

# Cod Inversions

KE Lotterhos

2024-03-18

`#install_github("jokergoo/ComplexHeatmap")` Didn't install

## Load packages and data

```
library(devtools)
```

```
## Loading required package: usethis
```

```
library(bigstatsr)
```

```
library(bigsnp)
```

```
library(ggplot2)
```

```
library(OutFLANK)
```

```
## Loading required package: qvalue
```

```
library(vegan)
```

```
## Loading required package: permute
```

```
##
```

```
## Attaching package: 'permute'
```

```
## The following object is masked from 'package:devtools':
```

```
##
```

```
##      check
```

```
## Loading required package: lattice
```

```
## This is vegan 2.6-6.1
```

```
library(pheatmap)
```

```
library(viridisLite)
```

```
library(heatmap3)
```

```
NCORES <- nb_cores()
```

```
datadir <- "/Users/lotterhos/Library/CloudStorage/GoogleDrive-k.lotterhos@gmail.com/.shortcut-targets-by-"
```

```
ggtheme <- theme_bw() + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(), p
```

```
sessionInfo()
```

```
## R version 4.4.0 (2024-04-24)
```

```
## Platform: x86_64-apple-darwin20
```

```
## Running under: macOS Sonoma 14.5
```

```
##
```

```
## Matrix products: default
```

```
## BLAS: /Library/Frameworks/R.framework/Versions/4.4-x86_64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.4-x86_64/Resources/lib/libRlapack.dylib; LAPACK
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] heatmap3_1.1.9    viridisLite_0.4.2 pheatmap_1.0.12    vegan_2.6-6.1
## [5] lattice_0.22-6    permute_0.9-7      OutFLANK_0.2       qvalue_2.36.0
## [9] ggplot2_3.5.1     bigsnpr_1.12.2     bigstatsr_1.5.12   devtools_2.4.5
## [13] usethis_2.2.3
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.2.1  dplyr_1.1.4        fastmap_1.2.0      bigassertr_0.1.6
## [5] flock_0.7          promises_1.3.0      digest_0.6.35      mime_0.12
## [9] lifecycle_1.0.4   cluster_2.1.6       ellipsis_0.3.2     magrittr_2.0.3
## [13] compiler_4.4.0    rlang_1.1.3         rngtools_1.5.2     tools_4.4.0
## [17] utf8_1.2.4        yaml_2.3.8          data.table_1.15.4  knitr_1.46
## [21] doRNG_1.8.6       htmlwidgets_1.6.4   pkgbuild_1.4.4     plyr_1.8.9
## [25] RColorBrewer_1.1-3 pkgload_1.3.4        miniUI_0.1.1.1     bigsparser_0.6.1
## [29] withr_3.0.0       purrr_1.0.2         grid_4.4.0         fansi_1.0.6
## [33] urlchecker_1.0.1  profvis_0.3.8       xtable_1.8-4       colorspace_2.1-0
## [37] fastcluster_1.2.6 scales_1.3.0         iterators_1.0.14   MASS_7.3-60.2
## [41] cli_3.6.2         rmarkdown_2.27      generics_0.1.3     remotes_2.5.0
## [45] rstudioapi_0.16.0 reshape2_1.4.4      sessioninfo_1.2.2  cachem_1.1.0
## [49] stringr_1.5.1     splines_4.4.0       parallel_4.4.0     vctrs_0.6.5
## [53] Matrix_1.7-0      foreach_1.5.2       bigparallelr_0.3.2 parallelly_1.37.1
## [57] glue_1.7.0        codetools_0.2-20    cowplot_1.1.3      stringi_1.8.4
## [61] gtable_0.3.5      later_1.3.2         munsell_0.5.1      tibble_3.2.1
## [65] pillar_1.9.0      htmltools_0.5.8.1   R6_2.5.1           doParallel_1.0.17
## [69] evaluate_0.23     shiny_1.8.1.1       memoise_2.0.1      httpuv_1.6.15
## [73] Rcpp_1.0.12       nlme_3.1-164        mgcv_1.9-1         xfun_0.44
## [77] fs_1.6.4          pkgconfig_2.0.3
```

## Read in files

```
breakpoints <- read.csv("../outputs/1-breakpoints.csv") # breakpoints output from script 1
head(breakpoints)
```

```
##      LG LGstartIndex LGendIndex      chrom LGstartPos LGendPos
## 1  LG1          33090      104012 NC_044048.1    11299038 28292263
## 2  LG2          111433      132871 NC_044049.1      481246 4466546
## 3  LG7          648745      696592 NC_044054.1    16828712 26336501
## 4 LG12          1098771      1141346 NC_044059.1     1809214 13690550
##      LGstartPos_minus1MB_index LGendPos_plus1MB_index
## 1                          29750                      106154
```

```
## 2          109720          135332
## 3          645345          700639
## 4          1096496          1142816
```

```
samp_full <- readRDS("../outputs/1-Samples.rds") # load sample info output from script 1
#load it as an RDS to keep the population order in analysis
str(samp_full)
```

```
## 'data.frame': 294 obs. of 15 variables:
## $ samp_full : int 16216 16220 16224 16225 16229 16231 16232 16233 16234 16236 ...
## $ Pop : Factor w/ 9 levels "Pop1","Pop2",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ PopID : Ord.factor w/ 9 levels "GoM.Mass.Olive.Winter"<...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Region : Factor w/ 2 levels "GoM","Iceland": 1 1 1 1 1 1 1 1 1 1 ...
## $ Ecotype : Factor w/ 2 levels "1-Nearshore",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ label : chr "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216" "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220" ...
## $ Region_color : chr "blue4" "blue4" "blue4" "blue4" ...
## $ Ecotype_color : chr "darkcyan" "darkcyan" "darkcyan" "darkcyan" ...
## $ PopID_color : chr "#006400FF" "#006400FF" "#006400FF" "#006400FF" ...
## $ Gsub_PC1 : num -0.0589 -0.067 -0.0651 -0.0614 -0.0538 ...
## $ Gsub_PC2 : num 0.02733 0.00134 0.00245 0.00487 0.0154 ...
## $ G_AllInv_PC1 : num 0.0353 0.065 0.0613 0.0515 0.0319 ...
## $ G_AllInv_PC2 : num -0.00791 -0.04903 -0.05377 -0.03753 0.01956 ...
## $ InvHclustOrder: int 146 258 230 147 172 238 243 142 164 156 ...
## $ InvHclustK5 : num 1 1 2 1 1 1 1 1 1 1 ...
```

```
pops <- read.table("../outputs/1-Pops.txt", header=TRUE)
head(pops)
```

```
## pop.names pop.IDs pop.colors region ecotype
## 1 Pop1 GoM.Mass.Olive.Winter #006400FF GoM Nearshore
## 2 Pop2 GoM.Mass.Red #FA8072FF GoM Nearshore
## 3 Pop3 GoM.Mass.Olive.Unknown #CD9B1D80 GoM Nearshore
## 4 Pop4 GoM.Cashes.Olive #006400BF GoM Nearshore
## 5 Pop5 GoM.Cashes.Red #FA8072BF GoM Nearshore
## 6 Pop6 Ice.Off.SW #9400D3FF Iceland Offshore
## pop.IDs2 pop2.colors
## 1 GoM Nearshore Olive - Winter Spawner #006400FF
## 2 GoM Nearshore Red #FA8072FF
## 3 GoM Nearshore Olive #00640080
## 4 GoM Nearshore Olive #00640080
## 5 GoM Nearshore Red #FA8072FF
## 6 Iceland Offshore #9400D3BF
```

```
str(pops)
```

```
## 'data.frame': 9 obs. of 7 variables:
## $ pop.names : chr "Pop1" "Pop2" "Pop3" "Pop4" ...
## $ pop.IDs : chr "GoM.Mass.Olive.Winter" "GoM.Mass.Red" "GoM.Mass.Olive.Unknown" "GoM.Cashes.Olive" ...
## $ pop.colors : chr "#006400FF" "#FA8072FF" "#CD9B1D80" "#006400BF" ...
## $ region : chr "GoM" "GoM" "GoM" "GoM" ...
## $ ecotype : chr "Nearshore" "Nearshore" "Nearshore" "Nearshore" ...
## $ pop.IDs2 : chr "GoM Nearshore Olive - Winter Spawner" "GoM Nearshore Red" "GoM Nearshore Olive" ...
## $ pop2.colors: chr "#006400FF" "#FA8072FF" "#00640080" "#00640080" ...
```

```
LGtext <- breakpoints$LG
LGtext
```

```
## [1] "LG1" "LG2" "LG7" "LG12"
```

## Loop through inversions

The following code loops through each inversion and does the same analysis. However, note there are some `if()` statements and `seeds` that were manually customized for nuances associated with each inversion, and that “hone-in” on specific regions of interest.

```
for (i in 1:4){

G_LGX <- readRDS(paste0("../outputs/1-G_",LGtext[i],".rds"))

pos <- colnames(G_LGX)
samp <- rownames(G_LGX)

## Step 1: Subset data to the inversion on Chromosome X and perform PCA. ####

Gx_snp <- add_code256(big_copy(G_LGX ,type="raw"),code=bigsnpr::CODE_012)
str(Gx_snp)
set.seed(123461356+i)
svd <- big_randomSVD(Gx_snp,
                      snp_scaleBinom(),
                      ncores = NCORES,
                      k=2)

samp_full[,paste0(LGtext[i],"_PC1")] <- svd$u[,1]
samp_full[,paste0(LGtext[i],"_PC2")] <- svd$u[,2]

str(samp_full)

g<- ggplot(samp_full[order(samp_full$Region,samp_full$Ecotype, decreasing=TRUE),]) + geom_point(aes(x=
    amount=0.001),
    y=jitter(get(paste0(LGtext[i],"_PC2")),
    amount=0.001),
    colour=PopID,
    pch=Ecotype,
    size=Ecotype)) +
  scale_colour_manual(values=pops$pop.colors) +
  xlab(paste0(LGtext[i],"_PC1")) +
  ylab(paste0(LGtext[i],"_PC2")) + ggtheme
ggsave(filename=paste0("../figures/", LGtext[i],"_PCA.pdf"), g, width=7, height=7)

# Step 2: Identify clusters in the PCA ####

# Note: kmeans sometimes does not find the correct # clusters, or it switches the order. Make sure to
if(i == 1){set.seed(9823487); n=5} # 5 clusters
if(i == 2){set.seed(6345219); n=6}
if(i == 3){set.seed(2234677); n=6} # LG7 seed 2234677
if(i == 4){set.seed(4923586); n=6}
clust <- kmeans(svd$u, n, nstart=20)
```

```

samp_full[,paste0(LGtext[i], "cluster")] <- as.character(unlist(clust$cluster))

head(samp_full)

g <- ggplot(samp_full) +
  geom_point(aes(x=get(paste0(LGtext[i], "_PC1")), y=get(paste0(LGtext[i], "_PC2")),
    col = get(paste0(LGtext[i], "cluster")))) +
  scale_colour_discrete() + ggtheme
ggsave(paste0("../figures/", LGtext[i], "_PCA_clusterIDs.pdf"), g, width=7, height=7)

# Step 3: Name inversion clusters ####

dim(G_LGX)
rownames(G_LGX) <- paste(samp_full[,paste0(LGtext[i], "cluster")],
  rownames(G_LGX),
  sep="_"
)

# just use a subset of 1000 SNPs to view heatmap
# order individuals by their cluster
# put SNPS in order after sampling with `sort`
G_heat <- G_LGX[order(samp_full[,paste0(LGtext[i],
  "cluster")]),
  sort(sample( 1:ncol(G_LGX), 1000))]

pdf(paste0("../figures/", LGtext[i], "_heatmap_1000KrandomSNPs_orderedByPCAcluster.pdf"), width=7, height=7)
par(oma=c(0,0,0,5))
heatmap(G_heat,
  scale="none",
  Rowv = NA,
  Colv = NA,
  cexRow = 0.3,
  cexCol=0.1)
dev.off()

# Step 4: How pops map to clusters ####
print(c("Mapping of samples to PCA cluster for", LGtext[i]))

print(table(samp_full[,paste0(LGtext[i], "cluster")],
  samp_full$PopID))

## Step 4: Manually map clusters to karyotype
# 0 indicates the reference allele and 1 indicates the alternative allele

# create a new variable from "cluster", which we will rename
# the levels for below
samp_full[,paste0(LGtext[i], "cluster_name")] <- factor(samp_full[,paste0(LGtext[i], "cluster")])

if (i == 1){

```

```

### LG1 cluster descriptions
# Cluster 1 is homozygous Alternate arrangement Iceland - "ALT-Iceland"
# Cluster 2 is heterokaryotype Iceland - "HET-Iceland"
# Cluster 3 is homozygous reference Iceland - "REF-Iceland"
# Cluster 4 is heterozygous GOM - "HET-GoM"
# Cluster 5 is homozygous reference GoM - "REF-GoM"
levels(samp_full[,paste0(LGtext[i], "cluster_name")]) <- c(
  "ALT-Iceland",
  "HET-Iceland",
  "REF-Iceland",
  "HET-GoM",
  "REF-GoM"
)
}
if (i == 2){
  ### LG2 cluster descriptions ####
  # Cluster 1 is homozygous for Alternate arrangement from GoM - "Gom-ALT"
  # Cluster 2 is heterokaryotypes from Iceland - "Iceland-HET"
  # Cluster 3 is homozygous for the reference arrangement from GoM - "GoM-REF"
  # Cluster 4 is homozygous for the reference arrangement from Iceland - "Iceland-REF"
  # Cluster 5 is homozygous for the alternate arrangement from Iceland - "Iceland-ALT" - only nearshore
  # Cluster 6 is heterokaryotypes from GoM - "GoM-HET"

  levels(samp_full[,paste0(LGtext[i], "cluster_name")]) <- c(
    "ALT-GoM",
    "HET-Iceland",
    "REF-GoM",
    "REF-Iceland",
    "ALT-Iceland",
    "HET-GoM"
  )
}

if (i == 3){
  ### LG7 Cluster descriptions ####
  # * Cluster 1 is heterokaryotypes from Iceland - "Iceland-HET"
  # * Cluster 2 is homozygous for the alternate arrangement from GoM - "GoM-ALT"
  # * Cluster 3 is heterokaryotypes from GoM - "GoM-HET"
  # * Cluster 4 is homozygous for the alternate arrangement from Iceland - "Iceland-ALT"
  # * Cluster 5 is homozygous for the reference arrangement from GoM - "GoM-REF"
  # * Cluster 6 is homozygous for the reference arrangement from Iceland - "Iceland-REF" (18 samples)

  levels(samp_full[,paste0(LGtext[i], "cluster_name")]) <- c(
    "HET-Iceland",
    "ALT-GoM",
    "HET-GoM",
    "ALT-Iceland",
    "REF-GoM",
    "REF-Iceland"
  )
}

if (i == 4){

```

```

#### LG12 Cluster descriptions ####
# * Cluster 1 is heterokaryotypes from Iceland - "HET-Iceland"
# * Cluster 2 is heterokaryotypes from Maine - "HET-GoM"
# * Cluster 3 is homozygous for the reference arrangement from Iceland "REF-Iceland"
# * Cluster 4 is homozygous for the alternate arrangement from Maine - "ALT-GoM"
# * Cluster 5 is homozygous for the reference arrangement from GoM - "REF-GoM"
# * Cluster 6 is homozygous for the alternate arrangement from Iceland - "ALT-ICELAND"

levels(samp_full[,paste0(LGtext[i], "cluster_name")]) <- c(
  "HET-Iceland",
  "HET-GoM",
  "REF-Iceland",
  "ALT-GoM",
  "REF-GoM",
  "ALT-Iceland"
)
}

#### sanity checks for cluster mapping####

print(c("Mapping of samples to named PCA cluster for", LGtext[i]))

print(table(samp_full[,paste0(LGtext[i], "cluster_name")],
  samp_full$PopID))

print(table(samp_full[,paste0(LGtext[i], "cluster")], samp_full[,paste0(LGtext[i], "cluster_name")]))

# Update rownames for the Genotype matrix
rownames(G_LGX) <- paste(samp_full[,paste0(LGtext[i], "cluster_name")], rownames(G_LGX), sep="_")

G_heat <- G_LGX[order(samp_full[,paste0(LGtext[i],
  "cluster")]),
  sort(sample( 1:ncol(G_LGX), 1000)))]

# One more sanity check with new row labels
pdf(paste0("../figures/", LGtext[i], "_heatmap_1000KrandomSNPs_orderedByPCAcluster_named.pdf"), width=7, height=7)
par(oma=c(0,0,0,5))
heatmap(G_heat,
  scale="none",
  Rowv = NA,
  Colv = NA,
  cexRow = 0.3,
  cexCol=0.1)
dev.off()

# Final PCA with clusters ID'd ####
g<- ggplot(samp_full[order(samp_full$Region, samp_full$Ecotype, decreasing=TRUE),]) + geom_point(aes(x=
  amount=0.001),
  y=jitter(get(paste0(LGtext[i], "_PC2")),
    amount=0.001),
  colour=PopID,

```

```

        pch=get(paste0(LGtext[i], "cluster_name")),
        size=Ecotype)) +
scale_colour_manual(values=pops$pop.colors) +
xlab(paste0(LGtext[i], "_PC1")) +
ylab(paste0(LGtext[i], "_PC2")) + ggtheme

ggsave(filename=paste0("../figures/", LGtext[i], "_PCA_final.pdf"), g, width=7, height=7)

# Step 5: Identify FST outliers and create heatmaps ####

# For a specific arrangement (reference homokaryotypes only or alternate homokaryotypes only), how is it
# Identify outliers between GOM and Iceland within the same arrangement

### Outliers between Reference arrangements ###
dim(G_LGX)
index_REF <- grep("REF", samp_full[, paste0(LGtext[i], "cluster_name")])

print(c("Outlier sample sizes for reference arrangement in", LGtext[i]))
t = table(samp_full[index_REF, paste0(LGtext[i], "cluster_name")],
          samp_full$PopID[index_REF])
print(t)
print(rowSums(t))

my_fst_REF <- MakeDiploidFSTMat(G_LGX[index_REF, ],
                               locusNames = colnames(G_LGX),
                               popNames = samp_full[index_REF, paste0(LGtext[i], "cluster_name")])
head(my_fst_REF)

mutts <- data.frame(chrom_pos = colnames(G_LGX))

mutts[, paste0(LGtext[i], "FST_betweenREFkaryotypes")] <- my_fst_REF$FST
mutts[, paste0(LGtext[i], "FST_betweenREFkaryotypes_outlier")] <- my_fst_REF$FST > 0.6

head(my_fst_REF)

head(mutts)
mutts$pos <- as.numeric(unlist(lapply(mutts$chrom_pos, function(x){strsplit(x, split="__")[[1]][2]})))

# SNP summary - distance between SNPs
snp_dist <- mutts$pos[2:(length(mutts$pos))] - mutts$pos[1:(length(mutts$pos)-1)]
summary(snp_dist)
gaps <- which(snp_dist > quantile(snp_dist, prob=0.999))

### Outliers between Alternate arrangements ###

# get indexes of ALT homokaryotypes
index_ALT <- grep("ALT", samp_full[, paste0(LGtext[i], "cluster_name")])

print(c("Outlier sample sizes for alternate arrangement in", LGtext[i]))
t = table(samp_full[index_ALT, paste0(LGtext[i], "cluster_name")],
          samp_full$PopID[index_ALT])

```



```

print(t )
print(rowSums(t))
# for LG7 note low sample size - 5 individuals in GoM

if (i == 1){
  # we don't run this on the 1st LG because
  # the altnerate is not present in GoM
  my_fst_ALT <- NA
  muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes")] <- NA
  muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier")] <-NA
}else{
  my_fst_ALT <- MakeDiploidFSTMat(G_LGX[index_ALT,],
    locusNames = colnames(G_LGX),
    popNames = samp_full[index_ALT,paste0(LGtext[i], "cluster_name")])

head(my_fst_ALT)

muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes")] <- my_fst_ALT$FST
muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier")] <- my_fst_ALT$FST > 0.6
}

### Plot the FST values ###
pdf(paste0("../figures/", LGtext[i], "_FST_betweenREFandALTkaryotypes.pdf"), width=20, height=6)
options(scipen = 999)
plot(muts$pos, muts[,paste0(LGtext[i], "FST_betweenREFkaryotypes")], pch=19, col=adjustcolor("blue", 0.5),
  ylab="FST", ylim=c(0,1), main=paste(LGtext[i], "FST between REF karyotypes blue, ALT orange"),
  xaxp =c(min(muts$pos), max(muts$pos), n=100), bty="l")
points(muts$pos, muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes")], pch=19, col=adjustcolor("orange", 0.5))
abline(h=0.6, col="grey", lwd=4)
#abline(v=muts$pos[gaps], col="cornflowerblue")
#abline(v=muts$pos[gaps+1], col="cornflowerblue")
abline(v=breakpoints$LGstartPos[i], col="cornflowerblue", lwd=4)
abline(v=breakpoints$LGendPos[i], col="cornflowerblue", lwd=4)
dev.off()

### Number of FST outliers ###
print(LGtext[i])
print("Number of loci in REF outliers")
print(
sum(muts[,paste0(LGtext[i], "FST_betweenREFkaryotypes_outlier")], na.rm=TRUE)
)

print("Number of loci in ALT outliers")
print(sum(
muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier")], na.rm=TRUE)
)

print("Number of loci in both REF and ALT outliers")
print(
sum(muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier")] &
  muts[,paste0(LGtext[i], "FST_betweenREFkaryotypes_outlier")],
  na.rm=TRUE)
)

```

```

    )
  )

  print("Total number of loci in LG window")
  print(nrow(muts))

# Step 6: heatmaps for FST outliers ####

### Colors for heatmap ###
  muts[,paste0(LGtext[i], "FST_betweenREFkaryotypes_outlier_color")] = "white" # non-outliers are white

  muts[which(muts[,paste0(LGtext[i], "FST_betweenREFkaryotypes_outlier")]
    ), # get indexes of outliers
    paste0(LGtext[i], "FST_betweenREFkaryotypes_outlier_color")
  ] <- "darkblue" # REF outliers are darkblue

  muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier_color")] = "white" # non-outliers are white
  muts[which(muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier")]
    ), # get indexes of outliers
    paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier_color")] <- "darkred" # REF outliers are darkred

### subset of loci for heatmap - outliers and some random loci ###

#### get the indexes (with reference to this LG) of the outliers ####
  heat_outliers <- sort(unique(which(
    muts[,paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier")] |
    muts[,paste0(LGtext[i], "FST_betweenREFkaryotypes_outlier")]
  )))

  # Run a quick heatmap just to get a clustering order based on outlier clustering for the subsequent p
  hc <- hclust(dist(G_LGX[,heat_outliers]))
  #do not run str() on this object
  samp_full[,paste0(LGtext[i], "heatmapFSToutliersOrder")] <- (hc$order)

  # final ordering will be on Region, then cluster name, then the order based on outlier clustering
  heatOrder <- order(samp_full$Region,
    samp_full[,paste0(LGtext[i], "cluster_name")],
    samp_full[,paste0(LGtext[i], "heatmapFSToutliersOrder")])

### Heatmap function ###
  make_heatmap <- function(mut_sub, heatOrder, filename){
    # mut_sub is the indexes of the mutations to use in the heatmap
    # heatOrder is the index and order to plot the individuals in rows

    G_LG_sub2 <- G_LGX[heatOrder, mut_sub]

    length(mut_sub)

    # create annotation table for heatmap columns (muts)
    muts_ann <- cbind(

```

```

    REF_outlier = muts[mut_sub,paste0(LGtext[i], "FST_betweenREFkaryotypes_outlier_color")],
    ALT_outlier = muts[mut_sub,paste0(LGtext[i], "FST_betweenALTkaryotypes_outlier_color")]
  )
  head(muts_ann)

  # for annotation in heatmap, the rownames of annotation table has to have the same colnames as the ma
  rownames(muts_ann) <- colnames(G_LG_sub2)

  # create annotation table for heatmap rows (samples)
  samp_ann <- cbind(
    Region = samp_full$Region_color[heatOrder],
    Pop = samp_full$PopID_color[heatOrder],
    Ecotype = samp_full$Ecotype_color[heatOrder]
  )

  # for annotation in heatmap, the rownames of the annotation table has to have the same rownames as th
  rownames(samp_ann) <- rownames(G_LG_sub2)
  head(samp_ann)
  dim(samp_ann)

  heatmap_colors <- colorRampPalette(c("cornflowerblue", "grey", "firebrick3"))(3)

  pdf(paste0("../figures/",filename), width=9, height=9)
  par(oma=c(0,0,0, 10))
  heatmap3(G_LG_sub2,
    Rowv=NA,
    Colv = NA,
    showColDendro = F,
    showRowDendro = F,
    scale = "none",
    ColSideColors=muts_ann,
    RowSideColors = samp_ann,
    col = heatmap_colors,
    cexCol = 0.1,
    cexRow=0.3,
    ColSideWidth = 2
  )

  dev.off()
} # end function

## Make full heatmap for LG7 with 4000 random loci, 100 loci at each end, and outliers
## indexes are used here
mut_sub <- sort(unique(c(1:100,
                        (ncol(G_LGX)-100):ncol(G_LGX),
                        sample(1:ncol(G_LGX), 4000),
                        heat_outliers))))

make_heatmap(mut_sub, heatOrder, paste0(LGtext[i], "_heatmap_outliers_plus4Ksnps.pdf"))

```

```

## Make heatmap for outliers
make_heatmap(heat_outliers, heatOrder, paste0(LGtext[i], "_heatmap_outliers.pdf"))

if (i ==1){
  # there are so many!
  # large linkage block on right side
  start_ind <- which(muts$pos==24230725)
  end_ind <- which(muts$pos== 24455605)
  mut_sub_a <- start_ind:end_ind
  make_heatmap(mut_sub_a, heatOrder, paste0(LGtext[i], "_heatmap_subLG_a_allSNPs.pdf"))
}
if (i ==2){
  ### LG2 zoom in to sub LGs ###
  # LG2_a on the left side of what looks like the breakpoint
  start_ind <- which(muts$pos==2672)
  end_ind <- which(muts$pos==344034)
  mut_sub_a <- start_ind:end_ind
  make_heatmap(mut_sub_a, heatOrder, paste0(LGtext[i], "_heatmap_subLG_a_allSNPs.pdf"))

  # LG2_b on the right side of the inversion
  start_ind <- which(muts$pos==4439467)
  end_ind <- which(muts$pos==4444636)
  mut_sub_b <- start_ind:end_ind
  make_heatmap(mut_sub_b, heatOrder, paste0(LGtext[i], "_heatmap_subLG_b_allSNPs.pdf"))
} # end if (i==2) statement

if (i ==3){ ### LG7 zoom in to sub LGs ###
  ### LG7_a ####
  # in the center of LG7 there is a small LG in ICE reference homokaryotypes
  start_ind <- grep("21008004", muts$chrom_pos)
  end_ind <- grep("21211273", muts$chrom_pos)
  mut_sub_a <- start_ind:end_ind
  make_heatmap(mut_sub_a, heatOrder, paste0(LGtext[i], "_heatmap_subLG_a_allSNPs.pdf"))

  ### LG7_b ####
  # On the right side of LG7 there appears to be
  # a small LG right after the breakpoint "LG7sub_b"
  # 26075924 -> before LG7 breakpoint
  # 26662926 -> approx start LG7sub_b
  # 26787191 -> approx end LG7sub_b
  # 26974475 -> ~100KB past breakpoint
  start_ind <- grep("26662926", muts$chrom_pos)
  end_ind <- grep("26787191", muts$chrom_pos)
  mut_sub_b <- start_ind:end_ind
  make_heatmap(mut_sub_b, heatOrder, paste0(LGtext[i], "_heatmap_subLG_b_allSNPs.pdf"))

  }#end if i=3 statement

saveRDS(muts,paste0("../outputs/2-",LGtext[i], "-mut.s.rds"))

} # end loop through i

```

```

## Reference class 'FBM.code256' [package "bigstatsr"] with 16 fields
## $ extptr      :<externalptr>

```

```

## $ extptr_rw      :<externalptr>
## $ nrow           : int 294
## $ ncol           : int 76405
## $ type           : Named int 1
## ..- attr(*, "names")= chr "unsigned char"
## $ backingfile    : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfyfyn80000gn/T/RtmpxsU1s/filebd0d3f"
## $ is_read_only   : logi FALSE
## $ address        :<externalptr>
## $ address_rw     :<externalptr>
## $ bk             : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfyfyn80000gn/T/RtmpxsU1s/filebd0d3f"
## $ rds            : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfyfyn80000gn/T/RtmpxsU1s/filebd0d3f"
## $ is_saved       : logi FALSE
## $ type_chr       : chr "unsigned char"
## $ type_size      : int 1
## $ file_size      : num 22463070
## $ code256        : num [1:256] 0 1 2 NA NA NA NA NA NA NA ...
## and 26 methods, of which 12 are possibly relevant:
##   add_columns, as.FBM, bm, bm.desc, check_dimensions, check_write_permissions,
##   copy#envRefClass, initialize, initialize#FBM, save, show#envRefClass,
##   show#FBM
## 'data.frame':   294 obs. of  17 variables:
## $ samp_full      : int  16216 16220 16224 16225 16229 16231 16232 16233 16234 16236 ...
## $ Pop            : Factor w/ 9 levels "Pop1","Pop2",...: 1 1 1 1 1 1 1 1 1 ...
## $ PopID          : Ord.factor w/ 9 levels "GoM.Mass.Olive.Winter"<...: 1 1 1 1 1 1 1 1 1 ...
## $ Region         : Factor w/ 2 levels "GoM","Iceland": 1 1 1 1 1 1 1 1 1 ...
## $ Ecotype        : Factor w/ 2 levels "1-Nearshore",...: 1 1 1 1 1 1 1 1 1 ...
## $ label          : chr   "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216" "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220" ...
## $ Region_color   : chr   "blue4" "blue4" "blue4" "blue4" ...
## $ Ecotype_color  : chr   "darkcyan" "darkcyan" "darkcyan" "darkcyan" ...
## $ PopID_color    : chr   "#006400FF" "#006400FF" "#006400FF" "#006400FF" ...
## $ Gsub_PC1       : num  -0.0589 -0.067 -0.0651 -0.0614 -0.0538 ...
## $ Gsub_PC2       : num   0.02733 0.00134 0.00245 0.00487 0.0154 ...
## $ G_AllInv_PC1   : num   0.0353 0.065 0.0613 0.0515 0.0319 ...
## $ G_AllInv_PC2   : num  -0.00791 -0.04903 -0.05377 -0.03753 0.01956 ...
## $ InvHclustOrder: int   146 258 230 147 172 238 243 142 164 156 ...
## $ InvHclustK5    : num    1 1 2 1 1 1 1 1 1 1 ...
## $ LG1_PC1        : num   0.0413 0.0391 0.0411 0.0405 0.0408 ...
## $ LG1_PC2        : num   0.0588 0.0571 0.0469 0.0492 0.0547 ...

## Warning: Using size for a discrete variable is not advised.

## [1] "Mapping of samples to PCA cluster for"
## [2] "LG1"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown GoM.Cashes.Olive
## 1              0              0              0              0
## 2              0              0              0              0
## 3              0              0              0              0
## 4              0              0              0              2
## 5             34             35             17             23
##
##      GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE Ice.Near.NE
## 1              0             25              1             27              1
## 2              0             12              7             11             14
## 3              0              2             32              1             25

```

```

##      4      0      0      0      0      0
##      5     25      0      0      0      0
## [1] "Mapping of samples to named PCA cluster for"
## [2] "LG1"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## ALT-Iceland      0      0      0
## HET-Iceland      0      0      0
## REF-Iceland      0      0      0
## HET-GoM          0      0      0
## REF-GoM         34     35     17
##
##      GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## ALT-Iceland      0      0     25      1     27
## HET-Iceland      0      0     12      7     11
## REF-Iceland      0      0      2     32      1
## HET-GoM          2      0      0      0      0
## REF-GoM         23     25      0      0      0
##
##      Ice.Near.NE
## ALT-Iceland      1
## HET-Iceland     14
## REF-Iceland     25
## HET-GoM          0
## REF-GoM          0
##
##      ALT-Iceland HET-Iceland REF-Iceland HET-GoM REF-GoM
## 1      54      0      0      0      0
## 2      0     44      0      0      0
## 3      0      0     60      0      0
## 4      0      0      0      2      0
## 5      0      0      0      0    134
##
## Warning: Using size for a discrete variable is not advised.
## [1] "Outlier sample sizes for reference arrangement in"
## [2] "LG1"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## ALT-Iceland      0      0      0
## HET-Iceland      0      0      0
## REF-Iceland      0      0      0
## HET-GoM          0      0      0
## REF-GoM         34     35     17
##
##      GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## ALT-Iceland      0      0      0      0      0
## HET-Iceland      0      0      0      0      0
## REF-Iceland      0      0      2     32      1
## HET-GoM          0      0      0      0      0
## REF-GoM         23     25      0      0      0
##
##      Ice.Near.NE
## ALT-Iceland      0
## HET-Iceland      0

```

```

## REF-Iceland      25
## HET-GoM          0
## REF-GoM          0
## ALT-Iceland HET-Iceland REF-Iceland      HET-GoM      REF-GoM
##           0           0           60           0           134
## Calculating FSTs, may take a few minutes...
## [1] "10000 done of 76405"
## [1] "20000 done of 76405"
## [1] "30000 done of 76405"
## [1] "40000 done of 76405"
## [1] "50000 done of 76405"
## [1] "60000 done of 76405"
## [1] "70000 done of 76405"
## [1] "Outlier sample sizes for alternate arrangement in"
## [2] "LG1"
##
##           GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## ALT-Iceland           0           0           0
## HET-Iceland           0           0           0
## REF-Iceland           0           0           0
## HET-GoM               0           0           0
## REF-GoM               0           0           0
##
##           GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## ALT-Iceland           0           0           25           1           27
## HET-Iceland           0           0           0           0           0
## REF-Iceland           0           0           0           0           0
## HET-GoM               0           0           0           0           0
## REF-GoM               0           0           0           0           0
##
##           Ice.Near.NE
## ALT-Iceland           1
## HET-Iceland           0
## REF-Iceland           0
## HET-GoM               0
## REF-GoM               0
## ALT-Iceland HET-Iceland REF-Iceland      HET-GoM      REF-GoM
##           54           0           0           0           0

## [1] "LG1"
## [1] "Number of loci in REF outliers"
## [1] 951
## [1] "Number of loci in ALT outliers"
## [1] 0
## [1] "Number of loci in both REF and ALT outliers"
## [1] 0
## [1] "Total number of loci in LG window"
## [1] 76405

## Reference class 'FBM.code256' [package "bigstatsr"] with 16 fields
## $ extptr      :<externalptr>
## $ extptr_rw   :<externalptr>
## $ nrow        : int 294
## $ ncol        : int 25613
## $ type        : Named int 1

```

```

##   .- attr(*, "names")= chr "unsigned char"
## $ backingfile : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfygn80000gn/T/RtmpxszU1s/filebd0d711"
## $ is_read_only: logi FALSE
## $ address      :<externalptr>
## $ address_rw   :<externalptr>
## $ bk           : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfygn80000gn/T/RtmpxszU1s/filebd0d711"
## $ rds          : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfygn80000gn/T/RtmpxszU1s/filebd0d711"
## $ is_saved     : logi FALSE
## $ type_chr     : chr "unsigned char"
## $ type_size    : int 1
## $ file_size    : num 7530222
## $ code256      : num [1:256] 0 1 2 NA NA NA NA NA NA NA ...
## and 26 methods, of which 12 are possibly relevant:
##   add_columns, as.FBM, bm, bm.desc, check_dimensions, check_write_permissions,
##   copy#envRefClass, initialize, initialize#FBM, save, show#envRefClass,
##   show#FBM
## 'data.frame':   294 obs. of  22 variables:
## $ samp_full      : int  16216 16220 16224 16225 16229 16231 16232 16233 16234 16236 ...
## $ Pop            : Factor w/ 9 levels "Pop1","Pop2",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ PopID          : Ord.factor w/ 9 levels "GoM.Mass.Olive.Winter"<...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Region         : Factor w/ 2 levels "GoM","Iceland": 1 1 1 1 1 1 1 1 1 1 ...
## $ Ecotype        : Factor w/ 2 levels "1-Nearshore",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ label          : chr   "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216" "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220" ...
## $ Region_color   : chr   "blue4" "blue4" "blue4" "blue4" ...
## $ Ecotype_color  : chr   "darkcyan" "darkcyan" "darkcyan" "darkcyan" ...
## $ PopID_color    : chr   "#006400FF" "#006400FF" "#006400FF" "#006400FF" ...
## $ Gsub_PC1       : num  -0.0589 -0.067 -0.0651 -0.0614 -0.0538 ...
## $ Gsub_PC2       : num  0.02733 0.00134 0.00245 0.00487 0.0154 ...
## $ G_AllInv_PC1   : num  0.0353 0.065 0.0613 0.0515 0.0319 ...
## $ G_AllInv_PC2   : num  -0.00791 -0.04903 -0.05377 -0.03753 0.01956 ...
## $ InvHclustOrder : int   146 258 230 147 172 238 243 142 164 156 ...
## $ InvHclustK5    : num   1 1 2 1 1 1 1 1 1 1 ...
## $ LG1_PC1        : num  0.0413 0.0391 0.0411 0.0405 0.0408 ...
## $ LG1_PC2        : num  0.0588 0.0571 0.0469 0.0492 0.0547 ...
## $ LG1cluster     : chr    "5" "5" "5" "5" ...
## $ LG1cluster_name : Factor w/ 5 levels "ALT-Iceland",...: 5 5 5 5 5 5 5 5 5 5 ...
## $ LG1heatmapFSToutliersOrder: int   44 16 101 82 127 12 3 77 9 103 ...
## $ LG2_PC1        : num  0.0125 0.0802 0.012 0.0132 0.0868 ...
## $ LG2_PC2        : num  0.0712 0.016 0.0473 0.0726 0.0244 ...

## Warning: Using size for a discrete variable is not advised.

## [1] "Mapping of samples to PCA cluster for"
## [2] "LG2"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown GoM.Cashes.Olive
## 1                13              7                8                10
## 2                 0              0                0                0
## 3                 5              9                3                3
## 4                 0              0                0                0
## 5                 0              0                0                0
## 6                16             19                6                12
##
##      GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE Ice.Near.NE
## 1                11              0              0              0              0

```



```

##      2          0          7          10          9          19
##      3          4          0          0          0          0
##      4          0         32         15         30         11
##      5          0          0         15          0         10
##      6         10          0          0          0          0
## [1] "Mapping of samples to named PCA cluster for"
## [2] "LG2"
##
##           GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## ALT-GoM                13          7          8
## HET-Iceland            0          0          0
## REF-GoM                5          9          3
## REF-Iceland            0          0          0
## ALT-Iceland            0          0          0
## HET-GoM               16         19          6
##
##           GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## ALT-GoM                10          11          0          0          0
## HET-Iceland            0          0          7         10          9
## REF-GoM                3          4          0          0          0
## REF-Iceland            0          0         32         15         30
## ALT-Iceland            0          0          0         15          0
## HET-GoM               12         10          0          0          0
##
##           Ice.Near.NE
## ALT-GoM                0
## HET-Iceland            19
## REF-GoM                0
## REF-Iceland            11
## ALT-Iceland            10
## HET-GoM                0
##
##      ALT-GoM HET-Iceland REF-GoM REF-Iceland ALT-Iceland HET-GoM
## 1      49          0          0          0          0          0
## 2       0         45          0          0          0          0
## 3       0          0         24          0          0          0
## 4       0          0          0         88          0          0
## 5       0          0          0          0         25          0
## 6       0          0          0          0          0         63
##
## Warning: Using size for a discrete variable is not advised.
## [1] "Outlier sample sizes for reference arrangement in"
## [2] "LG2"
##
##           GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## ALT-GoM                0          0          0
## HET-Iceland            0          0          0
## REF-GoM                5          9          3
## REF-Iceland            0          0          0
## ALT-Iceland            0          0          0
## HET-GoM                0          0          0
##
##           GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## ALT-GoM                0          0          0          0          0

```

```

## HET-Iceland          0          0          0          0          0
## REF-GoM              3          4          0          0          0
## REF-Iceland          0          0         32         15         30
## ALT-Iceland          0          0          0          0          0
## HET-GoM              0          0          0          0          0
##
##           Ice.Near.NE
## ALT-GoM              0
## HET-Iceland          0
## REF-GoM              0
## REF-Iceland          11
## ALT-Iceland          0
## HET-GoM              0
## ALT-GoM HET-Iceland  REF-GoM REF-Iceland ALT-Iceland  HET-GoM
##      0      0      24      88      0      0
## Calculating FSTs, may take a few minutes...
## [1] "10000 done of 25613"
## [1] "20000 done of 25613"
## [1] "Outlier sample sizes for alternate arrangement in"
## [2] "LG2"
##
##           GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## ALT-GoM              13          7          8
## HET-Iceland          0          0          0
## REF-GoM              0          0          0
## REF-Iceland          0          0          0
## ALT-Iceland          0          0          0
## HET-GoM              0          0          0
##
##           GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## ALT-GoM              10          11          0          0          0
## HET-Iceland          0          0          0          0          0
## REF-GoM              0          0          0          0          0
## REF-Iceland          0          0          0          0          0
## ALT-Iceland          0          0          0         15          0
## HET-GoM              0          0          0          0          0
##
##           Ice.Near.NE
## ALT-GoM              0
## HET-Iceland          0
## REF-GoM              0
## REF-Iceland          0
## ALT-Iceland          10
## HET-GoM              0
## ALT-GoM HET-Iceland  REF-GoM REF-Iceland ALT-Iceland  HET-GoM
##      49      0      0      0      25      0
## Calculating FSTs, may take a few minutes...
## [1] "10000 done of 25613"
## [1] "20000 done of 25613"
##
## [1] "LG2"
## [1] "Number of loci in REF outliers"
## [1] 106
## [1] "Number of loci in ALT outliers"

```

```

## [1] 30
## [1] "Number of loci in both REF and ALT outliers"
## [1] 28
## [1] "Total number of loci in LG window"
## [1] 25613

## Reference class 'FBM.code256' [package "bigstatsr"] with 16 fields
## $ extptr      :<externalptr>
## $ extptr_rw   :<externalptr>
## $ nrow        : int 294
## $ ncol        : int 55295
## $ type        : Named int 1
## ..- attr(*, "names")= chr "unsigned char"
## $ backingfile : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfyf80000gn/T/RtmpxsZU1s/filebd0d2f"
## $ is_read_only: logi FALSE
## $ address     :<externalptr>
## $ address_rw  :<externalptr>
## $ bk         : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfyf80000gn/T/RtmpxsZU1s/filebd0d2f"
## $ rds        : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfyf80000gn/T/RtmpxsZU1s/filebd0d2f"
## $ is_saved    : logi FALSE
## $ type_chr    : chr "unsigned char"
## $ type_size   : int 1
## $ file_size   : num 16256730
## $ code256     : num [1:256] 0 1 2 NA NA NA NA NA NA ...
## and 26 methods, of which 12 are possibly relevant:
##   add_columns, as.FBM, bm, bm.desc, check_dimensions, check_write_permissions,
##   copy#envRefClass, initialize, initialize#FBM, save, show#envRefClass,
##   show#FBM
## 'data.frame':   294 obs. of  27 variables:
## $ samp_full      : int  16216 16220 16224 16225 16229 16231 16232 16233 16234 16236 ...
## $ Pop            : Factor w/ 9 levels "Pop1","Pop2",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ PopID         : Ord.factor w/ 9 levels "GoM.Mass.Olive.Winter"<...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Region        : Factor w/ 2 levels "GoM","Iceland": 1 1 1 1 1 1 1 1 1 1 ...
## $ Ecotype       : Factor w/ 2 levels "1-Nearshore",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ label         : chr  "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216" "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220" ...
## $ Region_color   : chr  "blue4" "blue4" "blue4" "blue4" ...
## $ Ecotype_color  : chr  "darkcyan" "darkcyan" "darkcyan" "darkcyan" ...
## $ PopID_color    : chr  "#006400FF" "#006400FF" "#006400FF" "#006400FF" ...
## $ Gsub_PC1       : num  -0.0589 -0.067 -0.0651 -0.0614 -0.0538 ...
## $ Gsub_PC2       : num  0.02733 0.00134 0.00245 0.00487 0.0154 ...
## $ G_AllInv_PC1   : num  0.0353 0.065 0.0613 0.0515 0.0319 ...
## $ G_AllInv_PC2   : num  -0.00791 -0.04903 -0.05377 -0.03753 0.01956 ...
## $ InvHclustOrder : int   146 258 230 147 172 238 243 142 164 156 ...
## $ InvHclustK5    : num   1 1 2 1 1 1 1 1 1 1 ...
## $ LG1_PC1        : num  0.0413 0.0391 0.0411 0.0405 0.0408 ...
## $ LG1_PC2        : num  0.0588 0.0571 0.0469 0.0492 0.0547 ...
## $ LG1cluster     : chr   "5" "5" "5" "5" ...
## $ LG1cluster_name : Factor w/ 5 levels "ALT-Iceland",...: 5 5 5 5 5 5 5 5 5 5 ...
## $ LG1heatmapFSToutliersOrder: int   44 16 101 82 127 12 3 77 9 103 ...
## $ LG2_PC1        : num  0.0125 0.0802 0.012 0.0132 0.0868 ...
## $ LG2_PC2        : num  0.0712 0.016 0.0473 0.0726 0.0244 ...
## $ LG2cluster     : chr   "6" "1" "6" "6" ...
## $ LG2cluster_name : Factor w/ 6 levels "ALT-GoM","HET-Iceland",...: 6 1 6 6 1 1 6 6 1 6 ...
## $ LG2heatmapFSToutliersOrder: int   21 70 76 131 46 95 23 92 24 40 ...

```

```
## $ LG7_PC1 : num -0.000253 0.066013 0.067383 0.067273 -0.000549 ...
## $ LG7_PC2 : num 0.059 0.0272 0.016 0.0227 0.0676 ...
```

```
## Warning: Using size for a discrete variable is not advised.
```

```
## [1] "Mapping of samples to PCA cluster for"
```

```
## [2] "LG7"
```

```
##
```

	GoM.Mass.Olive.Winter	GoM.Mass.Red	GoM.Mass.Olive.Unknown	GoM.Cashes.Olive
## 1	0	0	0	0
## 2	2	2	0	0
## 3	10	5	5	11
## 4	0	0	0	0
## 5	22	28	12	14
## 6	0	0	0	0

```
##
```

	GoM.Cashes.Red	Ice.Off.SW	Ice.Near.SW	Ice.Off.NE	Ice.Near.NE
## 1	0	4	16	4	24
## 2	1	0	0	0	0
## 3	6	0	0	0	0
## 4	0	35	11	35	11
## 5	18	0	0	0	0
## 6	0	0	13	0	5

```
## [1] "Mapping of samples to named PCA cluster for"
```

```
## [2] "LG7"
```

```
##
```

	GoM.Mass.Olive.Winter	GoM.Mass.Red	GoM.Mass.Olive.Unknown
## HET-Iceland	0	0	0
## ALT-GoM	2	2	0
## HET-GoM	10	5	5
## ALT-Iceland	0	0	0
## REF-GoM	22	28	12
## REF-Iceland	0	0	0

```
##
```

	GoM.Cashes.Olive	GoM.Cashes.Red	Ice.Off.SW	Ice.Near.SW	Ice.Off.NE
## HET-Iceland	0	0	4	16	4
## ALT-GoM	0	1	0	0	0
## HET-GoM	11	6	0	0	0
## ALT-Iceland	0	0	35	11	35
## REF-GoM	14	18	0	0	0
## REF-Iceland	0	0	0	13	0

```
##
```

	Ice.Near.NE
## HET-Iceland	24
## ALT-GoM	0
## HET-GoM	0
## ALT-Iceland	11
## REF-GoM	0
## REF-Iceland	5

```
##
```

	HET-Iceland	ALT-GoM	HET-GoM	ALT-Iceland	REF-GoM	REF-Iceland
## 1	48	0	0	0	0	0
## 2	0	5	0	0	0	0
## 3	0	0	37	0	0	0
## 4	0	0	0	92	0	0

```

##      5      0      0      0      0      94      0
##      6      0      0      0      0      0      18

## Warning: Using size for a discrete variable is not advised.

## [1] "Outlier sample sizes for reference arrangement in"
## [2] "LG7"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## HET-Iceland              0              0              0
## ALT-GoM                  0              0              0
## HET-GoM                  0              0              0
## ALT-Iceland              0              0              0
## REF-GoM                  22             28             12
## REF-Iceland              0              0              0
##
##      GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## HET-Iceland              0              0              0              0
## ALT-GoM                  0              0              0              0
## HET-GoM                  0              0              0              0
## ALT-Iceland              0              0              0              0
## REF-GoM                  14             18              0              0
## REF-Iceland              0              0              0             13              0
##
##      Ice.Near.NE
## HET-Iceland              0
## ALT-GoM                  0
## HET-GoM                  0
## ALT-Iceland              0
## REF-GoM                  0
## REF-Iceland              5
## HET-Iceland      ALT-GoM      HET-GoM ALT-Iceland      REF-GoM REF-Iceland
##      0              0              0              0              94              18
## Calculating FSTs, may take a few minutes...
## [1] "10000 done of 55295"
## [1] "20000 done of 55295"
## [1] "30000 done of 55295"
## [1] "40000 done of 55295"
## [1] "50000 done of 55295"
## [1] "Outlier sample sizes for alternate arrangement in"
## [2] "LG7"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## HET-Iceland              0              0              0
## ALT-GoM                  2              2              0
## HET-GoM                  0              0              0
## ALT-Iceland              0              0              0
## REF-GoM                  0              0              0
## REF-Iceland              0              0              0
##
##      GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## HET-Iceland              0              0              0              0
## ALT-GoM                  0              1              0              0
## HET-GoM                  0              0              0              0
## ALT-Iceland              0              0             35             11             35

```

```

## REF-GoM          0          0          0          0          0
## REF-Iceland      0          0          0          0          0
##
##           Ice.Near.NE
## HET-Iceland      0
## ALT-GoM          0
## HET-GoM          0
## ALT-Iceland      11
## REF-GoM          0
## REF-Iceland      0
## HET-Iceland      ALT-GoM      HET-GoM ALT-Iceland      REF-GoM REF-Iceland
##           0           5           0           92           0           0
## Calculating FSTs, may take a few minutes...
## [1] "10000 done of 55295"
## [1] "20000 done of 55295"
## [1] "30000 done of 55295"
## [1] "40000 done of 55295"
## [1] "50000 done of 55295"
##
## [1] "LG7"
## [1] "Number of loci in REF outliers"
## [1] 149
## [1] "Number of loci in ALT outliers"
## [1] 1112
## [1] "Number of loci in both REF and ALT outliers"
## [1] 5
## [1] "Total number of loci in LG window"
## [1] 55295
##
## Reference class 'FBM.code256' [package "bigstatsr"] with 16 fields
## $ extptr          :<externalptr>
## $ extptr_rw       :<externalptr>
## $ nrow            : int 294
## $ ncol            : int 46321
## $ type            : Named int 1
## ..- attr(*, "names")= chr "unsigned char"
## $ backingfile     : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfygn80000gn/T/RtmpxsZU1s/filebd0d10..."
## $ is_read_only    : logi FALSE
## $ address         :<externalptr>
## $ address_rw      :<externalptr>
## $ bk              : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfygn80000gn/T/RtmpxsZU1s/filebd0d10..."
## $ rds             : chr "/private/var/folders/_c/r1wlxvm958b147mw5b1zfygn80000gn/T/RtmpxsZU1s/filebd0d10..."
## $ is_saved        : logi FALSE
## $ type_chr        : chr "unsigned char"
## $ type_size       : int 1
## $ file_size       : num 13618374
## $ code256         : num [1:256] 0 1 2 NA NA NA NA NA NA NA ...
## and 26 methods, of which 12 are possibly relevant:
##   add_columns, as.FBM, bm, bm.desc, check_dimensions, check_write_permissions,
##   copy#envRefClass, initialize, initialize#FBM, save, show#envRefClass,
##   show#FBM
## 'data.frame':   294 obs. of  32 variables:
## $ samp_full       : int  16216 16220 16224 16225 16229 16231 16232 16233 16234 16236 ...
## $ Pop             : Factor w/ 9 levels "Pop1","Pop2",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ PopID           : Ord.factor w/ 9 levels "GoM.Mass.Olive.Winter"<...: 1 1 1 1 1 1 1 1 1 1 ...

```

```

## $ Region : Factor w/ 2 levels "GoM","Iceland": 1 1 1 1 1 1 1 1 1 1 ...
## $ Ecotype : Factor w/ 2 levels "1-Nearshore",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ label : chr "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216" "GoM_1-Nearshore_GoM.Mass.Red_16216" ...
## $ Region_color : chr "blue4" "blue4" "blue4" "blue4" ...
## $ Ecotype_color : chr "darkcyan" "darkcyan" "darkcyan" "darkcyan" ...
## $ PopID_color : chr "#006400FF" "#006400FF" "#006400FF" "#006400FF" ...
## $ Gsub_PC1 : num -0.0589 -0.067 -0.0651 -0.0614 -0.0538 ...
## $ Gsub_PC2 : num 0.02733 0.00134 0.00245 0.00487 0.0154 ...
## $ G_AllInv_PC1 : num 0.0353 0.065 0.0613 0.0515 0.0319 ...
## $ G_AllInv_PC2 : num -0.00791 -0.04903 -0.05377 -0.03753 0.01956 ...
## $ InvHclustOrder : int 146 258 230 147 172 238 243 142 164 156 ...
## $ InvHclustK5 : num 1 1 2 1 1 1 1 1 1 1 ...
## $ LG1_PC1 : num 0.0413 0.0391 0.0411 0.0405 0.0408 ...
## $ LG1_PC2 : num 0.0588 0.0571 0.0469 0.0492 0.0547 ...
## $ LG1cluster : chr "5" "5" "5" "5" ...
## $ LG1cluster_name : Factor w/ 5 levels "ALT-Iceland",...: 5 5 5 5 5 5 5 5 5 5 ...
## $ LG1heatmapFSToutliersOrder: int 44 16 101 82 127 12 3 77 9 103 ...
## $ LG2_PC1 : num 0.0125 0.0802 0.012 0.0132 0.0868 ...
## $ LG2_PC2 : num 0.0712 0.016 0.0473 0.0726 0.0244 ...
## $ LG2cluster : chr "6" "1" "6" "6" ...
## $ LG2cluster_name : Factor w/ 6 levels "ALT-GoM","HET-Iceland",...: 6 1 6 6 1 1 6 6 1 6 ..
## $ LG2heatmapFSToutliersOrder: int 21 70 76 131 46 95 23 92 24 40 ...
## $ LG7_PC1 : num -0.000253 0.066013 0.067383 0.067273 -0.000549 ...
## $ LG7_PC2 : num 0.059 0.0272 0.016 0.0227 0.0676 ...
## $ LG7cluster : chr "3" "5" "5" "5" ...
## $ LG7cluster_name : Factor w/ 6 levels "HET-Iceland",...: 3 5 5 5 3 3 5 5 3 5 ...
## $ LG7heatmapFSToutliersOrder: int 18 44 49 13 136 280 209 174 185 137 ...
## $ LG12_PC1 : num -0.0889 -0.0829 -0.0894 -0.0173 -0.0178 ...
## $ LG12_PC2 : num 0.0202 0.0204 0.0197 0.0223 0.0212 ...

## Warning: Using size for a discrete variable is not advised.

## [1] "Mapping of samples to PCA cluster for"
## [2] "LG12"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown GoM.Cashes.Olive
## 1              0              0              0              0
## 2             18             19             12             12
## 3              0              0              0              0
## 4             13             15              5             12
## 5              3              1              0              1
## 6              0              0              0              0
##
##      GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE Ice.Near.NE
## 1              0              0             13              2             17
## 2             12              0              0              0              0
## 3              0             39             22             37             20
## 4             11              0              0              0              0
## 5              2              0              1              0              0
## 6              0              0              4              0              3

## [1] "Mapping of samples to named PCA cluster for"
## [2] "LG12"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## HET-Iceland              0              0              0

```

```

##      HET-GoM              18              19              12
##      REF-Iceland          0              0              0
##      ALT-GoM              13              15              5
##      REF-GoM              3              1              0
##      ALT-Iceland          0              0              0
##
##      GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
##      HET-Iceland      0              0              0              13              2
##      HET-GoM          12              12              0              0              0
##      REF-Iceland      0              0              39              22              37
##      ALT-GoM          12              11              0              0              0
##      REF-GoM          1              2              0              1              0
##      ALT-Iceland      0              0              0              4              0
##
##      Ice.Near.NE
##      HET-Iceland      17
##      HET-GoM          0
##      REF-Iceland      20
##      ALT-GoM          0
##      REF-GoM          0
##      ALT-Iceland      3
##
##      HET-Iceland HET-GoM REF-Iceland ALT-GoM REF-GoM ALT-Iceland
##      1          32      0          0          0          0          0
##      2          0      73          0          0          0          0
##      3          0      0          118         0          0          0
##      4          0      0          0          56         0          0
##      5          0      0          0          0          8          0
##      6          0      0          0          0          0          7
##
## Warning: Using size for a discrete variable is not advised.
## [1] "Outlier sample sizes for reference arrangement in"
## [2] "LG12"
##
##      GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
##      HET-Iceland          0              0              0
##      HET-GoM              0              0              0
##      REF-Iceland          0              0              0
##      ALT-GoM              0              0              0
##      REF-GoM              3              1              0
##      ALT-Iceland          0              0              0
##
##      GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
##      HET-Iceland      0              0              0              0              0
##      HET-GoM          0              0              0              0              0
##      REF-Iceland      0              0              39              22              37
##      ALT-GoM          0              0              0              0              0
##      REF-GoM          1              2              0              1              0
##      ALT-Iceland      0              0              0              0              0
##
##      Ice.Near.NE
##      HET-Iceland      0
##      HET-GoM          0
##      REF-Iceland      20

```



```

## ALT-GoM 0
## REF-GoM 0
## ALT-Iceland 0
## HET-Iceland HET-GoM REF-Iceland ALT-GoM REF-GoM ALT-Iceland
## 0 0 118 0 8 0
## Calculating FSTs, may take a few minutes...
## [1] "10000 done of 46321"
## [1] "20000 done of 46321"
## [1] "30000 done of 46321"
## [1] "40000 done of 46321"
## [1] "Outlier sample sizes for alternate arrangement in"
## [2] "LG12"
##
## GoM.Mass.Olive.Winter GoM.Mass.Red GoM.Mass.Olive.Unknown
## HET-Iceland 0 0 0
## HET-GoM 0 0 0
## REF-Iceland 0 0 0
## ALT-GoM 13 15 5
## REF-GoM 0 0 0
## ALT-Iceland 0 0 0
##
## GoM.Cashes.Olive GoM.Cashes.Red Ice.Off.SW Ice.Near.SW Ice.Off.NE
## HET-Iceland 0 0 0 0 0
## HET-GoM 0 0 0 0 0
## REF-Iceland 0 0 0 0 0
## ALT-GoM 12 11 0 0 0
## REF-GoM 0 0 0 0 0
## ALT-Iceland 0 0 0 4 0
##
## Ice.Near.NE
## HET-Iceland 0
## HET-GoM 0
## REF-Iceland 0
## ALT-GoM 0
## REF-GoM 0
## ALT-Iceland 3
## HET-Iceland HET-GoM REF-Iceland ALT-GoM REF-GoM ALT-Iceland
## 0 0 0 56 0 7
## Calculating FSTs, may take a few minutes...
## [1] "10000 done of 46321"
## [1] "20000 done of 46321"
## [1] "30000 done of 46321"
## [1] "40000 done of 46321"
##
## [1] "LG12"
## [1] "Number of loci in REF outliers"
## [1] 147
## [1] "Number of loci in ALT outliers"
## [1] 7683
## [1] "Number of loci in both REF and ALT outliers"
## [1] 9
## [1] "Total number of loci in LG window"
## [1] 46321

```

## Plot multi-inversion genotypes

Make a heatmap that shows the inversions in each ecotype

```
head(samp_full)
```

```
##      samp_full  Pop      PopID Region      Ecotype
## 1      16216 Pop1 GoM.Mass.Olive.Winter      GoM 1-Nearshore
## 2      16220 Pop1 GoM.Mass.Olive.Winter      GoM 1-Nearshore
## 3      16224 Pop1 GoM.Mass.Olive.Winter      GoM 1-Nearshore
## 4      16225 Pop1 GoM.Mass.Olive.Winter      GoM 1-Nearshore
## 5      16229 Pop1 GoM.Mass.Olive.Winter      GoM 1-Nearshore
## 6      16231 Pop1 GoM.Mass.Olive.Winter      GoM 1-Nearshore
##
##                                label Region_color Ecotype_color
## 1 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216      blue4      darkcyan
## 2 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220      blue4      darkcyan
## 3 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16224      blue4      darkcyan
## 4 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16225      blue4      darkcyan
## 5 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16229      blue4      darkcyan
## 6 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16231      blue4      darkcyan
##      PopID_color      Gsub_PC1      Gsub_PC2 G_AllInv_PC1 G_AllInv_PC2 InvHclustOrder
## 1      #006400FF -0.05893059 0.027334412 0.03531921 -0.007908553      146
## 2      #006400FF -0.06696424 0.001338206 0.06496404 -0.049032057      258
## 3      #006400FF -0.06511353 0.002449571 0.06134477 -0.053768027      230
## 4      #006400FF -0.06136966 0.004868653 0.05153129 -0.037532901      147
## 5      #006400FF -0.05381484 0.015398091 0.03185495 0.019557487      172
## 6      #006400FF -0.05952850 0.018067312 0.04147050 0.003228003      238
##      InvHclustK5      LG1_PC1      LG1_PC2 LG1cluster LG1cluster_name
## 1      1 0.04134799 0.05882174      5      REF-GoM
## 2      1 0.03906286 0.05708717      5      REF-GoM
## 3      2 0.04106403 0.04690637      5      REF-GoM
## 4      1 0.04051841 0.04919281      5      REF-GoM
## 5      1 0.04075889 0.05465377      5      REF-GoM
## 6      1 0.04174522 0.05062409      5      REF-GoM
##      LG1heatmapFSToutliersOrder      LG2_PC1      LG2_PC2 LG2cluster LG2cluster_name
## 1      44 0.01252752 0.07120433      6      HET-GoM
## 2      16 0.08018932 0.01597913      1      ALT-GoM
## 3      101 0.01202102 0.04732224      6      HET-GoM
## 4      82 0.01324539 0.07259462      6      HET-GoM
## 5      127 0.08684689 0.02435808      1      ALT-GoM
## 6      12 0.08514757 0.01490878      1      ALT-GoM
##      LG2heatmapFSToutliersOrder      LG7_PC1      LG7_PC2 LG7cluster
## 1      21 -0.0002525006 0.05899135      3
## 2      70 0.0660129234 0.02716254      5
## 3      76 0.0673826325 0.01602092      5
## 4      131 0.0672730111 0.02273250      5
## 5      46 -0.0005488408 0.06758172      3
## 6      95 0.0004475725 0.07562338      3
##      LG7cluster_name LG7heatmapFSToutliersOrder      LG12_PC1      LG12_PC2 LG12cluster
## 1      HET-GoM      18 -0.08888445 0.02019992      2
## 2      REF-GoM      44 -0.08290678 0.02036495      2
## 3      REF-GoM      49 -0.08941798 0.01966069      2
## 4      REF-GoM      13 -0.01732930 0.02229741      4
## 5      HET-GoM      136 -0.01782624 0.02120944      4
## 6      HET-GoM      280 -0.08656489 0.01913641      2
```

```

##   LG12cluster_name LG12heatmapFSToutliersOrder
## 1      HET-GoM                285
## 2      HET-GoM                210
## 3      HET-GoM                257
## 4      ALT-GoM                188
## 5      ALT-GoM                272
## 6      HET-GoM                192

samp_full <- samp_full[order(samp_full$Ecotype,
                             samp_full$Region,
                             samp_full$PopID),]

samp_inv <- samp_full[,grep("cluster_name", colnames(samp_full))]
head(samp_inv)

##   LG1cluster_name LG2cluster_name LG7cluster_name LG12cluster_name
## 1      REF-GoM      HET-GoM      HET-GoM      HET-GoM
## 2      REF-GoM      ALT-GoM      REF-GoM      HET-GoM
## 3      REF-GoM      HET-GoM      REF-GoM      HET-GoM
## 4      REF-GoM      HET-GoM      REF-GoM      ALT-GoM
## 5      REF-GoM      ALT-GoM      HET-GoM      ALT-GoM
## 6      REF-GoM      ALT-GoM      HET-GoM      HET-GoM

rownames(samp_inv) <- samp_full$label
samp_inv$label <- samp_full$label

library(ggplot2); library(reshape2)
samp_inv_melt <- melt(samp_inv, id.var = 'label')

## Warning: attributes are not identical across measure variables; they will be
## dropped

dim(samp_inv)

## [1] 294    5

dim(samp_inv_melt)

## [1] 1176    3
294*4

## [1] 1176

head(samp_inv_melt)

##               label      variable  value
## 1 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216 LG1cluster_name REF-GoM
## 2 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220 LG1cluster_name REF-GoM
## 3 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16224 LG1cluster_name REF-GoM
## 4 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16225 LG1cluster_name REF-GoM
## 5 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16229 LG1cluster_name REF-GoM
## 6 GoM_1-Nearshore_GoM.Mass.Olive.Winter_16231 LG1cluster_name REF-GoM

pdf("../figures/multi-inv-heatmap1.pdf", height=10,
     width=8)
ggplot(samp_inv_melt, aes(variable, label)) +
  geom_tile(aes(fill = value)) +
  scale_fill_manual(values=c("firebrick3", "red", "grey40", "grey60", "cornflowerblue", "blue" )) +

```

```

    theme(axis.text.y=element_text(size=2))
dev.off()

## pdf
## 2

samp_full <- samp_full[order(samp_full$Region,
                             samp_full$Ecotype,
                             samp_full$PopID,
                             samp_full$label),]

samp_inv <- samp_full[,grep("cluster_name", colnames(samp_full))]
head(samp_inv)

##   LG1cluster_name LG2cluster_name LG7cluster_name LG12cluster_name
## 1      REF-GoM      HET-GoM      HET-GoM      HET-GoM
## 2      REF-GoM      ALT-GoM      REF-GoM      HET-GoM
## 3      REF-GoM      HET-GoM      REF-GoM      HET-GoM
## 4      REF-GoM      HET-GoM      REF-GoM      ALT-GoM
## 5      REF-GoM      ALT-GoM      HET-GoM      ALT-GoM
## 6      REF-GoM      ALT-GoM      HET-GoM      HET-GoM

rownames(samp_inv) <- samp_full$label

colnames(samp_inv) <- gsub("cluster_name","",colnames(samp_inv))
str(samp_inv)

## 'data.frame': 294 obs. of 4 variables:
## $ LG1 : Factor w/ 5 levels "ALT-Iceland",...: 5 5 5 5 5 5 5 5 5 ...
## $ LG2 : Factor w/ 6 levels "ALT-GoM","HET-Iceland",...: 6 1 6 6 1 1 6 6 1 6 ...
## $ LG7 : Factor w/ 6 levels "HET-Iceland",...: 3 5 5 5 3 3 5 5 3 5 ...
## $ LG12: Factor w/ 6 levels "HET-Iceland",...: 2 2 2 4 4 2 4 4 2 4 ...

samp_inv_num <- samp_inv
# Let's order them:
# REF-Iceland = 1
# REF-GoM = 2
# HET-Iceland = 4
# HET-GOM = 5
# ALT-Iceland = 7
# ALT-GOM = 8

levels(samp_inv_num$LG1)

## [1] "ALT-Iceland" "HET-Iceland" "REF-Iceland" "HET-GoM"      "REF-GoM"

levels(samp_inv_num$LG1) <- c(7,4,1,5,2)

levels(samp_inv_num$LG2)

## [1] "ALT-GoM"      "HET-Iceland" "REF-GoM"      "REF-Iceland" "ALT-Iceland"
## [6] "HET-GoM"

levels(samp_inv_num$LG2) <- c(8, 4, 2, 1, 7, 5)

levels(samp_inv_num$LG7)

```

```
## [1] "HET-Iceland" "ALT-GoM"      "HET-GoM"      "ALT-Iceland" "REF-GoM"
## [6] "REF-Iceland"
```

```
levels(samp_inv_num$LG7) <- c(4, 8, 5, 7, 2, 1)
```

```
levels(samp_inv_num$LG12)
```

```
## [1] "HET-Iceland" "HET-GoM"      "REF-Iceland" "ALT-GoM"      "REF-GoM"
## [6] "ALT-Iceland"
```

```
levels(samp_inv_num$LG12) <- c(4, 5, 1, 8, 2, 7)
```

```
head(samp_inv_num)
```

```
##
##           LG1 LG2 LG7 LG12
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216  2  5  5  5
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220  2  8  2  5
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16224  2  5  2  5
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16225  2  5  2  8
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16229  2  8  5  8
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16231  2  8  5  5
```

```
samp_inv_num <- matrix(as.numeric(as.matrix(samp_inv_num)), ncol=4)
```

```
rownames(samp_inv_num) <- samp_full$label
colnames(samp_inv_num) <- colnames(samp_inv)
str(samp_inv_num)
```

```
## num [1:294, 1:4] 2 2 2 2 2 2 2 2 2 2 ...
## - attr(*, "dimnames")=List of 2
## ..$ : chr [1:294] "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216" "GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220" ...
## ..$ : chr [1:4] "LG1" "LG2" "LG7" "LG12"
```

```
samp_ann <- cbind(
  Region = samp_full$Region_color,
  Pop = samp_full$PopID_color,
  Ecotype = samp_full$Ecotype_color
)
```

*# for annotation in heatmap, the rownames of the annotation table has to have the same rownames as the samp\_inv\_num*

```
rownames(samp_ann) <- rownames(samp_inv_num)
head(samp_ann)
```

```
##
##           Region Pop      Ecotype
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16216 "blue4" "#006400FF" "darkcyan"
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16220 "blue4" "#006400FF" "darkcyan"
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16224 "blue4" "#006400FF" "darkcyan"
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16225 "blue4" "#006400FF" "darkcyan"
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16229 "blue4" "#006400FF" "darkcyan"
## GoM_1-Nearshore_GoM.Mass.Olive.Winter_16231 "blue4" "#006400FF" "darkcyan"
```

```
dim(samp_ann)
```

```
## [1] 294  3
```

```
heatmap_colors <- colorRampPalette(c("cornflowerblue", "grey", "firebrick3"))(8)
```

```
pdf("../figures/multi-inv-plot-v2.pdf", width=9, height=9)
par(oma=c(0,0,0, 10))
heatmap3(samp_inv_num,
         Rowv=NA,
         Colv = NA,
         showColDendro = F,
         showRowDendro = T,
         scale = "none",
         RowSideColors = samp_ann,
         col = heatmap_colors,
         cexCol = 1,
         cexRow=0.4,
         ColSideWidth = 2
)

dev.off()
```

```
## pdf
## 2
```

**Write outputs**