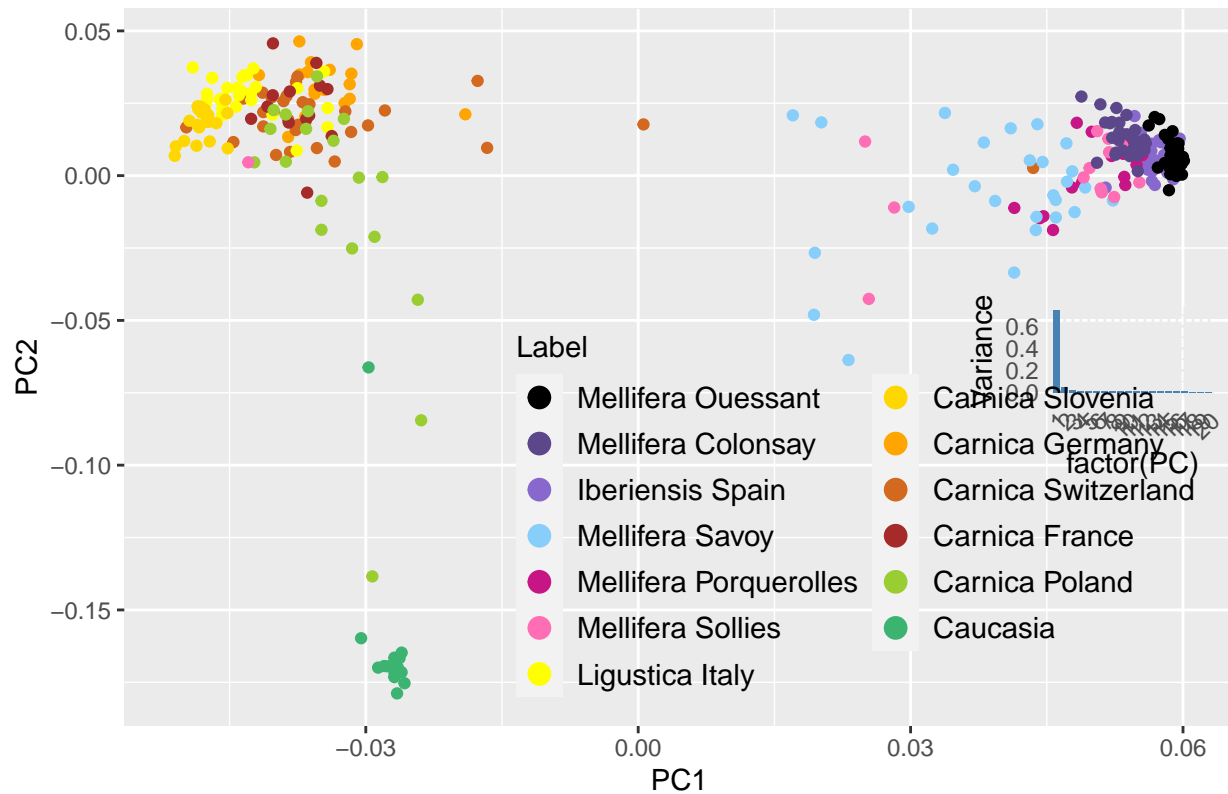
```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.035, xmax = 0.065, ymin = -0.04, ymax = -0.11)

```

PCA Plot – reference populations



```

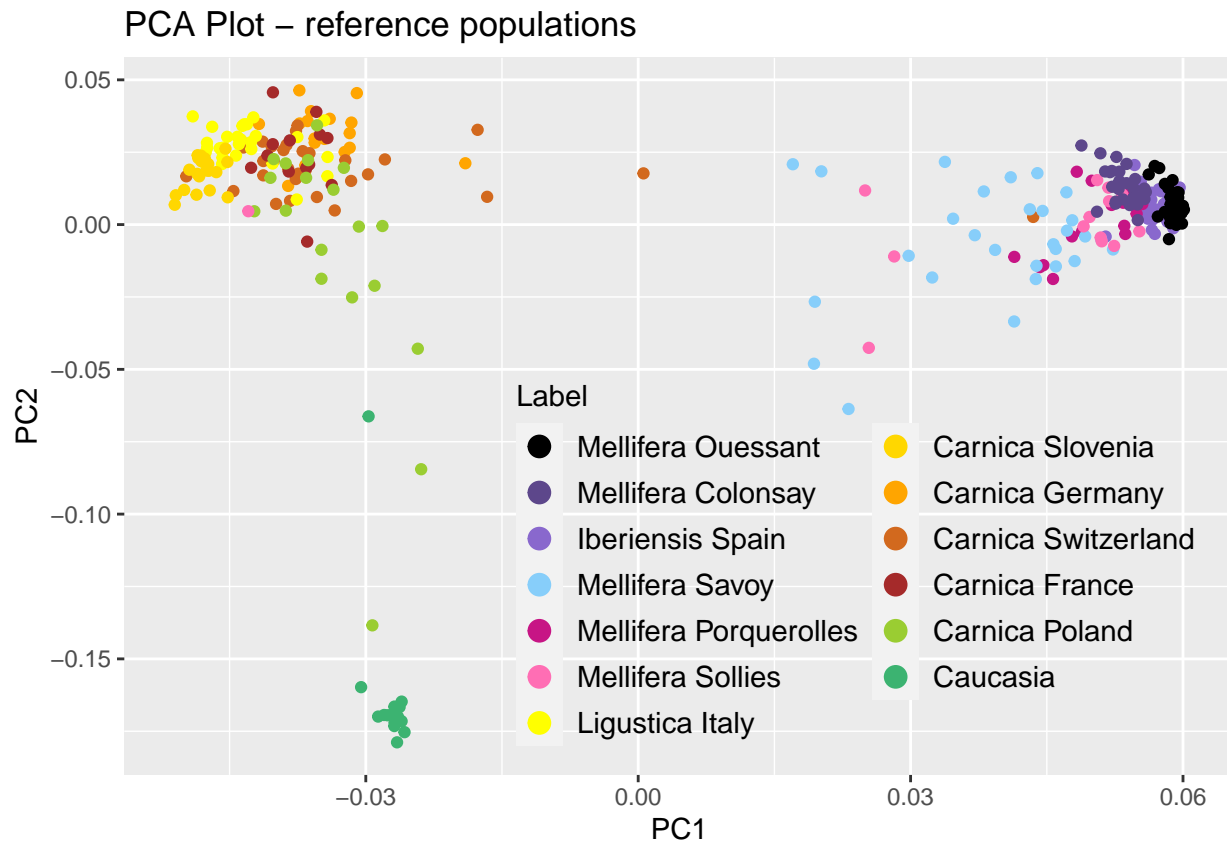
ggplot(data = eigenvec_LD001_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))

```

```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```



```

#ellipses autour des points selon Label
ggplot(data = eigenvec_LD001_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
                   breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                   labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),

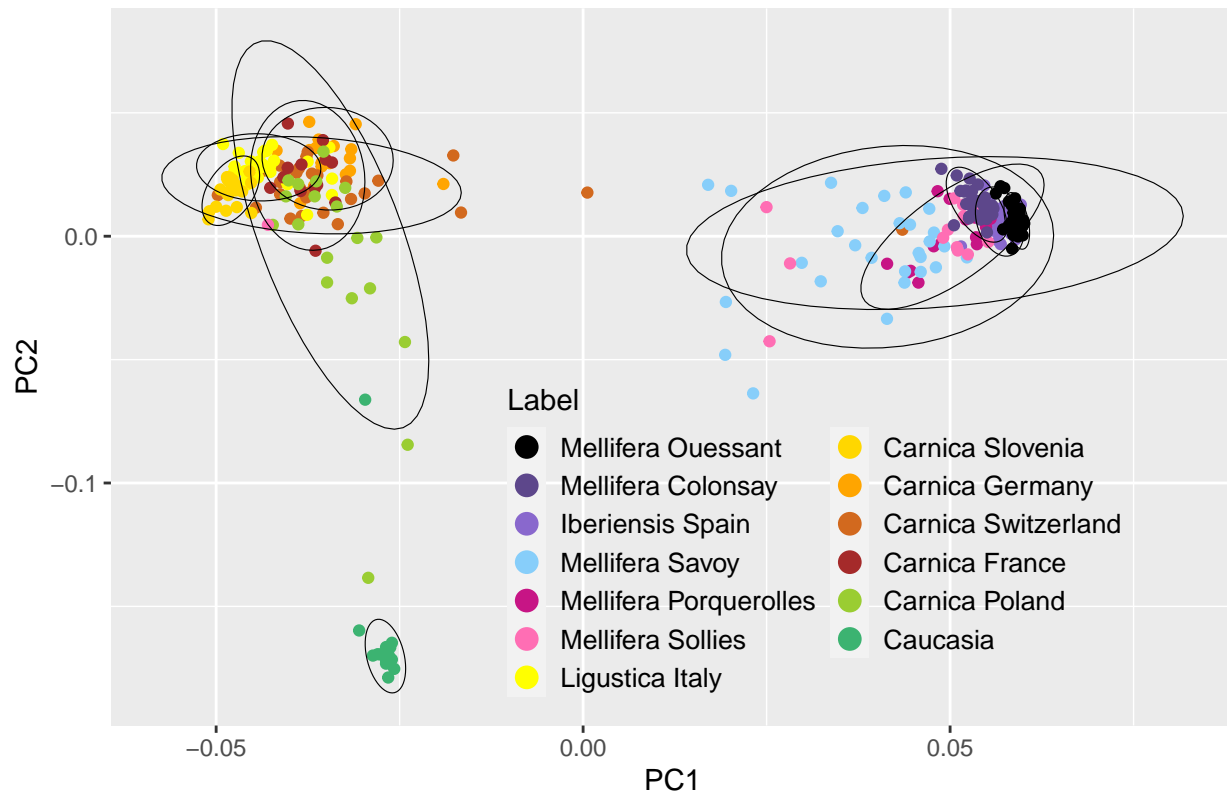
```

```

legend.background = element_rect(fill = "transparent"),
legend.key.size = unit(1, "lines"),
legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses avec couleur
ggplot(data = eigenvec_LD001_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensus Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensus Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France")) +
  scale_fill_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensus Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensus Spain",

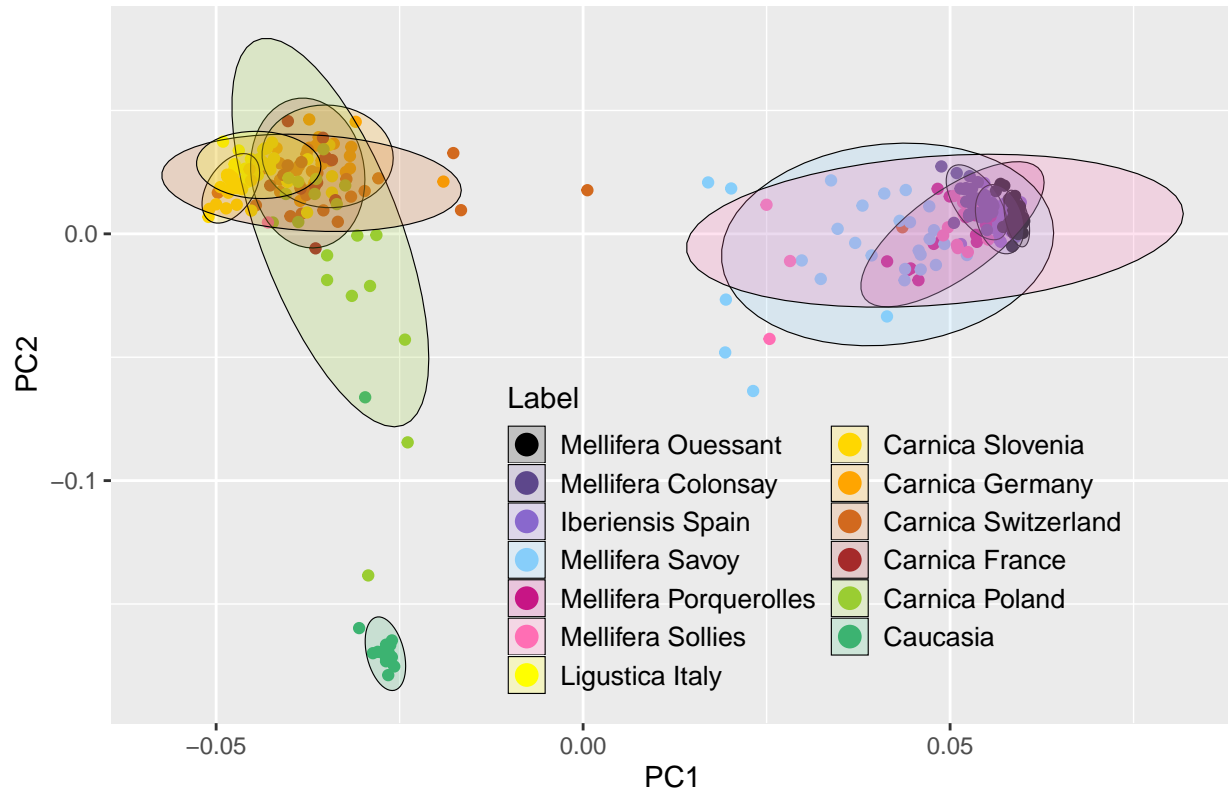
```

```

    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia")) +
  theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
        legend.background = element_rect(fill = "transparent"),
        legend.key.size = unit(1, "lines"),
        legend.text = element_text(size = 10)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

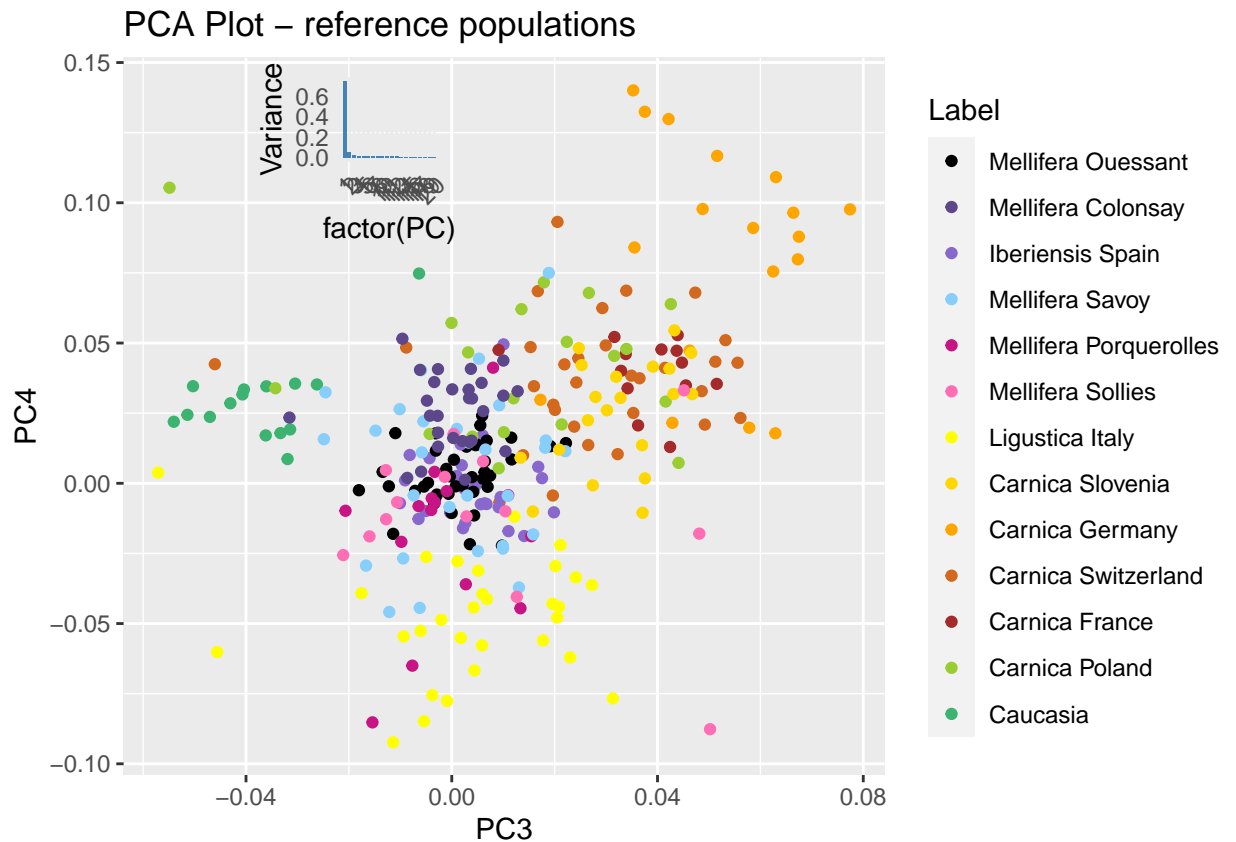
#ACP - plot variance - PC3 - PC4
ggplot(data = eigenvec_LD001_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
                     breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr"),
                     labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

```

```

annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = -0.04, xmax = 0, ymin = 0.08, ymax = 0.15)

```



ACP - SeqApiPop - 561 échantillons - MAF > 0.01 - LD pruning = 0.3 (fenêtre de 1749 SNPS et pas de 175 bp)

```

setwd("~/Documents/Stage_NB/data/SeqApiPop_561_maf001_LD03")

eigenvec_refpop <- read.table("SeqApiPop_561_maf001_LD03_acp.eigenvec", header = F)
eigenval_refpop <- read.table("SeqApiPop_561_maf001_LD03_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_refpop)[colnames(eigenvec_refpop) == "V2"] <- "name"
eigenvec_refpop_seq_api_labels <- merge(eigenvec_refpop, seq_api_labels, by = "name")

eigen_percent_refpop <- round((eigenval_refpop / (sum(eigenval_refpop) * 100), 2)

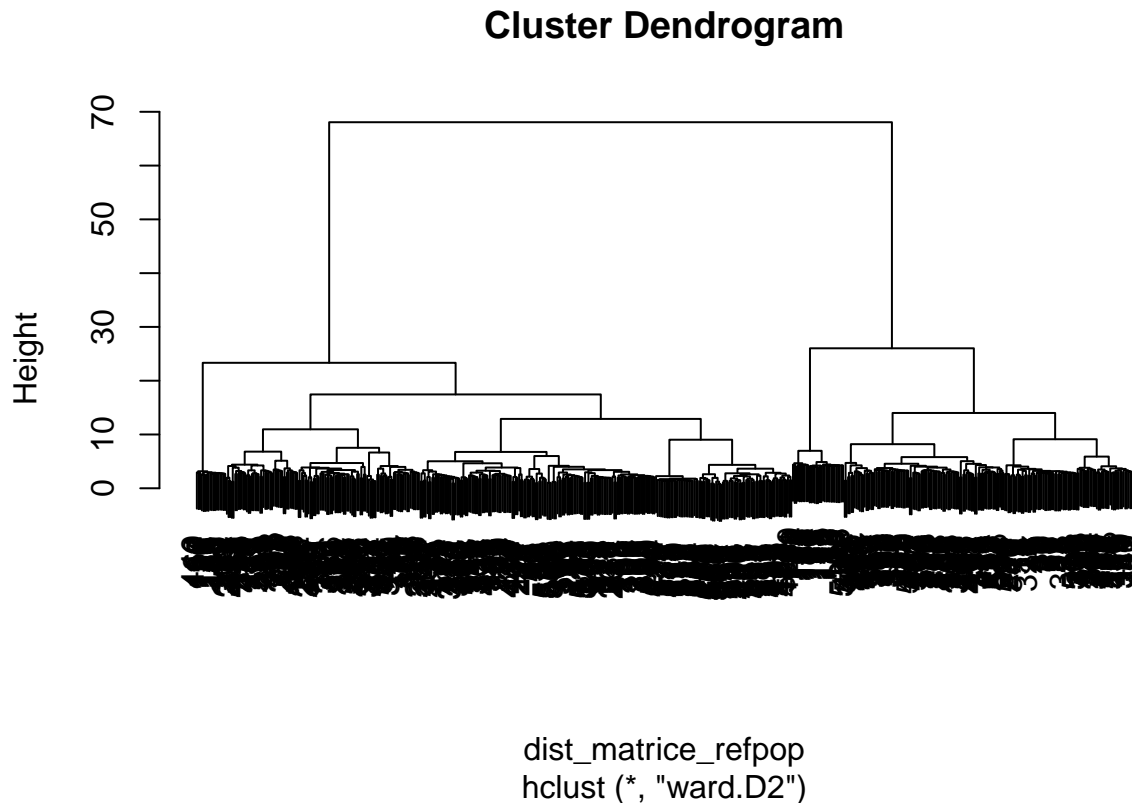
```

```

#Clustering hiérarchique
#Tree
setwd("~/Documents/Stage_NB/data/SeqApiPop_561_maf001_LD03")
matrice_app_refpop <- read.table("SeqApiPop_561_maf001_LD03_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)

```



```

#heatmap
#heatmap(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')

```

```

#ACP
#filter 629 -> 301 RefPop
eigenvec_refpop_seq_api_labels <- eigenvec_refpop_seq_api_labels[eigenvec_refpop_seq_api_labels$Genetic
eigenvec_refpop_seq_api_labels$Label
eigenvec_refpop_seq_api_labels$Label
eigenvec_refpop_seq_api_labels$Unique
eigenvec_refpop_seq_api_labels$Unique
eigenvec_refpop_seq_api_labels$Genet.

custom_colors_label2 <- c("mediumpurple3", "lightskyblue", "mediumvioletred", "hotpink1", "yellow", "go

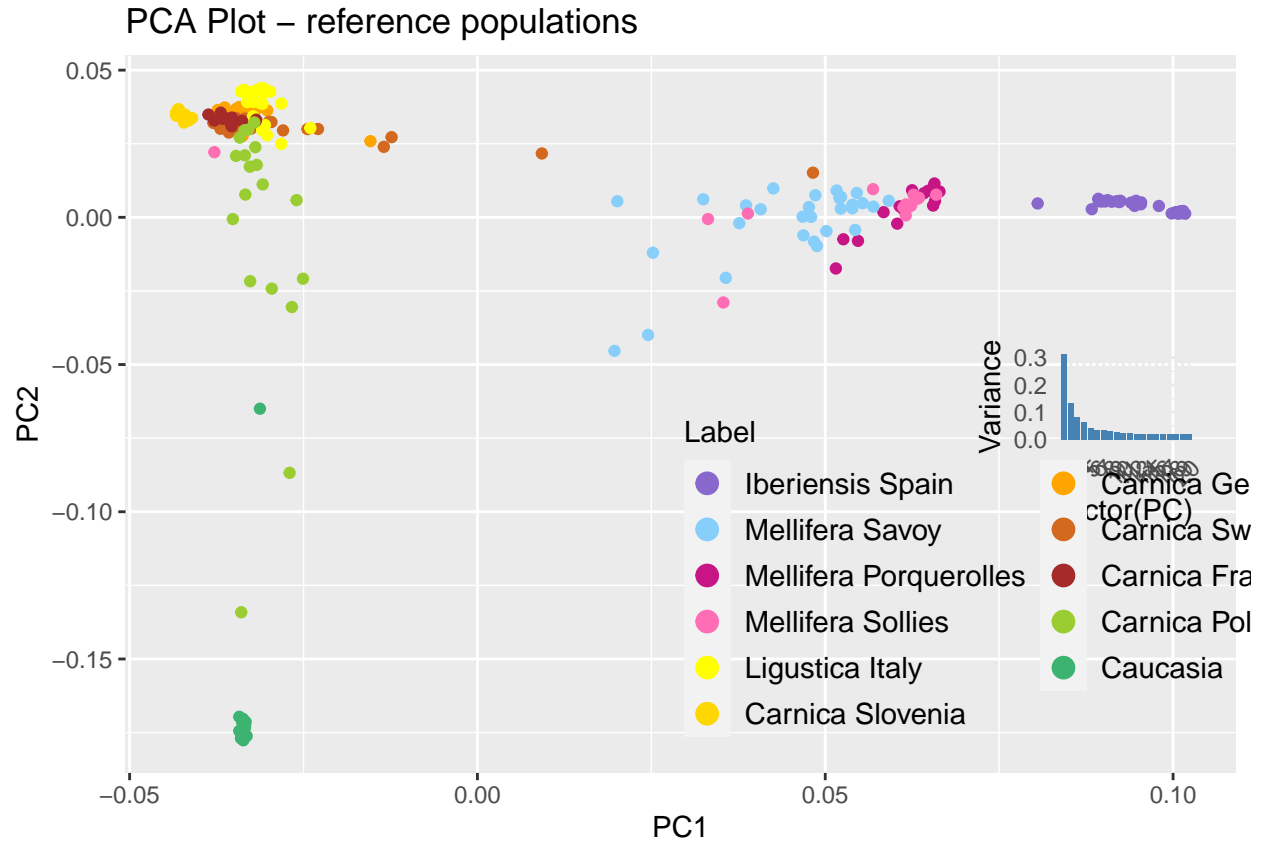
lambda <- eigenval_refpop$V1
variance_proportion <- lambda / sum(lambda)

```



```
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)
```

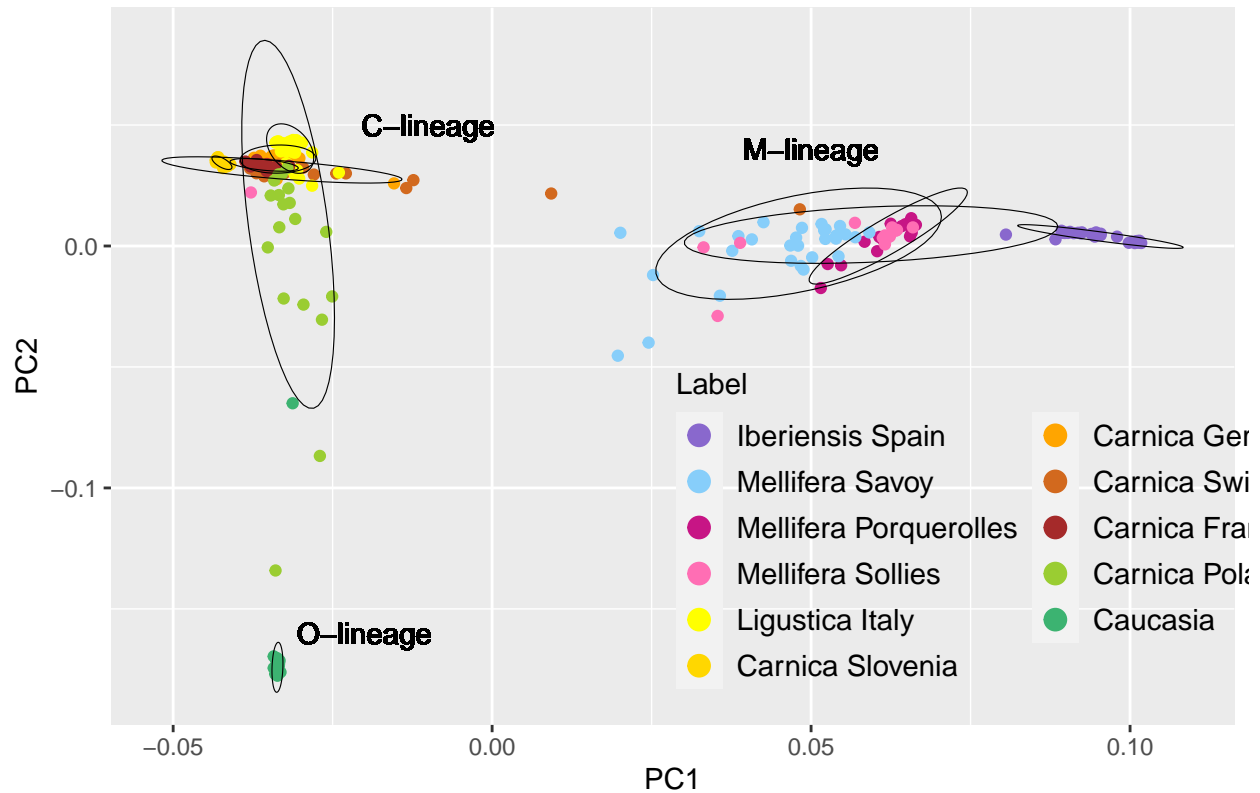
```
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.49, 0.03), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),
    legend.key.size = unit(1.2, "lines"),
    legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.105, xmax = 0.07, ymin = -0.11, ymax = -0.04)
```



PC1/PC2

```
#ellipses autour des points selon Label
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black") +
  geom_text(aes(x = 0.05, y = 0.04, label = "M-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.01, y = 0.05, label = "C-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.02, y = -0.16, label = "O-lineage"), size = 4, color = "black") +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c( "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c( "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.49, 0.03), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),
    legend.key.size = unit(1.2, "lines"),
    legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))
```

PCA Plot – reference populations



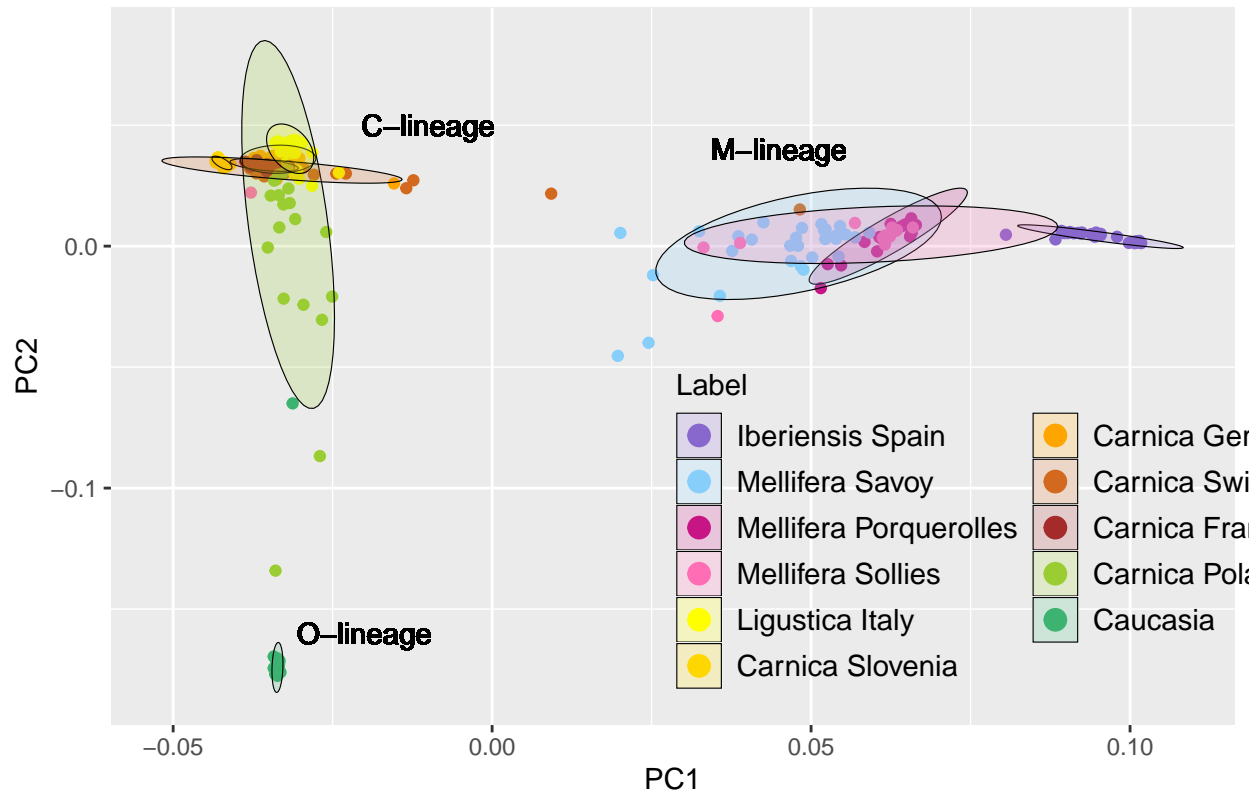
```
# ellipses avec couleur
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  geom_text(aes(x = 0.045, y = 0.04, label = "M-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.01, y = 0.05, label = "C-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.02, y = -0.16, label = "O-lineage"), size = 4, color = "black") +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c( "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
  scale_fill_manual(values = custom_colors_label2,
    breaks = c( "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c( "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
```

```

theme(legend.position = c(0.49, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#ACP - plot variance
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
                     breaks = c("Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                     labels = c("Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.05, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),

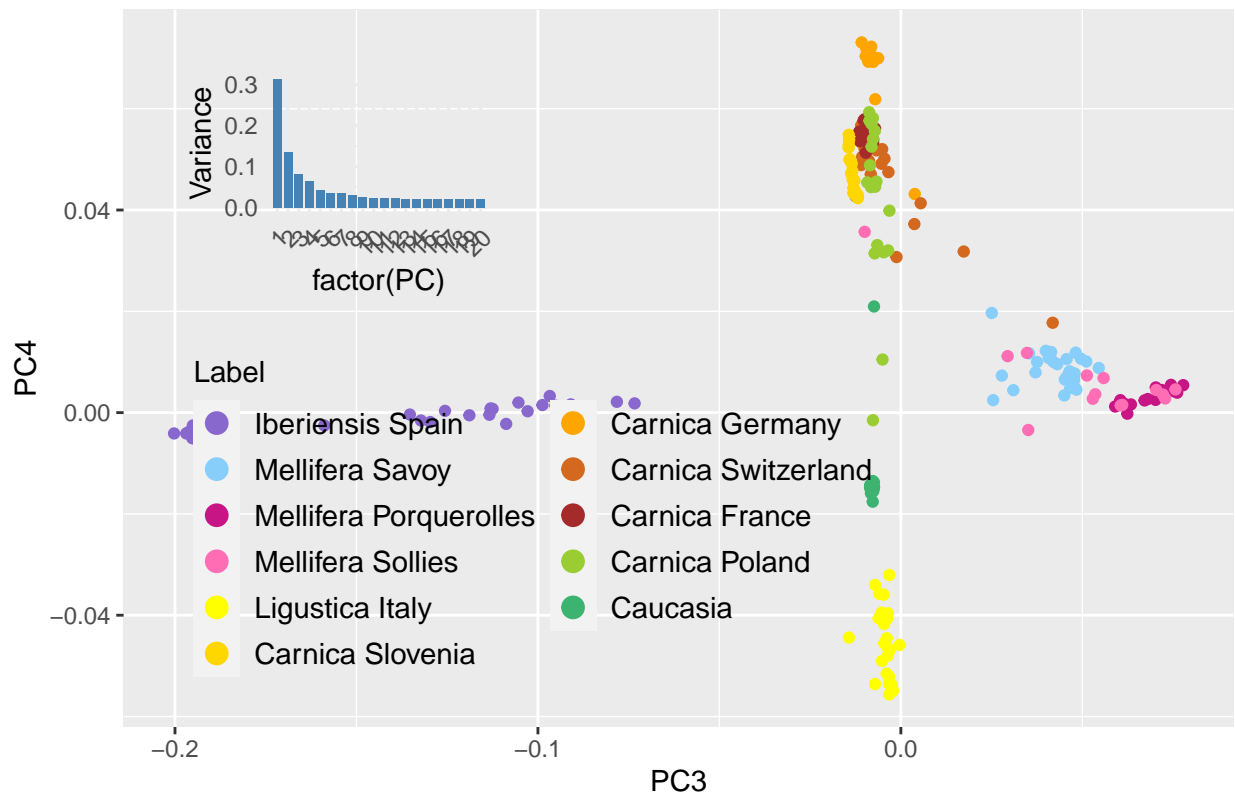
```

```

    legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
      geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
      theme_minimal() +
      theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = -0.11, xmax = -0.2, ymin = 0.07, ymax = 0.02)

```

PCA Plot – reference populations



PC3/PC4

```

#ellipses autour des points selon Label
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c( "Iberiensis Spain",
                "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c( "Iberiensis Spain",
                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.05, 0.05), legend.justification = c(0, 0),

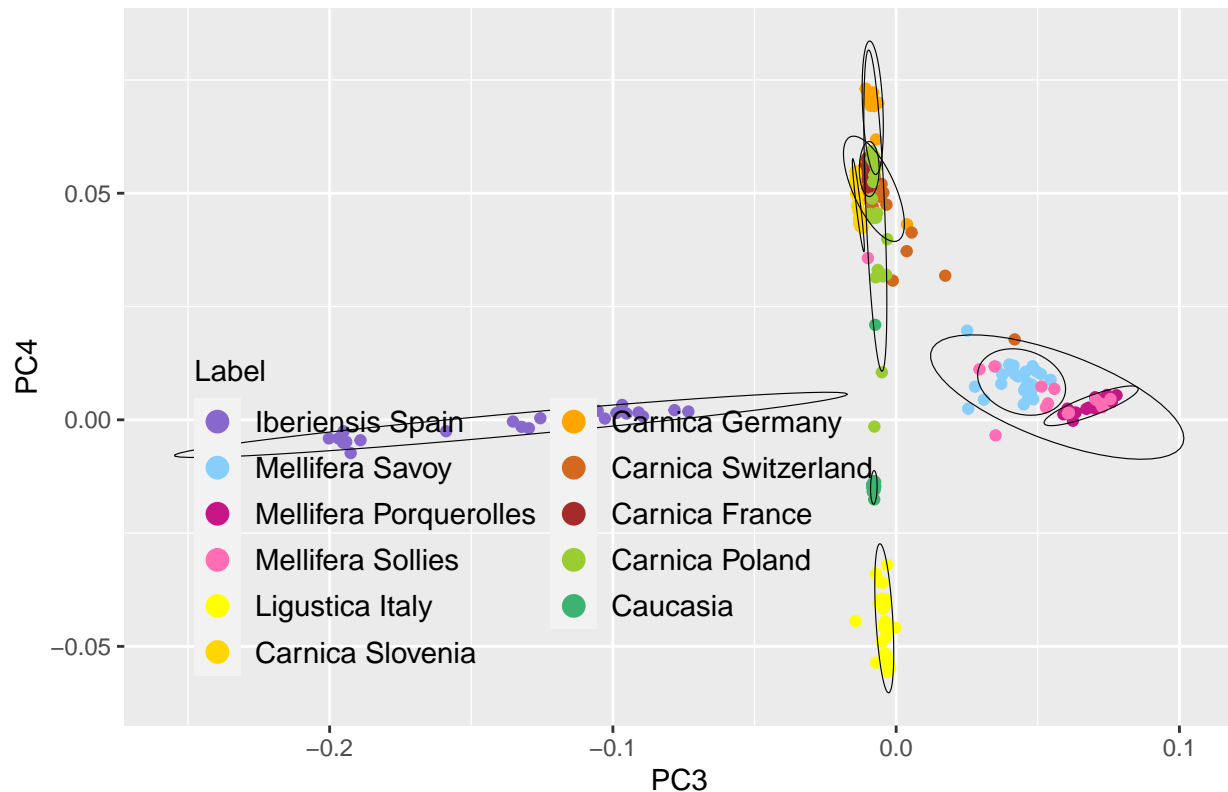
```

```

legend.background = element_rect(fill = "transparent"),
legend.key.size = unit(1.2, "lines"),
legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses avec couleur
ggplot(data = eigenvec_refpop_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  labs(title = "PCA Plot – reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c( "Iberiensis Spain",
                 "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                 "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                 "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Iberiensis Spain",
               "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
               "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
               "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France")) +
  scale_fill_manual(values = custom_colors_label2,
    breaks = c( "Iberiensis Spain",
                 "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                 "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                 "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c( "Iberiensis Spain",

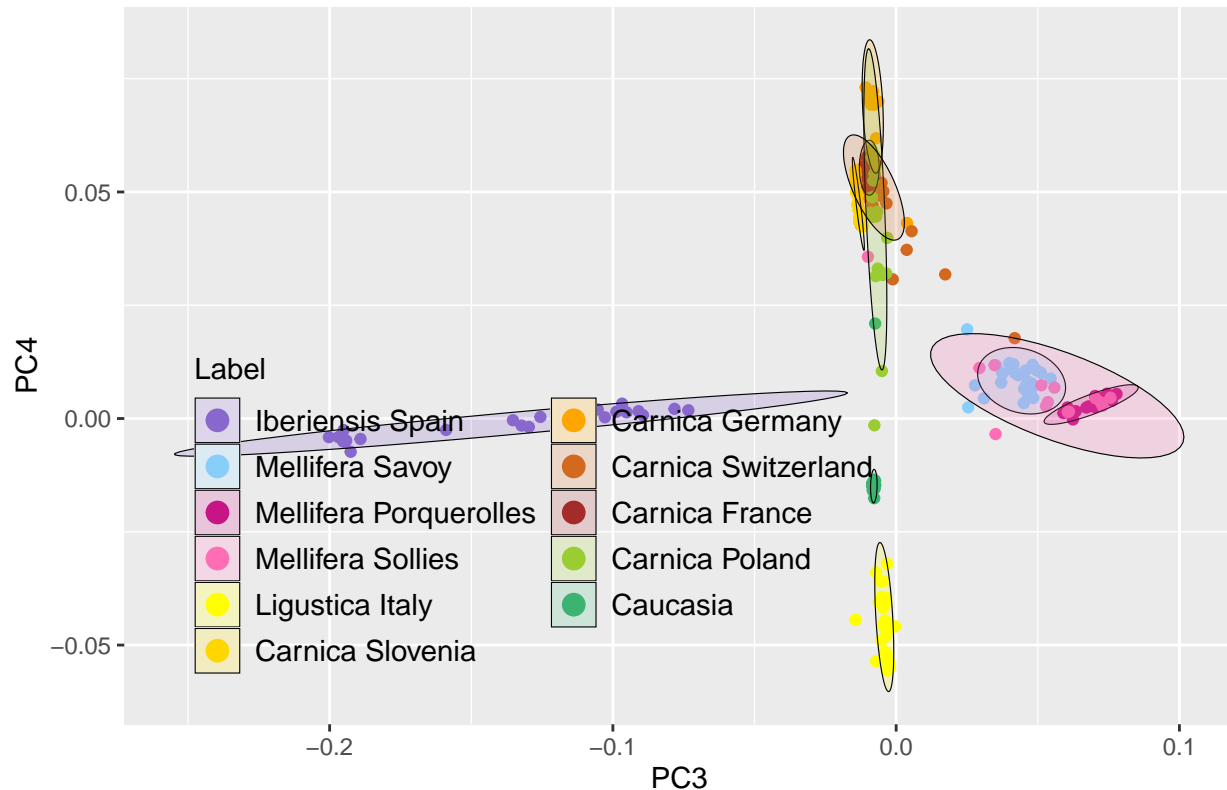
```

```

      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.05, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



ACP - SeqApiPop - 629 échantillons - SNPsBeeMuSe filtered

ACP - SNPsBeeMuSe filtered

```

setwd("~/Documents/Stage_NB/data/SeqApiPop_629_SNPsBeeMuSe")

# SNPsBeeMuSe filtered
eigenvec_SNPsBeeMuSe <- read.table("SeqApiPop_629_SNPsBeeMuSe_filtered_acp.eigenvec", header = F)
eigenval_SNPsBeeMuSe <- read.table("SeqApiPop_629_SNPsBeeMuSe_filtered_acp.eigenval", header = F)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_SNPsBeeMuSe)[colnames(eigenvec_SNPsBeeMuSe) == "V2"] <- "name"

```

```
eigenvec_SNPsBeeMuSe_seq_api_labels <- merge(eigenvec_SNPsBeeMuSe, seq_api_labels, by = "name")

eigen_percent_SNPsBeeMuSe <- round((eigenval_SNPsBeeMuSe / (sum(eigenval_SNPsBeeMuSe) )*100),2)
eigen_percent_SNPsBeeMuSe
```

PC1/PC2

```
##          V1
## 1  63.29
## 2  18.18
## 3   2.34
## 4   1.44
## 5   1.32
## 6   1.15
## 7   1.12
## 8   1.07
## 9   1.03
## 10  1.00
## 11  0.93
## 12  0.84
## 13  0.83
## 14  0.83
## 15  0.81
## 16  0.80
## 17  0.78
## 18  0.76
## 19  0.75
## 20  0.74
```

```
#Clustering hiérarchique
#Tree
matrice_app_refpop <- read.table("SeqApiPop_629_SNPsBeeMuSe_filtered_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)
```


Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

```
#heatmap
#heatplot(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')

#ACP
#filter 629 -> 301 RefPop
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label == "Quessant Conservatory",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Iberiensis Spain",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Savoy Conservatory",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Porquerolles Conservatory",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Solilies Conservatory",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Ligustica Italy",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Carnica Slovenia",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Carnica Germany",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Carnica Switzerland",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Carnica France",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Carnica Poland",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "Caucasia France",]

custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred", "darkslategray", "firebrick", "brown")

lambda <- eigenval_SNPdBeeMuSe$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

ggplot(data = eigenvec_SNPdBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Quessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solilies Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
```

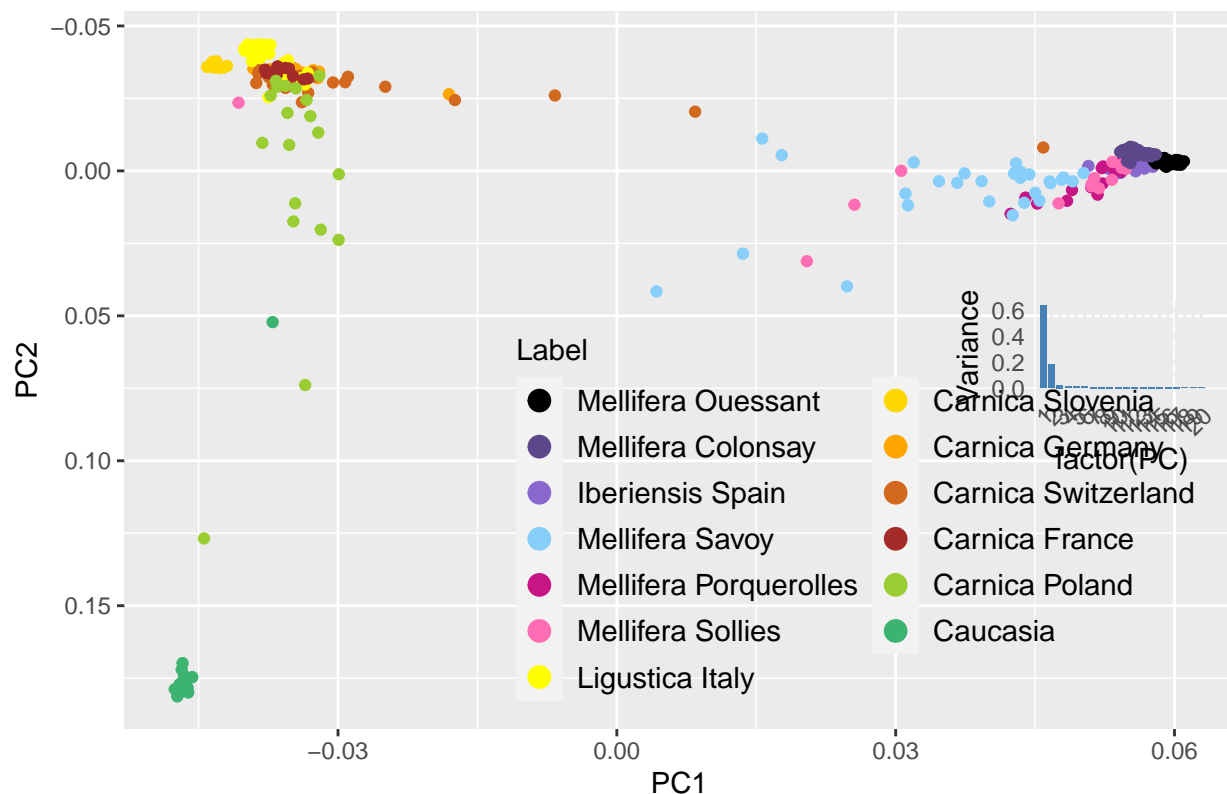
```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.035, xmax = 0.065, ymin = -0.04, ymax = -0.11)

```

PCA Plot – reference populations



```

#ellipses autour des points selon Label
ggplot(data = eigenvec_SNPBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",

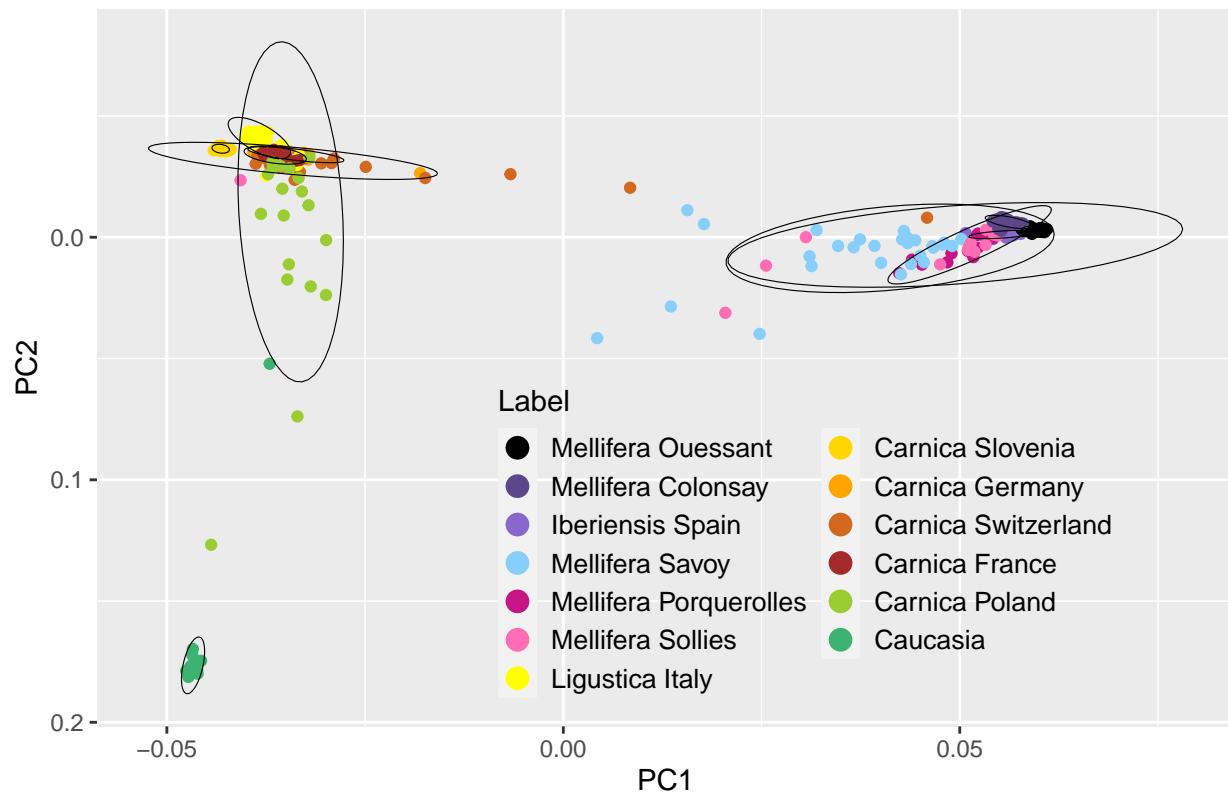
```

```

" Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
" Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
" Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
" Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
" Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
" Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
legend.background = element_rect(fill = "transparent"),
legend.key.size = unit(1, "lines"),
legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses avec couleur
ggplot(data = eigenvec_SNPBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
" Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
" Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
" Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",

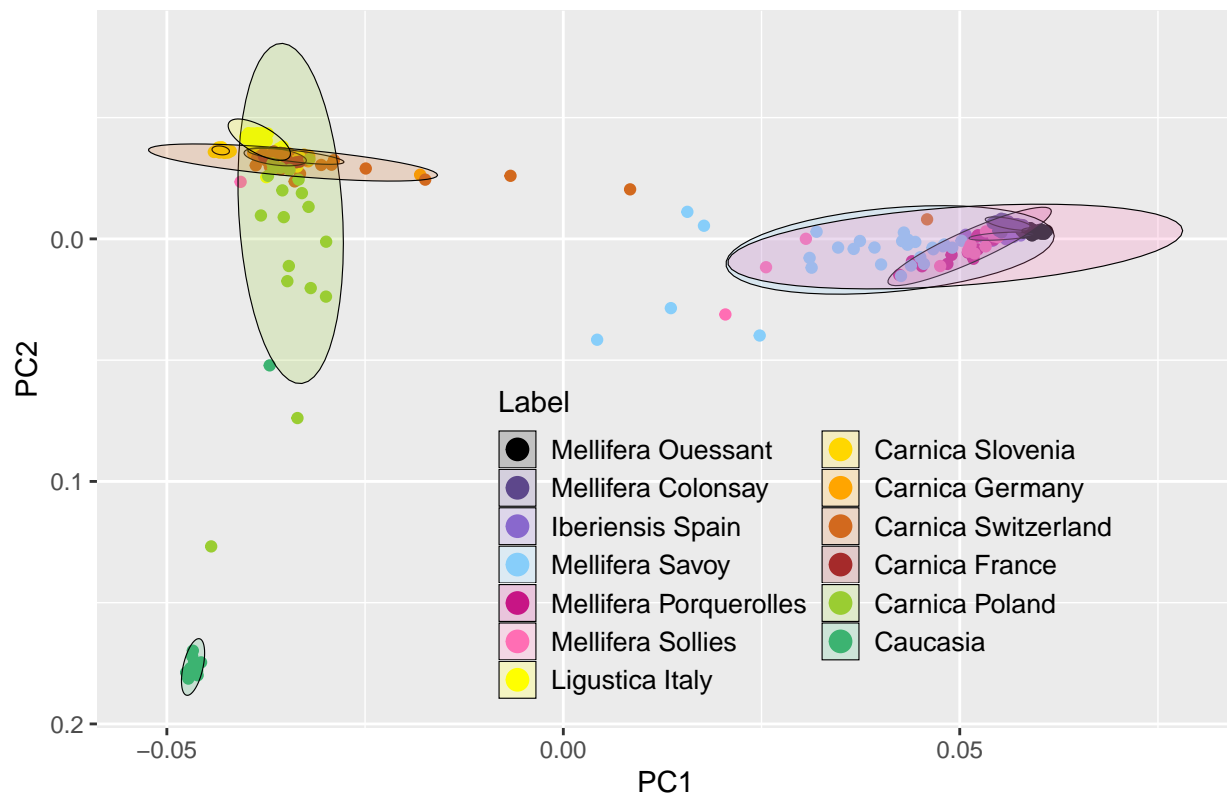
```

```

      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_fill_manual(values = custom_colors_label2,
  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
  labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
  legend.background = element_rect(fill = "transparent"),
  legend.key.size = unit(1, "lines"),
  legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

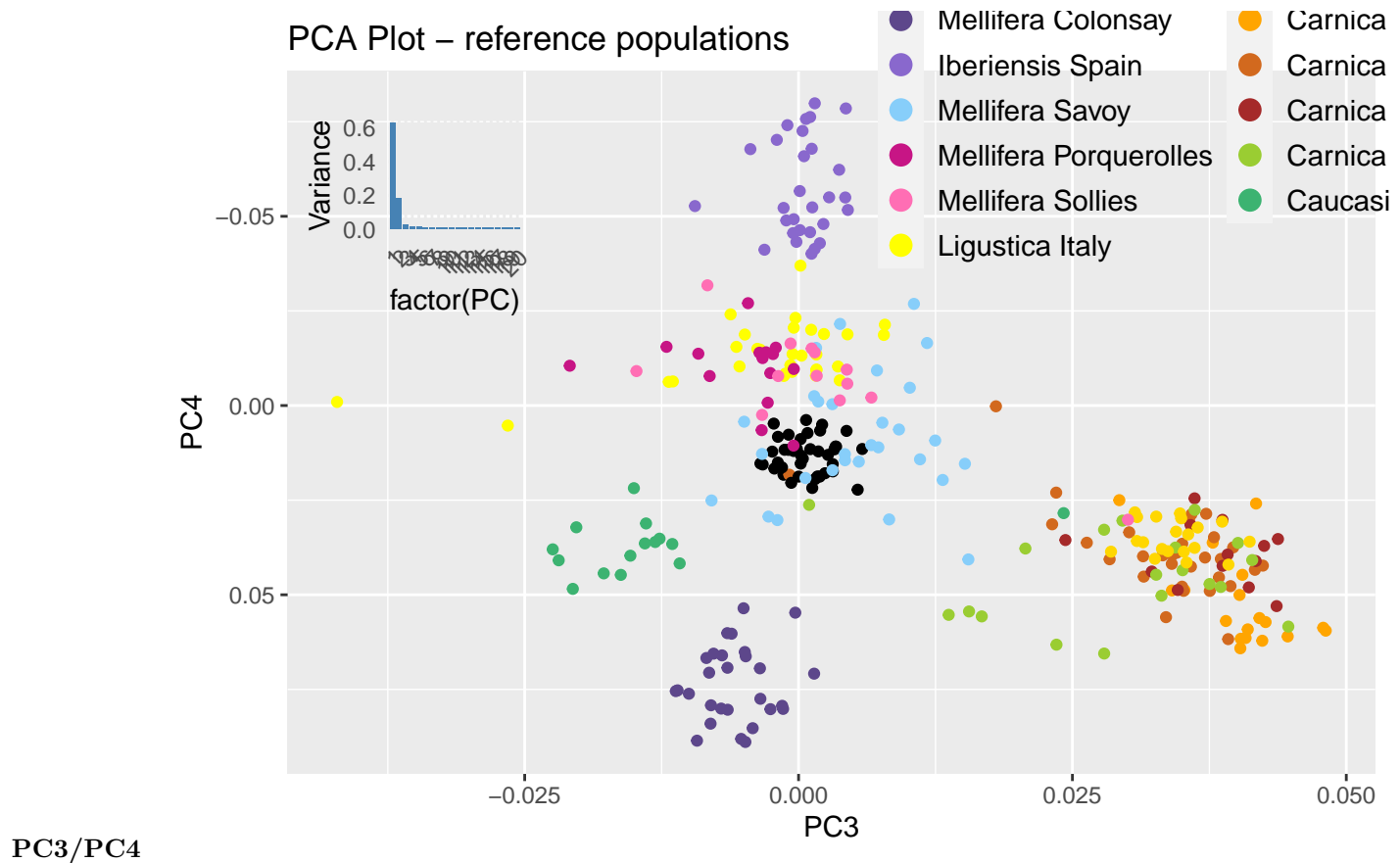
#PC3/PC4
ggplot(data = eigenvec_SNPBeeMuSe_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +

```

```

labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
scale_color_manual(values = custom_colors_label2,
  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
  labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
scale_y_reverse() + # Reverse the y-axis
theme(legend.position = c(0.53, 0.7), legend.justification = c(0, 0),
  legend.background = element_rect(fill = "transparent"),
  legend.key.size = unit(1.2, "lines"),
  legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = -0.046, xmax = -0.024, ymin = 0.08, ymax = 0.02)

```



```

# SNPsBeeMuSe filtered maf001 LD03
setwd("~/Documents/Stage_NB/data/SeqApiPop_629_SNPsBeeMuSe")

eigenvec_SNPsBeeMuSe <- read.table("SeqApiPop_629_SNPsBeeMuSe_filtered_maf001_LD03_acp.eigenvec", header = FALSE)
eigenval_SNPsBeeMuSe <- read.table("SeqApiPop_629_SNPsBeeMuSe_filtered_maf001_LD03_acp.eigenval", header = FALSE)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_SNPsBeeMuSe)[colnames(eigenvec_SNPsBeeMuSe) == "V2"] <- "name"
eigenvec_SNPsBeeMuSe_seq_api_labels <- merge(eigenvec_SNPsBeeMuSe, seq_api_labels, by = "name")

eigen_percent_SNPsBeeMuSe <- round((eigenval_SNPsBeeMuSe / (sum(eigenval_SNPsBeeMuSe) )*100),2)
eigen_percent_SNPsBeeMuSe

```

ACP - MAF > 0.01 - LD pruning = 0.3 (fenêtre de 1749 SNPs et pas de 175 bp)

```

##      V1
## 1  45.69
## 2  22.35
## 3   4.32
## 4   3.38
## 5   2.63
## 6   2.40
## 7   2.18
## 8   1.79
## 9   1.71
## 10  1.37
## 11  1.34
## 12  1.33
## 13  1.29
## 14  1.23
## 15  1.22
## 16  1.17
## 17  1.17
## 18  1.15
## 19  1.15
## 20  1.13

```

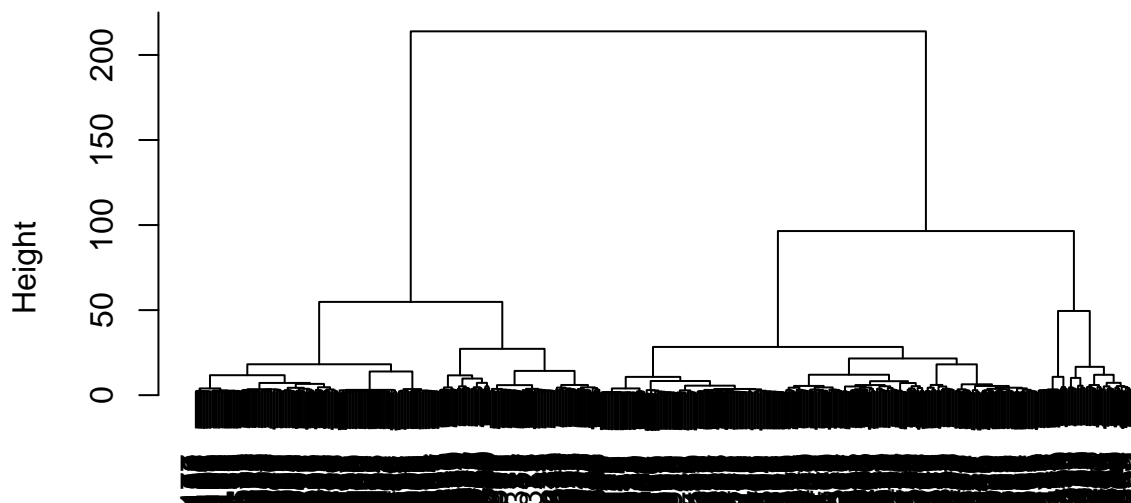
```

#Clustering hiérarchique
#Tree
matrice_app_refpop <- read.table("SeqApiPop_629_SNPsBeeMuSe_filtered_maf001_LD03_acp.rel", header = FALSE)

dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)

```

Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

```
#heatmap
#heatmap(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')

#ACP
#filter 629 -> 301 RefPop
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label == "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]

custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred", "darkgreen")

lambda <- eigenval_SNPdBeeMuSe$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)
```

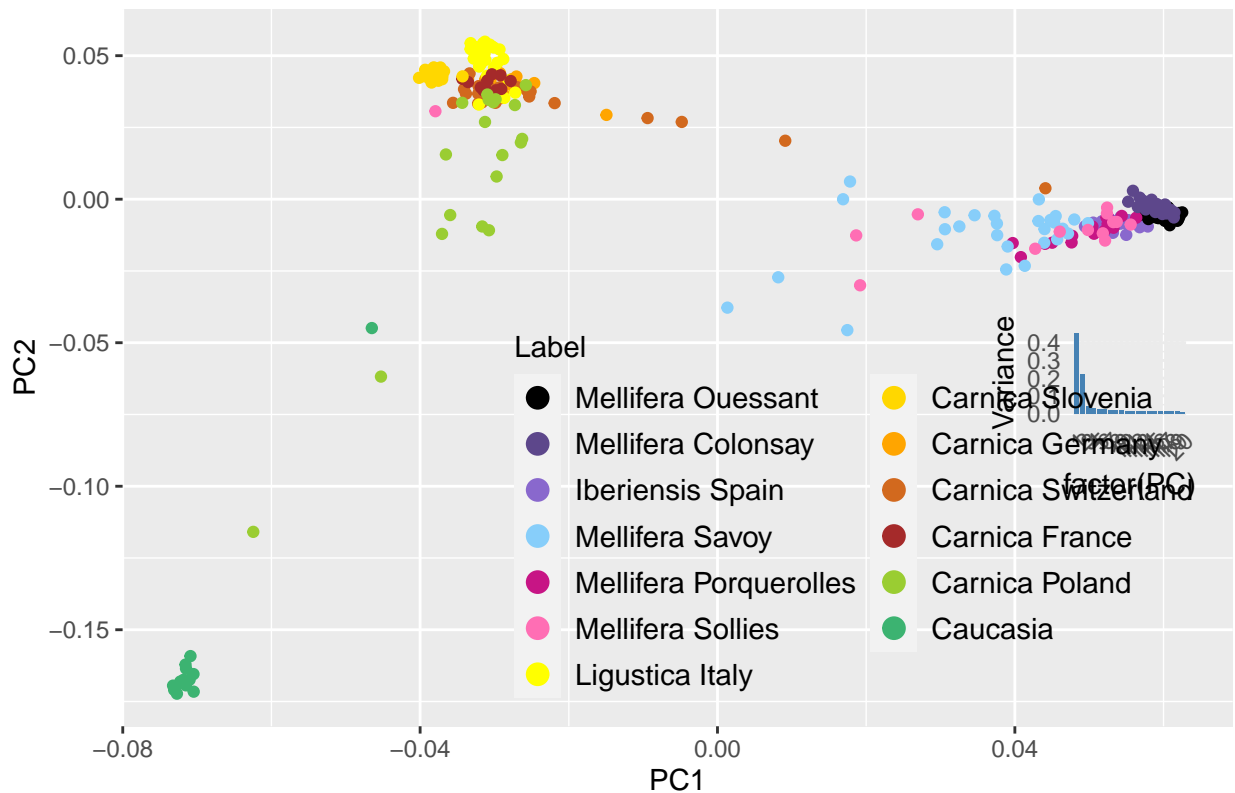
```
ggplot(data = eigenvec_SNPsBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                     breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
```

```

      "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France")
  labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
            "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
            "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
            "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
        legend.background = element_rect(fill = "transparent"),
        legend.key.size = unit(1.2, "lines"),
        legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.035, xmax = 0.065, ymin = -0.04, ymax = -0.11)

```

PCA Plot – reference populations



PC1/PC2

```

ggplot(data = eigenvec_SNPsBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",

```

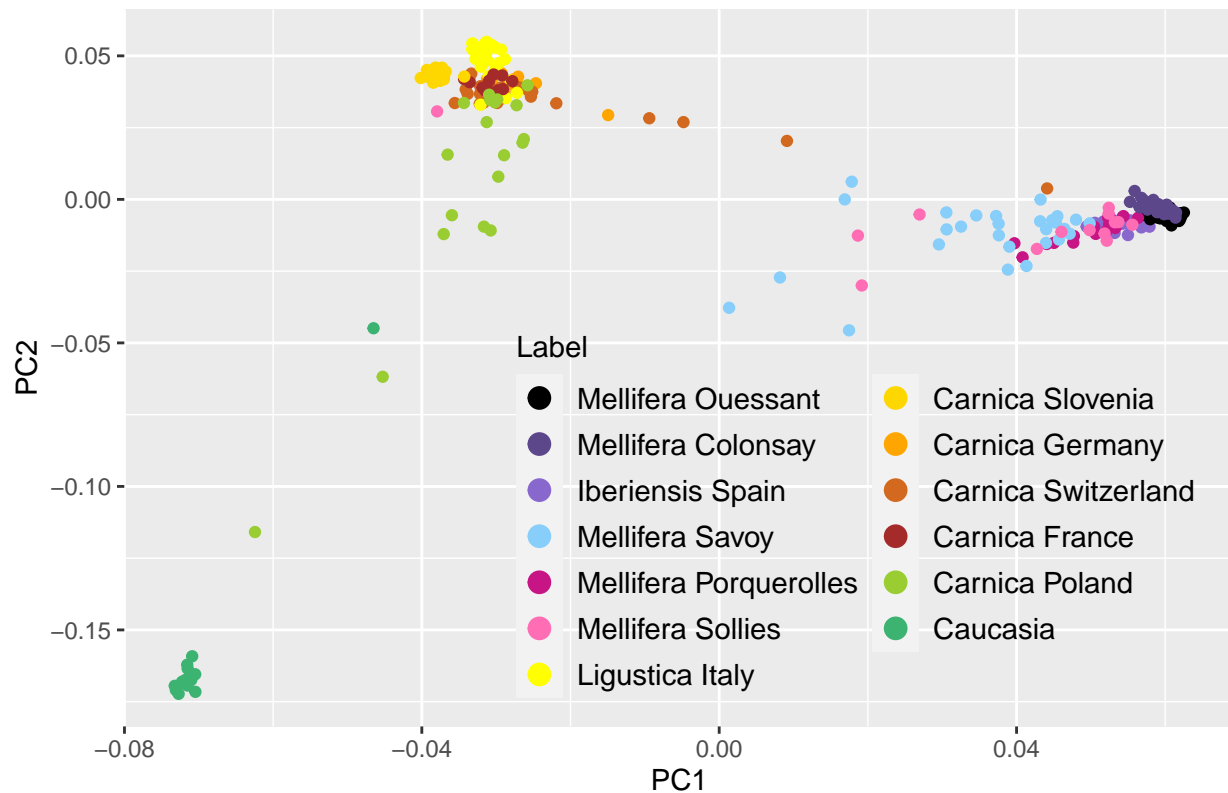


```

    "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#ellipses autour des points selon Label
ggplot(data = eigenvec_SNPBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
  breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
  labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",

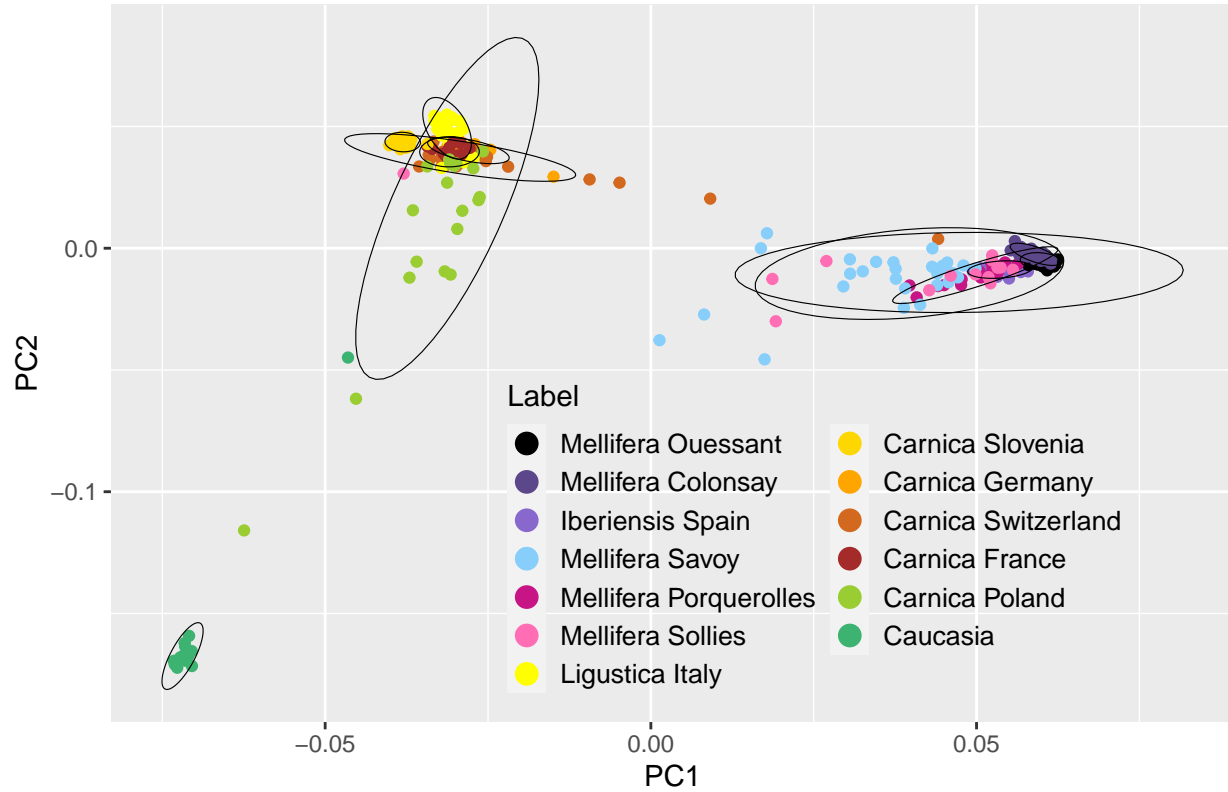
```

```

    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1, "lines"),
      legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses avec couleur
ggplot(data = eigenvec_SNPSeBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia")) +
  scale_fill_manual(values = custom_colors_label2,
                   breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

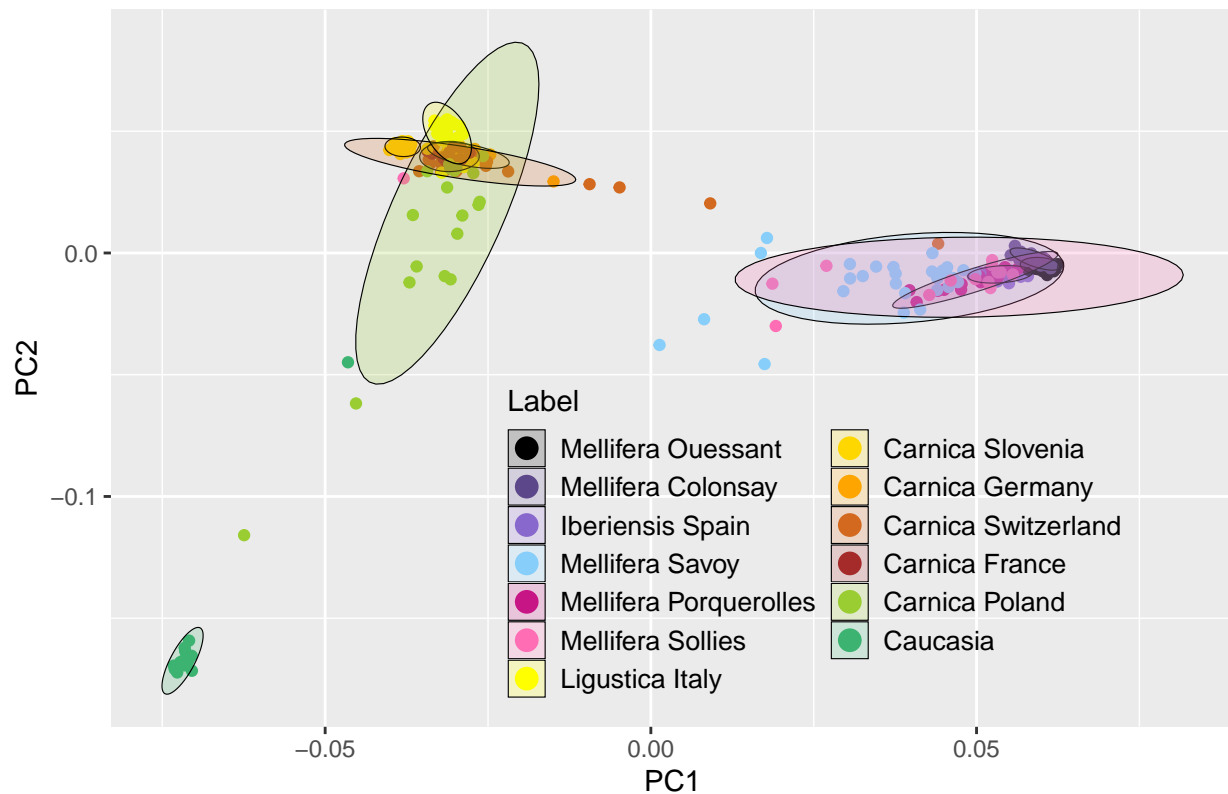
```

```

    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1, "lines"),
      legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#PC3/PC4
ggplot(data = eigenvec_SNPSeBeeMuSe_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
    "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France",
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",

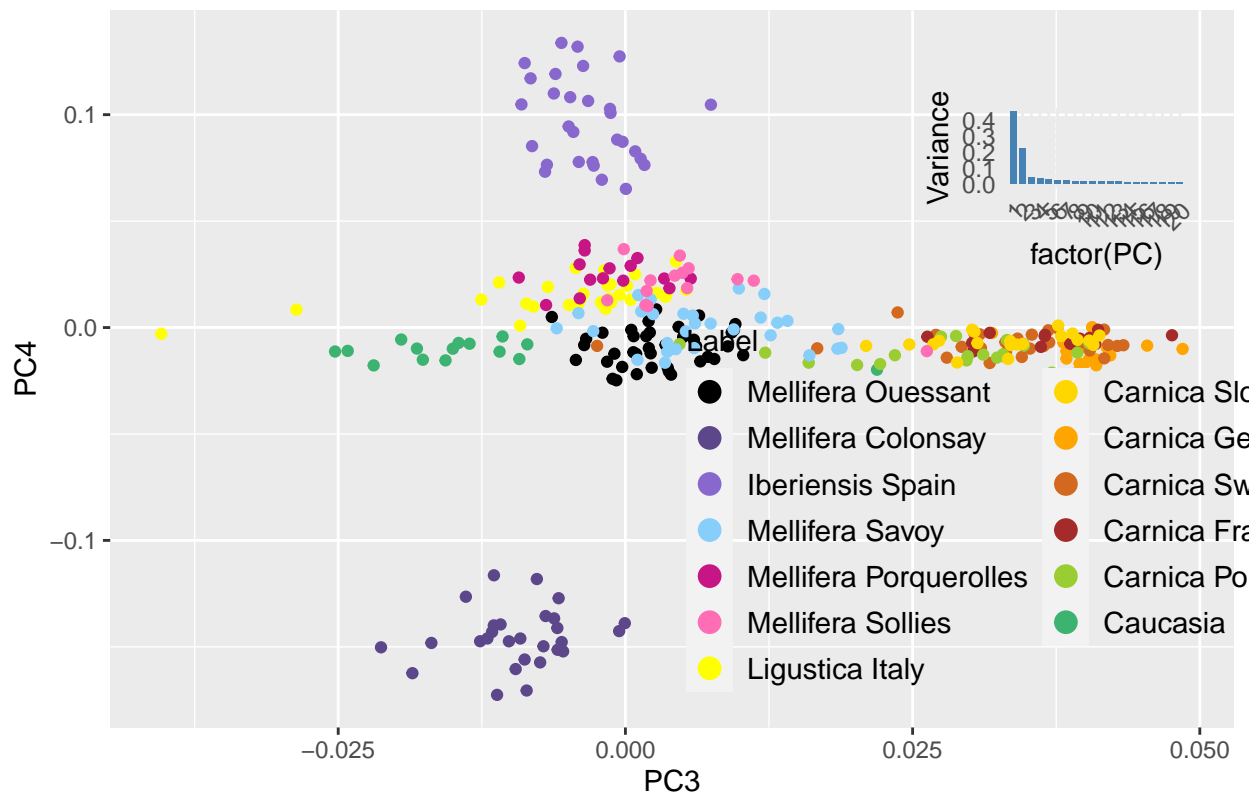
```

```

      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.5, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.025, xmax = 0.05, ymin = 0.02, ymax = 0.11)

```

PCA Plot – reference populations



PC3/PC4

```

setwd("~/Documents/Stage_NB/data/SeqApiPop_629_SNPsBeeMuSe")

eigenvec_SNPsBeeMuSe <- read.table("SeqApiPop_629_SNPsBeeMuSe_filtered_maf001_LD03_default_acp.eigenvec")
eigenval_SNPsBeeMuSe <- read.table("SeqApiPop_629_SNPsBeeMuSe_filtered_maf001_LD03_default_acp.eigenval")

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

```

```
colnames(eigenvec_SNPBeeMuSe)[colnames(eigenvec_SNPBeeMuSe) == "V2"] <- "name"
eigenvec_SNPBeeMuSe_seq_api_labels <- merge(eigenvec_SNPBeeMuSe, seq_api_labels, by = "name")

eigen_percent_SNPBeeMuSe <- round((eigenval_SNPBeeMuSe / (sum(eigenval_SNPBeeMuSe) )*100),2)
eigen_percent_SNPBeeMuSe
```

ACP - MAF > 0.01 - LD pruning = 0.1 (fenêtre de 50 SNPs et pas de 10 bp)

```
##      V1
## 1  35.37
## 2  18.39
## 3   6.17
## 4   5.22
## 5   3.95
## 6   3.20
## 7   3.06
## 8   2.57
## 9   2.17
## 10  2.01
## 11  1.95
## 12  1.90
## 13  1.87
## 14  1.80
## 15  1.78
## 16  1.76
## 17  1.75
## 18  1.71
## 19  1.69
## 20  1.67
```

#Clustering hiérarchique

#Tree

```
matrice_app_refpop <- read.table("SeqApiPop_629_SNPBeeMuSe_filtered_maf001_LD03_default_acp.rel", head
```

```
dist_matrice_refpop <- dist(matrice_app_refpop)
hc_refpop <- hclust(dist_matrice_refpop, method = "ward.D2")
plot(hc_refpop)
```

Cluster Dendrogram



```
dist_matrice_refpop
hclust (*, "ward.D2")
```

```
#ACP
#filter 629 -> 301 RefPop
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label == "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]
eigenvec_SNPdBeeMuSe_seq_api_labels <- eigenvec_SNPdBeeMuSe_seq_api_labels[eigenvec_SNPdBeeMuSe_seq_api_labels$Label != "RefPop",]

custom_colors_label2 <- c("black", "mediumpurple4", "mediumpurple3", "lightskyblue", "mediumvioletred", "darkgreen")

lambda <- eigenval_SNPdBeeMuSe$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)
```

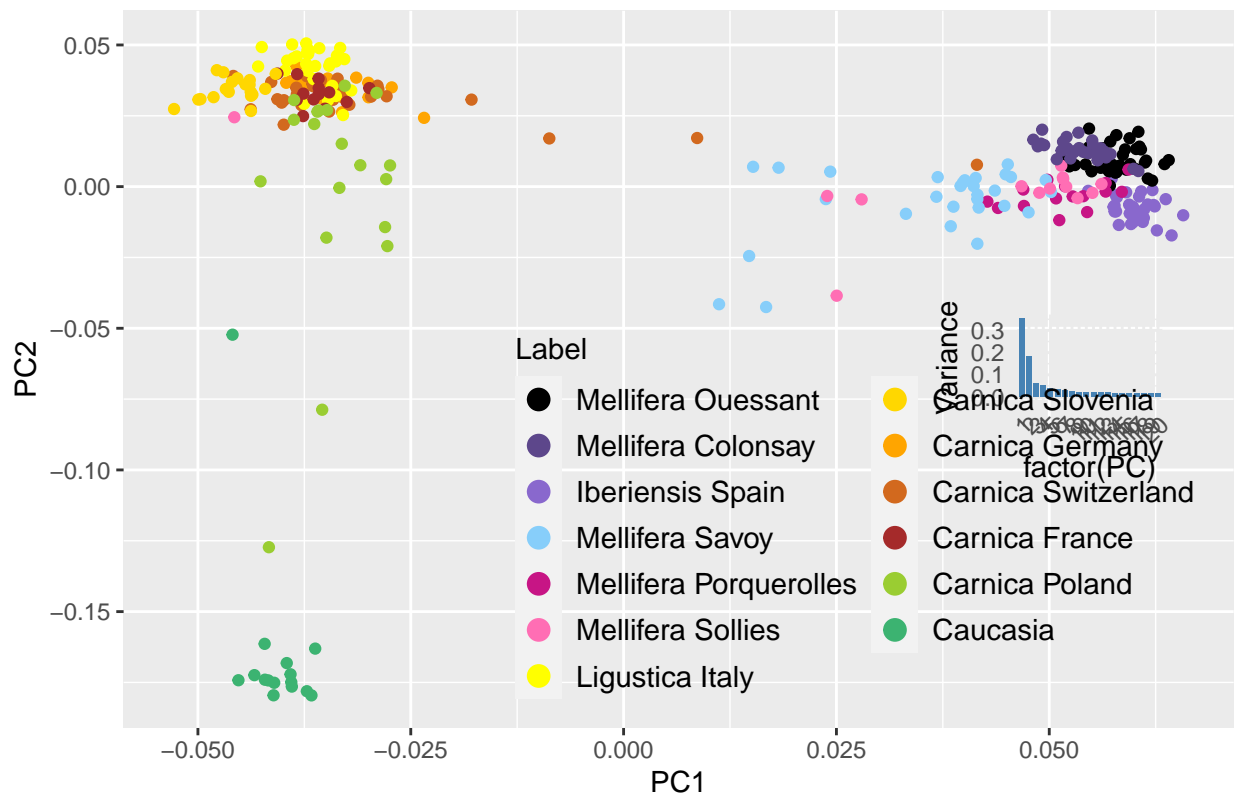
```
ggplot(data = eigenvec_SNPbBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Quessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
      "Liguria Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
```

```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.035, xmax = 0.065, ymin = -0.04, ymax = -0.11)

```

PCA Plot – reference populations



PC1/PC2

```

ggplot(data = eigenvec_SNPsBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
                    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr

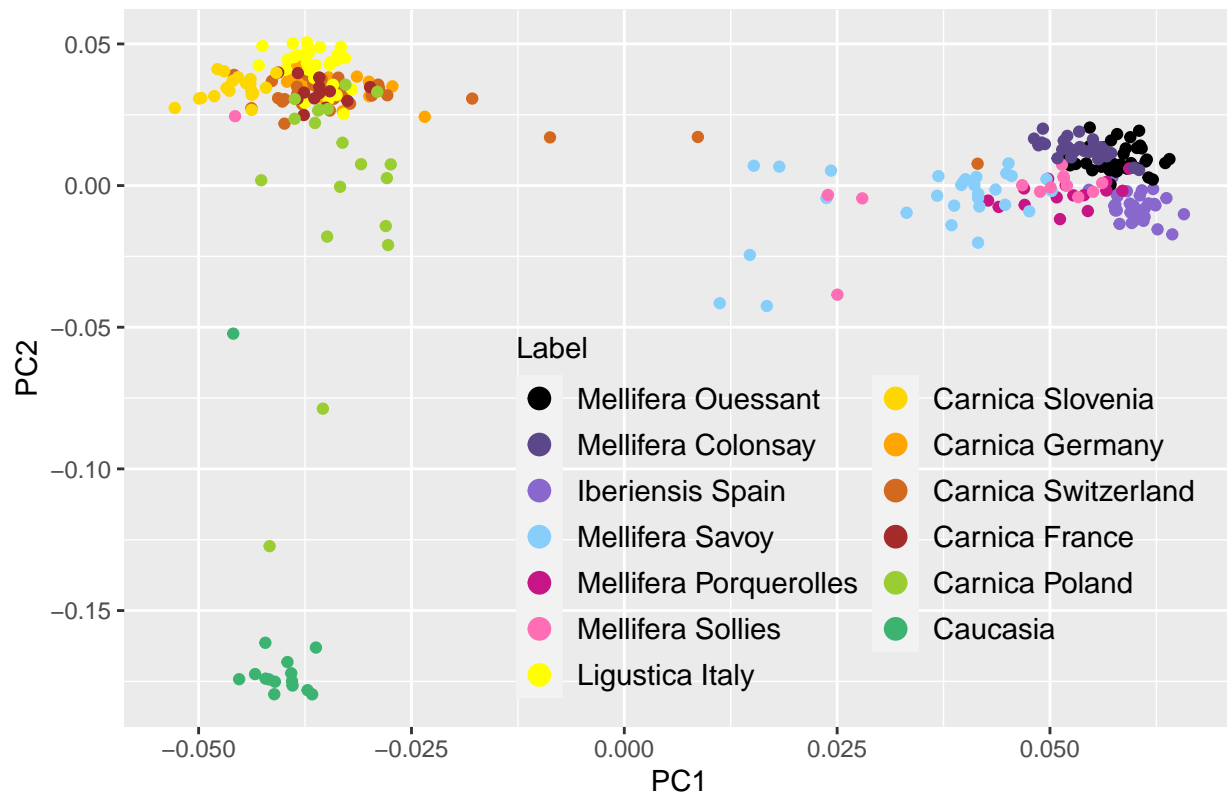
```

```

labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
           "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
           "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
           "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.34, 0.02), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#ellipses autour des points selon Label
ggplot(data = eigenvec_SNPBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black")
labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
scale_color_manual(values = custom_colors_label2,
                   breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                   labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.5, 0.04), legend.justification = c(0, 0),

```

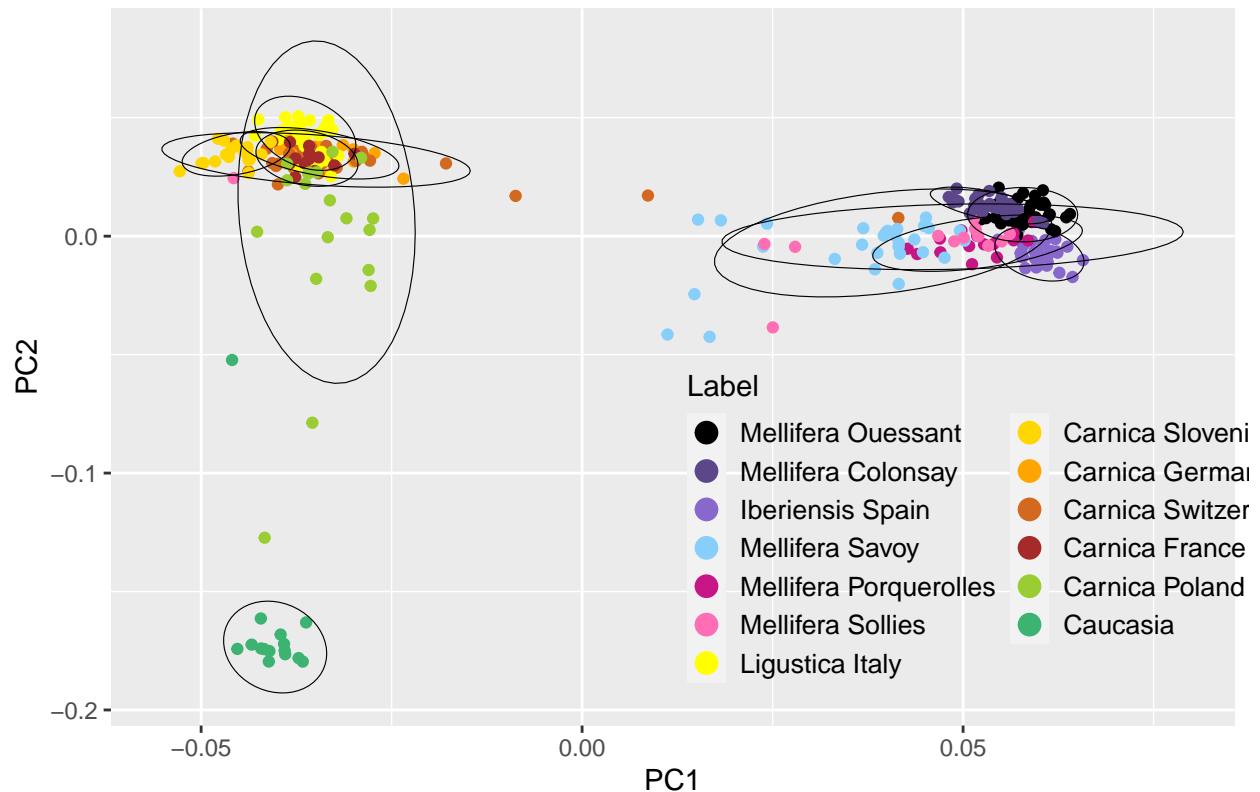


```

legend.background = element_rect(fill = "transparent"),
legend.key.size = unit(1, "lines"),
legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

# ellipses avec couleur
ggplot(data = eigenvec_SNPBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  labs(title = "PCA Plot – reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France")) +
  scale_fill_manual(values = custom_colors_label2,
    breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",

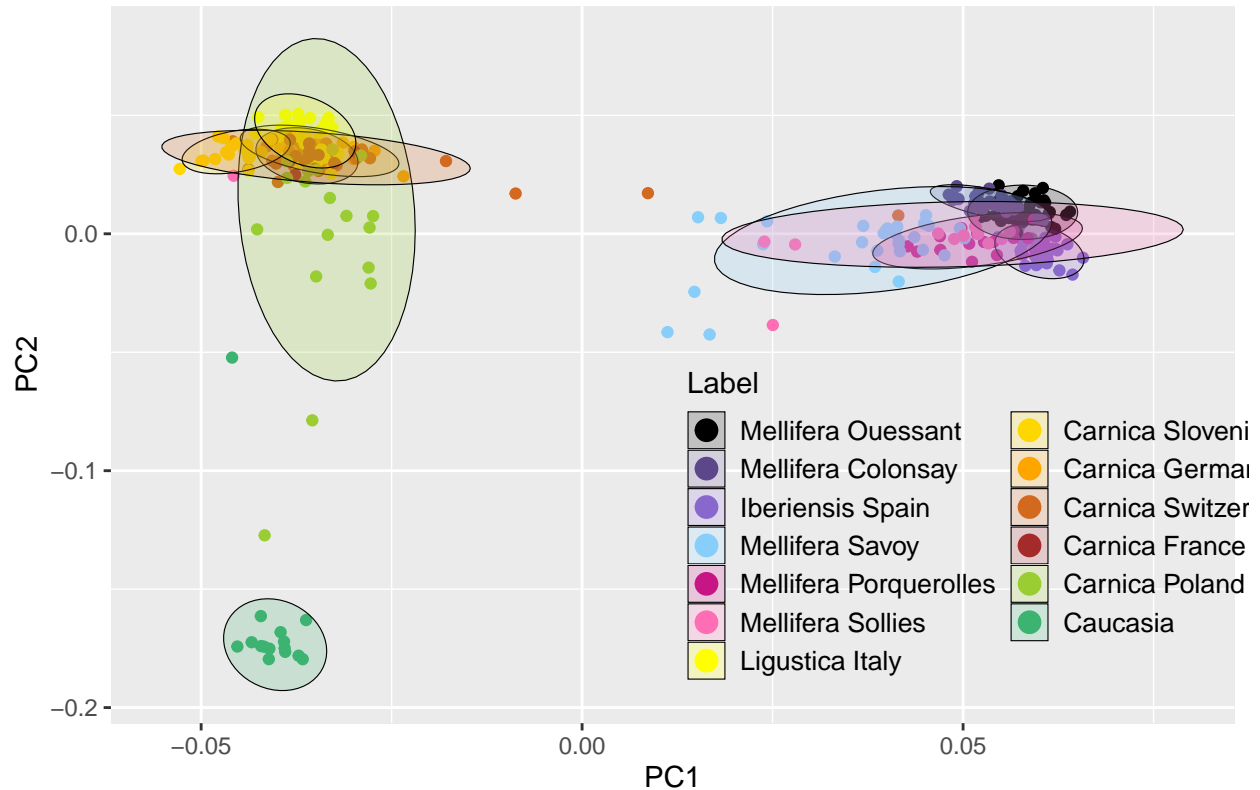
```

```

    "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia")) +
theme(legend.position = c(0.5, 0.04), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1, "lines"),
      legend.text = element_text(size = 10)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

#PC3/PC4
ggplot(data = eigenvec_SNPBeeMuSe_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot – reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label2,
                     breaks = c("Ouessant Conservatory", "Colonsay Conservatory", "Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia Fr"),
                     labels = c("Mellifera Ouessant", "Mellifera Colonsay", "Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))

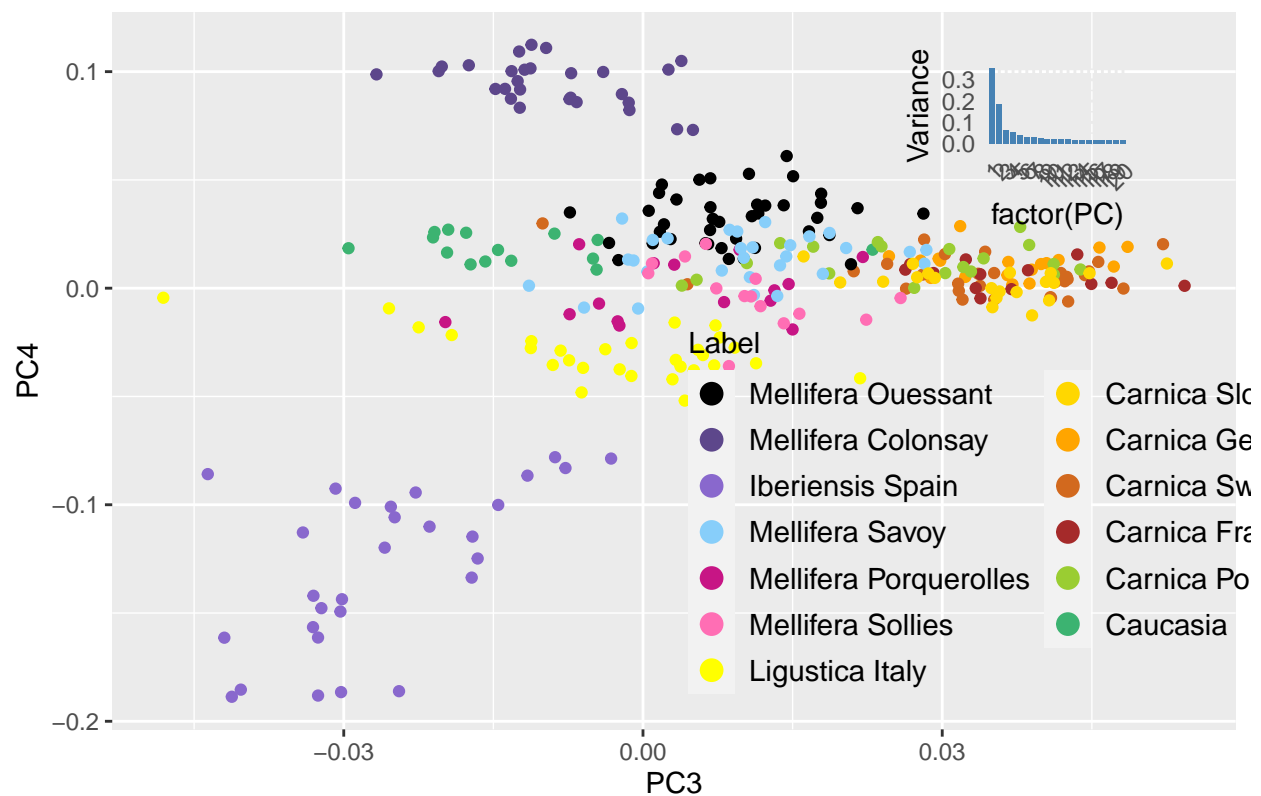
```

```

theme(legend.position = c(0.5, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.025, xmax = 0.05, ymin = 0.02, ymax = 0.11)

```

PCA Plot – reference populations



PC3/PC4

ACP - SeqApiPop - 561 échantillons - SNPsBeeMuSe filtered

```

setwd("~/Documents/Stage_NB/data/SeqApiPop_561_SNPsBeeMuse_LD03")

eigenvec_561_SNPsBeeMuSe <- read.table("SeqApiPop_561_SNPsBeeMuSe_filtered_maf001_LD03_acp.eigenvec", header = TRUE, as.is = TRUE)
eigenval_561_SNPsBeeMuSe <- read.table("SeqApiPop_561_SNPsBeeMuSe_filtered_maf001_LD03_acp.eigenval", header = TRUE, as.is = TRUE)

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

```

```
colnames(eigenvec_561_SNPBeeMuSe)[colnames(eigenvec_561_SNPBeeMuSe) == "V2"] <- "name"
eigenvec_561_SNPBeeMuSe_seq_api_labels <- merge(eigenvec_561_SNPBeeMuSe, seq_api_labels, by = "name")

eigen_percent_561_SNPBeeMuSe <- round((eigenval_561_SNPBeeMuSe / (sum(eigenval_561_SNPBeeMuSe) ) * 100)
```

```
#Clustering hiérarchique
#Tree
setwd("~/Documents/Stage_NB/data/SeqApiPop_561_SNPBeeMuse_LD03")
matrice_app_561_default <- read.table("SeqApiPop_561_SNPBeeMuSe_filtered_maf001_LD03_acp.rel", header = F)

dist_matrice_561_default <- dist(matrice_app_561_default)
hc_561_d <- hclust(dist_matrice_561_default, method = "ward.D2")
plot(hc_561_d)
```

ACP - MAF > 0.01 - LD pruning = 0.3 (fenêtre de 1749 SNPs et pas de 175 bp)

Cluster Dendrogram



```
dist_matrice_561_default
hclust (*, "ward.D2")
```

```
#heatmap
#heatmap(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')
```

```
#ACP
eigenvec_561_SNPBeeMuSe_seq_api_labels <- eigenvec_561_SNPBeeMuSe_seq_api_labels[eigenvec_561_SNPBeeMuSe_seq_api_labels[, "name"] == "V2", ]
eigenvec_561_SNPBeeMuSe_seq_api_labels <- eigenvec_561_SNPBeeMuSe_seq_api_labels[eigenvec_561_SNPBeeMuSe_seq_api_labels[, "name"] == "V2", ]
eigenvec_561_SNPBeeMuSe_seq_api_labels <- eigenvec_561_SNPBeeMuSe_seq_api_labels[eigenvec_561_SNPBeeMuSe_seq_api_labels[, "name"] == "V2", ]
eigenvec_561_SNPBeeMuSe_seq_api_labels <- eigenvec_561_SNPBeeMuSe_seq_api_labels[eigenvec_561_SNPBeeMuSe_seq_api_labels[, "name"] == "V2", ]
```

```

eigenvec_561_SNPsBeeMuSe_seq_api_labels
eigenvec_561_SNPsBeeMuSe_seq_api_labels

custom_colors_label <- c("mediumpurple3", "lightskyblue", "mediumvioletred", "hotpink1", "yellow", "gold")

lambda <- eigenval_561_SNPsBeeMuSe$V1
variance_proportion <- lambda / sum(lambda)
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)

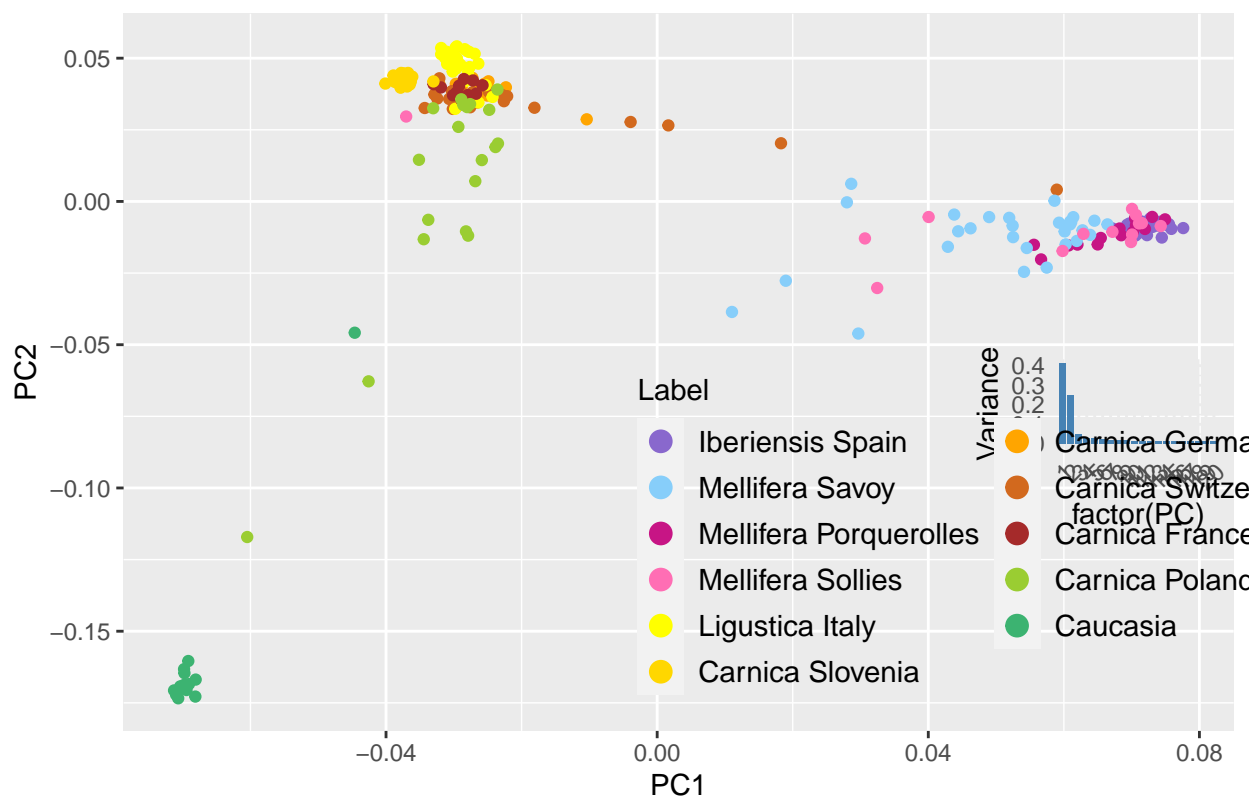
```

```

#ACP - plot variance
ggplot(data = eigenvec_561_SNPsBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label,
                     breaks = c("Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                     labels = c("Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.45, 0.03), legend.justification = c(0, 0),
        legend.background = element_rect(fill = "transparent"),
        legend.key.size = unit(1.2, "lines"),
        legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.085, xmax = 0.045, ymin = -0.12, ymax = -0.05)

```

PCA Plot – reference populations

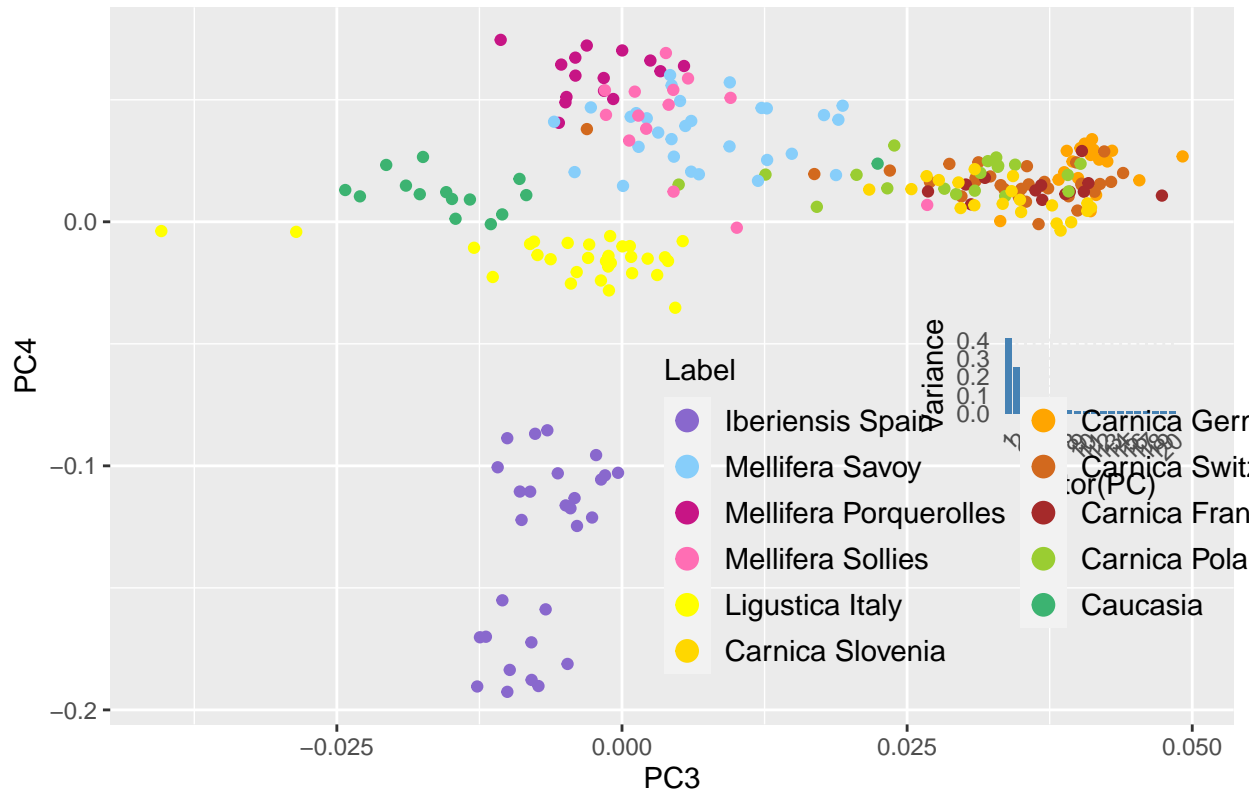


PC1/PC2

```
#ACP - plot variance
ggplot(data = eigenvec_561_SNPsBeeMuSe_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label,
    breaks = c( "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.48, 0.05), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),
    legend.key.size = unit(1.2, "lines"),
    legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
```

```
theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
), xmin = 0.025, xmax = 0.05, ymin = -0.04, ymax = -0.12)
```

PCA Plot – reference populations



PC3/PC4

```
setwd("~/Documents/Stage_NB/data/SeqApiPop_561_SNPBeeMuSe_LD_default")

eigenvec_561_SNPBeeMuSe <- read.table("SeqApiPop_561_SNPBeeMuSe_filtered_maf001_LD03_default_pruned_a
eigenval_561_SNPBeeMuSe <- read.table("SeqApiPop_561_SNPBeeMuSe_filtered_maf001_LD03_default_pruned_a

seq_api_labels <- read.csv("~/Documents/Stage_NB/data/SeqApiPop_labels.csv")

colnames(eigenvec_561_SNPBeeMuSe)[colnames(eigenvec_561_SNPBeeMuSe) == "V2"] <- "name"
eigenvec_561_SNPBeeMuSe_seq_api_labels <- merge(eigenvec_561_SNPBeeMuSe, seq_api_labels, by = "name")

eigen_percent_561_SNPBeeMuSe <- round((eigenval_561_SNPBeeMuSe / (sum(eigenval_561_SNPBeeMuSe) ) * 100)
eigen_percent_561_SNPBeeMuSe
```

ACP - MAF > 0.01 - LD pruning = 0.1 (fenêtre de 50 SNPs et pas de 10 bp)

```
##      V1
## 1  32.47
## 2  20.10
## 3   6.99
```

```
## 4 4.98
## 5 3.65
## 6 3.10
## 7 2.43
## 8 2.36
## 9 2.20
## 10 2.17
## 11 2.11
## 12 2.04
## 13 2.02
## 14 2.00
## 15 1.97
## 16 1.92
## 17 1.91
## 18 1.88
## 19 1.86
## 20 1.85
```

```
#Clustering hiérarchique
```

```
#Tree
```

```
setwd("~/Documents/Stage_NB/data/SeqApiPop_561_SNPsBeeMuSe_LD_default")
```

```
matrice_app_561_default <- read.table("SeqApiPop_561_SNPsBeeMuSe_filtered_maf001_LD03_default_pruned_acj")
```

```
dist_matrice_561_default <- dist(matrice_app_561_default)
```

```
hc_561_d <- hclust(dist_matrice_561_default, method = "ward.D2")
```

```
plot(hc_561_d)
```

Cluster Dendrogram



```
dist_matrice_561_default
hclust (*, "ward.D2")
```



```
#heatmap
#heatmap(as.matrix(dist(matrice_app_refpop,diag=T)), cols.default = F, lowcol = 'blue', highcol='yellow')
```

```
#ACP
```

```
eigenvec_561_SNPsBeeMuSe_seq_api_labels <- eigenvec_561_SNPsBeeMuSe_seq_api_labels[eigenvec_561_SNPsBeeMuSe_seq_api_labels[,1] > 0,]
eigenvec_561_SNPsBeeMuSe_seq_api_labels[eigenvec_561_SNPsBeeMuSe_seq_api_labels[,1] < 0,]
eigenvec_561_SNPsBeeMuSe_seq_api_labels[eigenvec_561_SNPsBeeMuSe_seq_api_labels[,1] > 0,]
eigenvec_561_SNPsBeeMuSe_seq_api_labels[eigenvec_561_SNPsBeeMuSe_seq_api_labels[,1] < 0,]
eigenvec_561_SNPsBeeMuSe_seq_api_labels[eigenvec_561_SNPsBeeMuSe_seq_api_labels[,1] > 0,]
eigenvec_561_SNPsBeeMuSe_seq_api_labels[eigenvec_561_SNPsBeeMuSe_seq_api_labels[,1] < 0,]
```

```
custom_colors_label <- c("mediumpurple3", "lightskyblue", "mediumvioletred", "hotpink1", "yellow", "gold")
```

```
lambda <- eigenval_561_SNPsBeeMuSe$V1
```

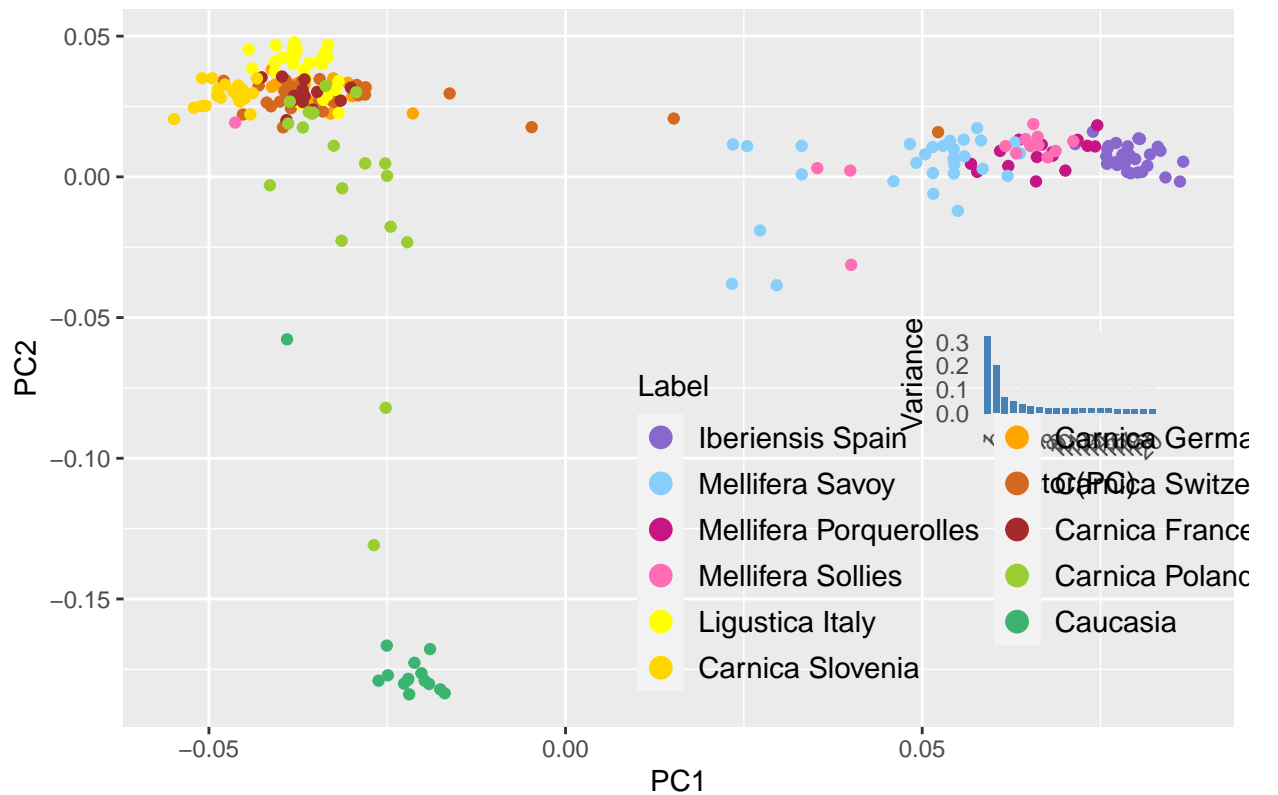
```
variance_proportion <- lambda / sum(lambda)
```

```
variance_df <- data.frame(PC = seq_along(variance_proportion), Variance = variance_proportion)
```

```
# ACP avec variance expliquée
```

```
ggplot(data = eigenvec_561_SNPsBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label,
                     breaks = c("Iberiensis Spain",
                                "Savoy Conservatory", "Porquerolles Conservatory", "Solliès Conservatory",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                     labels = c("Iberiensis Spain",
                                "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solliès",
                                "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                                "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.45, 0.03), legend.justification = c(0, 0),
        legend.background = element_rect(fill = "transparent"),
        legend.key.size = unit(1.2, "lines"),
        legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
  annotation_custom(
    ggplotGrob(
      ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
        geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
        theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
    ), xmin = 0.085, xmax = 0.045, ymin = -0.12, ymax = -0.05)
```

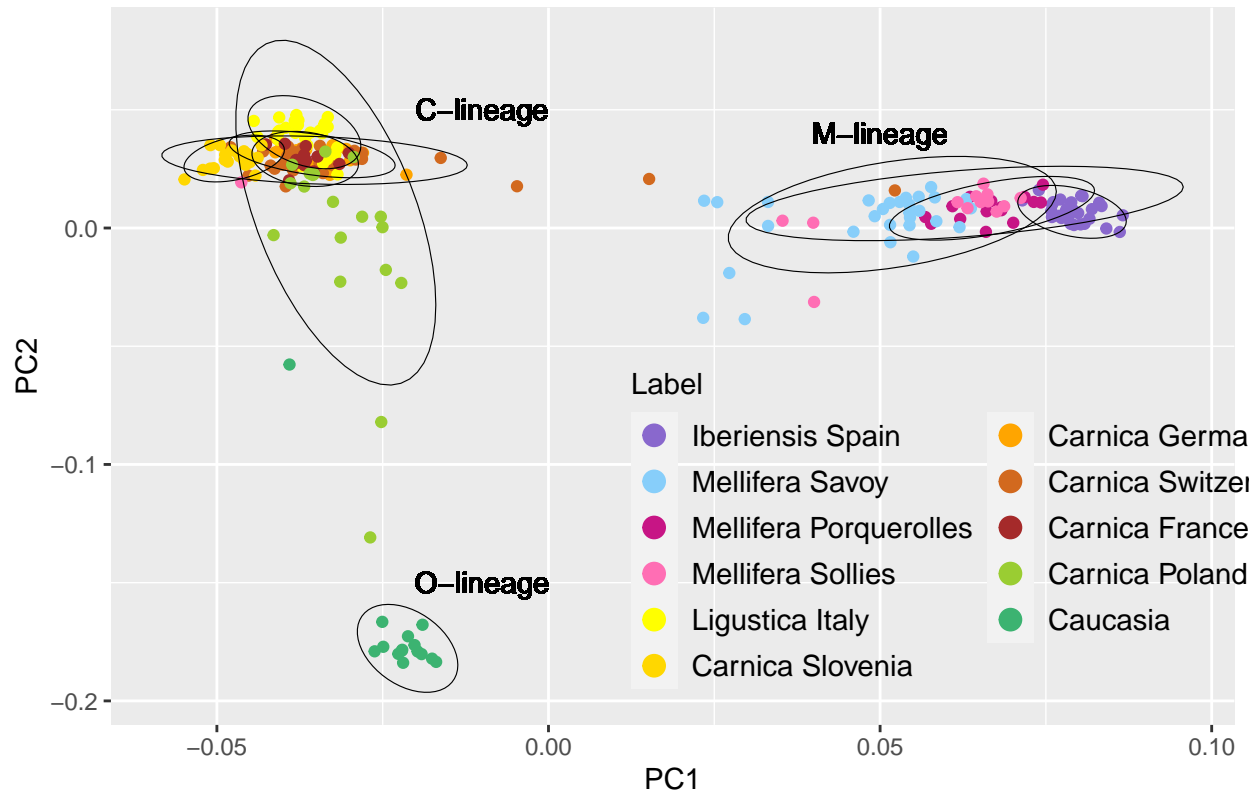
PCA Plot – reference populations



PC1/PC2

```
#ellipses autour des points selon Label
ggplot(data = eigenvec_561_SNPBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black") +
  geom_text(aes(x = 0.05, y = 0.04, label = "M-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.01, y = 0.05, label = "C-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.01, y = -0.15, label = "O-lineage"), size = 4, color = "black") +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label,
    breaks = c( "Iberiennis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c( "Iberiennis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.45, 0.03), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),
    legend.key.size = unit(1.2, "lines"),
    legend.text = element_text(size = 11)) +
  guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))
```

PCA Plot – reference populations



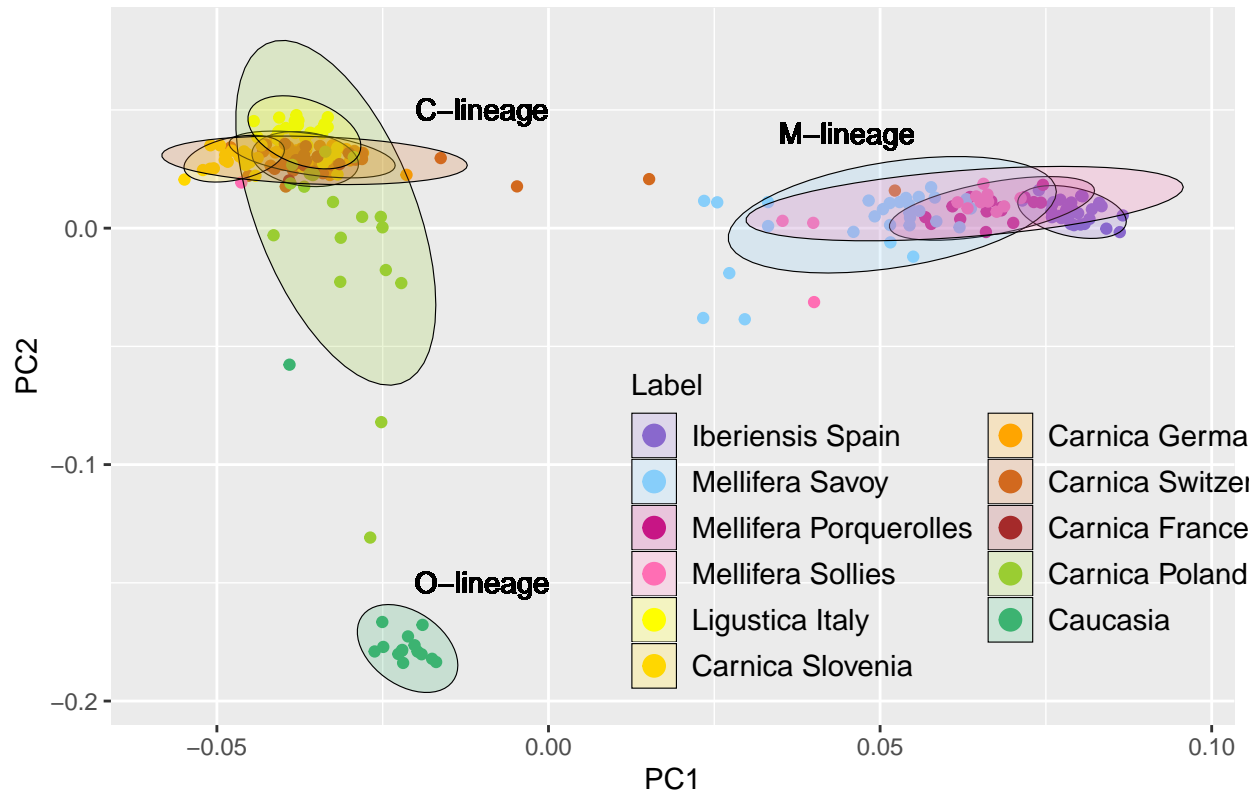
```
# ellipses avec couleur
ggplot(data = eigenvec_561_SNPsBeeMuSe_seq_api_labels, aes(x = V3, y = V4, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  geom_text(aes(x = 0.045, y = 0.04, label = "M-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.01, y = 0.05, label = "C-lineage"), size = 4, color = "black") +
  geom_text(aes(x = -0.01, y = -0.15, label = "O-lineage"), size = 4, color = "black") +
  labs(title = "PCA Plot - reference populations", x = "PC1", y = "PC2") +
  scale_color_manual(values = custom_colors_label,
    breaks = c( "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
  scale_fill_manual(values = custom_colors_label,
    breaks = c( "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c( "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Solles",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
```

```

theme(legend.position = c(0.45, 0.03), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

PCA Plot – reference populations



```

ggplot(data = eigenvec_561_SNPsBeeMuSe_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  labs(title = "PCA Plot - reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label,
                    breaks = c("Iberiensis Spain",
                              "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
                    labels = c("Iberiensis Spain",
                              "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
                              "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                              "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.45, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +

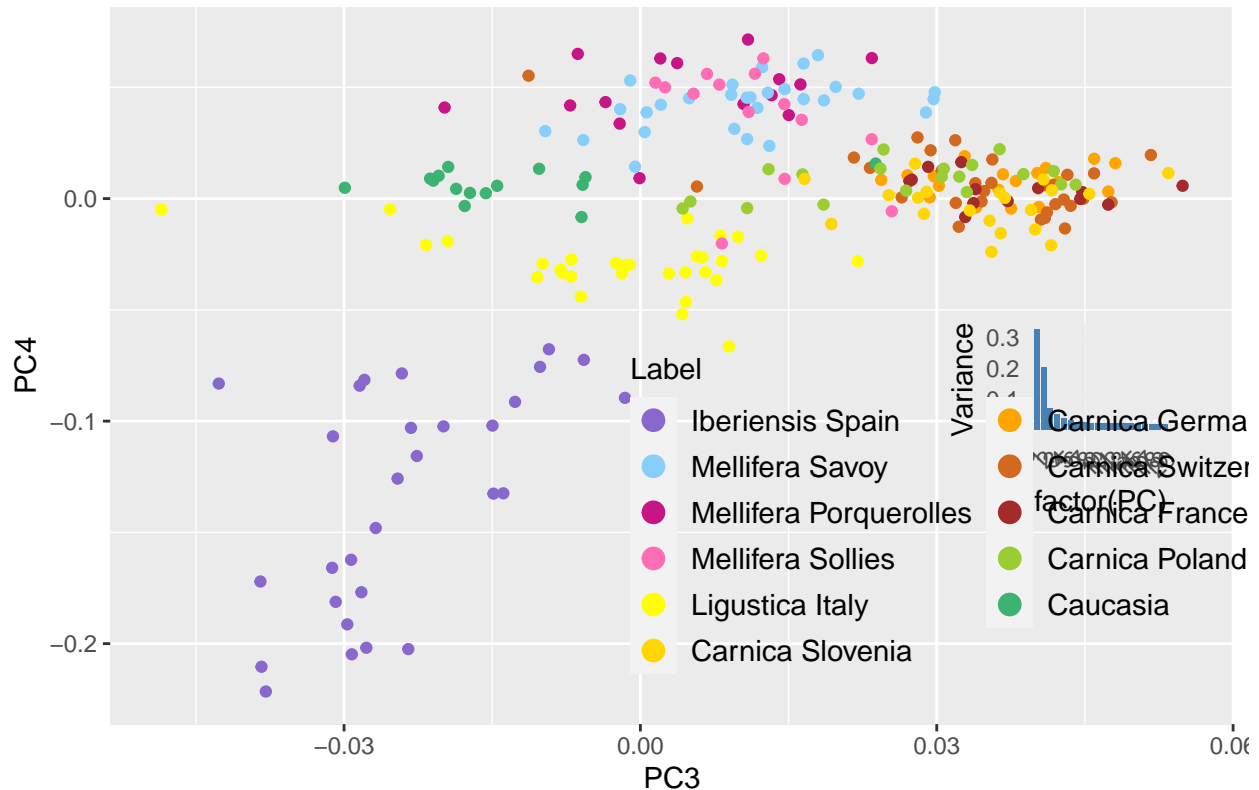
```

```

guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2)) +
annotation_custom(
  ggplotGrob(
    ggplot(variance_df, aes(x = factor(PC), y = Variance)) +
    geom_bar(stat = "identity", fill = "steelblue", width = 0.8) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
  ), xmin = 0.03, xmax = 0.055, ymin = -0.05, ymax = -0.15)

```

PCA Plot – reference populations



PC3/PC4

```

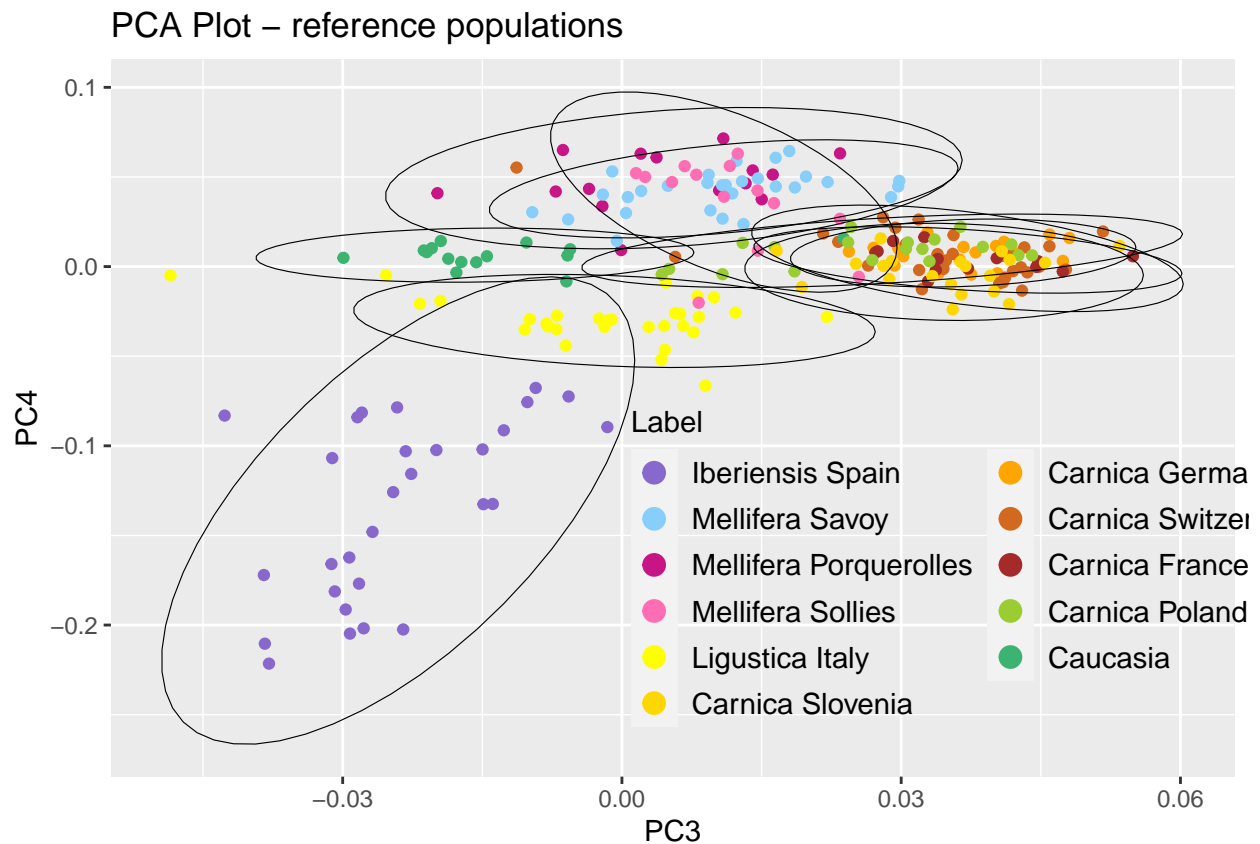
#ellipses autour des points selon Label
ggplot(data = eigenvec_561_SNPsBeeMuSe_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label), geom = "polygon", level = 0.97, alpha = 0, size = 0.2, color = "black") +
  labs(title = "PCA Plot – reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label,
    breaks = c( "Iberiensis Spain",
      "Savoy Conservatory", "Porquerolles Conservatory", "Sollies Conservatory",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c( "Iberiensis Spain",
      "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
      "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
      "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
  theme(legend.position = c(0.45, 0.05), legend.justification = c(0, 0),
    legend.background = element_rect(fill = "transparent"),

```

```

legend.key.size = unit(1.2, "lines"),
legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```



```

# ellipses avec couleur
ggplot(data = eigenvec_561_SNPsBeeMuSe_seq_api_labels, aes(x = V5, y = V6, color = Label)) +
  geom_point() +
  stat_ellipse(aes(group = Label, fill = Label), geom = "polygon", level = 0.97, alpha = 0.2, size = 0.5) +
  labs(title = "PCA Plot – reference populations", x = "PC3", y = "PC4") +
  scale_color_manual(values = custom_colors_label,
    breaks = c( "Iberiensis Spain",
                 "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                 "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                 "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c("Iberiensis Spain",
               "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",
               "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
               "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"))
  scale_fill_manual(values = custom_colors_label,
    breaks = c( "Iberiensis Spain",
                 "Savoy Conservatory", "Porquerolles Conservatory", "Solles Conservatory",
                 "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
                 "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia France"),
    labels = c( "Iberiensis Spain",
               "Mellifera Savoy", "Mellifera Porquerolles", "Mellifera Sollies",

```

```

    "Ligustica Italy", "Carnica Slovenia", "Carnica Germany",
    "Carnica Switzerland", "Carnica France", "Carnica Poland", "Caucasia"))
theme(legend.position = c(0.45, 0.05), legend.justification = c(0, 0),
      legend.background = element_rect(fill = "transparent"),
      legend.key.size = unit(1.2, "lines"),
      legend.text = element_text(size = 11)) +
guides(color = guide_legend(override.aes = list(size = 3.5), ncol = 2))

```

