The role of oceanographic connectivity in population differentiation
coastalNet Package

—

2024-03-23

Oceanographic connectivity driven by the direction and intensity of ocean currents can shape the distribution of intraspecific biodiversity (i.e., the genetic differentiaiton levels).

This code example explores how oceanographic connectivity influences genetic differentiation in a kelp species (Laminaria ochroleuca). It starts by loading data on kelp sampling locations and their genetic differences. Then, it retrieves oceanographic connectivity information from a database for the study region. The code calculates pairwise connectivity probabilities between sampling sites, considering multigeneration stepping-stone connections. Next, it builds a statistical model to see if higher connectivity is linked to greater genetic differentiation among kelp populations. The code visualizes this relationship with a scatterplot and creates a map to show the connections between sampling sites, with thicker lines indicating stronger connectivity.

By combining oceanographic connectivity information derived from coastalNet package withh empirical genetic data, this script demonstrates that intraspecific biodiversity can be highly structured by connectivity driven by oceanographic transport and barriers.

Here's a summary of the key steps and functionalities encapsulated in the code:

Environment Preparation and Package Loading

Cleans the R environment and forces garbage collection to ensure a clean workspace. Loads necessary R packages for the analysis, which include coastalNet package.

```
1  # Clean environment and load packages
2  rm(list = ls())
3  gc(reset=TRUE)
4
5  library(coastalNet)
6
7  library(ggplot2)
8  library(sf)
9  library(lme4)
10 library(rnaturalearth)
11 library(viridis)
12 sf_use_s2(FALSE)
```

Data Loading

Imports two CSV files from a GitHub repository: laminariaRecords: Contains geographical coordinates (longitude and latitude, WGS84) of sampled sites for Laminaria ochroleuca. laminariaPopDifferentiation: Contains pairwise genetic differentiation estimates between the sampled sites.

```
1
2 # Load data.frame containing coordinates (as longitude and longitude, WGS84) of sites
      sampled for the marine species Laminaria ochroleuca.
3 laminariaRecords <- read.csv("https://raw.githubusercontent.com/jorgeassis/coastalNet/main/
      vignettes/data/Laminaria-ochroleuca-Coords.csv", sep=";", header = TRUE)
4
5 # Load data.frame containing pairwise genetic differentiation estimates between coordinate
      sites
6 laminariaPopDifferentiation <- read.csv("https://raw.githubusercontent.com/jorgeassis/
      coastalNet/main/vignettes/data/Laminaria-ochroleuca-JostD.csv", sep=";", header = FALSE)
```

Connectivity Analysis

Initializes a local database for storing analysis results (if not already present). Determines hexagon IDs defining the study region based on the extent of the sampled sites, with a specified buffer. Calculates connectivity events within the study region for a specified period (120 days), considering all years, months, and days. Obtains pairwise connectivity estimates between the sampled sites using a forward connectivity type, considering the probability of connectivity and allowing for stepping-stone connections.

```
1 # Load database
2 getDataBase(myFolder="Database", overwrite=FALSE)
3
4 # Get hexagon IDs that define the study region
5 hexagonIDRegion <- getHexagonID(obj=laminariaRecords, level="extent", buffer=5, print=TRUE)
```


Figure 1: Hexagon IDs (in black) defining the study region

```
1 # Get connectivity events for the study region (all years, all months, all days, 120 days
      period)
2 connectivityEvents <- getConnectivityEvents(hexagonID=hexagonIDRegion, period=120 )
3
4 # Get hexagon IDs of the sampling sites
5 hexagonIDSites <- getHexagonID(obj=laminariaRecords, level="site", buffer=0, print=FALSE)
6
7 # Get pairwise connectivity estimates between coordinate sites
```

```
8  pairwiseConnectivity <- getPairwiseConnectivity(connectivityEvents, hexagonIDFrom=
       hexagonIDSites, connType="Forward", value="Probability", steppingStone=TRUE)
```

Data Preparation for Modeling

Constructs a data frame to match pairs of oceanographic connectivity with pairs of population differentiation
for all site pairs, excluding self-comparisons. Cleans the data frame by removing zero connectivity values
and missing data, and applies a negative log transformation to the connectivity values.

```
1  # Produce and data.frame matching pairs of oceanographic connectivity and pairs of
       population differentiation
2  modelDataFrame <- data.frame()
3  for( from in 1:nrow(laminariaRecords)) {
4    for( to in 1:nrow(laminariaRecords)) {
5      if( from == to ) { next }
6      modelDataFrame <- rbind(modelDataFrame,data.frame(from = from, to = to, connectivity =
           mean(pairwiseConnectivity$connectivityMatrix[from,to], pairwiseConnectivity$
           connectivityMatrix[to,from], na.rm=T), differentiation = laminariaPopDifferentiation
           [from,to]))
7    }
8  }
9
10 # Remove zero connectivity values, missing values and Log-transform connectivity
11 modelDataFrame <- modelDataFrame[modelDataFrame$connectivity != 0 ,]
12 modelDataFrame <- modelDataFrame[complete.cases(modelDataFrame),]
13 modelDataFrame$connectivity <- -log(modelDataFrame$connectivity)
```

Statistical Modeling

Fits a mixed-effects linear model (using lmer) to examine the relationship between oceanographic connectivity
and population differentiation, with random effects for the originating site. Generates a linear model to
compare observed versus predicted population differentiation, calculating the adjusted R-squared value and
Pearson's correlation coefficient to assess model fit.

```
1  # Model oceanographic connectivity to population differentiation
2  model <- lmer(differentiation ~ connectivity + (1|from), modelDataFrame, REML=F)
3  observedPredictedDF <- data.frame(Observed=(modelDataFrame$differentiation), Predicted=
       predict(model))
4  model <- lm(Observed ~ Predicted, data = observedPredictedDF)
5  r2 <- summary(model)$adj.r.squared
6  Pearson <- cor(observedPredictedDF)[1,2]
```

Visualization

Produces two plots: A scatter plot showing observed versus predicted population differentiation, including
a regression line, and displaying the adjusted R-squared and Pearson's correlation coefficient. A map visu-
alizing the stepping-stone oceanographic connectivity between populations, overlaying the sampled sites on
a world map and depicting connectivity paths.

```
1  plot1 <- ggplot() +
2    geom_point(data = observedPredictedDF, aes(x=Observed, y=Predicted), color="#000000", fill
         ="#000000", size=2 ) +
3    geom_point(data = observedPredictedDF, aes(x=Observed, y=Predicted), color="white", fill="
         white", size=1 ) +
4    geom_smooth(data = observedPredictedDF, method=lm, aes(x=Observed, y=Predicted), linetype
         = "dashed", fill="#c5593c", col='black', size=0.5, alpha = 0.5) +
```

```
5   xlab(paste0("Observed population differentiation")) + ylab("Predicted population
        differentiation") +
6   theme_minimal() +
7   theme( panel.grid.major = element_blank() ,
8          text = element_text(size=12) ,
9          axis.title.y = element_text(margin = margin(t = 0, r = 18, b = 0, l = 0)) ,
10         axis.title.x = element_text(margin = margin(t = 18, r = 0, b = 0, l = 0)) ,
11         legend.title = element_blank()) +
12         annotate("label", alpha = 0.5, label.padding=unit(0.5, "lines"), x = 0, y = 0.735,
                hjust=0,vjust=1 , label = paste0("Adjusted R2: ", format(round(r2, 3), nsmall =
                    3),"\nPearson's Corr.: ",format(round(Pearson, 3), nsmall = 3)))
13
14 plot1
```
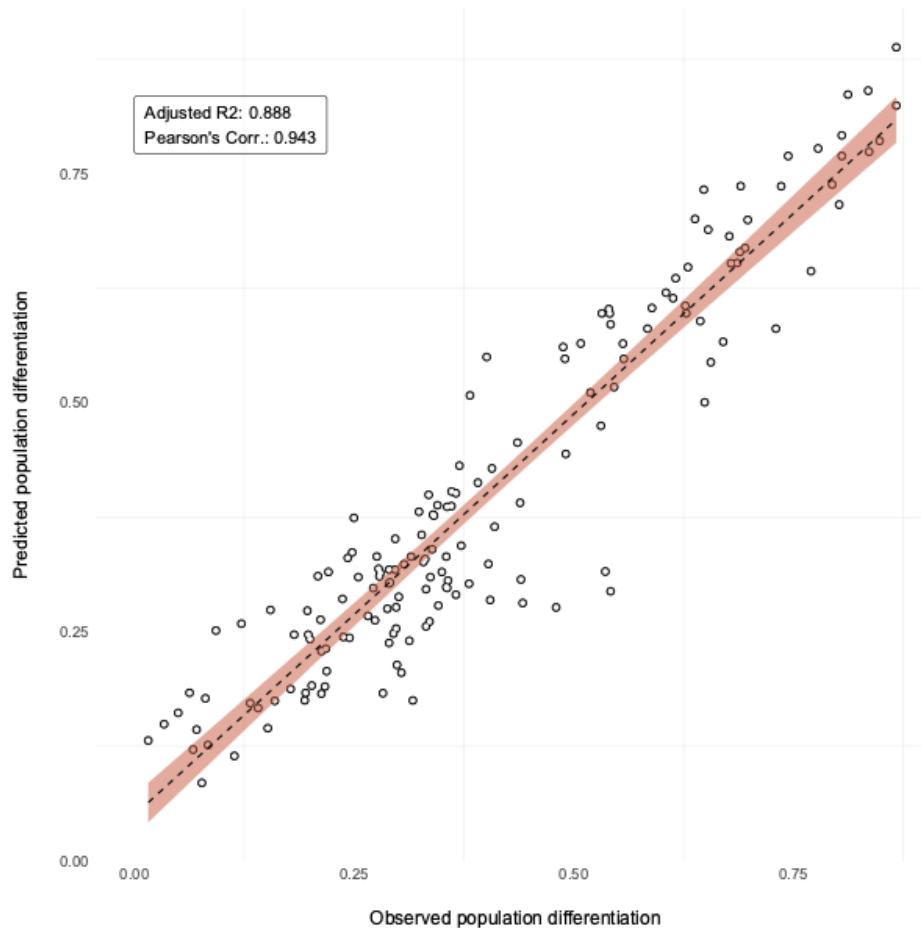

Figure 2: Role of oceanographic connectivity to population differentiation

```
1 # Map oceanographic connectivity between populations
2 mappedConnectivity <- mapConnectivity(connectivityPairs=pairwiseConnectivity$
      connectivityPairs)
3
4 # Load the worldmap and crop to the atudy region
5 worldMap <- ne_countries(scale = "medium", returnclass = "sf")
6 worldMap <- st_crop(worldMap,c(xmin=min(laminariaRecords[,1])-5,xmax=max(laminariaRecords
      [,1])+5,ymin=min(laminariaRecords[,2])-2.5,ymax=max(laminariaRecords[,2])+2.5))
7
8 # Get hexagon IDs that retrieved oceanographic connectivity estimates
9 hexagonIDConnected <- unique(c(mappedConnectivity$mappingData$FromHexagon,mappedConnectivity
```

```
        $mappingData$FromHexagon))
10
11  # Get a data.frame of the location of hexagons that retrieved oceanographic connectivity
        estimates
12  data("hexagonCells")
13  hexagonCellsConnected <- hexagonCells[hexagonCells$ID %in% hexagonIDConnected ,1]
14  hexagonCellsConnected <- st_coordinates(st_centroid(hexagonCellsConnected))
15
16  # Make a plot of the oceanographic connectivity between populations
17  plot2 <- ggplot() +
18    geom_sf(data = worldMap , fill="#CDCDCD", colour = "#9E9E9E" , size=0.25) +
19    geom_point(data = hexagonCellsConnected , aes(x = X, y = Y), colour = "#000000",size=2.5) +
20    geom_point(data = hexagonCellsConnected , aes(x = X, y = Y), colour = "#FFFFFF",size=1.25)
        +
21    geom_sf(data = mappedConnectivity$lineConnections , linewidth = 0.35 , aes(colour = Value)
        , alpha=0.75) +
22    scale_color_gradientn(colours=rev(magma(6)),na.value = NA, trans = "log") +
23    theme_minimal() + theme(axis.title.x=element_blank(),
24                            axis.ticks.x=element_blank(),
25                            axis.title.y=element_blank(),
26                            axis.ticks.y=element_blank(), legend.position = "none") +
27    coord_sf()
28
29  plot2
```
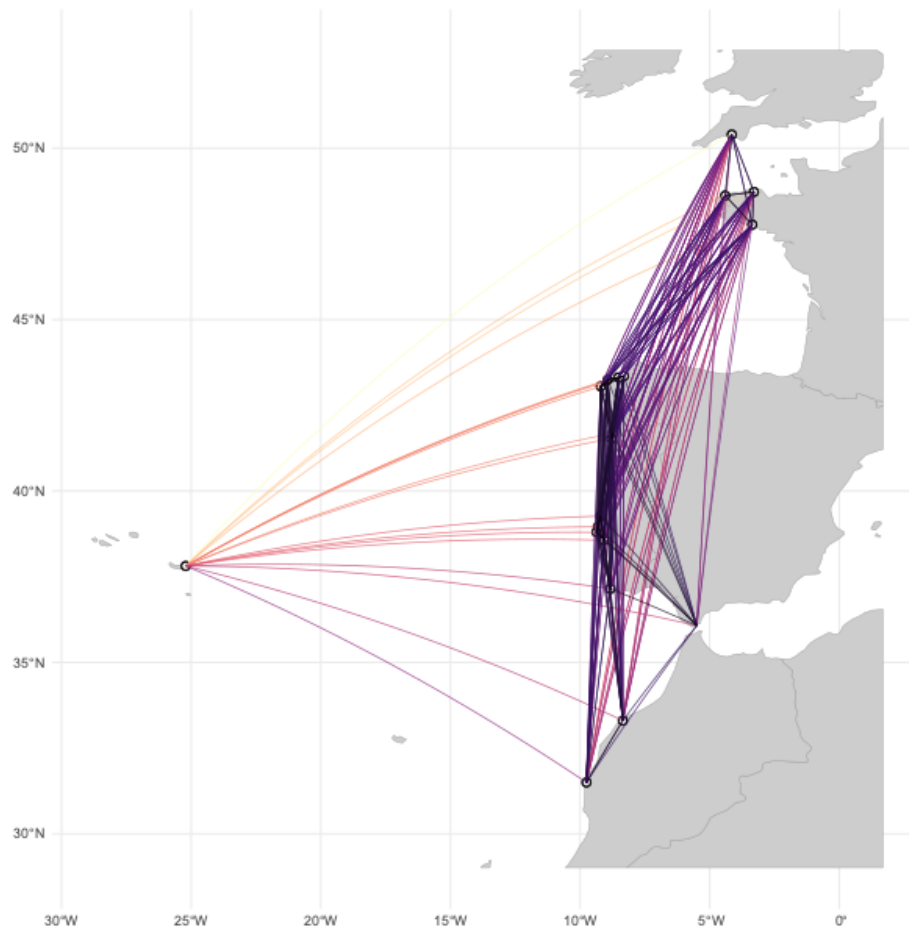

Figure 3: Stepping-stone oceanographic connectivity between populations