# 2024

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

**Genomika ewolucyjna i populacyjna**

**LAB 1**

## Cel ćwiczen:

Na kolejnych trzech spotkaniach będziemy pracować nad zadaniem problemowym. Nadrzędnym celem tych spotkań będzie: ocena **zmienności genetycznej (LAB\_1)** sosny błotnej w porównaniu **do taksonów referencyjnych**, określenie **struktury genetycznej i poziomu zróżnicowania (LAB\_2)** populacji tego gatunku w kontekście ochrony jego zasobów genetycznych. Dowiemy się również o metodach wnioskowania o historii demograficznej taksonów **(LAB\_3).**

## Opis danych do wykonania ćwiczeń

Plik do ćwiczeń: **Genotypy.csv** - zawierają genotypy sekwencji mikrosatelitarnych (**SSR**) loci jądrowego DNA badanych osobników oraz **współrzędne geograficzne badanych populacji**. Analizowane dane pochodzą z 4 izolowanych populacji rzadkiej sosny błotnej (*Pinus uliginosa*) oraz populacji referencyjnych, reprezentujących trzy spokrewnione gatunki sosen (*Pinus sylvestris*, *Pinus mugo*, *Pinus uncinata)*.

Obraz zawierający mapa

Opis wygenerowany automatycznie   
Pinus uliginosa - sosna błotna ▲

Pinus mugo - kosodrzewina ■

Pinus uncinata - sosna hakowata◆

Pinus sylvestris - sosna zwyczajna ●

Z uwagi na charakter zajęć i większą wagę poświęconą interpretacji wyników, **kod w R będzie częściowo podany i wymagający jedynie niewielkich modyfikacji lub krótkiego dopisania własnego kodu**. Funkcje należące do danych pakietów będą oznaczone w kodzie z użyciem operatora „**::**”, np. **poppr::read.genalex()**. Każde zadanie kończy się **pytaniami**, które należy zaadresować w oparciu o wyniki w danej części. Podsumowanie wyników (**ostatnie zadanie w protokole LAB\_3**) jest najistotniejszą częścią trzech zajęć – w której na podstawie wyników analiz odniosą się Państwo do postawionego problemu w formie **sprawozdania**.

## Zadanie I - Wczytanie danych do R i ich pierwszy ogląd

Załaduj wymagane paczki w R (jeśli istnieje konieczność zainstaluj je)

**Zapoznaj się z informacjami zapisanymi w obiekcie genind (genind$...) i oraz użyj funkcji z pakietu adegenet: summary(), popNames(), pop(), IndNames().**

## Pytania:

1. Ilu loci SSR użyto w tym zestawie danych? **9**

2. Ilu badanych osobników dotyczą te dane? **300**

3. Ile populacji sosen przebadano? **length(unique(pines\_data@pop)) = 12**

4. Jakimi akronimami oznaczono cztery badane gatunki sosen?

* **PUG - *Pinus uliginosa***
* **PM – *Pinus mugo***
* **PUN – *Pinus uncinata***
* **PS – *Pinus sylvestris***

5. W jaki sposób zakodowana była informacja o allelach SSR w formacie genalex, jak zakodowana jest w genind? **W genalex mieliśmy dla każdego osobnika długość loci w parze jego chromosomów, w genid mamy macierz liczności wystąpień każdego locus**

**(możliwe wartości 0, 1, 2)**

6. Które loci odznacza się największą, a które najmniejszą liczbą alleli?

**loci8 największą, a loci2 najmniejszą**

7. Podaj motyw (tj. powtórzenie 2, 3, 4 nukleotydowe) dla każdego z loci:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Loci1 | Loci2 | Loci3 | Loci4 | Loci5 | Loci6 | Loci7 | Loci8 | Loci9 |
| 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 |

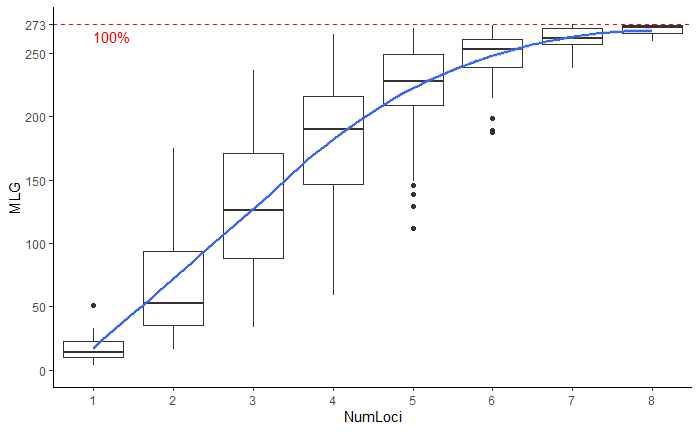
8. Jaki jest ogólny poziom missing data?

**loci1 loci2 loci3 loci4 loci5 loci6 loci7 loci8 loci9 Mean**

**0.0100 . 0.0033 0.0033 0.0033 0.0033 . . 0.0033 0.0030**

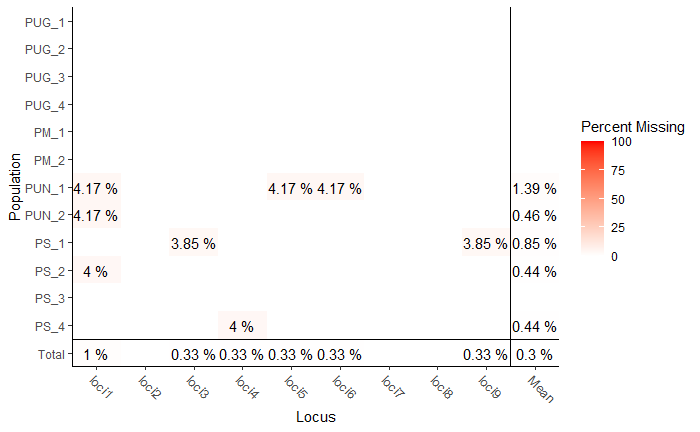
## Zadanie II- Sprawdzenie kompletności danych

1. Sprawdź przy pomocy krzywej akumulacji genotypów, czy dysponujemy odpowiednią rozdzielczością danych SSR .**Wklej do protokołu powstały wykres**.



**Na podstawie wykresu stwierdzam, że dysponujemy odpowiednią liczbą loci (o 1 większą niż wymagana).**

Sprawdź w których loci i w których populacjach występują brakujące allele **wklej do protokołu wykres** dla missing data podanych jako **wartość procentowa** i jako **liczba brakujących alleli**.

**Obraz zawierający tekst, diagram, zrzut ekranu, linia

Opis wygenerowany automatycznie**

**poppr::info\_table(gen = …,type = …, plot = …,low = …, scaled = …,percent = …)**

### # wykonaj tę analizę dla liczby i procent brakujących alleli

Wykorzystaj kod podany poniżej do wyłączenia loci i/lub osobników z wysokim poziomem missing data (przyjmując dopuszczalny poziom missing data = 10%)

#### XXX <- missingno(pop = …, type = …, cutoff = …)

## Pytania:

1. Czy w badaniu użyto wystarczającą ilość loci SSR?

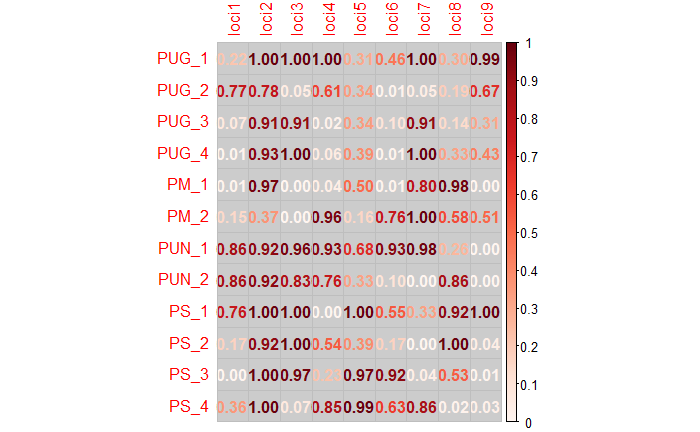
**Fajnie byłoby więcej, ale jest jak jest… w sumie, to czy nie było by też więcej missing data, gdybyśmy użyli większej liczby loci?**

2. Czy poziom brakujących danych przemawia za **wyłączeniem loci lub osobników z analiz**? Jeśli tak, to jakie to loci / osobniki? **NIE – bo poziom brakujących danych to po jednym osobniku w populacji któremu brakuje któregoś z 9 alleli**

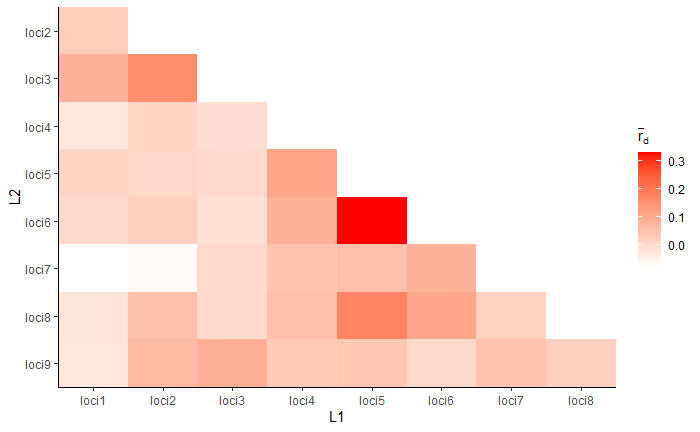
3. Czy istnieje inny sposób na braki w danych? **Tak – można imputować z występowania w danej populacji**

## Zadanie III- Oszacowanie zmienności wewnątrz populacji i na poziomie gatunków

1. Przeprowadź test równowagi HW **globalnie** (wszystkie osobniki z wszystkich populacji) oraz w **każdej z populacji z osobna**. **Wklej wykres do protokołu**



1. Sprawdź, czy nasze loci wykazują nierównowagę sprzężeń (LD). **Wklej wykres do protokołu**.

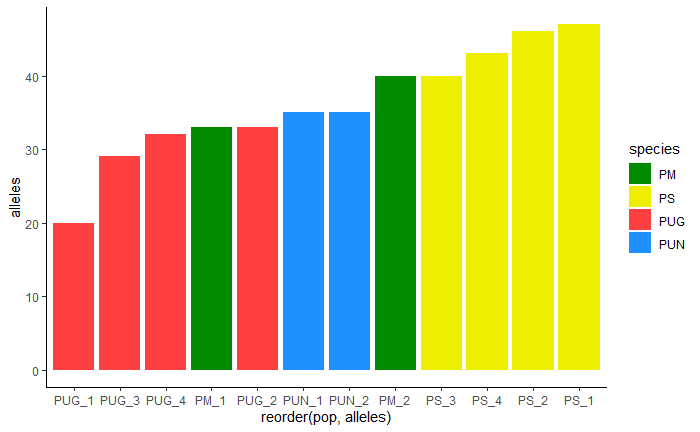
**Wygląda i to na to, że mamy tę nierównowagę sprzężeń – locus5 jest silnie sprzężone z locus6, prawda?**

**W samouczku zastosowanej funkcji jest napisane:**

**Obraz zawierający zrzut ekranu, tekst, Czcionka, linia

Opis wygenerowany automatycznie**

**Skoro indeks sprzężenia jest dla tej pary loci znacznie większy od zera, to znaczy, że istotnie są sprzężone.**

* 1. Wykonaj porównanie liczby alleli w populacjach i **przedstaw wyniki na wykresie słupkowym**. **Wklej wykres do protokołu**. ****

1. Oblicz **różnorodność alleliczną** (**mean allelic richness, boxplot over species**) **Wklej wykresy do protokołu**.

Obraz zawierający tekst, zrzut ekranu, Wielobarwność, diagram

Opis wygenerowany automatycznie

**# dodatkowo, korzystając z danych allelic.richness(XXX)$Ar**

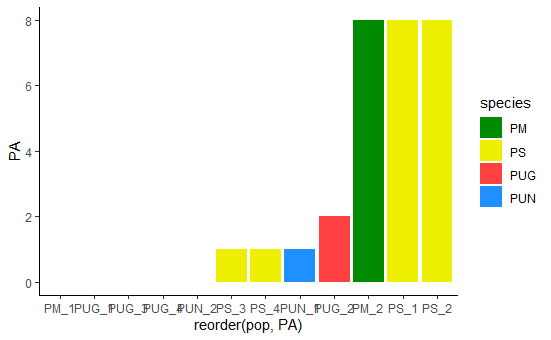
# możemy zrobić boxplot pokazujący AR na poziomie gatunkówObraz zawierający diagram, Prostokąt, kwadrat, zrzut ekranu

Opis wygenerowany automatycznie

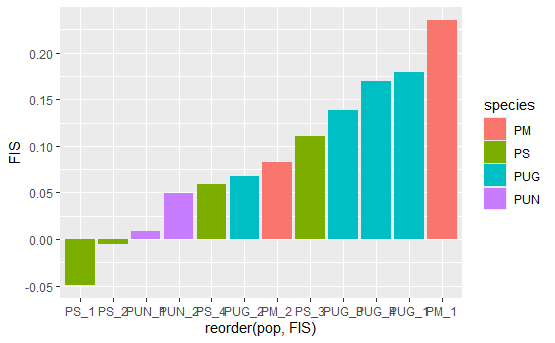
1. Oblicz **liczbę alleli prywatnych dla populacji i przedstaw ją na wykresie słupkowym**

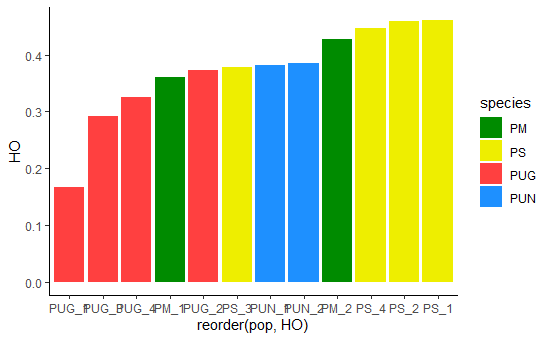
#### XXX <- poppr::private\_alleles(XXX) %>% apply(MARGIN = 1, FUN = sum)

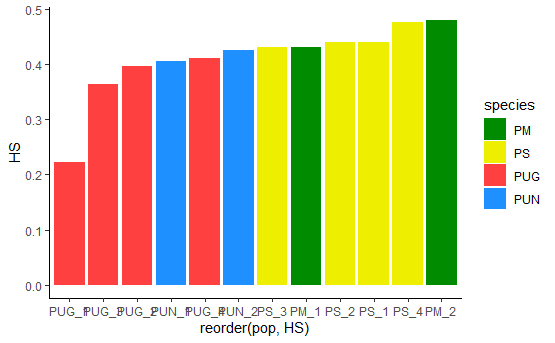
# do wizualizacji barplotu użyj analogicznego kodu jak w przypadku częstości alleli i AR



1. Oblicz heterozygotyczność obserwowaną **(Ho)** i spodziewaną **(He)** oraz współczynnik wsobności **(F)** i **przedstaw je na wykresie słupkowym**

# analogiczny kod użyj do obliczenia heterozygotyczności spodziewanej i współczynnika wsobności





1. Wykonaj **wizualizację miar zmienności genetycznej na mapie**. Wykorzystaj podany niżej kod:

##### Pytania:

1. Czy nasze loci wykazują odstępstwa od równowagi Hardego- Weinberga?

2. Uwzględniając wyniki **testu HW** i przyjmując próg istotności p = 0.05 odpowiedz na pytanie, czy któreś z loci należy wykluczyć **z analiz na podstawie testu HW** (czy wykazuje systematyczny wzór odstępstwa od HW dla wszystkich populacji)?

3. Czy **poziom korelacji** pomiędzy loci w teście LD jest **wysoki**?

4. Wymień populacje, które charakteryzują się najmniejszą oraz największą obserwowaną i spodziewaną heterozygotycznością.

5. Który gatunek odznacza się najmniejszą różnorodnością alleli?

6. Czy populacje sosny błotnej **odznaczają się wsobnością**?

7. Czy nasz dane wskazują na obniżoną zmienność u sosny błotnej?