**#A2024**

**dr Bartosz Łabiszak, Zakład Ekologii Roślin i Ochrony Środowiska**

**Genomika ewolucyjna i populacyjna**

**LAB 2**

**Zadanie IV. Analiza struktury populacji sosny błotnej**

1. Wykonaj analizę PCA na danych SSR na poziomie osobników. **Wklej ryciny do protokołu**.

# analiza PCA wymaga kompletności danych – nasze analizy wykazały

# niewielkie ilości missing data – dlatego wykonamy imputację braków

# danych wstawiając zamiast nich wartości średnie

**x = adegenet::tab(Genind\_object, NA.method = "mean")**

# Wykonanie PCA

**pca1 = ade4::dudi.pca(x, scannf = FALSE, scale = T, nf = 3, center = T)**

# Ile procent zmienności wyjaśniają główne składowe

**factoextra::get\_eigenvalue(pca1)**

**factoextra::fviz\_eig(pca1, addlabels = TRUE, ylim = c(0, 10))**

# wizualizacja danych

# stworzenie wektora kolorów

**kolory <- c(rep("brown1",4),rep("green4",2), rep("dodgerblue",2),rep("yellow2", 4) )**

**ind.pca <- factoextra::fviz\_pca\_ind( pca1, geom = "point", geom.ind = "point",**

**pointshape = 21,fill.ind = pop(Genind\_object),mean.point = F,pointsize = 2,addEllipses = F,palette = kolory)**

**ind.p +theme\_classic**

# wykonaj również PCA usuwając argument palette – tak by każda z

# populacji otrzymała swój kolor

1. Wykonaj analizę PCoA na poziomie **populacji i gatunków**. **Wklej wynik analizy do protokołu**

**# Analiza na poziomie populacji**

# W pierwszym kroku zamieniamy obiekt genind na genpop

**genpop\_object <- genind2genpop(x = XXX)**

# obliczamy dystans Nei (method = 1)

**d0<- dist.genpop(genpop\_object,method = 1 )**

# czy dystans jest euklidesowy? Wymaga tego analiza PcOA.

**is.euclid(d0)**

# zamiana na dystans euklidesowy

**d1 <- cailliez(d0, TRUE)**

# wykonujemy analizę pcoa

**pcoa <- wcmdscale(d1,k = 2,eig = T)**

# oblicz procent wyjaśnionej zmienności przez główne składowe

# (niestety, funkcje z pakietu factoextra nie działają z obiektem

# utworzonym funkcją wcmdscale()

# wykorzystamy zatem kod

**percent\_pcoa <- pcoa$eig/sum(pcoa$eig) \*100 %>% round(digits = 2)**

**barplot(percent\_pcoa)**

# wykorzystaj podany poniżej kod do wizualizacji wyników PCoA (pcoa$li)

# stwórz ramkę danych zawierającą dane z obiektu pcoa oraz informacje o # populacji i gatunku do wizualizacji wyników

**pcoa\_df <- …**

**pcoa\_plot <- ggplot(pcoa\_df, aes(x = PC1, y = PC2,**

**color = spec, label = pop))+**

**geom\_point(size = 4)+**

**geom\_text(nudge\_x = 0.04)+**

**theme\_classic()+**

**scale\_color\_manual(values = c("green4", "yellow2", "brown1", "dodgerblue"))**

**# Analiza na poziomie gatunków**

# utwórz obiekt genpop, dodając do funkcji genind2genpop argument **pop**

# który zawiera przyporządkowanie każdego osobnika do gatunku

**spec\_ind <- c(rep("PUG", 94), rep("PM", 50), rep("PUN", 48), rep("PS", 108))**

**genpop\_object\_spec <- genind2genpop(x = Genind\_object, pop = spec\_ind)**

# kolejne kroki wykonaj analogicznie jak w poprzednim punkcie,

# modyfikując w odpowiedni sposób kod

1. Wykonaj drzewo filogenetyczne metodą UPGMA dla populacji i gatunków, bazując na dystansie Nei. **Wklej oba drzewa do protokołu.**

# wykorzystaj dystans z poprzedniego zadania

**plot(upgma(d1))**

# analogicznie dla dystansu na poziomie gatunków

1. Wykonaj analizę klastrowania do sprawdzenia struktury populacji. Wykorzystaj w tym celu metody podobnej do **STRUCTURE**, opartej na sparse negative matrix factorization (**SNMF**) w paczce **R LEA**. Metoda ta pozwoli na oszacowanie **liczby klastrów genetycznych (K)** które tworzą nasze osobniki. Więcej o tej metodzie przeczytasz tu:

https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.12382

Format pliku wykorzystywany przez LEA **to .geno i lfmm** najprostszą i polecaną przez twórców drogą do stworzenia tych plików jest z użyciem funkcji: **struct2geno()** – która wymaga jednak plików wejściowego w formacie STRUCTURE. Najprostsza droga do tego pliku to wykorzystanie pakietu hierfstat i przejście przez drogę: **genind -> hierfstat -> STRUCTURE**

# genind do hierfstat

**hierfstat\_object <- hierfstat::genind2hierfstat(genind\_object)**

# zapisujemy hierfstat to pliku w formacie .str – do working directory!

**hierfstat::write.struct(dat = … ,ilab =… ,pop = …,fname = XXX.str)**

# ilab to wektor z oznaczeniem osobnika, pop przynależność do populacji

# fname nazwa pliku

# zamina STRUCTURE na geno

**LEA::struct2geno(input.file = ,ploidy = 2,FORMAT = 2,extra.column = 2)**

# wykonanie analizy LEA

**project = NULL**

**project = LEA::snmf("XXX.geno",**

**K = 1:15,**

**entropy = TRUE,**

**repetitions = 10,**

**project = "new",**

**alpha = 100)**

# sprawdzenie optymalnej ilości klastrów

**plot(project, col = "blue", pch = 19, cex = 1.2, type = "b")**

Im niższa wartość cross entropy, tym lepiej nasz model uwzględnia strukturę populacji. Czasami jednak występuje dalszy spadek, więc możemy wybrać wartość K, dla której pierwszy raz zanotowaliśmy największy spadek.

Wykonaj wykres dla najlepszego K z użyciem webservera <http://pophelper.com/> Plikiem wsadowym jest w tym wypadku plik z **rozszerzeniem .Q**. Znajdziesz go w swoim folderze roboczym, jego nazwa jest tożsama z nazwą nadaną plikowi .**geno**. Dla każdego K utworzony został folder z nazwą **run1-run10**, załaduj pliki dla wszystkich runów w obrębie best K jako input. Następnie w zakładce plot zaznacz wszystkie pliki, wykonaj opcję **Align cluster**, a następnie **Merge repets.** Wybierz schemat kolorystyczny i dodaj informację o populacji dla każdego osobnika.

**write.table(x = pops,file = "pops.txt",quote = F,col.names = F, row.names = F)**

Następnie powtórz powyższą analizę wyłącznie dla populacji sosny błotnej.

**sub\_genind <- poppr::popsub(genind\_object,sublist = c(1:4))**

# sublist to wektor z populacjami, które chcemy zachować

# wykonaj analogiczne kroki jak w wypadku analiz dla wszystkich populacji

**Wklej oba wykresy do protokołu.**

**Pytania:**

1. Jaki procent zmienności zawartej w oryginalnych danych **oddaje analiza PCA w układzie dwóch głównych składowych**?

2. Czy analiza PCA na poziomie osobników pozwala na rozróżnienie **czterech gatunków sosen**?

3. Czy analiza PCoA przeprowadzona na poszczególnych populacjach pozwala na rozróżnienie czterech badanych gatunków sosen **– czy populacje jednego gatunku tworzą grupy odrębne od populacji pozostałych gatunków**?

4. Który gatunek sosny jest **najbardziej podobny** (najbliżej spokrewniony) z sosną błotną? Który gatunek sosny jest **najdalszym krewnym** sosny błotnej?

5. Jaki procent zmienności zawartej w oryginalnych danych oddaje **analiza PCoA** w układzie dwóch głównych składowych na poziomie gatunków?

6. Ile głównych **grup genetycznych można wydzielić na otrzymanym drzewie**? Czy grupy te są zbieżne z przynależnością gatunkową badanych sosen?

8. Czy obraz wzajemnych relacji genetycznych na drzewie filogenetycznym pomiędzy populacjami sosny błotnej **jest zbieżny z wynikami analizy PcOA**?

9.Ile klastrów genetycznych wydziela analiza LEA w badanej próbie przy uzwglednienieniu wszystkich populacji? Czy odpowiadają one badanym gatunkom?

10. Czy wyniki analizy LEA przemawiają za istnieniem **substruktury w populacjach sosny błotnej**? Która populacja sosny błotnej wykazuje **największą odrębność genetyczną** względem pozostałych?

**Zadanie V. Analiza zróżnicowania populacji sosny błotnej**

1. Oblicz poziom zróżnicowania genetycznego mierzonego **miarą FST** dla wszystkich populacji badanych gatunków oraz na poziomie badanych gatunków populacji. **Wklej** **rycinę przedstawiającą FST dla par porównań do protokołu**.

# obliczenie pairwise FST dla wszystkich populacji

**pops\_fst <- hierfstat::genet.dist(Genind\_object, method = "WC84") %>% round(digits = 3)**

# obliczenie fst między gatunkami

# najpierw zmodyfikujemy obiekt Genind, nadając mu inny wektor populacji

**pop(Genind\_object)**

# stwórz kopie obiektu Genind

**Genind\_object2 <- Genind\_object**

**adegenet::pop(Genind\_object2) <- spec\_ind** #wektor z poprzedniego zadania

**spec\_fst <- …**

# wizualizacja fst

# zamień obiekt zawierający FST na matrycę

**fst\_matrix\_pop <- …**

**corrplot(corr = fst\_matrix\_pop,is.corr = F,method = "color",bg = "grey80",type = "lower",diag = F)**

# wykonaj corrplot również dla spec\_fst

1. Wykonaj [test Mantela](https://en.wikipedia.org/wiki/Mantel_test) w celu przetestowania następujących hipotez dotyczących związku pomiędzy dystansem geograficznym a genetycznym dzielącym badane populacje:

Ho - brak zróżnicowania genetycznego na badanym obszarze zasięgu gatunku

H1 – występuje korelacja dodatnia pomiędzy odległością genetyczną a geograficzną (ang. **IBD *isolation by distance***)

Wykonaj wariant tej analizy z uwzględnieniem **wszystkich populacji i wyłącznie dla populacji *P. uliginosa***

# jako oszacowanie dystansu genetycznego wykorzystaj miarę fst z #poprzedniego zadania

**gendist <- …**

#map\_df utworzyłeś w zadaniu III w pubkcie g)

**geodist <- pegas::geod(lon = map\_df$lon, lat = map\_df$lat)**

**mantel <- mantel.randtest(m1 = geodist,m2 = gendist, nrepet = 9999)**

**plot(mantel)**

**mantel**

# wykres testu mantela

**plot(as.vector(genet\_dist)~as.vector(geodist), pch=20)**

**abline(lm(as.vector(genet\_dist)~as.vector(geodist)))**

#sprawdzenie siły korelacji

**model1 <- (lm(as.vector(gendist)~as.vector(geodist)))**

**summary(model1)**

# wykonanaj analogicznie testu mantela dla populacji *P.uliginosa*

**geodist\_PUG <- … # użyj subettingu na obiekcie geodist**

**gendist\_PUG <- … # analogicznie jak wyżej**

**Wklej wykresy z wynikami testu Mantela do protokołu**.

**Pytania:**

1.Które gatunek wykazuje najmniejsze podobieństwo genetycznie do sosny błotnej?

1. Która para populacji w obrębie sosny błotnej wykazuje największą dywergencję?

2. Porównaj wartości Fst pomiędzy populacjami sosny błotnej z wartościami FST pomiędzy gatunkami, spójrz również na wartości Fst w porównaniach populacji w obrębie innych gatunków. Czy poziom zróżnicowania populacji sosny błotnej jest wysoki?

3. Jaki jest współczynnik korelacji między dystansem genetycznym a geograficznym dla badanych populacji sosny błotnej?

4. Czy wyniki testu Mantela są istotne statystycznie? Za przyjęciem której z hipotez przemawiają?