

WYKORZYSTANIE WYSOKOWYDAJNYCH TECHNIK ANALIZ GENOMU W BADANIACH NAUKOWYCH I HODOWLI ZWIERZĄT GOSPODARSKICH*

Artur Gurgul*, Tomasz Ząbek, Monika Bugno-Poniewierska

Instytut Zootechniki Państwowy Instytut Badawczy, Dział Genomiki i Biologii Molekularnej Zwierząt,
Laboratorium Genomiki, 32-083 Balice k. Krakowa

*E-mail: artur.gurgul@izoo.krakow.pl

Rozwój technologii oraz redukcja kosztów wykorzystania wysokowydajnych technik analizy genomu umożliwiły implementację nowoczesnych metod analitycznych do badań i praktyki hodowlanej zwierząt gospodarskich. Do najpopularniejszych technik molekularnych wykorzystywanych w badaniach na skalę genomu należą: sekwencjonowanie następnej generacji oraz analizy z wykorzystaniem mikromacierzy genotypowych, pozwalające na stosunkowo szybkie i tanie prowadzenie szeroko zakrojonych badań populacyjnych. Wyniki otrzymane z wysokowydajnych systemów genotypowania mogą zostać wykorzystane do szeregu badań o charakterze poznawczym i aplikacyjnym. W związku z tym, celem niniejszego opracowania było omówienie zasad działania i najpopularniejszych zastosowań podstawowych narzędzi analizy genomu, ze szczególnym uwzględnieniem mikromacierzy, badań asocjacyjnych w skali genomu i wybranych zagadnień z dziedziny genetyki populacyjnej oraz podstaw selekcji genomowej – jako praktycznego aspektu wykorzystania danych genomowych.

Słowa kluczowe: GWAS, mikromacierze, sekwencjonowanie, selekcja genomowa, SNP

Na przestrzeni ostatniego dziesięciolecia dokonał się ogromny postęp w dziedzinie badań genetycznych, związany przede wszystkim z opracowaniem metod masowego sekwencjonowania DNA oraz rozwojem technologii wysokoprzepustowego genotypowania. Zwiększenie skali eksperymentów i przeniesienie ich na poziom analizy genomu spowodowało istotną redukcję kosztów oznaczeń w przeliczeniu na marker i zapoczątkowało lawinowy przyrost ilości danych, dotyczących zmienności i struktury genomów zwierząt gospodarskich (Morozova i Marra, 2008). W ślad za wzrastającą ilością danych dotyczących polimorfizmu DNA oraz wzajemnych związków i zależności pomiędzy markerami genetycznymi a cechami fenotypowymi, nastąpił

*Praca finansowana z zadania badawczego 04-010.1.

dynamiczny rozwój metod statystycznych i obliczeniowych, pozwalających na uporządkowanie ogromu informacji i wychwycenie rzeczywistych zależności z morza szumu statystycznego (Andrews i Luikart, 2014). Synergiczny rozwój metod molekularnych i obliczeniowych zaowocował także wdrożeniem do praktyki hodowlanej metod genomiki – w postaci selekcji genomowej, która jest obecnie obiecującym narzędziem stymulowania postępu genetycznego w obrębie cech produkcyjnych i funkcjonalnych (Hayes i in., 2009 b). Te nowe techniki analityczne otworzyły także spektrum nowatorskich obszarów badawczych obejmujących pełne rozpoznanie struktury, polimorfizmu i mechanizmów regulacji ekspresji genomu, jak również identyfikację możliwie najszerszego spektrum *loci* cech ilościowych (QTL) warunkujących cenne dla hodowców cechy fenotypowe.

Celem niniejszej pracy było omówienie najpopularniejszych metod analiz genomu oraz przykładów ich zastosowania w badaniach genetycznych prowadzonych u różnych gatunków zwierząt gospodarskich.

Wysokowydajne sekwencjonowanie DNA – przełom w badaniach nad genomami

Impulsem do poszukiwania nowych, wysokowydajnych metod sekwencjonowania DNA był przebieg projektu sekwencjonowania genomu człowieka, który pomimo wysiłków licznych zespołów badawczych pochłonął znacznie więcej środków oraz czasu niż początkowo przewidywano. Pomimo dużej dokładności odczytów i wysokiej wiarygodności sekwencji, zastosowane metody sekwencjonowania okazały się zbyt mało wydajne dla szybkiego odczytu ogromnej, liczącej około 3 mld par zasad sekwencji genomowej (Hert i in., 2008). Obecnie dostępne techniki sekwencjonowania określane wspólnym mianem „sekwencjonowania następnej generacji” lub „masowego równoległego sekwencjonowania”, z których najpopularniejsza jest metoda sekwencjonowania przez syntezę (SBS) opracowana przez firmę Illumina, polegająca na równoczesnym odczytywaniu setek milionów klonów pojedynczych cząstek DNA związanych ze stałym podłożem. W przypadku reakcji SBS, do uniwersalnego startera połączonego z adapterem (oligonukleotydem o znanej sekwencji) poprzedzającym badaną matrycę przyłączane są kolejno znakowane fluorescencyjnie nukleotydy, komplementarne do odczytywanej sekwencji. Fluorescencja nukleotydów jest sczytywana w kolejnych cyklach reakcji, co umożliwia ustalenie sekwencji wszystkich badanych cząsteczek jednocześnie (Ansorge, 2009). Rozwiązanie to pozwala na uzyskanie wydajności reakcji sięgającej 1,5 biliona par zasad w pojedynczym eksperymencie, wystarczających na kilkusetkrotne pokrycie ludzkiego genomu. Pomimo wysokiej wydajności istotną wadą tego podejścia jest stosunkowo mała długość odczytywanej sekwencji wynosząca standardowo 100 par zasad (pz) lub w przypadku niektórych systemów 250 pz ciągłej sekwencji (Frey i in., 2014). W związku z tym, że sekwencje te są losowymi fragmentami pochodzącymi z genomu, muszą one zostać uporządkowane w kontigi lub zmapowane względem znanej sekwencji referencyjnej (przy resekwencjonowaniu znanych genomów). Zadanie to jest wymagającym procesem obliczeniowym i może być obciążone pewnym błędem ze względu na istnienie w genomie sekwencji powtarzalnych i sekwencji homologicznych (Li i in., 2008). Niemniej jednak, sekwencjonowanie następnej generacji

dostarcza wiarygodnej informacji zarówno na temat sekwencji, jak i polimorfizmu badanej cząsteczki DNA, dając możliwość poznania ogromnej liczby polimorfizmów, mających potencjał jako markery DNA lub mutacje funkcjonalne (Xu i in., 2012). Dane te mogą zostać wykorzystane do konstrukcji mikromacierzy genotypowych, które jako ekonomiczny sposób oznaczania ogromnej liczby polimorfizmów DNA są najczęściej wykorzystywane w badaniach genetycznych prowadzonych u zwierząt gospodarskich.

Mikromacierze genotypowe

Najpopularniejszą wysokowydajną techniką badania zmienności genomów są analizy z użyciem mikromacierzy genotypowych. Pozwalają one na jednoczesną identyfikację tysięcy lub nawet setek tysięcy polimorfizmów pojedynczych nukleotydów (SNP). Najszerzej wykorzystywanymi i cechującymi się najwyższą wiarygodnością i powtarzalnością wyników są mikromacierze typu BeadChip opracowane przez firmę Illumina. Pod względem konstrukcyjnym (w pewnym uproszczeniu) są one płytkami szklanymi, na których w odpowiednio przygotowanych dołkach o średnicy 3 μm umieszczono krzemowe koraliki opłaszczone sondami molekularnymi, mającymi zdolność (poprzez hybrydyzację) wykrywania komplementarnych do siebie sekwencji nukleotydowych. Sekwencje sond sąsiadują z miejscami polimorficznymi genomu, które w wyniku reakcji wydłużania sondy o jeden znakowany fluorescencyjnie nukleotyd mogą zostać oznaczone w wyniku skanowania macierzy w skanerach o wysokiej rozdzielczości (<http://www.illumina.com/technology/beadarray-technology.html>). Standardowe macierze genotypowe zawierają sondy dla tzw. genomowych paneli markerów SNP. Oznacza to, że pozwalają na genotypowanie SNP równomiernie rozmieszczonych w genomie, przeważnie niekodujących i polimorficznych w możliwie największej liczbie ras i populacji (Syvänen, 2005). Te niekodujące SNP mają zdolność opisywania zmienności genomu oraz wykrywania przybliżonej lokalizacji QTL poprzez wykorzystanie zjawiska nierównowagi sprzężeniowej (ang. linkage disequilibrium; LD). Nierównowaga sprzężeniowa jest nielosową asocjacją dwóch lub więcej markerów genetycznych wynikającą z ich segregacji we wspólnym haplocyocie, czyli istnienia segmentów genomu o wspólnym pochodzeniu filogenetycznym, w obrębie których rzadko zachodzi rekombinacja (Pritchard i Przeworski, 2001; Espigolan i in., 2013). Istnieje co najmniej kilka miar nierównowagi sprzężeniowej, z których najpopularniejsza oznaczana jest symbolem r^2 . Miara ta opisuje korelację wystąpień alleli w dwóch różnych *loci* i przyjmuje wartość 1 przy pełnym sprzężeniu markerów oraz 0 przy ich niezależnej segregacji. Zaletą tej miary jest znacznie mniejsza wrażliwość na różnice we frekwencji alleli pomiędzy markerami niż w przypadku innych miar LD (Du i in., 2007; Khatkar i in., 2008). Lokalna częstość rekombinacji jest jednym z głównych czynników kształtujących LD w poszczególnych regionach genomu. Frakcje DNA o niskiej częstości rekombinacji, takie jak np. cały chromosom Y lub regiony zlokalizowane w pobliżu centromerów autosomów, charakteryzują się wysokim poziomem LD na długich odcinkach sekwencji genomowej. Natomiast szybki rozpad LD obserwuje się w obrębie tak zwanych gorących miejsc rekombinacji (Jeffreys i in., 2001). Zasięg nierównowagi sprzężeniowej, a więc dystans pomiędzy markerami na jakim wykrywalne jest użyteczne LD, jest

najistotniejszym czynnikiem warunkującym przydatność mikromacierzy do identyfikacji QTL czy poszukiwania mutacji przyczynowych chorób genetycznych (Ogawa i in., 2014). Im niższa wartość średniej nierównowagi sprzężeń na danym odcinku sekwencji genomowej, tym większa gęstość markerów jest potrzebna do precyzyjnej identyfikacji QTL (Pritchard i Przeworski, 2001; Du i in., 2007). W przypadku identyfikacji QTL, miara LD pomiędzy markerem i QTL opisuje także proporcję wariancji genetycznej powodowanej przez QTL, jaką można wykryć za pomocą markera genetycznego (Ogawa i in., 2014). Oznacza to, że przy nierównowadze sprzężeń na poziomie $r^2=0,5$ jedynie 50% zmienności genetycznej warunkowanej przez QTL będzie mogło zostać wykryte za pomocą markera. W populacji bydła holsztyńskiego obliczono, że dla utrzymania średniej nierównowagi sprzężeń pomiędzy markerami umieszczonymi na macierzy a potencjalnym QTL na poziomie 0,5, konieczne jest rozmieszczenie markerów w genomie ze średnim dystansem około 10 tysięcy par zasad (kpz) (Gautier i in., 2007). Gęstość ta jest i tak znacząco niższa niż konieczna do badań u ludzi (Shifman i in., 2003) ze względu na fakt, że zasięg nierównowagi sprzężeń w populacjach zwierząt hodowlanych jest znacznie wyższy, co można tłumaczyć wyraźnie mniejszą efektywną wielkością populacji zwierząt hodowlanych oraz silną presją selekcyjną ukierunkowaną na pożądane haplotypy (McRae i in., 2002). Przykładowo, u niemieckiego bydła holsztyńskiego stwierdzono, że nierównowaga sprzężeń pomiędzy poszczególnymi parami markerów na dystansie do 25 kpz wyrażona miarą r^2 wynosi 0,3 ($\pm 0,32$) i obniża się do 0,20 ($\pm 0,24$) na dystansie 50–75 kpz (Qanbari i in., 2010 b), natomiast u jednej z hiszpańskich ras owiec, średnie LD na dystansie do 10 kpz wyniosło 0,33, a dla markerów o dystansie pomiędzy 200–500 kpz zaobserwowano niską wartość LD równą 0,061 (García-Gómez i in., 2012). Zasięg i struktura nierównowagi sprzężeń różni się pomiędzy poszczególnymi rasami i gatunkami i w dużym stopniu zależy od historii populacji, zdarzeń demograficznych, ostrości selekcji, mutacji i innych mechanizmów powodujących ewolucję frekwencji alleli (Slatkin, 2008).

Innym czynnikiem warunkującym przydatność mikromacierzy do badań asocjacyjnych czy populacyjnych jest frekwencja alleli markerów umieszczonych na macierzy (Gurgul i in., 2013). Im wyższa frekwencja rzadszego allelu (ang. minor allele frequency; MAF) tym większa szansa, że dany marker będzie opisywał zjawiska charakterystyczne dla całej populacji, a nie powiązane z rzadką zmianą mutacyjną, występującą w pojedynczym haplocyocie ancestralnym. Badania przeprowadzone przy konstrukcji najpopularniejszej macierzy przeznaczonej do badań u bydła (BovineSNP50) zawierającej sondy dla około 54 tys. markerów wykazały średni MAF u 19 badanych ras na poziomie 0,24 (http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp50.pdf).

Badania asocjacyjne

Z punktu widzenia genetyki i hodowli zwierząt gospodarskich największą popularnością cieszą się badania prowadzone w kierunku identyfikacji molekularnego podłoża obserwowanych, dziedzicznych różnic fenotypowych w obrębie cech produkcyjnych i funkcjonalnych. Badania te zmierzają do precyzyjnej lokalizacji *loci* cech ilościowych, a w późniejszej fazie wykrycia polimorfizmów funkcjonalnych,

które bezpośrednio modyfikują mechanizmy ekspresji genów lub strukturę kodowanych białek. Znaczące miejsce wśród tego typu badań zajmują badania typu GWAS (ang. genome-wide association study), które polegają na poszukiwaniu asocjacji pomiędzy markerami umieszczonymi na macierzy a zmiennością danej cechy. Tego typu badania wykorzystują zjawisko nierównowagi sprzężeniowej pomiędzy markerem a właściwym QTL, dzięki czemu możliwe jest wykrycie efektu QTL za pomocą potencjalnie neutralnego markera (Tarres i in., 2009). Markery o istotnych efektach mogą wskazywać miejsca genomu, w których znajduje się właściwy polimorfizm funkcjonalny. Przydatność tej metody do wykrywania QTL jest zależna od wielu czynników, spośród których najważniejsze to: odziedziczalność cechy, część całkowitej zmienności genetycznej warunkowanej przez poszczególne QTL (ilość QTL i ich siła), zasięg nierównowagi sprzężeniowej, liczba dostępnych obserwacji (badanych zwierząt) oraz liczba wykorzystywanych markerów genetycznych (Hong i Park, 2012). Bardziej złożone modele statystyczne wykorzystują także poprawkę na strukturę oraz spokrewnienie populacji, pozwalające poprawić efekty markerów o efekty linii genetycznej (Yang i in., 2014). Pomimo że ten typ badań wydaje się być potężnym narzędziem w badaniach genetycznych, należy mieć świadomość, że większość cech ilościowych warunkowana jest przez dużą liczbę rozsianych w genomie polimorfizmów funkcjonalnych o stosunkowo niewielkim efekcie (Lango i in., 2010), co znacznie utrudnia ich wiarygodną identyfikację. Przykładem może być praca Lango i in. (2010), w której stwierdzono, że aż około 180 *loci* powiązanych ze wzrostem u ludzi wyjaśnia tylko 10% całkowitej zmienności genetycznej. Z wykorzystaniem podejścia GWAS przeprowadzono szereg badań zmierzających do wykrycia głównych QTL dla cech produkcyjnych. Przykładowo, dla wydajności mleka u bydła przeprowadzono badania, w których zidentyfikowano 734 SNP o istotnym wpływie na badaną cechę (Bolormaa i in., 2010; Jiang i in., 2010; Mai i in., 2010). Markery te zlokalizowane były głównie w obrębie chromosomów: 8, 9, 10, 11, 13, 25 i 29, jak również w sąsiedztwie genu *DGATI*. Natomiast u kurcząt brojlerów zidentyfikowano 9 markerów SNP, istotnie powiązanych z masą ciała w 7–12 tygodniu życia. Markery te zlokalizowane były głównie w obszarze chromosomu 4, na odcinku pomiędzy 71,6 a 80,2 Mbp sekwencji genomowej (Gu i in., 2011).

Selekcja genomowa

Dalszym rozwinięciem podejścia GWAS oraz selekcji wspomaganej markerami (Williams, 2005) jest koncepcja selekcji genomowej. Nie zakłada ona poszukiwania pojedynczych markerów powiązanych z poszczególnymi QTL, lecz dopuszcza możliwość, że każdy badany marker jest w nierównowadze sprzężeń z potencjalnym QTL, a więc wyraża pewną część wariacji genetycznej (Meuwissen i in., 2001). Oryginalnie, koncepcja selekcji genomowej zamyka się w dwóch krokach: oszacowanie efektów markerów w oparciu o populację referencyjną (o znanym genotypie i wartości hodowlanej) oraz oszacowanie genomowej wartości hodowlanej u zwierząt kandydujących wyłącznie na podstawie genotypów. Zaproponowano co najmniej kilka metod pozwalających na oszacowanie efektów markerów dla celów selekcji genomowej (Wang i in., 2012). Różnice pomiędzy nimi polegają przede wszystkim na pierwotnym założeniu odnośnie dystrybucji efektów w genomie oraz wykorzystaniu

różnych źródeł informacji odnośnie spokrewnienia zwierząt. Najczęściej stosowane metody (oparte o metodę BLUP) zakładają, że efekty markerów podlegają rozkładowi normalnemu o równej wariancji, a więc nie dopuszczają istnienia bardzo silnych QTL oraz zakładają, że niemal każdy marker będzie zasocjowany z QTL (Hayes i in., 2009 b; VanRaden i in., 2009). Opcjonalne metody (BayesA, BayesB, BayesC, BayesR) dopuszczają rozkład t-Studenta z różną liczbą stopni swobody, lub rozkłady normalne o różnej wariancji w zależności od wielkości efektu markera (Meuwissen i in., 2001; Habier i in., 2011; Wang i in., 2012). Różnica pomiędzy podejściami opartymi o BLUP polega przede wszystkim na wykorzystaniu macierzy spokrewnień obliczonej na podstawie rodowodów lub macierzy spokrewnień genomowych – reprezentującej proporcję genomu identyczną przez stan (IBS) pomiędzy zwierzętami (tzw. G-BLUP). Oszacowane efekty markerów są później wykorzystywane do obliczenia bezpośredniej wartości genomowej, poprzez zsumowanie efektów wszystkich markerów genetycznych (SNP-BLUP), lub bezpośrednio jak w tradycyjnej metodzie BLUP (G-BLUP) (Koivula i in., 2012). Genomowe wartości hodowlane są ostatecznie prezentowane jako kombinowana wartość hodowlana, która jest obliczana jako indeks złożony z konwencjonalnej wartości hodowlanej (lub indeksu rodowodowego dla młodych zwierząt) i bezpośredniej oceny genomowej, wyważonych przez ich dokładność (http://wycena.izoo.krakow.pl/doc/metody_oceny_2014_2.pdf).

Biorąc pod uwagę elementy składające się na postęp hodowlany w jednostce czasu, selekcja genomowa może przynieść wymierne korzyści przede wszystkim w zakresie zwiększenia dokładności oceny (uzupełnienie indeksów rodowodowych o bezpośrednią wartość genomową) oraz skrócenia odstępu międzypokoleniowego (wykorzystanie młodych zwierząt w hodowli, bez oczekiwania na ostateczną ocenę na potomstwie) (de Roos i in., 2011). Selekcja genomowa może także przyczynić się do lepszego zarządzania inbredem i zmiennością genetyczną, poprzez wybór najbardziej wartościowych osobników spośród grupy pełnego rodzeństwa, jednak wymaga to dodatkowej kontroli inbrodu na poziomie danych pochodzących z genomu (Soneson i in., 2012).

Badania populacyjne

Dane genotypowe pozyskane z mikromacierzy mogą także zostać wykorzystane do szeregu badań dotyczących struktury genomu, wzajemnych zależności istniejących w genomie oraz zmian zachodzących w genomie pod wpływem różnorodnych procesów populacyjnych czy też sztucznej selekcji. Do najpopularniejszych zagadnień badawczych należą: identyfikacja sygnatur kierunkowej selekcji, analiza ciągów homozygotyczności oraz zmienności i dystansu genetycznego pomiędzy populacjami (Gurgul i in., 2014).

Ciekawym zagadnieniem wydają się być badania ukierunkowanie na identyfikację zmian zachodzących we frekwencji alleli genomowych paneli markerów pod wpływem kierunkowej selekcji. Presja selekcyjna ukierunkowana na QTL powoduje stosunkowo szybkie zmiany we frekwencji alleli w danym *locus*, zmierzające do utrwalenia korzystnego wariantu w populacji (Wilkinson i in., 2013). Zmiany te odzwierciedlane są przez sąsiadujące SNP będące w nierównowadze sprzężeń z korzystnym allelem QTL i pozwalają na stosunkowo precyzyjne wskazanie jego lokalizacji.

Generalnie, większość metod wykorzystywanych do identyfikacji sygnatur selekcji polega na porównaniu dystrybucji frekwencji alleli w populacjach o różnym poziomie selekcionowanych cech, poprzez obliczenie statystyk populacyjnych będących funkcją frekwencji alleli (Weir i in., 2005; Wilkinson i in., 2013). Zaproponowano również specyficzne testy istotności wykrytych regionów podlegających presji selekcyjnej (Kim and Stephan, 2002; Voight i in., 2006; Stella i in., 2010). Jedną z bardziej zaawansowanych metod umożliwiającą analizę pojedynczych populacji jest zaproponowany przez Sabeti i in. (2002) test REHH (ang. relative extended haplotype homozygosity) polegający na identyfikacji regionów genomu obejmujących wyjątkowo długie haplotypy o wysokiej frekwencji w populacji. Dodatkowo, test ten pozwala na uwzględnienie częstości rekombinacji w obrębie sąsiadujących haplotypów i wykluczanie błędu spowodowanego różnicą w częstości rekombinacji w poszczególnych regionach genomu (Qanbari i in., 2011).

Z wykorzystaniem jednego z dostępnych podejść metodycznych Hayes i in. (2009 a) zidentyfikowali 15 regionów genomu wykazujących sygnatury różnicowej selekcji pomiędzy bydłem mlecznym i mięsnym. Większość z tych regionów została zlokalizowana na chromosomie 20, w pobliżu *locus* genu *GHR* (receptora hormonu wzrostu) o silnym wpływie na zawartość białka w mleku krów mlecznych (Blott i in., 2003). Natomiast, w oparciu o test REHH, Qanbari i in. (2010 a) zidentyfikowali 12 haplotypów będących pod wpływem presji selekcyjnej u bydła holenderskiego, które obejmowały geny *FABP3*, *CLPN3*, *SPERT*, *HTR2A5*, *ABCE1*, *BMP4* i *PTGER2*, które stanowią potencjalne QTL dla cech mlecznych i reprodukcyjnych.

Inną ciekawą aplikacją danych genotypowych uzyskanych z mikromacierzy jest identyfikacja ciągów homozygotyczności – ROH (ang. runs of homozygosity). ROH są długimi, ciągłymi homozygotycznymi segmentami genomu, powstającymi w wyniku identyczności obydwu alleli odziedziczonych od wspólnego przodka, a więc segmentami autozygotycznymi (Purfield i in., 2012). Istnienie w genomie zwierzęcia długich homozygotycznych segmentów wskazuje na stosunkowo duże i niedawne spokrewnienie rodziców (Kirin i in., 2010), natomiast krótkie segmenty ROH powstają w wyniku powtarzających się zdarzeń rekombinacyjnych i odzwierciedlają dawne spokrewnienie zwierząt (McQuillan i in., 2008). Procent genomu będącego w ROH może posłużyć do obliczenia współczynnika inbrodu obrazującego proporcję genomu osobnika rzeczywiście identyczną dzięki pochodzeniu (odziedziczeniu od wspólnego przodka), a więc uwzględniającego zdarzenie rekombinacyjne i przypadkową identyczność poszczególnych alleli (Ron i in., 1996; Carothers i in., 2006). Współczynnik ten wykazuje umiarkowaną korelację (0,4–0,7) z klasycznym współczynnikiem inbrodu obliczonym na podstawie rodowodów (Ferenčakovic i in., 2011). Ciągi homozygotyczności mogą także wskazywać na istnienie w populacji częstych i długich haplotypów, które podlegają presji selekcyjnej i zmierzają do utrwalenia w populacji. Analiza częstości poszczególnych ROH może więc także posłużyć do identyfikacji sygnatur selekcji (Kim i in., 2013), po uwzględnieniu poprawki na zjawisko dryfu genetycznego.

Podsumowując, sekwencjonowanie następnej generacji w połączeniu ze stosunkowo tanimi i wydajnymi metodami genotypowania w oparciu o mikromacierze są potężnym narzędziem w ręku genetyków, dającym szczegółowy wgląd w genomy

badanych gatunków i pozwalającym na precyzyjne mapowanie *loci* cech ilościowych. Niemniej jednak, stopień złożoności genomu oraz złożoność dziedziczenia cech ilościowych sprawia, że pełne rozpoznanie zjawisk i zależności zachodzących w genomie jest ciągle wyzwaniem dla współczesnej genetyki. Pewne nadzieje w tym zakresie wiąże się z szerokim wykorzystaniem technik sekwencjonowania, które pozwalają na bezpośrednie wychwytywanie polimorfizmów funkcjonalnych bez konieczności polegania na nierównowadze sprzężeń. Dalsza redukcja kosztów sekwencjonowania i wzrost wydajności sekwenatorów będzie więc źródłem dalszego postępu w badaniach nad zależnościami pomiędzy genomem a obserwowanym fenotypem.

Piśmiennictwo

- Andrews K.R., Luikart G. (2014). Recent novel approaches for population genomics data analysis. *Molecul. Ecol.*, 23: 1661–1667.
- Ansorge W.J. (2009). Next-generation DNA sequencing techniques. *N. Biotechnol.*, 25: 195–203.
- Blott S., Kim J.J., Moision S. (2003). Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics*, 163: 253–266.
- Bolormaa S., Pryce J.E., Hayes B.J., Goddard M.E. (2010). Multivariate analysis of a genome-wide association study in dairy cattle. *J. Dairy Sci.*, 93: 3818–3833.
- Carothers A.D., Rudan I., Kolcic I., Polasek O., Hayward C., Wright A.F., Campbell H., Teague P., Hastie N.D., Weber J.L. (2006). Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Ann. Hum. Genet.*, 70: 666–676.
- de Roos A.P., Schrooten C., Veerkamp R.F., van Arendonk J.A. (2011). Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. *J. Dairy Sci.*, 94: 1559–1567.
- Du F.X., Clutter A.C., Lohuis M.M. (2007). Characterizing linkage disequilibrium in pig populations. *Int. J. Biol. Sci.*, 3: 166–178.
- Espigolan R., Baldi F., Boligon A.A., Souza F.R.P., Gordo D.G.M., Tonussi R.L., Cardoso D.F., Oliveira H.N., Tonhati H., Sargolzaei M., Schenkel F.S., Carneiro R., Ferro J.A., Albuquerque L.G. (2013). Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics*, 14, p. 305.
- Ferencakovic M., Hamzic E., Gredler B., Curik I., Sölkner J. (2011). Runs of homozygosity reveal genomewide autozygosity in the Austrian Fleckvieh cattle. *ACS*, 76: 325–328.
- Frey K.G., Herrera-Galeano J.E., Redden C.L., Luu T.V., Servetas S.L., Mateczun A.J., Mokashi V.P., Bishop-Lilly K.A. (2014). Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics*, 15, p. 96.
- García-Gómez E., Sahana G., Gutiérrez-Gil B., Arranz J.J. (2012). Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics*, 13, p. 43.
- Gautier M., Faraut T., Moazami-Goudarzi K., Navratil V., Foglio M., Grohs C., Bolland A., Garnier J.G., Boichard D., Lathrop G.M., Gut I.G., Eggen A. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics*, 177: 1059–1070.
- Gu X., Feng C., Ma L., Song C., Wang Y., Da Y., Li H., Chen K., Ye S., Ge C., Hu X., Li N. (2011). Genome-wide association study of body weight in chicken F2 resource population. *PLoS One*, 6: e21872.
- Gurgul A., Zukowski K., Pawlina K., Ząbek T., Semik E., Bugno-Poniewierska M. (2013). The evaluation of bovine SNP50 beadchip assay performance in Polish red cattle breed. *Folia Biol. (Krakow)*, 61:173–176.

- Gurgul A., Semik E., Pawlina K., Szmatoła T., Jasielczuk I., Bugno-Poniewierska M. (2014). The application of genome-wide SNP genotyping methods in studies on livestock genomes. *J. Appl. Genet.*, 55: 197–208.
- Habier D., Fernando R.L., Kizilkaya K., Garrick D.J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12, p. 186.
- Hayes B.J., Chamberlain A.J., Maceachern S., Savin K., McPartlan H., MacLeod I., Sethuraman L., Goddard M.E. (2009 a). A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Anim. Genet.*, 40: 176–184.
- Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. (2009 b). Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.*, 92: 433–443.
- Hert D.G., Fredlake C.P., Barron A.E. (2008). Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis*, 29: 4618–4626.
- Hong E.P., Park J.W. (2012). Sample size and statistical power calculation in genetic association studies. *Genomics & Informatics*, 10: 117–122.
- Jeffreys A.J., Kauppi L., Neumann R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.*, 29: 217–222.
- Jiang L., Liu J., Sun D., Ma P., Ding X., Yu Y., Zhang Q. (2010). Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS One*, 5: e13661.
- Khatkar M.S., Nicholas F.W., Collins A.R., Zenger K.R., Cavanagh J.A.L., Barris W., Schnabel R.D., Taylor J.F., Raadsma H.W. (2008). Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics*, 9, p. 187.
- Kim E.S., Cole J.B., Huson H., Wiggans G.R., Van Tassell C.P., Crooker B.A., Liu G., Da Y., Sonstegard T.S. (2013). Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. *PLoS One*, 8: e80813.
- Kim Y., Stephan W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160: 765–777.
- Kirin M., McQuillan R., Franklin C.S., Campbell H., McKeigue P.M., Wilson J.F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5: e13996.
- Koivula M., Strandén I., Su G., Mäntysaari E.A. (2012). Different methods to calculate genomic predictions – comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). *J. Dairy Sci.*, 95: 4065–4073.
- Lango A.H., Estrada K., Lettre G., Berndt S.I., Weedon M.N., Rivadeneira F., Willer C.J., Jackson A.U. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467: 832–838.
- Li H., Ruan J., Durbin R. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.*, 18: 1851–1858.
- Mai M.D., Sahana G., Christiansen F.B., Guldbbrandtsen B. (2010). A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *J. Anim. Sci.*, 88: 3522–3528.
- McQuillan R., Leutenegger A.L., Abdel-Rahman B., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.*, 83: 359–372.
- McRae A.F., McEwan J.C., Dodds K.G., Wilson T., Crawford A.M., Slate J. (2002). Linkage disequilibrium in domestic sheep. *Genetics*, 160: 1113–1122.
- Meuwissen T.H., Hayes B.J., Goddard M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157: 1819–1829.
- Morozova O., Marra M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92: 255–264.
- Ogawa S., Matsuda H., Taniguchi Y., Watanabe T., Nishimura S., Sugimoto Y., Iwaisaki H. (2014). Effects of single nucleotide polymorphism marker density on degree of genetic variance explained and genomic evaluation for carcass traits in Japanese Black beef cattle. *BMC Genet.*, 15, p. 15.

- Pritchard J.K., Przeworski M. (2001). Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.*, 69: 1–14
- Purfield D.C., Berry D.P., McParland S., Bradley D.G. (2012). Runs of homozygosity and population history in cattle. *BMC Genetics*, 13, p. 70.
- Qanbari S., Pimentel E.C., Tetens J., Thaller G., Lichtner P., Sharifi A.R., Simianer H. (2010 a). The pattern of linkage disequilibrium in German Holstein cattle. *Anim. Genet.*, 41: 346–356.
- Qanbari S., Pimentel E.C., Tetens J., Thaller G., Lichtner P., Sharifi A.R., Simianer H. (2010 b). A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.*, 41: 377–389.
- Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S., Moore S., Thaller G., Simianer H. (2011). Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics*, 12, p. 318.
- Ron M., Blanc Y., Band M., Ezra E., Weller J.I. (1996). Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. *J. Dairy Sci.*, 79: 676–681.
- Sabeti P.C., Reich D.E., Higgins J.M., Levine H.Z., Richter D.J., Schaffner S.F., Gabriel S.B., Platko J.V., Patterson N.J., McDonald G.J., Ackerman H.C., Campbell S.J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E.S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419: 832–837.
- Shifman S., Kuypers J., Kokoris M., Yakir B., Darvasi A. (2003). Linkage disequilibrium patterns of the human genome across populations. *Hum. Mol. Genet.*, 12: 771–776.
- Slatkin M. (2008). Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.*, 9: 447–485.
- Sonesson A.K., Woolliams J.A., Meuwissen T.H. (2012). Genomic selection requires genomic control of inbreeding. *Genet. Sel. Evol.*, 44, p. 27.
- Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P. (2010). Identification of selection signatures in cattle breeds selected for dairy production. *Genetics*, 185: 1451–1461.
- Syvänen A.C. (2005). Toward genome-wide SNP genotyping. *Nat. Genet.* 37, Suppl., S5–10.
- Tarres J., Guillaume F., Fritz S. (2009). A strategy for QTL fine-mapping using a dense SNP map. *BMC Proc.*, 3 Suppl. 1: S3.
- VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F., Schenkel F.S. (2009). Invited review: reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.*, 92: 16–24.
- Voight B.F., Kudaravalli S., Wen X., Pritchard J.K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.*, 4: e72.
- Wang C.L., Ma P.P., Zhang Z., (2012). Comparison of five methods for genomic breeding value estimation for the common dataset of the 15th QTL-MAS Workshop. *BMC Proc.*, 6 (Suppl. 2): S13.
- Weir B.S., Cardon L.R., Anderson A.D., Nielsen D.M., Hill W.G. (2005). Measures of human population structure show heterogeneity among genomic regions. *Genome Res.*, 15: 1468–1476.
- Wilkinson S., Lu Z.H., Megens H.J., Archibald A.L., Haley C., Jackson I.J., Groenen M.A., Crooijmans R.P., Ogden R., Wiener P. (2013). Signatures of diversifying selection in European pig breeds. *PLoS Genet.* 9: e1003453.
- Williams J.L. (2005). The use of marker-assisted selection in animal breeding and biotechnology. *Rev. Sci. Tech.*, 24: 379–391.
- Xu F., Wang W., Wang P., Jun Li M., Chung Sham P., Wang J. (2012). A fast and accurate SNP detection algorithm for next-generation sequencing data. *Nat. Commun.*, 3, p. 1258.
- Yang J., Zaitlen N.A., Goddard M.E., Visscher P.M., Price A.L. (2014). Mixed model association methods: advantages and pitfalls. *Nat. Genet.*, 46: 100–106.

ARTUR GURGUL, TOMASZ ZĄBEK, MONIKA BUGNO-PONIEWIERSKA

Using high-throughput genome analysis techniques in livestock research and breeding

SUMMARY

Technological developments and the reduced costs of high-throughput genome analysis techniques enabled modern analytical methods to be applied in livestock research and breeding practice. The most popular molecular techniques used in genome-scale research include next-generation sequencing and genotyping microarray analysis which allows large-scale population studies to be performed relatively quickly and at a low cost. The results obtained with high-throughput genotyping systems can be applied in many studies that are of cognitive significance and useful for application. Therefore, the aim of this paper was to discuss the operating principles and most popular applications of basic genome analysis tools, with special consideration of microarrays, genome-scale association studies, selected issues in population genetics and the fundamentals of genomic selection as a practical aspect of using genomic data.

Key words: GWAS, microarrays, sequencing, genomic selection, SNP