

1. Cel projektu i powody podjęcia danej tematyki badawczej

Najnowsze trendy w hodowli bydła mlecznego zacierają w kierunku wyselekcjonowania osobników o wysokiej wydajności, jednocześnie charakteryzujących się dobrą zdrowotnością oraz efektywnym wykorzystaniem paszy. Dzięki temu dysponujemy coraz większą liczbą danych opisujących dobrostan i zdrowotność zwierząt. Tego typu cechy są zaliczane do kategorii tzw. „nowych cech” (ang. *novel traits*), ponieważ są one gromadzone dopiero od niedawna, a co za tym idzie ich uwarunkowanie genetyczne nie zostało jeszcze poddane intensywnej analizie. Celem projektu jest analiza genetycznego podłoża cech opisujących jakość nóg i racic w kontekście przyczyn kulawizny bydła mlecznego ze szczególnym uwzględnieniem efektów epistatycznych – czyli interakcji pomiędzy parami genów.

2. Schemat prowadzonych badań

Dane dostępne do analizy dotyczą 985 krów i obejmują cechy opisujące jakość nóg i racic oraz genotypy 76 934 polimorfizmów pojedynczego nukleotydu opisujące zmienność genetyczną analizowanych osobników. Dodatkowo posiadamy informacje o dokładnej lokalizacji polimorfizmów w genomie, o pochodzeniu krów, o przypadkach eliminacji krów ze stada i przyczynach tych eliminacji oraz informacje dodatkowe takie jak m. in. data urodzenia osobnika, czy liczba ukończonych laktacji.

Pierwszy etap prac obejmuje edycję danych w celu zidentyfikowania i usunięcia błędnych obserwacji. Następnie zostaną obliczone korelacje pomiędzy poszczególnymi polimorfizmami wyrażone jako współczynnik zaburzenia równowagi Hardyego-Weinberga. Kolejnym etapem będzie opracowanie funkcji, która opisuje zależność współczynnika zaburzenia równowagi Hardyego-Weinberga od odległości pomiędzy polimorfizmami zlokalizowanymi na tym samym chromosomie. Te polimorfizmy, które odbiegają od standardowej zależności zostaną wybrane do dalszych badań w myśl założenia, że między takimi polimorfizmami zachodzi zjawisko epistazy genetycznej. Następnym etapem prac polega na obliczeniu efektów genetycznych addytywnych i dominacyjnych dla wszystkich 76 934 polimorfizmów przy zastosowaniu dwu odmiennych podejść statystycznych: (a) wielu modeli, każdy uwzględniający jedynie efekt pojedynczego polimorfizmu połączony z późniejszym zastosowaniem korekty na wielokrotne testowanie uwzględniającej korelacje pomiędzy polimorfizmami oraz (b) pojedynczego modelu uwzględniającego efekty addytywne i dominacyjne wszystkich polimorfizmów równocześnie. Powyższe modele posłużą do wybrania polimorfizmów o efektach genetycznie addytywnych i dominacyjnych znaczących dla zmienności analizowanych cech jakości nóg i racic. Ostatnim etapem statycznego modelowania zmienności genetycznej jakości nóg i racic jest stworzenie modelu zawierającego polimorfizmy o znaczących efektach addytywnych, dominacyjnych i epistatycznych. W finalnej fazie projektu dla tych polimorfizmów przeprowadzimy adnotację genomową, czyli określimy ich położenie względem funkcjonalnych elementów genomu takich jak egzony, regiony UTR, czy regiony międzygenowe. Dla genów w pobliżu (maksymalnie 1 000 nukleotydów) lub wewnątrz których zlokalizowane są polimorfizmy o znaczących efektach wyznaczymy ścieżki metaboliczne zdefiniowane przez bazy danych: Kyoto Encyclopedia of Genes and Genomes oraz Reactome. Będzie to próba określenia jakie procesy metaboliczne mają wpływ na prawdopodobieństwo występowania chorób układu lokomocyjnego bydła.