Lottia.digitalis

Lilli Krier, Eric Crandall

10-21-19

Contents

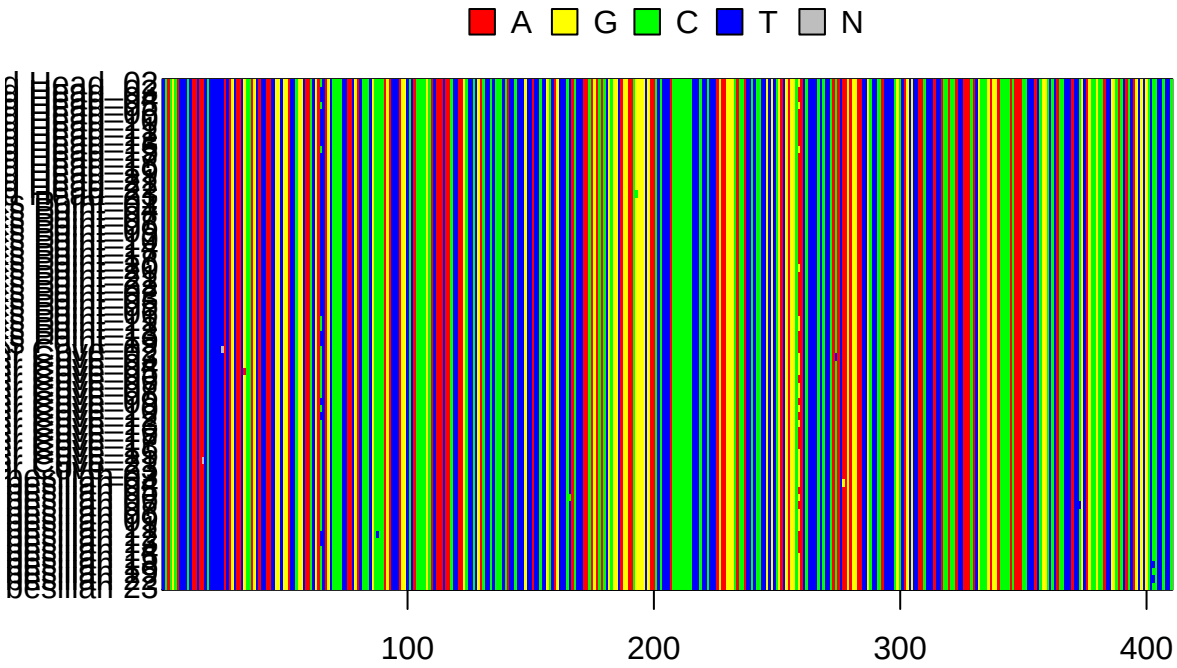
Setup

```
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

Ldigitalis <-read.FASTA("Final_Lottia_Alignments/Lottia_digitalis.fasta")
```

Initial look at data

Align the data and look at a quick image of it

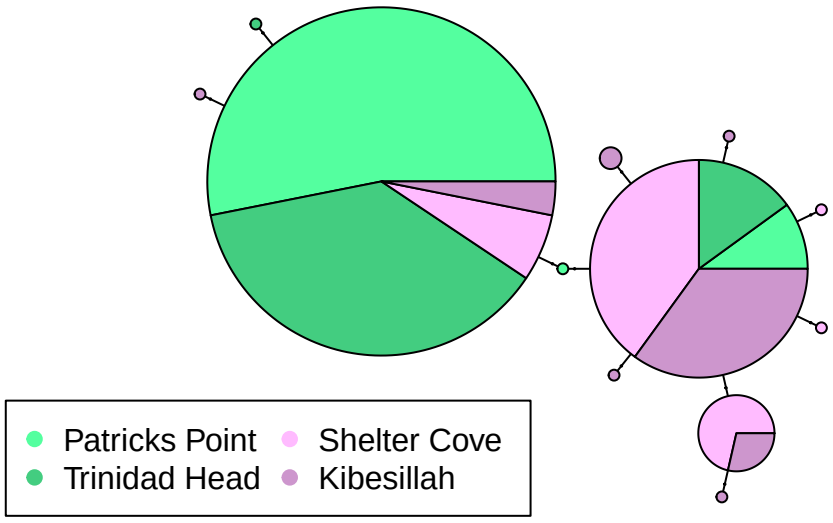
```
Ldigitalis <- muscle(Ldigitalis)

image.DNAbin(Ldigitalis)
```



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(Ldigitalis),"[A-Za-z]+\\s*[A-Za-z]+")

#Label the haplotypes and sort
h <- pegas::haplotype(Ldigitalis)
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","Kibesillah")]

#Pick some funky colors
mycolors <- c("seagreen1","seagreen3","plum1", "plum3")



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

```
Ldigitalis_g <- sequence2gtypes(Ldigitalis, strata = pop)

Ldigitalis_g <- labelHaplotypes(Ldigitalis_g)

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(Ldigitalis,pop=pop,fun=nuc.div)
Fs<-stratastat(Ldigitalis,pop=pop,fun=fusFs)

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

	Nucleotide.Diversity	Fu.s.Fs
Trinidad Head	0.0018902	0.4582495
Patricks Point	0.0011168	-0.3774818
Shelter Cove	0.0023789	-1.4136598
Kibesillah	0.0034756	-4.1917041

```
Ldigitalis_g
```

```
## $gtypes
##
```

```
## <<< gtypes created on 2019-10-25 11:28:40 >>>
##
## Contents: 69 samples, 1 locus, 4 strata
##
## Strata summary:
##                num.samples num.missing num.alleles prop.unique.alleles
## Kibesillah              16           0           8           0.6250000
## Patricks Point          20           0           3           0.3333333
## Shelter Cove            17           0           5           0.4000000
## Trinidad Head           16           0           3           0.3333333
##                heterozygosity
## Kibesillah          0.8083333
## Patricks Point      0.2789474
## Shelter Cove        0.7132353
## Trinidad Head       0.4250000
##
## $unassigned
## $unassigned$gene.1
## NULL
```

F-statistics

Took some hints from here, and here

Pairwise PhiST

Let's plot the pairwise PhiST statistics as a heatmap

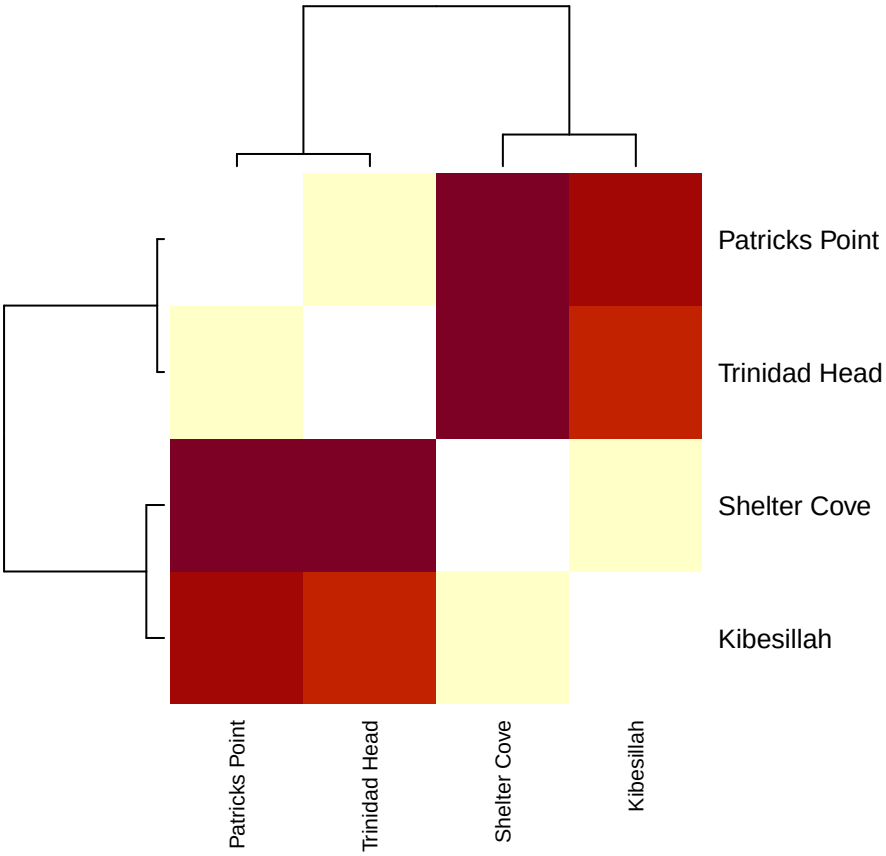
```
pairwise_phi <- pairwiseTest(Ldigitalis_g$gtypes, stats = "phist", nrep = 10000, model = "raw")
```

```
##
## <<< gtypes created on 2019-10-25 11:28:40 >>>
##  2019-10-25 11:28:40 : Pairwise tests : 10000 permutations
##    2019-10-25 11:28:40 : Kibesillah v. Patricks Point
##    2019-10-25 11:28:41 : Kibesillah v. Shelter Cove
##    2019-10-25 11:28:42 : Kibesillah v. Trinidad Head
##    2019-10-25 11:28:42 : Patricks Point v. Shelter Cove
##    2019-10-25 11:28:43 : Patricks Point v. Trinidad Head
##    2019-10-25 11:28:43 : Shelter Cove v. Trinidad Head
##
## Population structure results:
##                                 pair.label       PHIst PHIst.p.val
## 1    Kibesillah (16) v. Patricks Point (20)  0.36982491  0.00019998
## 2      Kibesillah (16) v. Shelter Cove (17) -0.01538677  0.56004400
## 3     Kibesillah (16) v. Trinidad Head (16)  0.33153153  0.00079992
## 4  Patricks Point (20) v. Shelter Cove (17)  0.41626979  0.00009999
## 5 Patricks Point (20) v. Trinidad Head (16) -0.02885108  0.74872513
## 6   Shelter Cove (17) v. Trinidad Head (16)  0.44285870  0.00019998
```

```
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(2,4,3,1),c(2,4,3,1)]

distance <- dist(phiST)
cluster <- hclust(distance, method="ward.D")
dendrogram <- as.dendrogram(cluster)
Rowv <- c(4,3,2,1)
dendrogram <- reorder(dendrogram, Rowv)
reorderfun <- function(d,w) { d }

heatmap(phiST,scale="none",Rowv=dendrogram, reorderfun=reorderfun, cexRow = 1, cexCol=0.8, legend = "co
```

```
## Warning in plot.window(...): "legend" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "legend" is not a graphical parameter

## Warning in title(...): "legend" is not a graphical parameter
```



AMOVA

```
#write.csv(pop,"amova_hypotheses_lottia.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_lottia.csv")
CapeMendocino<-amovahyps$CapeMendocino
```

```r
pop1<-as.factor(pop)

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

Test the Cape Mendocino hypothesis:

```r
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 1.979667e-04 1.979667e-04  1
## pop1          1.103392e-06 5.516958e-07  2
## Error         3.302693e-04 5.081065e-06 65
## Total         5.293393e-04 7.784402e-06 68
##
## Variance components:
##                     sigma2 P.value
## CapeMendocino  5.7346e-06  0.0000
## pop1          -2.6439e-07  0.9411
## Error          5.0811e-06
##
## Phi-statistics:
## CapeMendocino.in.GLOBAL (Phi_CT)          pop1.in.GLOBAL (Phi_ST)
##                       0.54349612                       0.51843824
##   pop1.in.CapeMendocino (Phi_SC)
##                      -0.05489084
##
## Variance coefficients:
##        a        b        c
## 17.13131 17.33158 34.43478
```

```r
# 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]

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

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