**z2024**

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

**Genomika ewolucyjna i populacyjna**

**LAB 3**

**Zadanie VI. Analiza procesów demograficznych sosny błotnej**

W tej części będziemy pracować z programami zewnętrznymi: **Bottleneck, NeEstimator i DIY-ABCRF.** Programy te wymagają pliku zapisanego w formacie **GENEPOP – który nie jest tożsamy z obiektem genpop w adegenet.** Niestety, żaden z programów nie ma implementacji w R, a zwłaszcza **Bottleneck** trąci myszką – choć są nadal użyteczne i wykorzystywane w analizach (**wyzwanie dla Was bioinformatycy – nowa implementacja w ładnej paczce! )**

GENEPOP ma dwa sposoby zapisywania danych: z wykorzystaniem **dwóch i trzech** znaków na allel. Niestety – potrzebować będziemy obu, a funkcja konwertująca genind do formatu GENEPOP w R dają tylko format 3 znakowy. Co więcej, konwerter radzi sobie z missing data w takim formacie:

NA/NA -> 000000

ale „wywala się” i w przypadku genotypu

**NA/209 -> 00209 (jeden znak mniej)**

**Musimy więc przygotować 2 wersje plików i naprawić ten błąd.**

Eksportuj dane z obiektu genind do formatu pliku**.**

**install.packages("graph4lg")**

**library(graph4lg)**

**genind\_to\_genepop(x = GENIND, output = "all\_pops.txt")**

# domyślnie plik w Twoim working directory

Skorzystaj ze strony: <https://genepop.curtin.edu.au/genepop_op7.html> i narzędzia konwersji na inne formaty. Zaznacz opcję wyników **plain text i 3 digits**. Program zwróci uwagę na ww. błąd ze źle zakodowanym allelem – napraw ten błąd i wykonaj konwersję raz jeszcze. Zapisz powstały plik do pliku tekstowego (zaznacz który plik jest wersją dwu, a który 3 cyfrową kodowania allelu).

Upewnij się, że plik wejściowy do **Bottleneck** wygląda tak:

Obraz zawierający tekst, elektronika, zrzut ekranu, wyświetlacz

Opis wygenerowany automatycznie

a) Otwórz program **Bottleneck** przeprowadź analizę, której celem jest **wykrycie śladu niedawnego zmniejszenia wielkości populacji (genetyczne wąskie gardło, ang. genetic bottlneck)** z wykorzystaniem danych genetycznych. Więcej informacji o działaniu programu **Bottleneck** uzyskasz tu:

https://www1.montpellier.inrae.fr/CBGP/software/Bottleneck/pub.html

Obraz zawierający tekst, zrzut ekranu, wyświetlacz, oprogramowanie

Opis wygenerowany automatycznie

Zinterpretuj uzyskane wyniki dla 3 testów (opcja *summary*) - porównaj uzyskane w teście statystycznym wartości *p* do wartości krytycznej *p* = 0.05 przy następujących hipotezach:

Ho - brak śladu niedawnego spadku zmienności genetycznej

H1 –ślad niedawnego *bottlenecku*

**Pytania:**

1. Które z populacji sosny błotnej najprawdopodobniej przeszły niedawno przez genetyczne wąskie gardło? Czy wszystkie testy wspierają tę hipotezę?

**Wygląda na to, że wsw**

1. Czy w innych gatunkach także zaobserwowano potencjalny wpływ genetycznego wąskiego gardła?
2. Czy Twoim zdaniem wynik ten przemawia za tym, że dane genetyczne odzwierciedlają niedawny spadek liczebności populacji sosny błotnej na skutek presji antorpogenicznej?

b) Otwórz program **NeEstimato**r i załaduj w formacie GENEPOP (program działa z wersją 2 i 3 znakową kodowania alleli). Wykonaj oszacowanie **efektywnej wielkości populacji** z użyciem danych genetycznych wybierając ustawienia jak poniżej. O działaniu programu przeczytasz tu: [10.1111/1755-0998.12157](https://doi.org/10.1111/1755-0998.12157)

**Obraz zawierający tekst, zrzut ekranu, oprogramowanie, numer

Opis wygenerowany automatycznie**

**Pytania:**

1. Czy populacje sosny błotnej odznaczają się niższą efektywną wielkością populacji względem populacji taksonów referencyjnych? **TAK**

2. Jak należy zinterpretować wartość Infinity uzyskaną dla niektórych oszacowań? (zobacz paragraf **Negative or infinite estimates of Ne)** w artykule [10.1111/1755-0998.12157](https://doi.org/10.1111/1755-0998.12157) (pdf w załaczeniu) **Było dzielenie przez zero**

3 Wyjaśnij termin Minimum viable population. Które z populacji wykazują efektywną wielkość populacji mniejszą niż zakładana w świetle tej koncepcji?

**Ile osobników musi być w populacji, żeby przetrwała np. 100 lat**

Model zerowy N0 == N1 populacja nie zmienia się w czasie

Model n0 w między czasie Ndb na koniec N1, czy w przeszłości był bottleneck?

Porównujemy te dwa modele

Symulacje scenariuszy demograficznych w programie **DIYABC-RF**

Przeprowadzimy modelowanie demograficzne, porównujące 2 scenariusze:

**S1** brak zmian wielkości populacji w czasie vs

**S2** zakładający wąskie gardło ewolucyjne

Wykonamy test wyłącznie na populacjach sosny błotnej (**PUG1-PUG4**) w dwóch wariantach:

1. Z pominięciem informacji o strukturze populacji (**wszystkie populacje -> jedna grupa**)

2. Dla każdej z populacji z osobna (PUG1 & PUG2 & PUG3 & PUG4)

**Przygotowanie plików wejściowych do analizy:**

Przygotuj na bazie pliku **GENEPOP 3 digits** plik wejściowy do analiz. Plik powinien mieć rozszerzenie **.mss**, którego modyfikacja polega na dodaniu informacji o proporcji płci oraz pochodzeniu naszych loci:

Obraz zawierający tekst, zrzut ekranu, numer, wyświetlacz

Opis wygenerowany automatycznie

**UWAGA! Zwróć uwagę, że po POP nie powinno być żadnych spacji!**

Włącz **DIYABC-RF\_GUI.bat** i przejdź do pola

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Strona internetowa

Opis wygenerowany automatycznie

**Project settings**

Nadaj unikalną nazwę projektowi, wybierz odpowiedni rodzaj danych wejściowych oraz załaduj plik w formacie .mss

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Opis wygenerowany automatycznie

**Trening set simulations**

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Strona internetowa

Opis wygenerowany automatycznie

To pole umożliwia zdefiniowanie scenariuszy demograficznych, które będziemy testować, wybór priorów do naszych symulacji oraz określenie modelu mutacji naszych markerów.

W polu **Historical models** zdefiniuj 2 badane przez nas scenariusze. Pierwszy to:

N1

0 sample 1

drugi:

N0

0 sample 1

t2-db VarNe 1 Nb

t2 VarNe 1 Na

Wybierz opcję **validate models**.

W polu **Priors and conditions** zatwiedź domyślne ustawienia priorów.

W polu **Condition settings** ustaw:

t2>=db

Na>Nb

Nb=<N0

W polu**: Setup Microsat locus configuration** ustaw motywy każdego z loci SSR **(LAB 5, Zadanie 1 pytanie 7)**

W polu **Microsat/Sequence group priors** zaakceptuj ustawienia domyślne.

**Wykonaj 20 000 symulacji traning set** – następnie wykonaj **Prior and scenario checking**

(to ustawienie jest opcjonalne, pozwoli na sprawdzenie, czy nasze symulowane scenariusze mają parametry podobne do danych empirycznych).

Przejdź do zakładki: **Random Forest Anlysis**

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Opis wygenerowany automatycznie

W pierwszej kolejności wykonaj **Model choice.** Wybierz nazwę modelu i zatwierdź ją pprzyciskiem **validate**. Wykonaj RF na domyślnych ustawieniach. **Sprawdź, który model został wskazany jako najlepszy.** Pobierz wyniki z zakładki **Project administration**. W pobranym folderze sprawdź pliki:

**XXX\_out.predictions i XXX\_out\_graph\_lda.png**

Wykonaj esytmację parametrów **(t2, db, Na, Nb, Nx)**. dla modelu wybranego przez **DIYABC**. Wyniki estymacji będą w plikach **estimparam\_out.predictions**. Podaj wyniki wraz z przedziałami ufności 95%.Parametry czasowe są **wyrażone w liczbie pokoleń, przeskaluj je na lata** – (czas przemiany pokoleń u sosny to 20 lat).

**Powtórz powyższe analizy dla każdej z populacji *P.uliginosa* z osobna (PUG1 & PUG2 & PUG3 & PUG4)**

**Pytania:**

**1.** Czy wyniki analizy DIYABC potwierdzają scenariusz niedawnego zmniejszenia wielkości populacji?

2. Czy wzór ten jest spójny dla wszystkich populacji tego gatunku?

3. Na kiedy datowany jest epizod Bottlneck?

**Zadanie VI. Podsumowanie wyników.**

Zastanów się nad wynikami wszystkich przeprowadzonych analiz **(LAB\_1 – LAB\_3)**. Jakie wnioski można wyciągnąć na ich podstawie, w kontekście ochrony zasobów genetycznych sosny błotnej – czy w populacji tego gatunku można dostrzec ślady istnienia struktury genetycznej, jak się ma jego zmienność do zmienności pozostałych taksonów, czy z danych genetycznych można wyczytać ślady zmian wielkości populacji w przeszłości? Czy uważasz za uzasadnione podjęcie działań ochronnych dla tego gatunku? Jeśli tak, jakie środki należałoby według Ciebie przedsięwziąć?