

**AUTOREFERAT
OPIS DOROBKU I OSIĄGNIĘĆ NAUKOWYCH**

dr inż. Tomasz Suchocki

**KATEDRA GENETYKI
WYDZIAŁ BIOLOGII I HODOWLI ZWIERZĄT
UNIwersytet PRZYRODNICZY WE WROCŁAWIU**

Wrocław 2019

Spis treści

I	Posiadane dyplomy i stopnie naukowe	1
II	Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych	1
III	Prace stanowiące szczególne osiągnięcie naukowe wynikające z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.) pod wspólnym tytułem: „Użycie markerów genetycznych do wyjaśnienia zmienności genetycznej cech o złożonej strukturze dziedziczenia u bydła mlecznego”	2
IV	Opis szczególnych osiągnięć naukowych pod tytułem: „Użycie markerów genetycznych do wyjaśnienia zmienności genetycznej cech o złożonej strukturze dziedziczenia u bydła mlecznego”	3
V	Bibliografia	17

I Posiadane dyplomy i stopnie naukowe

Urodziłem się 17 marca 1981 roku we Wrocławiu. W latach 1996-2000 byłem uczniem XIV Liceum Ogólnokształcącego im. Polonii Belgijskiej we Wrocławiu, w klasie matematyczno-fizycznej z poszerzonym programem informatyki i języka angielskiego. Egzamin maturalny zdałem w 2000 r.

W latach 2000-2006 studiowałem na Wydziale Podstawowych Problemów Techniki Politechniki Wrocławskiej, kierunek Matematyka, specjalność Statystyka matematyczna. W dniu 6 lipca 2006 roku obroniłem pracę magisterską pt. „Estymacja parametrów rozkładu na podstawie danych otrzymywanych w chwilach losowych” (promotor dr hab. Alicja Jokiel-Rokita) i uzyskałem dyplom magistra inżyniera matematyki o specjalności statystyka matematyczna.

W latach 2006-2010 byłem słuchaczem Studiów Doktoranckich na Wydziale Biologii i Hodowli Zwierząt Uniwersytetu Przyrodniczego we Wrocławiu. W dniu 6 kwietnia 2011 roku obroniłem pracę doktorską na Uniwersytecie Wrocławskim (Wydział Biotechnologii) pt. „Modelling QTL effect on Bos taurus autosome 6 using random regression test day models” (promotor prof. dr hab. Joanna Szyda) i uzyskałem tytuł doktora nauk biologicznych w zakresie biotechnologii.

II Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

Od 21 lutego 2011 r. do 31 grudnia 2011 r. byłem zatrudniony w Katedrze Genetyki Wydziału Biologii i Hodowli Zwierząt Uniwersytetu Przyrodniczego we Wrocławiu na stanowisku asystenta. Od 1 stycznia 2012 do chwili obecnej jestem zatrudniony w Katedrze Genetyki Wydziału Biologii i Hodowli Zwierząt Uniwersytetu Przyrodniczego we Wrocławiu na stanowisku adiunkta. Mój dorobek naukowy obejmuje 48 prac (łącznie z pracami stanowiącymi szczególne osiągnięcie naukowe), w tym 17 oryginalnych prac naukowych (16 w bazie Journal Citation Reports), 30 prac w materiałach z konferencji o zasięgu krajowym i międzynarodowym oraz jeden artykuł popularnonaukowy. Sumaryczny Im-

Impact Factor wszystkich prac opublikowanych w czasopiśmie znajdujących się w bazie JCR wynosi 36,191, a suma punktów wg list Ministerstwa Nauki i Szkolnictwa Wyższego z roku wydania pracy wynosi 430 (475 wg aktualnej listy czasopiśm MNiSW). Całkowita liczba cytowań wszystkich prac w bazie Web of Science wynosi 175, a indeks Hirscha = 6.

III Prace stanowiące szczególne osiągnięcie naukowe wynikające z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.) pod wspólnym tytułem: „Użycie markerów genetycznych do wyjaśnienia zmienności genetycznej cech o złożonej strukturze dziedziczenia u bydła mlecznego”

1. Szyda, J., **Suchocki, T.**, Qanbari, S., Liu, Z., and Simianer, H. 2017. Assessing the degree of stratification between closely related Holstein-Friesian populations. *Journal of Applied Genetics*, **58**:521–526.
2. **Suchocki, T.**, Szyda J. 2015. Genome-wide association study for semen production traits in Holstein-Friesian bulls. *Journal of Dairy Science*, **98**(8), 5774–5780.
3. **Suchocki, T.**, Wojdak-Maksymiec K., Szyda J. 2016. Using gene networks to identify genes and pathways involved in milk production traits in Polish Holstein dairy cattle. *Czech Journal of Animal Science*, **61**(11), 526–538.
4. **Suchocki, T.**, Liu Z., Żarnecki A., Szyda J. 2017. Dissecting an interplay between genomic and pedigree sources of information for the estimation of breeding values for milk yield in Polish Holstein-Friesian dairy cattle in a one-step approach based on a random regression test day model. *Animal Science Papers and Reports*, **35**, 193–198.

IV Opis szczególnych osiągnięć naukowych pod tytułem: „Użycie markerów genetycznych do wyjaśnienia zmienności genetycznej cech o złożonej strukturze dziedziczenia u bydła mlecznego”

Identyfikacja regionów w genomie, w szczególności genów związanych ze zmiennością fenotypów cech ilościowych, jest bardzo ważna dla wzbogacenia wiedzy o zmienności genetycznej gatunku. U bydła mlecznego dane pochodzące z mikromacierzy SNP (ang. Single Nucleotide Polymorphism, SNP) stały się popularne i stosunkowo tanie. Możemy użyć dziesiątek tysięcy polimorfizmów pojedynczego nukleotydu dostępnych dla każdego osobnika w celu wykrycia istotnych związków fenotyp-genotyp. Wiele krajów włączyło już informacje genomyczne do rutynowych systemów oceny wartości hodowlanej bydła mlecznego (Hayes et al. 2009; VanRaden et al. 2009). Stanowią one obecnie ważną część rutynowej oceny wartości hodowlanej bydła mlecznego zarówno na poziomie krajowym, jak i międzynarodowym (Liu et al. 2010). Oprócz predykcji genomowej wartości hodowlanej, takie dane są również cennym źródłem informacji dla genetyków i zootechników, ponieważ dostępność SNP dla dobrze zdefiniowanych populacji ze szczegółowo zmierzonymi wartościami fenotypowymi oraz dokładnymi informacjami o czynnikach środowiskowych i spokrewnieniu stanowi podstawę do badania zmienności genetycznej dla cech o złożonej strukturze dziedziczenia (ang. complex traits). Dotychczas przeprowadzono znaczną liczbę badań asocjacyjnych całego genomu (ang. Genome-Wide Association Study, GWAS) dla cech rutynowo mierzonych u bydła mlecznego, co najlepiej ilustruje zgłoszenie 34 754 genów cech ilościowych (ang. Quantitative trait loci, QTL) i 731 oryginalnych publikacji naukowych do bazy cattleQTLdb (<http://www.animalgenome.org/QTLdb/cattle>, wydanie 31). W konsekwencji zidentyfikowano kilkaset genów o dużym wpływie na fenotyp, określane jako geny kandydujące (np. PPARGC1 i APBB2, Suchocki et al. 2013) lub nawet w postaci mutacji przyczynowych (np. DGAT1, Grisart et al. 2002). Obecnie, w analizach genetycznych dla populacji bydła mlecznego, nacisk kładzie się na geny o średnim efekcie addytywnym oraz takich, których wpływ na zmienność cech przejawia się poprzez efekty nieaddytywne, takie jak dominacja lub epistaza (np.

Sun i in. 2014 ; Kemper i in. 2016). Te pierwsze cieszą się dużym zainteresowaniem w programach selekcji, zwłaszcza z uwagi na możliwy spadek zmienności genetycznej przypisywanej genom głównym, podczas gdy te drugie są ważne dla lepszego zrozumienia genetycznej determinacji cech. W celu zachowania błędu pierwszego oraz drugiego rodzaju na założonym poziomie wymagany jest duży rozmiar próby. Z jednej strony, dla bydła mlecznego duże rozmiary prób danych są obecnie stosunkowo łatwe do uzyskania dzięki powszechnemu stosowaniu mikromacierzy Illumina BovineSNP50 BeadChip w krajowych programach selekcji i wymianie danych genomowych, fenotypowych i rodowodowych pomiędzy krajami. Z drugiej strony, programy krajowe różnią się celami selekcji, a w konsekwencji również adaptacją biologiczną będącą odpowiedzią na selekcję. W związku z tym krajowe populacje różnią się pod względem częstości alleli SNP i wzorów nierównowagi sprzężeń (ang. Linkage disequilibrium, LD) (Rosenberg i in. 2010). Na przykład w niemieckiej populacji rasy holsztyńsko-fryzyjskiej selekcja opierała się na zrównoważonym celu hodowlanym obejmującym produkcję, reprodukcję i cechy funkcjonalne, podczas gdy w polskiej populacji przez wiele dziesięcioleci nacisk hodowlany był kładziony wyłącznie na wydajności białka i tłuszczu, z innymi nieprodukcyjnymi cechami włączonymi do indeksu selekcji dopiero w 2007 r. Dlatego ważnym jest poznanie różnic w strukturze genetycznej populacji przed wprowadzeniem danych dla danego kraju do wspólnej analizy.

Analizy określające różnice w strukturze genetycznej pomiędzy polską, a niemiecką populacją rasy holsztyńsko-fryzyjską została przeprowadzona w pierwszej pracy zgłoszonej jako szczególne osiągnięcie naukowe pt.: **„Assessing the degree of stratification between closely related Holstein-Friesian populations”**. Głównymi celami pracy było wykorzystanie konwencjonalnych wartości hodowlanych poddanych deregresji (ang. Deregressed proofs, DRP) reprezentujących pseudo-fenotypy cech produkcyjnych oraz indeksu liczby komórek somatycznych w mleku, oraz genotypów SNP do porównania wzorców różnorodności genetycznej między niemiecką i polską populacją rasy holsztyńsko-fryzyjskiej. W szczególności badane aspekty obejmowały: porównanie wzorców LD pomiędzy populacjami, ocenę różnic w efektach poszczególnych regionów genomu na

analizowane cechy, ocenę różnic w częstości alleli SNP oraz oszacowanie składników (ko)wariancji dla efektów poligenicznych. W analizie wykorzystano 2 234 polskich i 2 294 niemieckich buhajów, dla których były dostępne wartości DRP z rutynowej oceny wartości hodowlanej przeprowadzonej odpowiednio w Polsce i Niemczech. Analizowanymi cechami były wydajności mleka (MY), białka (PY), tłuszczu (FY) oraz indeks liczby komórek somatycznych w mleku (SSC). Każdy z buhajów był zgenotypowany przy pomocy mikromacierzy Illumina BovineSNP50 BeadChip. Oryginalne dane zostały poddane filtracji tak aby frekwencja rzadszego allelu (ang. Minor allele frequency, MAF) oraz jakość genotypowania (ang. Call rate, CR) były większe odpowiednio od 1% i 90%. Tylko markery, które po filtracji występowały w obydwu populacjach zostały wykorzystane w ostatecznej analizie. Liczba polimorfizmów, które spełniły ww. warunki wynosiła 39 557. Efekty poszczególnych SNP zostały obliczone zarówno dla populacji polskiej jak i niemieckiej, przy pomocy modelu mieszanego z losowym efektem SNP wykorzystywanego w rutynowej ocenie genomowej wartości hodowlanej w Polsce (Szyda i in. 2011). Model statystyczny miał następującą formę:

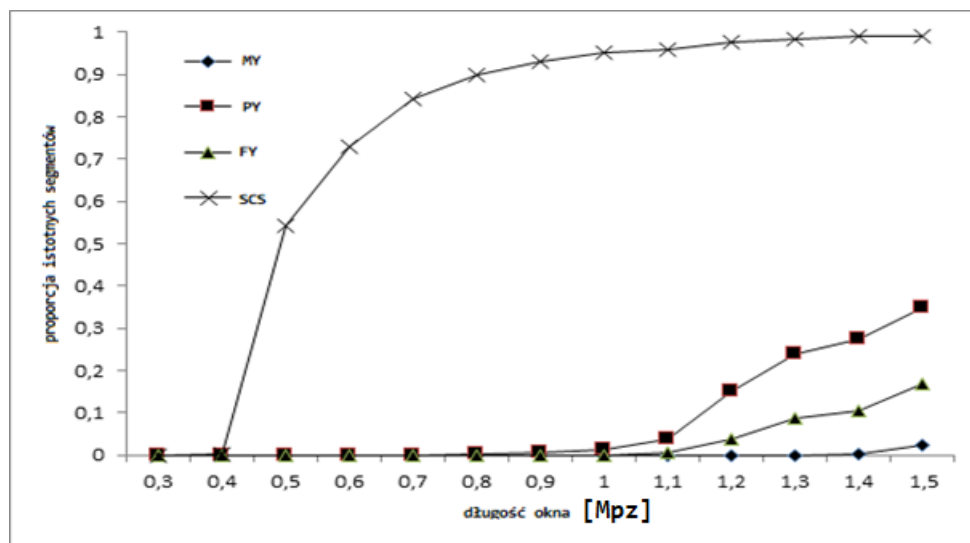
$$\mathbf{y} = \mu + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad (1)$$

gdzie \mathbf{y} jest wektorem DRP dla MY, FY, PY lub SCC; μ jest średnią ogólną; \mathbf{Z} jest macierzą wystąpień genotypów SNP z wartościami -1 , 0 lub 1 odpowiednio dla homozygoty referencyjnej, heterozygoty lub homozygoty alternatywnej; \mathbf{u} jest losowym efektem SNP natomiast $\boldsymbol{\epsilon}$ jest wektorem błędów. Zakładamy, że $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_{SNP}} \frac{\sigma_u^2}{N_{SNP}})$ i $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{D}\sigma_\epsilon^2)$, gdzie \mathbf{I} jest macierzą identycznościową; σ_u^2 jest wariancją addytywnie poligeniczną; N_{SNP} jest ilością markerów SNP; \mathbf{D} jest macierzą diagonalną z odwrotną wartością ilości efektywnych córek (ang. Effective Daughter Contribution, EDC) na przekątnej; σ_ϵ^2 jest wariancją błędu.

W celu zbadania różnic w zmienności analizowanych cech między obiema populacjami zdefiniowano regiony genomowe o różnych długościach, w obrębie których porównano efekty SNP obliczone dla niemieckiej i polskiej populacji. W szczególności, każdy region zdefiniowany był jako przedział o określonej długości wyrażonej w parach zasad. Dla każdego takiego okna testowano hipotezę zerową zakładającą równość średniego efektu SNP w regionie w populacji pol-

skiej i niemieckiej. Dodatkowo, osobno dla obu populacji, zostały wyznaczone współczynniki LD pomiędzy wszystkimi parami SNP przy pomocy programu fastPHASE (Scheet i Stephens, 2006). W rezultacie zaobserwowaliśmy znacząco niższą wartość współczynnika LD w populacji polskiej.

Dla każdego z rozważanych przedziałów średnia wartość współczynnika LD w populacji polskiej była znacznie niższa niż w populacji niemieckiej. Jednak wraz ze wzrostem odległości między markerami współczynnik LD spadał 1,7 razy szybciej ($P = 5 \times 10^{-13}$) w populacji niemieckiej niż w polskiej. Na przykład, dla krótkich odległości pomiędzy markerami wynoszącymi od 0 do 0,025 Mpz oraz od 0,050 do 0,075 Mpz współczynnik LD spadał średnio o 0,088316 w populacji niemieckiej i tylko o 0,04914 w populacji polskiej. Natomiast dla dużych odległości między SNP analizowanymi dla przedziałów od 1,5 do 3,0 oraz od 3,0 do 5,0 Mpz, współczynnik LD spadał o 0,01073 w populacji niemieckiej i o 0,00634 w populacji polskiej. Porównanie efektów genomowych dla regionów chromosomowych zależało od długości zdefiniowanego okna. Sytuacja ta uniemożliwiła identyfikację i opisanie regionów odpowiedzialnych za różnice w populacji. Z drugiej strony, porównanie wyników odpowiadających różnym długościom okna pojawiło się w układzie systematycznym: niezależnie od długości porównywanych regionów, stwierdzono tylko kilka istotnych różnic dla cech produkcyjnych, podczas gdy wiele różnic było istotnych dla SCS (Rysunek 1). Przykładowo, rozpatrując regiony o długościach 0,6 Mpz wyróżniono 3 819 segmentów, ze średnią liczbą 10,3 SNP na segment. Nie zidentyfikowano żadnych różnic dla MY i FY, cztery segmenty różniły się istotnie dla PY, ale aż 2 787 segmentów było różnych dla SCS.



Rysunek 1: Proporcja segmentów genomu o istotnie różnych efektach pomiędzy polską, a niemiecką populacją, dla różnych długości zdefiniowanych segmentów.

Oszacowane komponenty (ko)wariancji addytywnie genetycznej było spójne dla wszystkich cechach. Dla każdej cechy populację niemiecką charakteryzowały znacznie wyższe wariancje niż polską, przy czym stosunek wariancji wahał się od 1,8 dla SCS do 2,8 dla PY. Wszystkie kowariancje addytywnie genetyczne były pozytywne i wynosiły 0,22, 0,24, 0,30 i 0,39, odpowiednio dla MY, PY, FY i SCS. Porównanie polskiej i niemieckiej populacji wykazało, że obserwowane różnice w estymowanych regionach genomowych zależą od różnic w układzie nierównowagi sprzężeń (LD) między populacjami oraz modelu dziedziczenia cech. Uwzględnienie takich różnic ma bezpośredni wpływ na potencjalną międzynarodową ocenę genomowej wartości hodowlanej. Dlatego proponowanym rozwiązaniem byłoby używanie skumulowanego lub uśrednionego efektu dla grup SNP sklasyfikowanych według ich lokalizacji genomowej (Pz) lub wartości współczynnika LD, zamiast pojedynczych oszacowań SNP w modelu SNP-MACE.

W drugiej pracy zgłoszonej jako szczególne osiągnięcie naukowe pt.: „**Genome-wide association study for semen production traits in Holstein-Friesian bulls**” zajęliśmy się problemem znalezienia genów kandydujących dla nowych cech fe-

notypowych opisujących jakość nasienia buhajów. Analizowaliśmy pięć następujących cech: koncentrację plemników, objętość ejakulatu, liczbę plemników, ruchliwość plemników oraz wskaźnika ruchliwości plemników, dostępnych dla 1 212 buhajów hodowanych w czterech stacjach inseminacyjnych na terenie Polski oraz zgenotypowanych przy użyciu mikromacierzy Illumina BovineSNP50 BeadChip. Identyfikacja pozycji SNP w genomie oraz korespondujących z nimi genów kandydujących jest bardzo ważne dla poprawy płodności buhajów. Cechy te mogą również być w przyszłości poddawane selekcji w rutynowych ocenach genomowej wartości hodowlanej. Celem pracy było oszacowanie parametrów (ko)wariancji addytywnie genetycznej oraz identyfikacja istotnych statystycznie SNP dla każdej z ww. cech. Do oszacowania parametrów (ko)wariancji addytywnie genetycznej oraz odziedziczalności użyto wielocechowego modelu mieszanego o następującej postaci:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad (2)$$

gdzie \mathbf{y} jest macierzą złożoną ze wszystkich rozważanych cech o wymiarach $n \times t$, gdzie $n = 21\,906$ jest ilością wszystkich obserwacji fenotypowych dla danej cechy, a $t = 5$ jest ilością rozważanych cech; $\boldsymbol{\beta}$ jest wektorem efektów stałych złożonym ze średniej ogólnej, wieku buhaja w momencie przeprowadzania testu (w miesiącach), sezonu przeprowadzania testu (1 między październikiem, a marcem i 2 między kwietniem, a wrześniem) oraz numerem stacji inseminacyjnej; $\boldsymbol{\alpha}$ jest losowym efektem addytywnie poligenicznym buhaja; $\boldsymbol{\epsilon}$ jest wektorem błędów. Zakładamy, że $\boldsymbol{\alpha} \sim \mathcal{N}(0, \mathbf{A} \otimes \mathbf{G})$ i $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{D} \otimes \mathbf{R})$, gdzie \mathbf{A} jest macierzą spokrewnień; \mathbf{G} jest poligeniczną macierzą kowariancji pomiędzy cechami; \mathbf{D} jest macierzą diagonalną z odwrotną wartością ilości rekordów fenotypowych na przekątnej i \mathbf{R} jest macierzą kowariancji residualnej pomiędzy cechami.

Estymatory (ko)wariancji addytywnie genetycznych oraz błędu losowego zostały umieszczone w Tabeli 1. Wartości odziedziczalności wahały się od 0,26 do 0,34 odpowiednio dla objętości ejakulatu i koncentracji plemników. Odziedziczalności te są znacznie wyższe, niż w przypadku cech płodności krów, które w populacji polskiej wahały się od 0,02 dla współczynnika niepowtarzalności unasienniania jałówek do 0,08 dla przestoju poporodowego.

Uzyskane parametry zostały wykorzystane w jednocechowych modelach mieszanych, które posłużyły do identyfikacji istotnych statystycznie SNP. Modele te mają następującą formę:

$$\hat{\alpha} = \mu + \mathbf{Z}^* \mathbf{g} + \epsilon^*, \quad (3)$$

gdzie $\hat{\alpha}$ reprezentuje wektor wartości addytywnie genetycznych dla danej cechy uzyskanych przy pomocy Modelu 2; μ jest średnią ogólną; \mathbf{Z}^* jest macierzą wystąpień genotypów SNP z wartościami $-1, 0$ lub 1 odpowiednio dla homozygoty referencyjnej, heterozygoty lub homozygoty alternatywnej; \mathbf{g} jest losowym efektem SNP natomiast ϵ^* jest wektorem błędów. Zakładamy, że $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \frac{\sigma_g^2}{N_{SNP}})$ i $\epsilon \sim \mathcal{N}(0, \mathbf{D}\sigma_\epsilon^2)$, gdzie \mathbf{I} jest macierzą identycznościową; σ_g^2 jest wariancją addytywnie poligeniczną; N_{SNP} jest ilością markerów SNP; \mathbf{D} jest macierzą diagonalną z odwrotną ilością rekordów fenotypowych na przekątnej i σ_ϵ^2 jest wariancją błędu.

Dla cechy określającej ruchliwość plemników zostało zidentyfikowanych 20 istotnych SNP zlokalizowanych na 12 autosomach. Dla koncentracji plemników znaleźliśmy siedem istotnych SNP: trzy na chromosomie X i po jednym na chromosomach 1, 6, 23 i 24. Dla objętości ejakulatu oraz wskaźnika ruchliwości zidentyfikowaliśmy odpowiednio trzy i jeden istotny SNP. Wszystkie te markery były zlokalizowane na chromosomie X. Liczba plemników korespondowała z dwunastoma istotnymi SNP zlokalizowanymi na chromosomach 1, 2, 3, 7, 8, 16 i X. Badanie te wykazały kluczową rolę chromosomu X w determinacji zmienności cech związanych z jakością nasienia. Dodatkowo geny zlokalizowane w okolicy istotnych statystycznie SNP zostały wybrane jako geny kandydujące. Wcześniejsze badania prowadzone dla cech związanych z jakością nasienia przez Blashek i in. (2011) także potwierdziły naszą tezę o istotnym wpływie genów zlokalizowanych na chromosomie X. Jednakże, niektóre z genów kandydujących zidentyfikowanych w naszej pracy nie pojawiły się do tej pory w literaturze jako powiązane z płodnością buhajów. 43 istotne statystycznie regiony z naszej pracy zostały włączone do bazy cattleQTLdb . Były to trzy SNP dla objętości ejakulatu, 19 dla ilości i koncentracji plemników oraz 21 dla ruchliwości plemników.

W trzeciej pracy zgłoszonej jako szczególne osiągnięcie naukowe pt.: „Using

gene networks to identify genes and pathways involved in milk production traits in Polish Holstein dairy cattle” poszliśmy krok dalej i tworzyliśmy ścieżki genów (ang. gene pathways), które można określić jako jednostki funkcjonalne dla analizowanych cech o złożonej strukturze dziedziczenia. Wykorzystaliśmy sieci genów do określenia procesów fizjologicznych umożliwiających opis zmienności cech rejestrowanych u bydła mlecznego. Rozważaliśmy cechy produkcyjne tj. wydajności mleka (MY), białka (PY) i tłuszczu (FY). Główne aspekty tej pracy, obejmowały przeciwdziałanie utracie informacji wynikającej z ignorowania informacji na temat nierównowagi sprzężeń pomiędzy SNP zlokalizowanymi w obrębie lub pobliżu tego samego genu oraz braku możliwości biologicznej walidacji uzyskanej sieci genów. W analizie wykorzystano 2 601 buhajów rasy polskiej holsztyńsko-fryzyjskiej zgenotypowanych przy pomocy mikromacierzy Illumina BovineSNP50 BeadChip i posiadających jako predyktory konwencjonalne wartości hodowlanej ww. cech. Efekty SNP oszacowano za pomocą modelu mieszanego z losowymi efektami SNP. Model miał taką samą strukturę jak Model 1. Następnie obliczono wpływ każdego markera i przetestowano je pod kątem istotności dla każdej z cech. Następnie, na podstawie ścieżek metabolicznych oraz informacji z literatury przy użyciu programu KOBAS (Mao i in. 2005) utworzono sieć genów o bezpośrednim lub pośrednim wpływie na fenotyp każdej z cech. Sieci zostały poddane walidacji przy pomocy analizy permutacyjnej. Oszacowano wpływ 4 345 genów, co stanowi 16,4% wszystkich genów zmapowanych w genomie bydła (wersja UMD3.1). Zakładając maksymalną 10% frakcję błędu pierwszego rodzaju, w przypadku wydajności mleka, 50 różnych ontologii genów (ang. Gene ontology, GO) i trzy ścieżki genów zdefiniowane wg Encyklopedii Kioto Genów i Genomów (ang. Kyoto Encyclopedia of Genes and Genomes, KEGG) były istotnie nadreprezentowane w ścieżkach utworzonych na podstawie danych rzeczywistych. Dla wydajności tłuszczu dziewięć terminów GO było istotnie nadreprezentowanych. Z kolei dla wydajności białka 28 terminów GO i sześciu ścieżek KEGG było istotnie nadreprezentowanych w rzeczywistych danych. Na podstawie informacji fizjologicznych zidentyfikowaliśmy zestawy loci zaangażowanych w determinację wydajności mleka (224 geny), wydajność tłuszczu (72 geny) i wydajność białka (546 genów). Wykaz genów zawarto w Tabeli

2., natomiast w Tabeli 3. znajdują się istotne ścieżki KEGG. Niektóre z genów, wykazujące znaczny wpływ na analizowane cechy, zostały już zidentyfikowane we wcześniejszych badaniach, podczas gdy inne reprezentują nowe odkrycia, a zatem prawdopodobnie są to geny o średnim lub małym wpływie na zmienność cech produkcyjnych.

W przyszłości będzie możliwe zastosowanie podejścia obejmującego użycie fenotypów krów oraz genotypów dla krów i buhajów w ramach jednego modelu statystycznego. Najważniejszą zaletą takiego podejścia jest równoczesne wykorzystanie źródeł informacji genomowych, rodowodowych i fenotypowych. Takie podejście pozwala również na dokonanie predykcji jednocześnie dla zwierząt posiadających i nieposiadających genotypów. W ostatniej pracy pt.: **„Dissecting an interplay between genomic and pedigree sources of information to estimate breeding values for milk yield in Polish Holstein-Friesian dairy cattle in a one-step approach based on a random regression test day model”**, zgłoszonej jako szczególne osiągnięcie naukowe, badaliśmy możliwość użycia takiego modelu oraz próbę jego kalibracji poprzez określenie optymalnych parametrów wag dla kombinacji macierzy spokrewnień opartej na informacjach rodowodowych i genomowych. Niniejsze badanie pilotażowe miało na celu przetestowanie modelu jednokrokowego (ang. single step) w kontekście użycia go do jednoczesnej oceny wartości hodowlanej dla zgenotypowanych i niezgenotypowanych osobników. Testy były przeprowadzone dla populacji polskiej holsztyńskofryzyjskiej, z uwzględnieniem różnych macierzy spokrewnień. Dane obejmowały 890 zwierząt (10 zgenotypowych buhajów, 100 krów z danymi fenotypowymi oraz ich 780 przodków bez genotypów). Model mieszany, gdzie fenotypem były próbne udoje krów, z wielomianami Legendre’a drugiego rzędu do modelowania zmiennego w czasie efektywnie poligenicznego, został użyty do estymacji parametrów (ko)wariancji oraz do predykcji genomowych wartości hodowlanych (ang. genomically enhanced breeding values, GEBV). Model użyty do estymacji parametrów (ko)wariancji miał następującą formę:

$$\mathbf{y} = \mathbf{h}\mathbf{t}\mathbf{d} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{p} + \boldsymbol{\epsilon}, \quad (4)$$

gdzie \mathbf{y} jest wektorem z próbnymi udojami dla wydajności mlecznej (MY) dla

100 krów; \mathbf{htd} jest stałym efektem stado - próbny udój; \mathbf{a} jest losowym efektem addytywnie poligenicznym; \mathbf{p} jest losowym efektem trwałym środowiskowym natomiast ϵ jest błędem losowym. Efekt addytywnie poligeniczny oraz trwałe środowiskowy były modelowane przy pomocy wielomianów Legendre'a stopnia 2 jako zmienne w czasie. Zakładaliśmy, że $\mathbf{a} \sim \mathcal{N}(0, \mathbf{A} \otimes \mathbf{G}_a)$, $\mathbf{p} \sim \mathcal{N}(0, \mathbf{I}_n \otimes \mathbf{P})$ oraz $\epsilon \sim \mathcal{N}(0, \mathbf{I}_N \sigma_\epsilon^2)$, gdzie \mathbf{A} jest macierzą spokrewnień; \mathbf{G}_a jest macierzą (ko)wariancji dla współczynników regresji losowych dla efektu addytywnie poligenicznego; \mathbf{P} jest macierzą (ko)wariancji dla współczynników regresji losowych dla efektu środowiskowego natomiast \mathbf{I}_n oraz \mathbf{I}_N oznaczały odpowiednio macierze identycznościowe o wymiarach n (ilość krów) oraz N (ilość wszystkich próbnym udojów). \mathbf{Z}_1 i \mathbf{Z}_2 są macierzami wystąpienia odpowiednio dla efektów \mathbf{a} i \mathbf{p} . Model użyty do estymacji GEBV miał następującą formę:

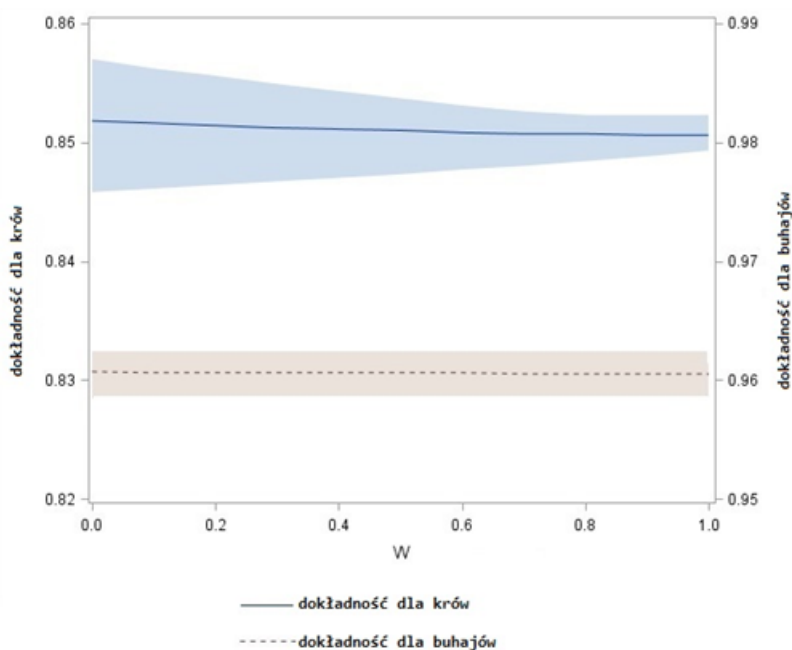
$$\mathbf{y} = \mathbf{htd} + \mathbf{Z}_1^* \mathbf{a}^* + \mathbf{Z}_2 \mathbf{p} + \epsilon^*, \quad (5)$$

gdzie \mathbf{a}^* jest losowym efektem addytywnie poligenicznym dla wszystkich zwierząt (zgenotypowanych oraz nie) oraz \mathbf{Z}_1^* jest macierzą z ortogonalnymi wielomianami zależnymi od czasu pobierania próbnego udoju (ang. days in milk, DIM). Zakładaliśmy, że $\mathbf{a}^* \sim \mathcal{N}(0, \mathbf{H} \otimes \mathbf{G}_a)$ oraz $\epsilon^* \sim \mathcal{N}(0, \mathbf{I}_N \sigma_{\epsilon^*}^2)$, gdzie \mathbf{H} jest macierzą o następującej postaci:

$$\mathbf{H} = \begin{pmatrix} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{G}_W \mathbf{A}_{22}^{-1} \mathbf{A}_{21} + \mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21} & \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{G}_W \\ \mathbf{G}_W \mathbf{A}_{22}^{-1} \mathbf{A}_{21} & \mathbf{G}_W \end{pmatrix}, \quad (6)$$

gdzie \mathbf{A}_{12} jest macierzą spokrewnień pomiędzy niezgenotypowanymi i zgenotypowanymi osobnikami; \mathbf{A}_{22} jest macierzą spokrewnień pomiędzy zgenotypowanymi osobnikami; \mathbf{A}_{11} jest macierzą spokrewnień pomiędzy niezgenotypowanymi osobnikami natomiast $\mathbf{G}_W = (1 - W)\mathbf{G} + W\mathbf{A}_{22}$, gdzie \mathbf{G} jest macierzą podobieństwa genetycznego skonstruowaną przy pomocy metody VanRadena (2008). W modelu tym zastosowano macierz łączącą informacje o spokrewnieniu na podstawie rodowodów oraz SNP zamiast tradycyjnej macierzy spokrewnień. W tej macierzy proporcje informacji pochodzących z rodowodów i SNP zostały określone przez wagi (W). Rozważano różne wartości wag. Na średnią dokładność GEBV nie miał znaczącego wpływu parametr W . Wahala się ona między 0,8506 a 0,8518 dla 100 krów z fenotypami oraz pomiędzy 0,9605 a 0,9607

dla zgenotypowanych buhajów (Rysunek 2). Nadal można zaobserwować stałą tendencję, że najwyższą wiarygodność oceny zawsze uzyskiwano dla parametru $W = 0$ odpowiadającego czystej ocenie genomowej, podczas gdy najniższe wiarygodności uzyskano dla $W = 1$ odpowiadającego czystej ocenie opartej na rodowodach. Średnia dokładność dla przodków wyniosła 0,2616 i była niezależna od parametru W .



Rysunek 2: Dokładność GEBV dla krów i buhajów w zależności od wagi (W). Zacięniowane części odpowiadają odchyleniu standardowemu dla dokładności GEBV.

Wyniki naszej pracy wskazują, że włączenie informacji genomowej do konwencjonalnej oceny genomowej poprawia dokładność predykcji GEBV. Jednak informacje o rodowodzie są ważne dla utrzymania stabilności oceny dla zwierząt niegenotypowanych, gdyż to źródło informacji redukuje odchylenie standardowe dokładności prognozowania. Chociaż szacowane różnice były bardzo małe ze względu na niewielką liczebność próby, to nadal trend dla współczynnika W i dokładność oceny został zachowany.

	objętość ejakulatu	wskaźnik ruchliwości	ruchliwość	koncentracja plemników	liczba plemników
objętość ejakulatu	0,26 ± 0,062	0,04 ± 0,029	0,07 ± 0,029	0,01 ± 0,029	0,82 ± 0,016
wskaźnik ruchliwości	-0,01 ± 0,029	0,26 ± 0,054	0,61 ± 0,023	0,51 ± 0,025	0,29 ± 0,028
ruchliwość	0,12 ± 0,029	0,64 ± 0,022	0,31 ± 0,064	0,33 ± 0,027	0,23 ± 0,028
koncentracja plemników	-0,06 ± 0,029	0,59 ± 0,023	0,56 ± 0,024	0,34 ± 0,068	0,50 ± 0,025
liczba plemników	0,82 ± 0,016	0,36 ± 0,027	0,41 ± 0,026	0,49 ± 0,025	0,27 ± 0,060

Tabela 1: Wartości (ko)wariancji addytywnie genetycznej (poniżej przekątnej) i błędu (powyżej przekątnej) oraz współczynnika odziedziczalności (przekątna).

Symbol genu	Pełna nazwa genu	MY	BTA	Efekt addytywny genu	P-wartość
DGAT1	diacylglycerol O-acyltransferase 1		14	7,52	0,00047
PLEC	Plectin		14	7,37	0,00059
C8orf33	chromosome 8 open reading frame 33		14	6,29	0,00335
MAF1	MAF1 homolog (<i>S. cerevisiae</i>)		14	4,94	0,02138
MAPK15	mitogen-activated protein kinase 15		14	3,95	0,06600
RHPN1	rhophilin, Rho GTPase binding protein 1		14	2,90	0,17846
LY6D	lymphocyte antigen 6 complex, locus D		14	2,79	0,19241
		FY			
DGAT1	diacylglycerol O-acyltransferase 1		14	0,39	< 0,00001
PLEC	Plectin		14	0,31	0,00022
MAF1	MAF1 homolog (<i>S. cerevisiae</i>)		14	0,28	0,00093
C8orf33	chromosome 8 open reading frame 33		14	0,22	0,00883
RHPN1	rhophilin, Rho GTPase binding protein 1		14	0,18	0,03115
LY6D	lymphocyte antigen 6 complex, locus D		14	0,17	0,03786
MAPK15	mitogen-activated protein kinase 15		14	0,17	0,04407
AGO2	argonaute RISC catalytic component 2		14	0,14	0,09502
GML	glycosylphosphatidylinositol anchored mol, Like		14	0,11	0,19682
		PY			
AP1B1	adaptor-related protein complex 1, beta 1 subunit		17	0,09	0,13446
HEPHL1	hephaestin-like 1		29	0,09	0,14508
LHX8	LIM homeobox 8		3	0,09	0,16909
FBP2	fructose-1,6-bisphosphatase 2		8	0,09	0,17416
TANC2	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2		19	0,08	0,18452
DHX34	DEAH (Asp-Glu-Ala-His) box polypeptide 34		18	0,08	0,18670

Tabela 2: Geny wraz z ich położeniem oraz efektem addytywnym i P-wartością.

Symbol ścieżki KEGG	Nazwa ścieżki	95% CI dla ilorazu szans	P-wartość
MY			
bta05412	kardiomiopatia arytmogenna prawej komory	2,2 – 11,9	0,04464
bta05414	kardiomiopatia rozstrzeniowa	2,0 – 11,2	0,07862
bta04530	połączenie ścisle (barierowe)	1,7 – 6,9	0,09542
FY			
bta03018	degradacja RNA	2,1 – 10,9	0,04594
PY			
bta04142	lizosom	8,8 – 51,7	0,00001
bta04110	cykl komórkowy	9 413 3,0 – 11,4	0,00005
bta00030	szlak pentozofosforanowy	7,5 – 245,4	0,00588
bta04144	endocytoza	2,2 – 8,8	0,00848
bta00051	metabolizm fruktozy i mannozy	6,8 – 216,6	0,00884
bta04721	cykl pęcherzyków synaptycznych	2,9 – 37,5	0,07768

Tabela 3: Istotne KEGG na poziomie istotności $\alpha = 0,01$, wraz z 95% przedziałami ufności dla ilorazu szans.

V Bibliografia

1. Blaschek M, Kaya A, Zwald N, Memili E, Kirkpatrick BW. 2011. A whole-genome association analysis of noncompensatory fertility in Holstein bulls. *J Dairy Sci*, 94: 4695 – 4699.
2. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M, Snell R. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res*, 12:222 – 231.
3. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. 2009. Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*, 92:433 – 443.
4. Kemper KE, Littlejohn MD, Lopdell T, Hayes BJ, Bennett LE, Williams RP, Xu XQ, Visscher PM, Carrick MJ, Goddard ME. 2016. Leveraging genetically simple traits to identify small-effect variants for complex phenotypes. *BMC Genomics*, 17:858.
5. Liu Z, Seefried F, Reinhardt F, Thaller G, Reents R. 2010. Dairy cattle genetic evaluation enhanced with genomic information. In: Proceedings of the 9th World Congress on Genetics Applied to Livestock Production (WCGALP), Leipzig, Germany, August 2010.
6. Mao X, Cai T, Olyarchuk JG, Wei L. 2005. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*, 21: 3787 – 3793.
7. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. 2010. Genome-wide association studies in diverse populations. *Nat Rev Genet*, 11:356 – 366.
8. Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet*, 78:629 – 644.

9. Suchocki T, Szyda J, Zhang Q. 2013. Modelling QTL effect on BTA06 using random regression test day models. *J Appl Genet*, 54:49 – 60.
10. Sun C, VanRaden PM, Cole JB, O'Connell JR. 2014. Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects. *PLoS One*, 9:e103934.
11. Szyda J, Żarnecki A, Suchocki T, Kamiński S. 2011. Fitting and validating the genomic evaluation model to Polish Holstein-Friesian cattle. *J Appl Genet*, textbf52:363 – 366.
12. VanRaden PM. 2008. Efficient methods to compute genomic predictions. *J Dairy Sci*, 91:4414 – 4423.
13. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci*, 92:16 – 24.

Wrocław, dn.25 kwietnia 2019 r.

Tomasz Suchocki

dr inż. Tomasz Suchocki