Nucella ostrina

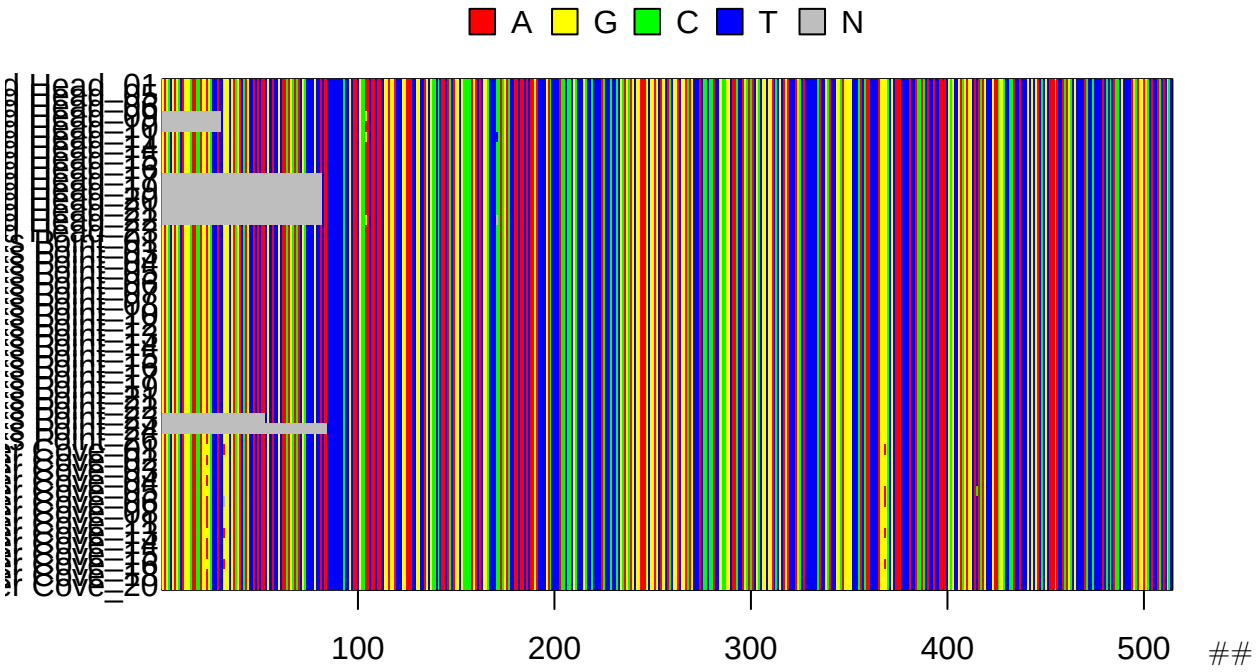
Lilli Krier, Eric Crandall

10-25-19

Contents

Setup

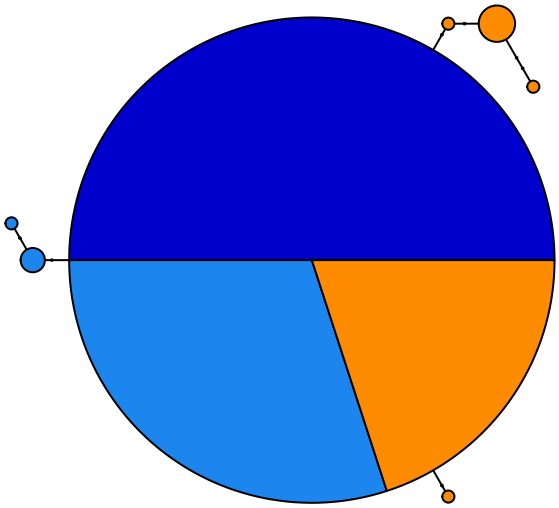
```
library(rentrez)
library(ape)
library(pegas)
library(strataG)
library(stringr)
library(knitr)
library(ggplot2)
library(reshape2)

#set the working directory
#opts_knit$set(root.dir = "~/Users/crandall_lab/Desktop/Lilli_Krier/CapeMendocino")



# read in the data to a DNAbin object

Nostrina <-read.FASTA("Final_Ostrina_Alignments/Nucella_ostrina.fasta")

# Initial look at data


Nostrina <- muscle(Nostrina)

image.DNAbin(Nostrina)
```



Haplotype network

First make a network that colors each population separately

```
#make a population map by stripping the population name off of the DNAbin labels
pop <- str_extract(rownames(Nostrina),"[A-Za-z]+\\s*[A-Za-z]+")

#Label the haplotypes and sort
h <- pegas::haplotype(Nostrina)
h <- sort(h, what = "label")
#Create the network
net <- pegas::haploNet(h)
#Funky code to make a table of which haplotypes occur in which populations
i<-stack(setNames(attr(h, "index"), rownames(h)))
i<-i[order(i$values),]
ind.hap<-table(hap=i$ind, pop=pop)
# put columns of ind.hap into North to South order
ind.hap<-ind.hap[,c("Patricks Point","Trinidad Head","Shelter Cove")]

#Pick some funky colors
mycolors <- c("blue3","dodgerblue2","darkorange", "darkorange4")



plot(net, size=attr(net,"freq"), scale.ratio=2, pie = ind.hap, legend = F, labels = F, threshold = 0, sh
```



```
#legend("bottomleft", colnames(ind.hap), col=mycolors, pch=19, ncol = 2)
```

Diversity Statistics

```
Nostrina_g <- sequence2gtypes(Nostrina, strata = pop)

Nostrina_g  <- labelHaplotypes(Nostrina_g)
```

```
## Warning in .removeIdsMissingAllLoci(g): The following samples are missing
## data for all loci and have been removed: Patricks Point_23, Patricks
## Point_24, Shelter Cove_06, Trinidad Head_09, Trinidad Head_10, Trinidad
## Head_17, Trinidad Head_19, Trinidad Head_20, Trinidad Head_21, Trinidad
## Head_22
```

```
stratastat<-function(x,pop=pop,fun=nuc.div){
  #this function will calculate stats for a DNAbin object (x), stratified across populations given in p
  # Some functions this will work with: nuc.div(), theta.s(), tajima.test() from pegas, FusFs(), exptdH
stats<-NULL
for(p in unique(pop)){
  stat<-fun(x[grep(p,rownames(x)),])
  stats<-c(stats,stat)
}
names(stats)<-unique(pop)
return(stats)
}


nucd<-stratastat(Nostrina,pop=pop,fun=nuc.div)
#Fs<-stratastat(Nostrina,pop=pop,fun=fusFs)

nucd
```

```
##  Trinidad Head Patricks Point   Shelter Cove
##   0.0007936508   0.0000000000   0.0022063706
```

```
#kable(data.frame("Nucleotide Diversity"=nucd, "Fu's Fs" = Fs ))
```

```
Nostrina_g
```

```
## $gtypes
##
## <<< gtypes created on 2019-10-27 09:12:05 >>>
##
## Contents: 39 samples, 1 locus, 3 strata
##
## Strata summary:
##                num.samples num.missing num.alleles prop.unique.alleles
## Patricks Point          18           0           1                 0.0
## Shelter Cove            13           0           4                 0.5
## Trinidad Head            8           0           2                 0.5
##                heterozygosity
## Patricks Point      0.0000000
## Shelter Cove        0.6025641
## Trinidad Head       0.2500000
##
## $unassigned
## $unassigned$gene.1
##                                        haplotype min.substitutions
## Trinidad Head_09                   Shelter Cove_03                20
## Trinidad Head_10                   Shelter Cove_03                19
## Trinidad Head_17                   Shelter Cove_03                59
## Trinidad Head_19                   Shelter Cove_03                59
## Trinidad Head_20                   Shelter Cove_03                59
## Trinidad Head_21                   Shelter Cove_03                59
## Trinidad Head_22  Shelter Cove_03, Trinidad Head_11                61
## Patricks Point_23                  Shelter Cove_03                35
## Patricks Point_24                  Shelter Cove_03                62
## Shelter Cove_06   Shelter Cove_01, Trinidad Head_01                 1
```

F-statistics

Took some hints from here, and here

Pairwise PhiST

Let's plot the pairwise PhiST statistics as a heatmap

```
pairwise_phi <- pairwiseTest(Nostrina_g$gtypes, stats = "phist", nrep = 10000, model = "raw")
```

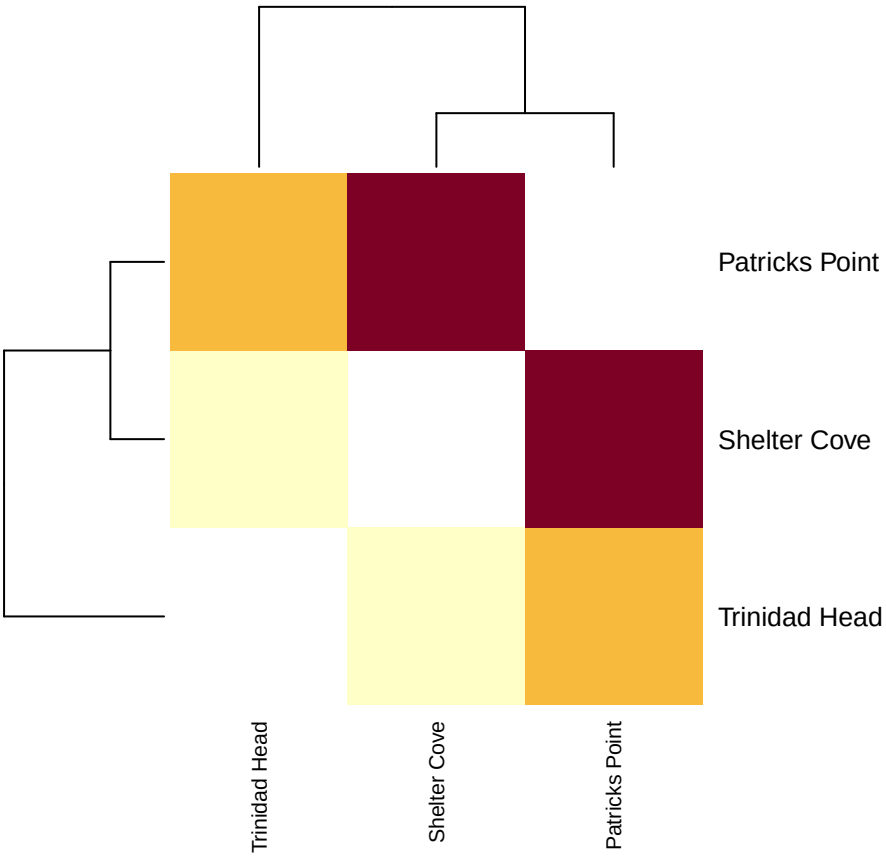
```
##
## <<< gtypes created on 2019-10-27 09:12:05 >>>
##  2019-10-27 09:12:05 : Pairwise tests : 10000 permutations
##    2019-10-27 09:12:05 : Patricks Point v. Shelter Cove
##    2019-10-27 09:12:06 : Patricks Point v. Trinidad Head
##    2019-10-27 09:12:06 : Shelter Cove v. Trinidad Head
##
## Population structure results:
##                                pair.label      PHIst PHIst.p.val
## 1 Patricks Point (18) v. Shelter Cove (13) 0.29370236   0.0079992
```

```
## 2 Patricks Point (18) v. Trinidad Head (8) 0.11032657   0.3030697
## 3   Shelter Cove (13) v. Trinidad Head (8) 0.01774471   0.3564644
```

```
pairwise_phi$pair.mat$PHIst
```

```
##                Patricks Point Shelter Cove Trinidad Head
## Patricks Point             NA   0.00799920     0.3030697
## Shelter Cove        0.2937024           NA     0.3564644
## Trinidad Head       0.1103266   0.01774471            NA
```

```
phiST <- pairwise_phi$pair.mat$PHIst

#symmetricize the matrix
phiST[upper.tri(phiST)] = t(phiST)[upper.tri(phiST)]
#put it in North to South Order
phiST<-phiST[c(1,3,2),c(1,3,2)]

distance <- dist(phiST)
cluster <- hclust(distance, method="ward.D")
dendrogram <- as.dendrogram(cluster)
Rowv <- c(1,2,3)
#dendrogram <- reorder(dendrogram, Rowv)
#reorderfun <- function(d,w) { d }

heatmap(phiST,scale="none", cexRow = 1, cexCol=0.8)
```



AMOVA

```
#write.csv(pop,"amova_hypotheses_nucella.csv",row.names = F, quote=F)
#Open this file up in a spreadsheet program and fill out your hypotheses for higher level hierarchies,

#read them back into R
amovahyps<-read.csv("amova_hypotheses_nucella2.csv")

CapeMendocino<-amovahyps$CapeMendocino
pop1<-as.factor(pop)

#Calculate the p-distance among all sequences
dists<-dist.dna(Nostrina,model="raw")
```

Test the Cape Mendocino hypothesis:

```
amova_out<-pegas::amova(formula=dists~CapeMendocino/pop1,nperm=1000)


amova_out
```

```
##
##  Analysis of Molecular Variance
##
## Call: pegas::amova(formula = dists ~ CapeMendocino/pop1, nperm = 1000)
##
##                          SSD          MSD df
## CapeMendocino 1.437124e-05 1.437124e-05  1
## pop1          1.862939e-06 1.862939e-06  1
## Error         4.253711e-05 9.247197e-07 46
## Total         5.877129e-05 1.224402e-06 48
##
## Variance components:
##                    sigma2 P.value
## CapeMendocino 6.3100e-07  0.3277
## pop1          5.4729e-08  0.0649
## Error         9.2472e-07
##
## Phi-statistics:
## CapeMendocino.in.GLOBAL (Phi_CT)          pop1.in.GLOBAL (Phi_ST)
##                       0.39181610                       0.42580008
##   pop1.in.CapeMendocino (Phi_SC)
##                       0.05587779
##
## Variance coefficients:
##        a        b        c
## 17.14286 15.10204 20.00000
```

```
# You can see the output above, but we can re-calculate it just to be sure.

FCT<-amova_out$varcomp[1,1]/sum(amova_out$varcomp[,1])
FCTp<-amova_out$varcomp[1,2]

FSC<-amova_out$varcomp[2,1]/(amova_out$varcomp[2,1]+amova_out$varcomp[3,1])
FSCp<-amova_out$varcomp[2,2]
```

```r
FST<-(amova_out$varcomp[1,1]+amova_out$varcomp[2,1])/sum(amova_out$varcomp[,1])
FSTp<-NA

result<-c(FCT,FSC,FST)
result
```

```
## [1] 0.39181610 0.05587779 0.42580008
```