Batis diops

Peggy Huss & Jacob C. Cooper

2024-10-10

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

PCA Results

Directions from this site.

```
pca <- read_table("~/Documents/06_plink/chamaetylas_nothin_plink_pca.eigenvec",
                  col_names = F)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_double(),
##   X4 = col_double(),
##   X5 = col_double(),
##   X6 = col_double(),
##   X7 = col_double(),
##   X8 = col_double(),
##   X9 = col_double(),
##   X10 = col_double(),
##   X11 = col_double()
## )
```

```
eigenval <- scan("~/Documents/06_plink/chamaetylas_nothin_plink_pca.eigenval")
```

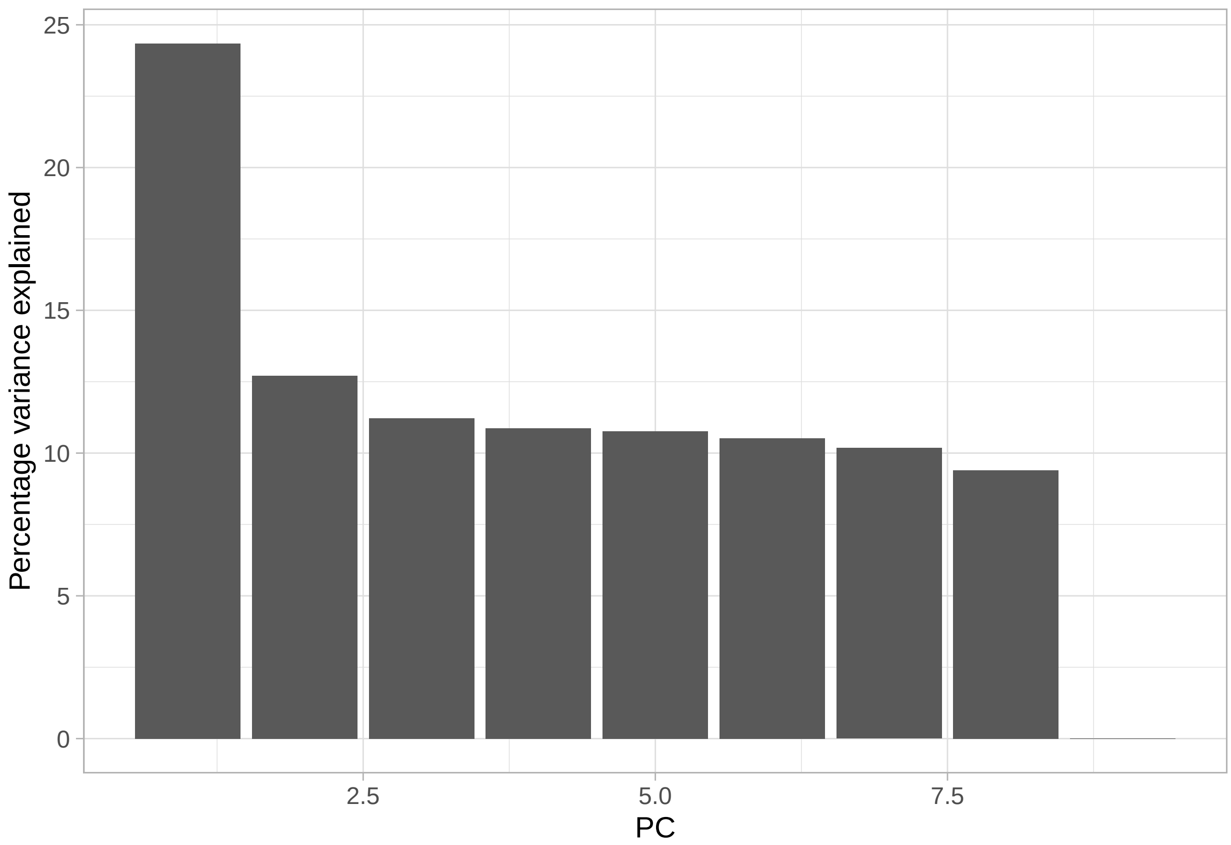
```
# sort out the pca data
# remove nuisance column
pca <- pca[,-1]
# set names
names(pca)[1] <- "ind"
names(pca)[2:ncol(pca)] <- paste0("PC", 1:(ncol(pca)-1))
```

```r
# sort out the individual species and pops
# region
region <- rep(NA, length(pca$ind))

region[grep("FMNH346401", pca$ind)] <- "East"
region[grep("FMNH355643", pca$ind)] <- "East"
region[grep("FMNH385038", pca$ind)] <- "East"
region[grep("FMNH443853", pca$ind)] <- "West"
region[grep("FMNH443854", pca$ind)] <- "West"
region[grep("FMNH443856", pca$ind)] <- "West"
region[grep("FMNH450510", pca$ind)] <- "West"
region[grep("FMNH450511", pca$ind)] <- "West"
region[grep("FMNH450512", pca$ind)] <- "West"

# location
loc <- rep(NA, length(pca$ind))

loc[grep("FMNH346401", pca$ind)] <- "Kibira-Nyungwe"
loc[grep("FMNH355643", pca$ind)] <- "Rwenzori"
loc[grep("FMNH385038", pca$ind)] <- "Bwindi"
loc[grep("FMNH443853", pca$ind)] <- "Kahuzi-Biega"
loc[grep("FMNH443854", pca$ind)] <- "Kahuzi-Biega"
loc[grep("FMNH443856", pca$ind)] <- "Kahuzi-Biega"
loc[grep("FMNH450510", pca$ind)] <- "Kabogo"
loc[grep("FMNH450511", pca$ind)] <- "Kabogo"
loc[grep("FMNH450512", pca$ind)] <- "Kabogo"

# from tutorial - species_location
# combine - if you want to plot each in different colours
# spp_loc <- paste0(spp, "_", loc)
```

```r
# remake data.frame
pca <- as_tibble(data.frame(pca, loc, region))
```

```r
# first convert to percentage variance explained
pve <- data.frame(PC = 1:length(eigenval), pve = eigenval/sum(eigenval)*100)
```

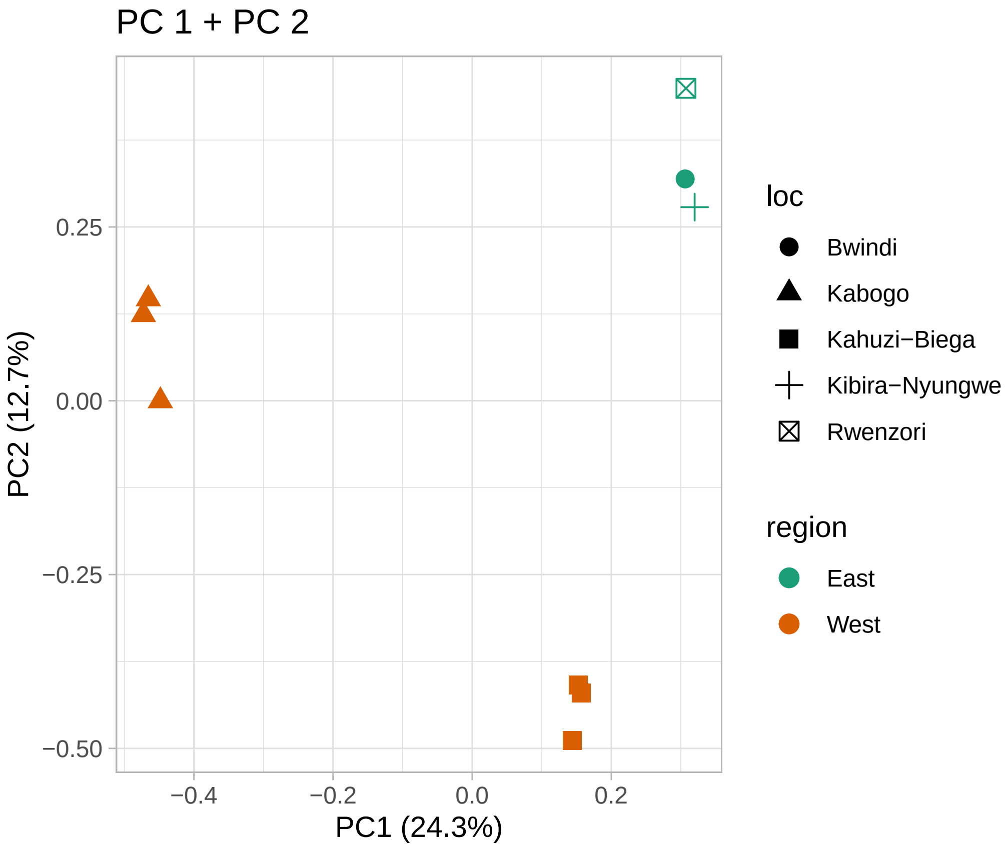
```r
# make plot
a <- ggplot(pve, aes(PC, pve)) + geom_bar(stat = "identity")
a + ylab("Percentage variance explained") + theme_light()
```



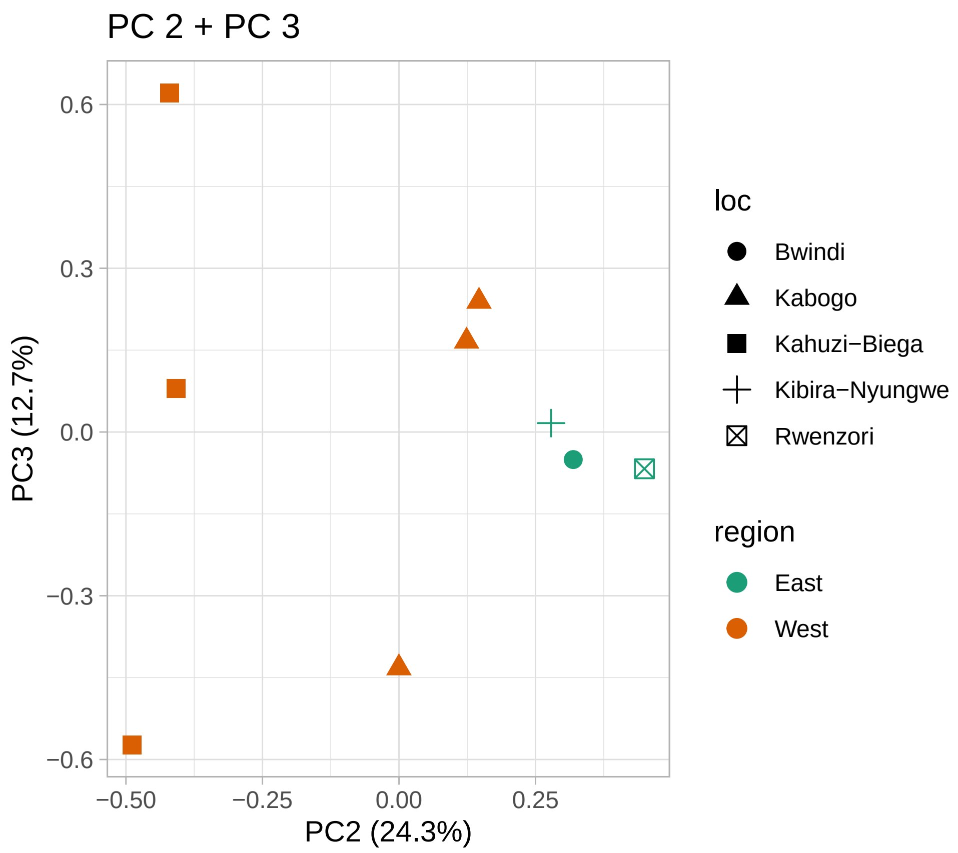
```r
# calculate the cumulative sum of the percentage variance explained
cumsum(pve$pve)
```

```
## [1]  24.34940  37.05314  48.27125  59.13935  69.90116  80.42408  90.60210
## [8] 100.00000 100.00000
```

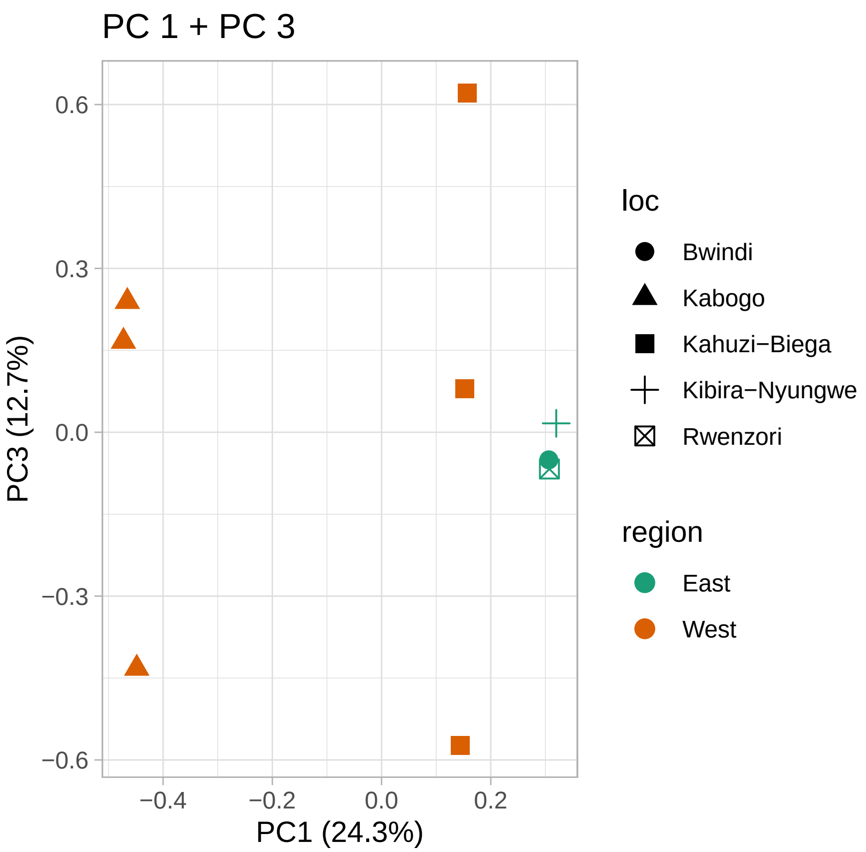
```r
# plot pca
# color brewer colorblind friendly palette
b <- ggplot(pca, aes(PC1, PC2, col = region, shape = loc)) +
  geom_point(size = 3) +
  scale_colour_manual(values = c("#1b9e77", "#d95f02", "#7570b3")) +
  coord_equal() +
  theme_light() +
  xlab(paste0("PC1 (", signif(pve$pve[1], 3), "%)")) +
  ylab(paste0("PC2 (", signif(pve$pve[2], 3), "%)")) +
  ggtitle("PC 1 + PC 2")

print(b)
```



```r
# plot pca
# color brewer colorblind friendly palette
b <- ggplot(pca, aes(PC2, PC3, col = region, shape = loc)) +
  geom_point(size = 3) +
  scale_colour_manual(values = c("#1b9e77", "#d95f02", "#7570b3")) +
  coord_equal() +
  theme_light() +
  xlab(paste0("PC2 (", signif(pve$pve[1], 3), "%)")) +
  ylab(paste0("PC3 (", signif(pve$pve[2], 3), "%)")) +
  ggtitle("PC 2 + PC 3")

print(b)
```



```
# plot pca
# color brewer colorblind friendly palette
b <- ggplot(pca, aes(PC1, PC3, col = region, shape = loc)) +
  geom_point(size = 3) +
  scale_colour_manual(values = c("#1b9e77", "#d95f02", "#7570b3")) +
  coord_equal() +
  theme_light() +
  xlab(paste0("PC1 (", signif(pve$pve[1], 3), "%)")) +
  ylab(paste0("PC3 (", signif(pve$pve[2], 3), "%)")) +
  ggtitle("PC 1 + PC 3")

print(b)
```



ADMIXTURE Results

Directions from this site.

```
samplelist <- read_tsv("~/Documents/GitHub/rift_birds/chamaetylas.txt",col_names = "sample")
```

```
## Rows: 9 Columns: 1
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (1): sample
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
read_delim("~/Documents/06_plink/chamaetylas_nothin_plink.5.Q",col_names = paste0("Q",seq(1:6)),
           delim = " ")
```

```
## Rows: 9 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: " "
## dbl (5): Q1, Q2, Q3, Q4, Q5
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

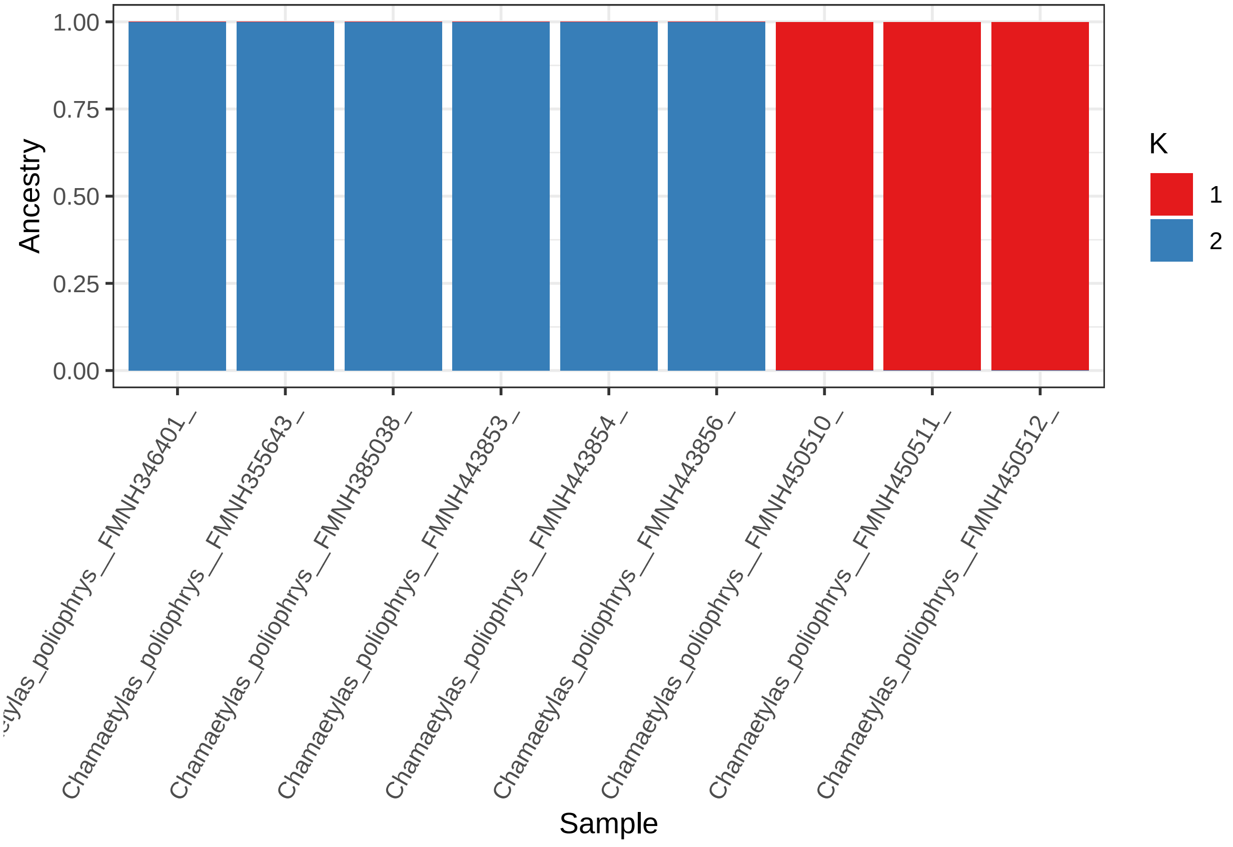
```
## # A tibble: 9 x 5
##        Q1       Q2      Q3      Q4      Q5
##     <dbl>    <dbl>   <dbl>   <dbl>   <dbl>
## 1 0.00001 1.00     0.00001 0.00001 0.00001
## 2 1.00    0.00001  0.00001 0.00001 0.00001
## 3 0.00001 1.00     0.00001 0.00001 0.00001
## 4 0.00001 0.000012 0.00001 0.00001 1.00
## 5 0.00001 0.00001  0.00001 0.00001 1.00
## 6 0.00001 0.00001  0.00001 0.00001 1.00
## 7 0.00001 0.00001  1.00    0.00001 0.00001
## 8 0.00001 0.00001  1.00    0.00001 0.00001
## 9 0.00001 0.00001  0.00001 1.00    0.00001
```

```
cv <- c(1.24030, 2.46934, 2.49640,
        2.78206, 0.92918, 0.35207)

k <- 1:6

cv_tab <- cbind(cv,k) %>% as.data.frame()

ggplot(cv_tab,aes(x = k, y = cv)) +
  geom_line()
```

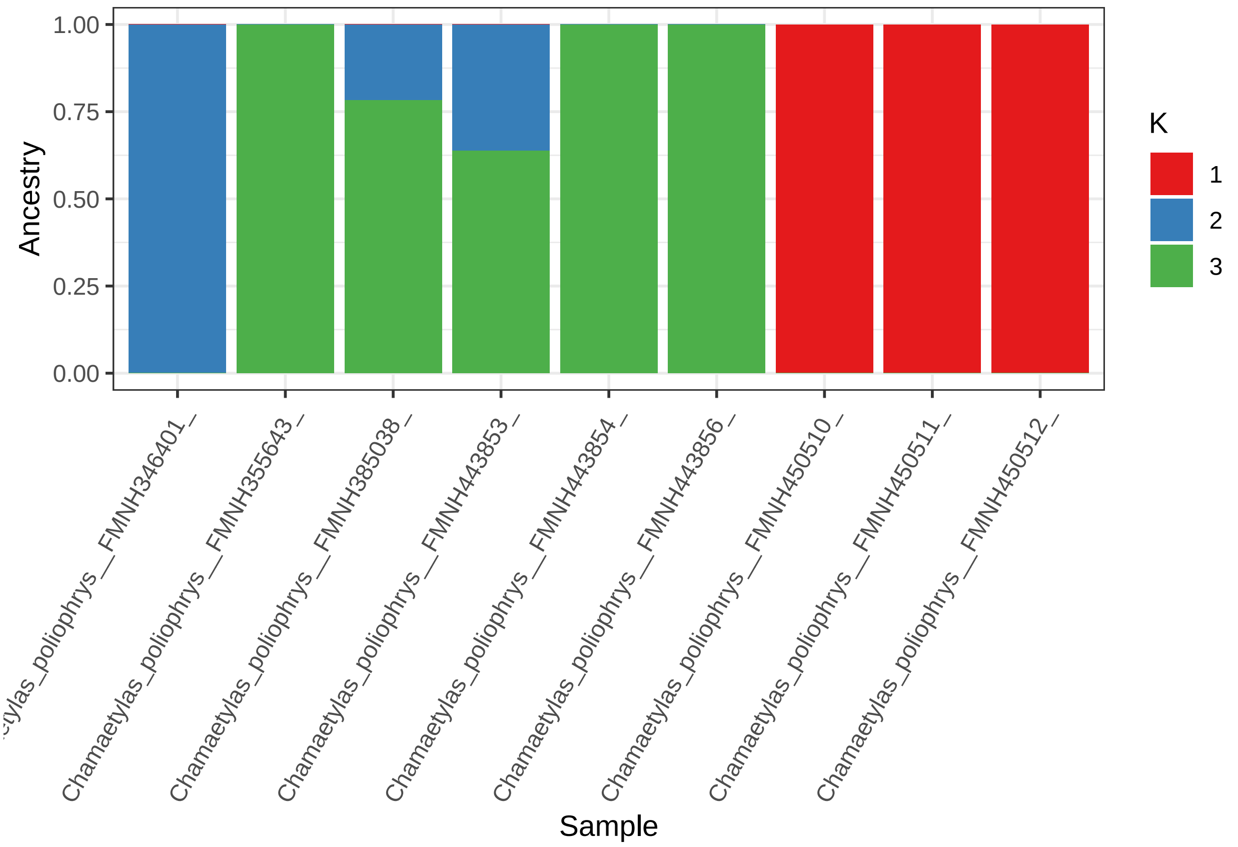
```
all_data <- tibble(sample=character(),
                   k=numeric(),
                   Q=character(),
                   value=numeric())

for (k in c(2:5)){
  data <- read_delim(paste0("~/Documents/06_plink/chamaetylas_nothin_plink.",k,".Q"),
                  col_names = paste0("Q",seq(1:k)),
                  delim=" ")
  data$sample <- samplelist$sample
  data$k <- k

  #This step converts from wide to long.
  data %>% gather(Q, value, -sample,-k) -> data
  all_data <- rbind(all_data,data)
}
```

```
## Rows: 9 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: " "
## dbl (2): Q1, Q2
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 9 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: " "
## dbl (3): Q1, Q2, Q3
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 9 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: " "
## dbl (4): Q1, Q2, Q3, Q4
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 9 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: " "
## dbl (5): Q1, Q2, Q3, Q4, Q5
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

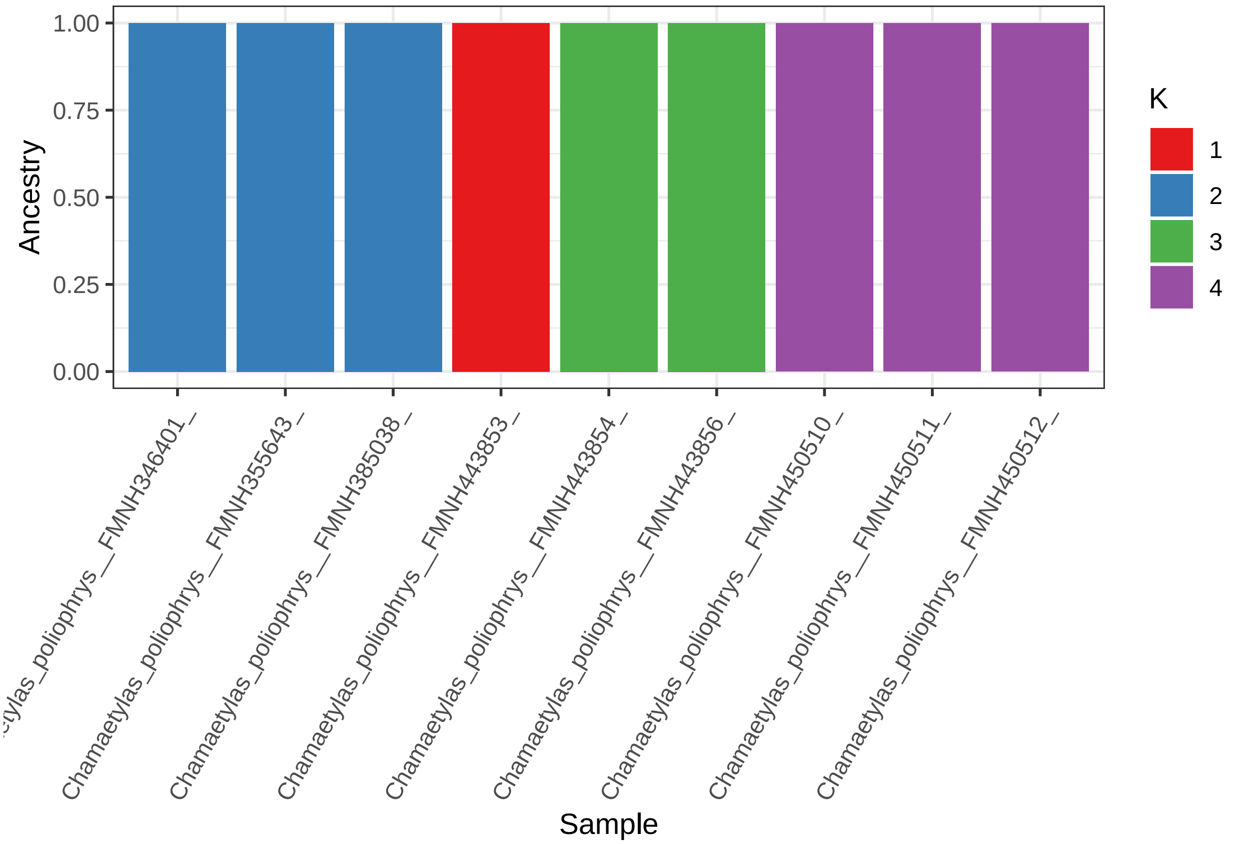
```
all_data %>%
  filter(k == 2) %>%
  ggplot(.,aes(x=sample,y=value,fill=factor(Q))) +
  geom_bar(stat="identity",position="stack") +
  xlab("Sample") + ylab("Ancestry") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  scale_fill_brewer(palette="Set1",name="K",
                    labels=c("1","2"))
```



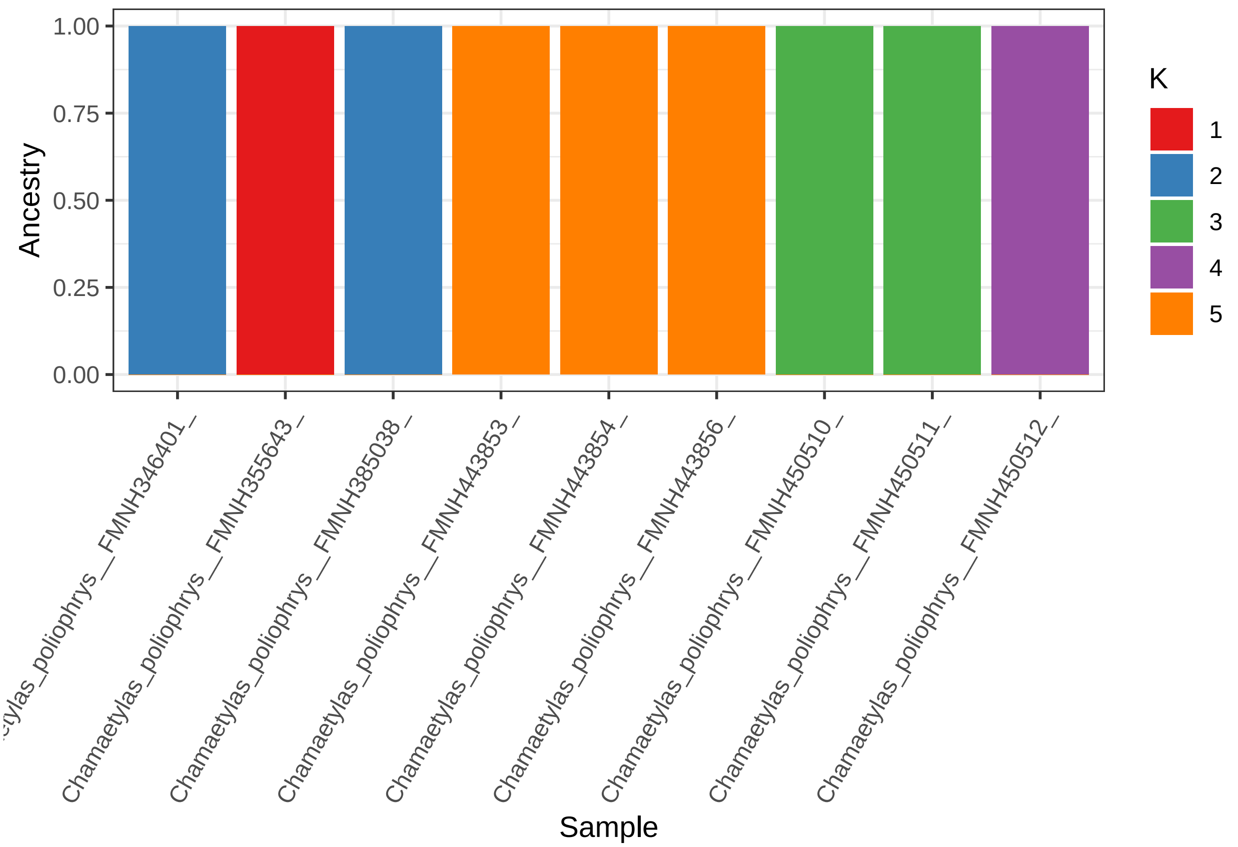
```
all_data %>%
  filter(k == 3) %>%
  ggplot(.,aes(x=sample,y=value,fill=factor(Q))) +
  geom_bar(stat="identity",position="stack") +
  xlab("Sample") + ylab("Ancestry") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  scale_fill_brewer(palette="Set1",name="K",
                    labels=c("1","2", "3"))
```



```
all_data %>%
  filter(k == 4) %>%
  ggplot(.,aes(x=sample,y=value,fill=factor(Q))) +
  geom_bar(stat="identity",position="stack") +
  xlab("Sample") + ylab("Ancestry") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  scale_fill_brewer(palette="Set1",name="K",
                    labels=c("1","2", "3", "4"))
```



```r
all_data %>%
  filter(k == 5) %>%
  ggplot(.,aes(x=sample,y=value,fill=factor(Q))) +
  geom_bar(stat="identity",position="stack") +
  xlab("Sample") + ylab("Ancestry") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  scale_fill_brewer(palette="Set1",name="K",
                    labels=c("1","2", "3", "4", "5"))
```



1.00
0.75
0.50
0.25
0.00
Ancestry
K
1
2
3
4
5
ɛtylas_poliophrys__FMNH346401_
Chamaetylas_poliophrys__FMNH355643_
Chamaetylas_poliophrys__FMNH385038_
Chamaetylas_poliophrys__FMNH443853_
Chamaetylas_poliophrys__FMNH443854_
Chamaetylas_poliophrys__FMNH443856_
Chamaetylas_poliophrys__FMNH450510_
Chamaetylas_poliophrys__FMNH450511_
Chamaetylas_poliophrys__FMNH450512_
Sample