

title: "Local PCA results"

date: "Mon Apr 13 10:15:25 2020"

Render this, for instance, like:

```
templater::render_template("summarize_run.Rmd",output="lostruct_results/type_snp_size_10000_jobid_324902/run_summary.html",char
```

This run had these parameters:

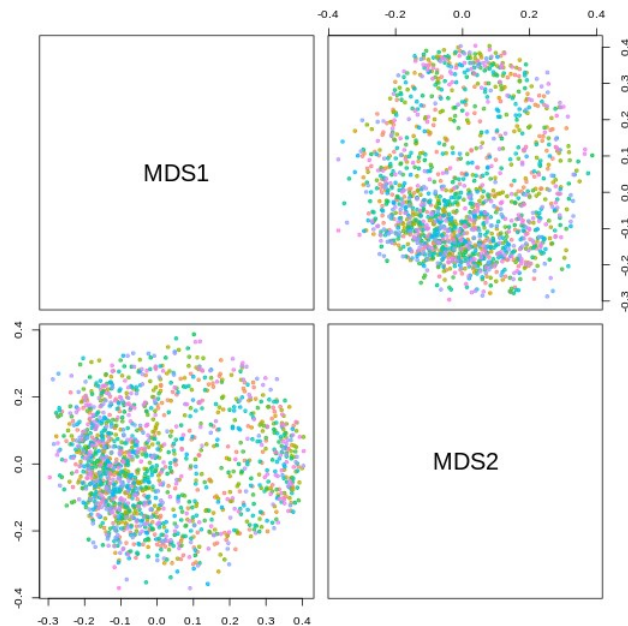
- *window type* : snp
- *window size* : 1000
- *number of pcs* : 2
- *number of MDS coordinates* : 2

Here are the number of windows per chromosome,

and the computed MDS coordinates, colored by chromosome:

```
table(regions$chrom)
##
## MRVK01000062.1 MRVK01000135.1 MRVK01000147.1 MRVK01000157.1 MRVK01000189.1
## 17 34 29 55 33
## MRVK01000312.1 MRVK01000345.1 MRVK01000360.1 MRVK01000381.1 MRVK01000490.1
## 26 21 10 15 25
## MRVK01000512.1 MRVK01000537.1 MRVK01000539.1 MRVK01000624.1 MRVK01000680.1
## 33 11 12 17 30
## MRVK01000730.1 MRVK01000747.1 MRVK01000773.1 MRVK01000817.1 MRVK01000826.1
## 42 61 41 24 33
## MRVK01000835.1 MRVK01000889.1 MRVK01000923.1 MRVK01000975.1 MRVK01000985.1
## 44 17 80 24 6
## MRVK01001060.1 MRVK01001149.1 MRVK01001295.1 MRVK01001299.1 MRVK01001315.1
## 65 16 11 21 58
## MRVK01001321.1 MRVK01001353.1 MRVK01001624.1 MRVK01001658.1 MRVK01001676.1
## 25 10 12 78 15
## MRVK01001759.1 MRVK01001797.1 MRVK01001817.1 MRVK01001827.1 MRVK01001862.1
## 19 8 19 24 8
## MRVK01001988.1 MRVK01002016.1 MRVK01002019.1 MRVK01002062.1 MRVK01002072.1
## 57 12 11 11 68
## MRVK01002239.1 MRVK01002250.1 MRVK01002279.1 MRVK01002320.1 MRVK01002330.1
## 21 17 44 13 15
## MRVK01002434.1 MRVK01002435.1 MRVK01002484.1 MRVK01002514.1 MRVK01002630.1
## 27 10 61 24 30
## MRVK01002634.1
## 8
```

```
pairs( mds.coords[,mds.cols], pch=20, col=adjustcolor(chrom.cols,0.75) )
```



```
if (do.pdfs) { pdf_copy() }
## pdf version at: figure/allscaffolds10Mbp_1000SNPc/mds_pairplot-1.pdf
## Error in device(...): cannot open file 'figure/allscaffolds10Mbp_1000SNPc/mds_pairplot-1.pdf'
```

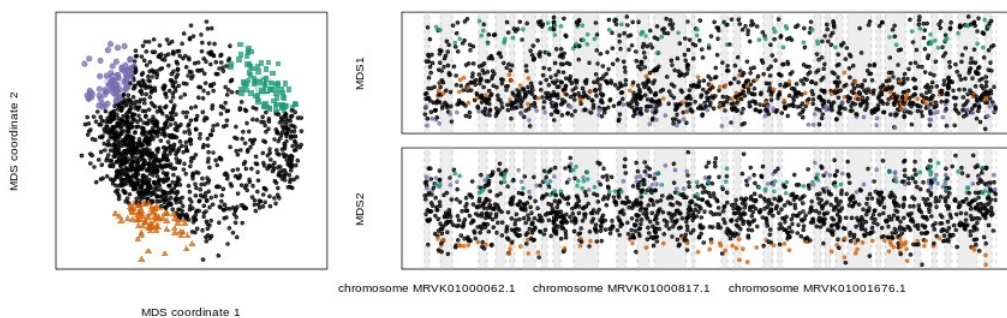
Here are the extreme windows in the MDS plot:

```
mds.corners <- corners( mds.coords[,mds.cols[1:2]], prop=.05 )
```

```

# set up colors and pchs for corners
corner.cols <- brewer.pal(3,"Dark2")
corner.pch <- c(15,17,19)
ccols <- rep("black",nrow(mds.coords))
cpch <- rep(20,nrow(mds.coords))
for (k in 1:ncol(mds.corners)) {
  ccols[ mds.corners[,k] ] <- corner.cols[k]
  cpch[ mds.corners[,k] ] <- corner.pch[k]
}
# centroids of the corners in MDS space
corner.mds <- do.call(rbind, lapply(1:ncol(mds.corners),
  function (ii){
    colMeans(mds.coords[mds.corners[,ii],-(1:2)])
  } ) )
# plot corners and MDS along the chromosome
spacing <- 1
opar <- par(mar=c(4,4,2,1)+.1,mgp=c(2.5,0.8,0))
layout(matrix(c(rep(1,length(mds.cols)),1+seq_along(mds.cols)),ncol=2),
  widths=c(1,2), heights=layout_heights(length(mds.cols),dl=spacing,ncol=2))
plot( mds.coords[,mds.cols[1:2]], pch=cpch,
  col=adjustcolor(ccols,0.75), asp=1,
  xaxt='n', yaxt='n',
  xlab="MDS coordinate 1", ylab="MDS coordinate 2" )
points( corner.mds, pch=20, cex=5,
  col=adjustcolor(corner.cols,0.25))
text( corner.mds, labels=seq_len(nrow(corner.mds)),
  col=corner.cols, cex=2, lwd=2 )
opar2 <- par(mar=c(par("mar"),spacing/2)[c(5,2,3,4)])
for (k in mds.cols) {
  lastone <- (k==mds.cols[length(mds.cols)])
  if (lastone) { par(mar=c(par("mar"),opar2$mar[1])[c(5,2,3,4)]) }
  chrom.plot( mds.coords[,k], pch=20,
    xlab=if (lastone) { "Position (Mb)" } else { "" }, # main=paste("MDS coordinate",match(k,mds.cols)),
    chrom.labels=lastone,
    ylab=colnames(mds.coords)[k],
    col=adjustcolor(ccols,0.75) )
  # do this for all but first
  par(mar=c(par("mar"),spacing/2)[c(1,2,5,4)])
}

```



```

par(opar)
if (do.pdfs) { pdf_copy() }
## pdf version at: figure/allscaffolds10Mbp_1000SNPc/plot_corners-1.pdf
## png
## 2

```

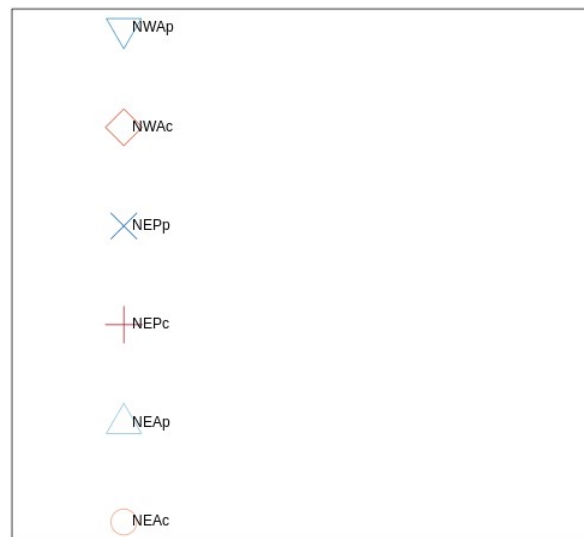
Now, we'll look at PCA plots from the extracted corners.

(this is done without a ton of memory by accumulating the covariance matrix in `running_cov`):

```
corner.npc <- 4
corner.regions <- vector(mode='list', length=ncol(mds.corners))
for (k in seq_along(corner.regions)) {
  corner.regions[[k]] <- regions[ mds.corners[,k], ]
}
corner.covmats <- vector(mode='list', length=ncol(mds.corners))
for (k in seq_along(corner.covmats)) {
  reg <- regions[ mds.corners[,k], ]
  qfun <- multi_vcf_query_fn( chrom.list=chroms, file=bcf.files, regions=reg )
  corner.covmats[[k]] <- running_cov(qfun,1:nrow(reg), normalize.rows=TRUE)
}
## Taking input= as a system command ('bcftools query -f '[ %GT]\n' -r MRVK01000135.1:19528169-20462752 NA') and a variable ha:
## Error in colMeans(x, na.rm = TRUE): 'x' must be an array of at least two dimensions
corner.pca <- vector(mode='list', length=ncol(mds.corners))
for (k in seq_along(corner.pca)) {
  corner.pca[[k]] <- cov_pca(covmat=corner.covmats[[k]], k=corner.npc, w=opt$weights)
}
## Error in rep_len(sqrt(w), ncol(covmat)): invalid 'length.out' value
```

Here is the color scheme:

```
pop.names <- levels(samps$population)
pop.cols <- c("#F4A582", "#92C5DE", "#B2182B", "#2166AC", "#D6604D", "#4393C3")
pop.pch <- seq_len(nlevels(samps$population))
plot( rep(1,length(pop.names)), seq_along(pop.names),
      cex=4, pch=pop.pch, col=pop.cols, xlim=c(0,length(pop.names)),
      xlab='', ylab='', xaxt='n', yaxt='n' )
text( rep(1,length(pop.names)), seq_along(pop.names), labels=pop.names, pos=4 )
```



Here are all pairwise plots of the first 4 PCs for each of the three corners:

```
layout(t(1:3))
for (i in 1:(corner.npc-1)) {
  for (j in (i+1):corner.npc) {
    for (k in 1:ncol(mds.corners)) {
      vectors <- matrix( corner.pca[[k]][-(1:(1+corner.npc))], ncol=corner.npc )[,c(i,j)]
```

```
colnames(vectors) <- paste("PC", c(i,j))
par(mgp=c(0.7,0.7,0), mar=c(2,2,2,0)+.1)
plot(vectors, pch=pop.pch[samps$population],
      col=pop.cols[samps$population],
      xaxt='n', yaxt='n' )
if (i==1 && j==2) {
  mtext(paste("corner",k),side=3)
}
}
if (do.pdfs) { pdf_copy(plot.id=paste(i,j,sep="_")) }
}
}
## Error in matrix(corner.pca[[k]][-(1:(1 + corner.npc))], ncol = corner.npc): 'data' must be of a vector type, was 'NULL'
```