

## Oszacowanie obciążenia genetycznego na podstawie danych genomowych – ewolucyjna perspektywa na genetykę konserwatorską

### Wstęp:

Celem ćwiczeń było zapoznanie się z metodyką wykorzystaną w pracy Speak i in. 2024: *Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations* (*Molecular Ecology Resources*). Metodyka została zastosowana w celu badania obciążenia genetycznego dwóch gatunków ptaków: biegusa łyżkodziobego (*C\_pyg*) oraz biegusa rdzawoszyjnego (*C\_ruf*).

Biegus łyżkodzioby (*Calidris pygmaea*) to mały, wędrowny ptak z rodziny bekasowatych. W sezonie lęgowym występuje na Półwyspie Czukockim, zimuje natomiast w Azji Południowo-Wschodniej. Jest jednym z najrzadziej występujących gatunków ptaków na naszej planecie – maksymalnie 200 par podchodzi corocznie do rozrodu, a gatunek zaklasyfikowany jest jako krytycznie zagrożony wyginięciem. W celu zwiększenia sukcesu reprodukcyjnego i ochrony populacji podejmowane są różnorodne działania. Jednym z takich działań było podjęcie prób zrozumienia w jakim stopniu ograniczona wielkość populacji, wpływa na erozję genetyczną i akumulację szkodliwych wariantów genetycznych w populacji.

Obciążenie genetyczne odnosi się do negatywnego wpływu szkodliwych mutacji na zdolność przetrwania i reprodukcji populacji. Mutacje te mogą prowadzić do obniżenia ogólnej wartości przystosowawczej (fitness) populacji, co w konsekwencji wpływa na jej dynamikę i zdolność do przetrwania w zmieniających się warunkach środowiskowych.

Użyte komendy do pracy z danymi oraz kod w programie Rstudio jest dostępny w repozytorium GitHub

## Materiały i metody:

W pracy został wykorzystany scaffold4 oraz pliki FASTQ po sekwencjonowaniu paired-end: *C\_pyg\_22* oraz *C\_ruf\_08*. Na surowych plikach wykonano kontrolę jakości i filtrowanie za pomocą FastQC oraz fastp. Następnym krokiem było indeksowanie genomu referencyjnego oraz mapowanie odczytów FASTQ do otrzymanego scaffoldu za pomocą bwa mem. Po indeksowaniu użyto bcftools call do wywoływania wariantów SNP w formacie VCF.

Kluczowym etapem pracy było znalezienie elementów ultrakonserwatywnych w badanych genomach. Mutacje w tych elementach mogą oznaczać wysoką szkodliwość i potencjalny wpływ na fitness osobnika. Po wyodrębnieniu takich elementów, z dodatkowym zakresem 1000 nukleotydów sąsiadujących, wykonano SNP calling oraz filtrację. W ostatnim etapie użyto zasobów genomowych kury domowej do przyrównania chromosomów i znalezienia takiego, który najlepiej odpowiada scaffoldowi użytego do mapowania. Do oszacowania szkodliwości mutacji wykorzystano wartości CADD (*Combined Annotation Dependent Depletion*), a do konwersji koordynatów pomiędzy różnymi wersjami genomu – narzędzie **CrossMap**. Ostatecznie przeprowadzono adnotację SNP.

W celu oszacowania całkowitego<sup>(1)</sup> i zrealizowanego<sup>(2)</sup> obciążenia genetycznego użyto poniższych wzorów.

$$\text{Genetic load (individual } k) = \sum_{i=1}^{L(\text{hom})} s_i + \sum_{j=1}^{L(\text{het})} 0.5s_j \quad (1)$$

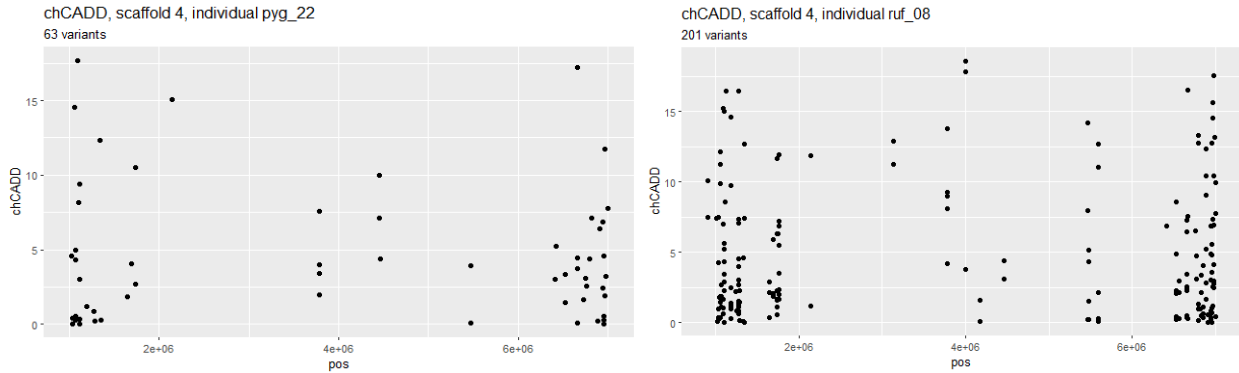
$$\text{Realized load (individual } k) = \sum_{i=1}^{L(\text{hom})} s_i + \sum_{j=1}^{L(\text{het})} h_j s_j \quad (2)$$

Do analizy statystycznej użyto m.in. testu korelacji Spearmana do zależności między pozycją SNP, a wartością chCADD oraz testu U Manna-Whitneya do porównania wartości chCADD między grupami. Do obliczeń oraz wykresu wykorzystano program Rstudio.

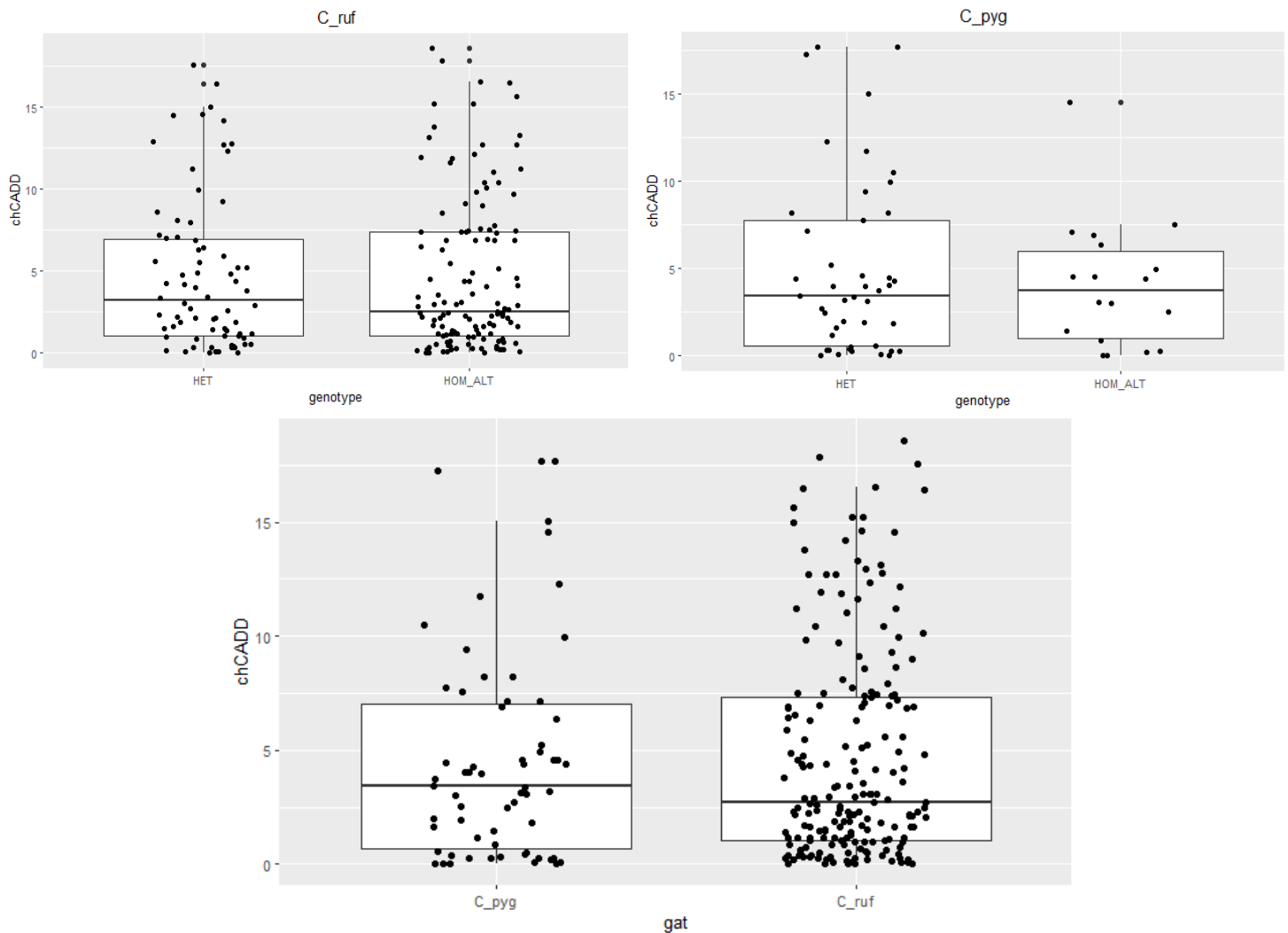
## Wyniki:

Dla obu gatunków policzono liczbę zidentyfikowanych wariantów oraz ich podział na homozygoty alternatywne i heterozygoty. Dla biegusa rdzawoszyjnego (*C\_ruf*) zidentyfikowano 201 wariantów, w tym 74 heterozygoty i 127 homozygot. Natomiast dla biegusa łyżkodziobego

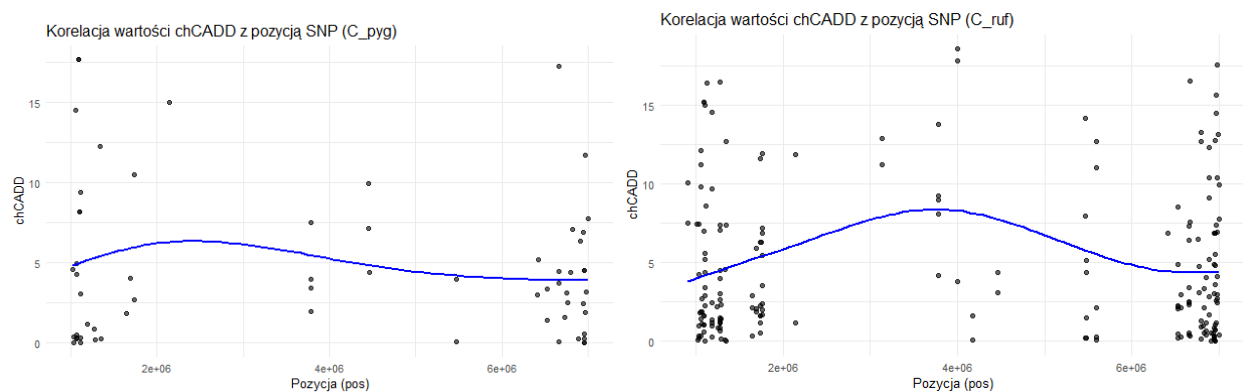
(*C\_pyg*) zidentyfikowano 63 warianty, w tym 45 heterozygot i 18 homozygot. Następnie wizualizowano wyniki wzdłuż scaffoldu dla obu gatunków.



Ze względu na brak normalności rozkładu danych, do analizy średnich wartości CADD między homozygotami a heterozygotami oraz między gatunkami zastosowano wykresy boxplot i test U Manna-Whitneya.



Kolejnym krokiem było wykorzystanie testu Spearmana oraz wizualizacja korelacji wartości CADD z pozycją SNP.



Główne badanie skupiło się na oszacowaniu całkowitego (GL) i zrealizowanego (RL) obciążenia genetycznego. Wyniki pokazano w tabeli, przedstawiając różnice między gatunkami.

Gatunek	CADD-hom	CADD-het	L-hom	L-het	GL	RL
C_pyg	72.49795	221.9137	18	45	6298.020675	2303.574615
C_ruf	592.83175	352.9381	127	74	88348.34343	77901.37449

## Dyskusja:

Na podstawie wyników można zauważyć, że wartości całkowitego i zrealizowanego obciążenia genetycznego znacznie różnią się między analizowanymi gatunkami. Biegus Rdzawoszyjny (*C\_ruf*) wykazuje znacznie wyższe obciążenie genetyczne niż biegus łyżkodzioby (*C\_pyg*). Wynik ten może wynikać z różnic w historii ewolucyjnej, wielkości populacji lub stopniu izolacji genetycznej tych gatunków. Wysoka wartość GL i RL dla *C\_ruf* sugeruje, że gatunek ten jest bardziej narażony na kumulację szkodliwych mutacji, co w dłuższej perspektywie może prowadzić do obniżenia zdolności adaptacyjnych populacji oraz zwiększenia ryzyka jej wyginięcia. Z kolei niższe obciążenie genetyczne u *C\_pyg* może wskazywać na lepsze zarządzanie populacją, mniejsze kumulowanie szkodliwych mutacji w przeszłości lub większy potencjał adaptacyjny tego gatunku.

Wyniki analizy zależności między wartościami CADD a pozycją SNP na scaffoldzie wykazały brak istotnej korelacji dla obu gatunków. Test Spearmana dla *C\_ruf* oraz *C\_pyg* dał

współczynniki korelacji bliskie zeru oraz wysokie p-wartości, co oznacza, że nie ma dowodów na zależność pomiędzy tymi zmiennymi. Sugeruje to losowe rozmieszczenie mutacji o wysokiej szkodliwości wzdłuż analizowanego skafolda, co jest zgodne z oczekiwaniami w przypadku braku specyficznego rozmieszczenia mutacji w genomie.

W trakcie analizy porównano również wartości CADD pomiędzy genotypami (homozygotami alternatywnymi i heterozygotami) oraz pomiędzy gatunkami. Test U Manna-Whitneya nie wykazał istotnych różnic w wartościach CADD między tymi grupami. Wynik ten sugeruje, że zarówno genotypy, jak i gatunki mają podobny rozkład szkodliwości mutacji, mimo że *C. ruf* wykazuje wyższe wartości całkowitego i zrealizowanego obciążenia genetycznego. Brak różnic w wartościach CADD między gatunkami może wynikać z podobnych mechanizmów wpływu mutacji na oba gatunki lub ograniczonego zestawu danych.

Wnioski z tej pracy pokazują, że analiza obciążenia genetycznego może dostarczyć cennych informacji o zdrowiu genetycznym populacji oraz potencjalnych zagrożeniach związanych z akumulacją mutacji. Wyniki te mogą być pomocne w opracowywaniu strategii ochrony obu gatunków, szczególnie w kontekście dynamicznie zmieniających się warunków środowiskowych i presji antropogenicznych.