

WERONIKA JARSKA
AGNIESZKA NIEDZIELA *
RENATA ORŁOWSKA
PIOTR T. BEDNAREK

Instytut Hodowli i Aklimatyzacji Roślin — Państwowy Instytut Badawczy, Radzików

Strategie molekularne w nowoczesnej hodowli roślin

Molecular strategies in modern plant breeding

Rozwój biologii molekularnej, a w szczególności technologii markerowych oraz narzędzi statystycznych umożliwiających analizę dużej liczby danych uzyskiwanych dla szerokiego typu populacji mapujących sprawia, że zmienia się podejście do selekcji materiałów roślinnych dla potrzeb hodowli. Coraz silniejszy nacisk kładzie się na selekcję wielu cech użytkowych przy jednoczesnym wykorzystaniu elitarnych, zwykle niespokrewnionych ze sobą lecz wyrównanych pod względem genetycznym materiałów roślinnych. Rozwijane obecnie metody umożliwiające prowadzenie selekcji za pomocą markerów DNA w oparciu o złożone populacje mapujące są znane tylko wąskiej grupie specjalistów natomiast metody wykorzystujące ściśle zdefiniowane modele genetyczne, ze względu na ich liczne ograniczenia, zdają się być negowane. Niniejsza praca poświęcona jest omówieniu możliwości metod selekcji wykorzystujących szeroki wachlarz metod biologii molekularnej.

Słowa kluczowe: mapowanie genetyczne, mapowanie asocjacyjne, selekcja genomowa

Advances in molecular biology, including the new marker technologies and statistical tools allow analysis of large data sets evaluated for different types of mapping populations and change approaches to selection of breeding materials. The emphasis is put on selecting many traits simultaneously based on elite, usually non-related but genetically uniform, plant materials. Currently available methods for selecting forms via DNA-based marker technologies using complex mapping populations are well known to a limited number of specialists whilst applications involving defined mendelian populations (due to their numerous limitations) are being neglected. Thus, the review is dedicated to describing wide range of selection approaches that could be applied to different breeding programs depending on experimental requirements.

Key words: association mapping, genetic mapping, genomic selection

*Autor korespondencyjny: Agnieszka Niedziela

Redaktor prowadzący: Barbara Zagdańska

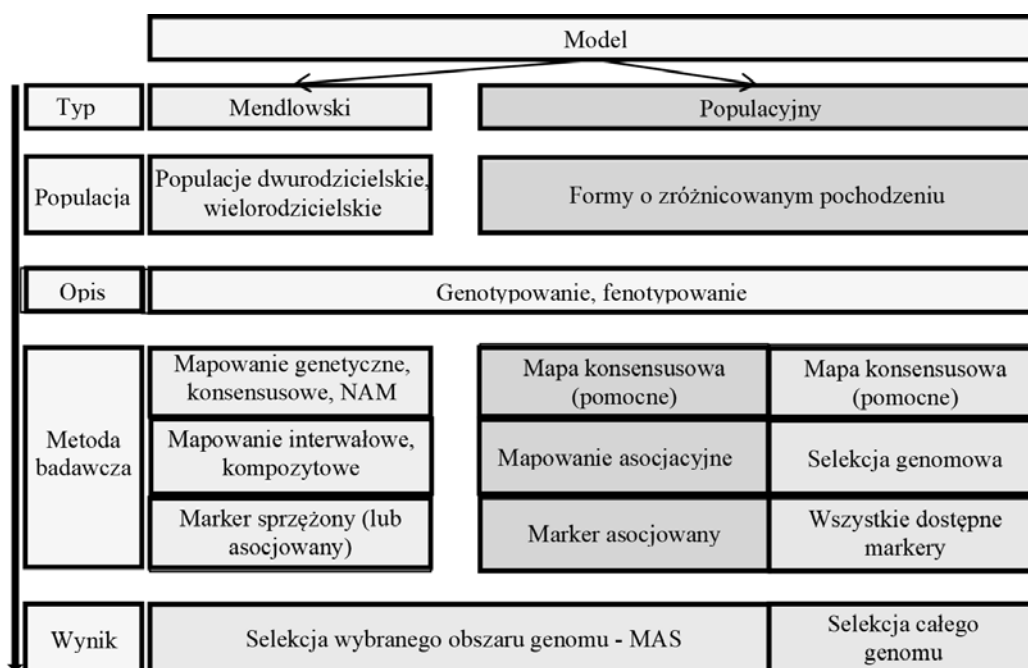
WSTĘP

Dynamiczny rozwój metod biologii molekularnej, a w szczególności technik bazujących na markerach DNA, coraz częściej znajduje odzwierciedlenie w hodowli roślin. Wraz z wykorzystaniem pierwszych markerów DNA do selekcji form o podwyższonej tolerancji na choroby (Young i in., 1988; McMullen i Louie, 1989) czy opracowaniem markerów wybranych cech ilościowych (Lander i Botstein, 1989), oczywistym stało się, że taka selekcja nie zawsze spełnia stawiane przed nią oczekiwania. Problemem stała się selekcja form o podwyższonych wartościach cech warunkowanych wieloma *loci*, czy szeregiem różnych alleli tylko częściowo tłumaczących obserwowaną zmienność (Kearsey i Farquhar, 1998). Konieczne było opracowanie takich narzędzi molekularnych i statystycznych, które podołałyby wymogom współczesnej hodowli. Wraz z rozwojem technologii umożliwiających uzyskanie dużej liczby użytecznych markerów DNA, włączeniem do selekcji narzędzi genetyki populacyjnej, rozwinięto takie metody jak mapowanie asocjacyjne (Pritchard i in., 2000) oraz selekcję genomową (Meuwissen i in., 2001). To właśnie te narzędzia mogą okazać się najbardziej użyteczne w selekcji. Ich potencjał nie jest jeszcze do końca poznany, a w przypadku selekcji genomowej wiele modeli znajduje się w fazie badań i wymaga weryfikacji eksperymentalnej (Nakaya i Isobe, 2012). Takie badania są już realizowane przez duże firmy hodowlane, jednak uzyskiwane wyniki nie są powszechnie dostępne. Celem niniejszej pracy jest przedstawienie możliwości oraz zastosowań narzędzi badawczych wykorzystywanych do selekcji materiałów hodowlanych bez wnikania w złożoności matematyczno-statystyczne oraz pokazanie obszarów ich wykorzystania w hodowli roślin uprawnych.

MODELE SELEKCJI BAZUJĄCE NA MARKERACH DNA

Można wyróżnić dwa zasadnicze modele selekcji wspartej markerami DNA, a mianowicie model bazujący na genetyce mendlowskiej i genetyce populacyjnej (rys. 1). Pierwszy z nich bazuje na populacjach roślinnych, których formy rodzicielskie są zdefiniowane. W tym przypadku wyróżnia się populacje dwurodzicielskie takie jak: pokolenie F_2 , linie podwojonych haploidów (DH, ang. Double Haploid), linie blisko-izogeniczne (NIL, ang. Near Isogenic Lines) i rekombinacyjne linie wsobne (RIL, ang. Recombinant Inbred Lines) oraz populacje wielurodzicielskie takie jak MAGIC (ang. Multi-parent Advanced Generation Inter-Cross populations), czy NAM (ang. Nested Association Mapping). Model bazujący na metodach genetyki populacyjnej nie wymaga znajomości pochodzenia badanych form. Zakłada jednak, że są one zróżnicowane, a w zależności od przyjętej strategii wykorzystuje się formy mniej lub bardziej wyrównane genetycznie. Niezależnie od konkretnego modelu badane populacje są genotypowane za pomocą markerów DNA oraz fenotypowane. W przypadku populacji mendlowskich zwykle budowane są mapy genetyczne, które następnie służą do identyfikacji obszarów genomu kodujących badaną cechę. Jeśli w tych obszarach znajdują się markery silnie sprzężone z cechą mogą one zostać poddane weryfikacji, a następnie wykorzystane w

procesie selekcji wspartej markerami molekularnymi (MAS, ang. Marker Assisted Selection). Przy takim podejściu użytecznym jest konstruowanie map konsensusowych, gdyż uwzględniają one zmienność genetyczną wielu form rodzicielskich. Nie zawsze jednak dostępne są populacje właściwe do mapowania genetycznego. Zwykle łatwiej dostępny jest zróżnicowany materiał hodowlany. W zależności od tego, czy celem jest selekcja cech warunkowanych silnymi, czy słabymi genami odpowiedzialnymi za ekspresję cechy wyróżnia się mapowanie asocjacyjne oraz selekcję genomową. Pierwsza, analogicznie do markerów identyfikowanych za pomocą modelu mendlowskiego, umożliwi identyfikację markerów określonych obszarów genomu, druga uznaje, że ważne są wszystkie markery gdyż odzwierciedlają złożoność genomu i interakcję poszczególnych jego obszarów, a tym samym również złożoność samej cechy.



Rys. 1. Modele selekcji bazujące na markerach DNA

Fig. 1. Selection models based on DNA markers

MODELE MENDLOWSKIE

Metody selekcji bazujące na modelu mendlowskim składają się z kilku etapów: wyboru populacji mapującej, fenotypowania, genotypowania, konstruowania mapy genetycznej, analizy statystycznej powiązań cecha-marker, typowania markerów do selekcji, a następnie ich walidacji. Istnieje wiele możliwych strategii w zależności od złożoności badanej cechy.

Cechy monogenowe

W selekcji cech monogenowych zakłada się, że marker jest silnie sprzężony z genem odpowiedzialnym za ekspresję cechy. Identyfikacja takiego markera zwykle wymaga wyprowadzenia populacji mapujących typu F_2 , BC lub DH. Następnie taka populacja jest fenotypowana i genotypowana. Do celów statystycznych wykorzystuje się metodę analizy pojedynczego markera (SMA, ang. Single-Marker Analysis) zwaną też metodą regresji (Sax, 1923). Metoda ta pozwala na wytypowanie markerów sprzężonych z cechą poprzez określenie korelacji fenotypu z segregującymi markerami DNA (Mauricio, 2001). Wykorzystuje ona trzy podstawowe metody statystyczne: t-test, analizę wariancji (ANOVA) oraz regresję liniową (Boopathi, 2013). Po raz pierwszy w genetyce roślin SMA została zastosowana przez Karla Saxa w 1923, który udowodnił za jej pomocą istnienie powiązań pomiędzy kolorem nasion fasoli a ich wagą (Sax, 1923). Zaletą SMA jest brak konieczności konstruowania map genetycznych. Marker wykryty tą metodą będzie użyteczny głównie na puli roślin, na której został opracowany i może się okazać nieefektywny w przypadku innych materiałów roślinnych (Xu i Crouch, 2008). Ze względu na prostotę eksperymentalną opisanego podejście znalazło zastosowanie w przypadku licznych chorób warunkowanych pojedynczymi genami oraz cech, których ekspresja zachodzi w późnych fazach rozwojowych bądź cech, które ujawniają się w zależności od warunków środowiskowo-klimatyczno-glebowych (Semagn i in., 2010; Boopathi, 2013).

Cechy wielogenowe

Identyfikacja markerów cech wielogenowych z wykorzystaniem genetyki mendelowskiej wymaga opracowania map genetycznych oraz dobrze fenotypowanych populacji, głównie dwurodzicielskich. Mapy genetyczne są zagęszczane markerami molekularnymi, a następnie analizowane statystycznie w celu wykrywania obszarów genomu warunkujących ekspresję cechy. Populacje dwurodzicielskie bazują na homozygotycznych roślinach rodzicielskich różniących się pod względem badanej cechy. Można je podzielić na dwie grupy: „śmiertelne” i „nieśmiertelne” (Boopathi, 2013). Pierwsze z nich mają ograniczone zastosowanie i służą zwykle do wstępnych badań. Przykładem takiej populacji są rośliny pokolenia F_2 . Natomiast takie populacje jak NIL, RIL, czy DH mogą być wykorzystywane wielokrotnie. Ich zaletą jest ustabilizowany genotyp i możliwość badania cechy w wielu powtórzeniach w różnych środowiskach bez konieczności prowadzenia wielokrotnych analiz molekularnych. Wybór populacji mapującej w znacznej mierze determinuje rozdzielczość uzyskiwanych map genetycznych (Soto-Cerda i Cloutier, 2012). Ze względu na liczbę procesów rekombinacyjnych najmniejszą rozdzielczością cechują się populacje pokolenia F_2 , a najwyższą linie RIL. W przypadku populacji DH częstość rekombinacji zostaje niezmienna z chwilą wytworzenia linii DH (Rafalski, 2010). Przy wyborze populacji mapującej należy uwzględnić czas potrzebny na jej wyprowadzenie. Najdłużej wyprowadza się linie RIL, najkrócej populacje DH. Kolejnym czynnikiem decydującym o wyborze populacji jest jej liczebność niezbędna do uzyskania pożądanej rozdzielczości przy mapowaniu oraz związane z tym koszty genotypowania i fenotypowania. Zależnie od wymaganego stopnia zagęszczenia map genetycznych do genotypowania wykorzystuje się różne systemy markerowe. Do niedawna były to markery kodominujące, takie jak markery polimorfizmu długości fragmentów restrykcyjnych

(RFLP, ang. Restriction Fragments Length Polymorphism), następnie markery mikrosatelitarnego polimorfizmu krótkich tandemowych powtórzeń (SSR, ang. Simple Sequence Repeats) oraz markery dominujące polimorfizmu losowo powielonych fragmentów DNA (RAPD, ang. Random Amplification Polymorphic DNA), markery amplifikacji polimorficznego DNA z użyciem arbitralnie dobranego startera (AP-PCR, ang. DNA fingerprinting by arbitrarily primed polymerase chain reaction), markery polimorfizmu długości amplifikowanego fragmentu (AFLP, ang. Amplified Fragments Length Polymorphism) czy DArT (ang. Diversity Arrays Technology). Obecnie, ze względu na malejące koszty sekwencjonowania oraz możliwość uzyskania nieograniczonej liczby markerów, zastosowanie znajdują markery GBS (Genotyping by Sequencing) czy DArTseq bazujące na sekwencjonowaniu nowej generacji (Next Generation Sequencing). Uzyskane markery DNA służą do opracowania map genetycznych niezbędnych do mapowania cech ilościowych.

Istotnym elementem mapowania cech ilościowych było rozwinięcie odpowiednich metod statystycznych. Pierwszą z takich metod było mapowanie interwałowe (IM, ang. Interval Mapping), które opracowano w 1989 roku (Lander i Botstein, 1989). Mapowanie interwałowe wykorzystuje informacje od dwóch brzegowych markerów i polega na ustaleniu sprzężenia między *locus* cechy ilościowej (QTL, ang. Quantitative Trait Loci), a markerami ograniczającymi wyznaczony przedział na mapie genetycznej (Zeng, 1993; Broman, 2001). Prawdopodobieństwo wystąpienia *locus* cechy ilościowej jest określane na podstawie krzywej prawdopodobieństwa rozkładu cechy (LOD, ang. logarithm of the odds ratio), a wiarygodność takiego wyniku jest testowana permutacyjnie (Lander i Botstein, 1989). Jeśli maksimum krzywej LOD przewyższa wartość krytyczną (wartość testu permutacji), to dany obszar genomu najprawdopodobniej jest odpowiedzialny za ekspresję badanej cechy. Praktyka pokazuje jednak, że uzyskiwany wynik niekoniecznie odzwierciedla rzeczywistą sytuację. Wielokrotnie obserwowano, że identyfikowany QTL nie był stabilny i zmieniał swoją lokalizację (Zeng, 1993). W niektórych przypadkach obserwowano występowanie silnych, ale całkowicie nierzeczywistych QTL-i (tzw. ghost QTL) (Jansen, 1993; Mauricio, 2001). Poza tym, jeśli cecha jest warunkowana kilkoma QTL-ami występującymi w bliskim sąsiedztwie, to uzyskiwane wyniki zwykle są silnie zniekształcone (Boopathi, 2013; Broman, 2001). Mimo ograniczeń metodę z powodzeniem wykorzystano do identyfikacji QTL-i związanych z cechami morfologicznymi soi takimi jak wielkość liści czy waga nasion (Mansur i in., 1993), do identyfikacji QTL związanych z cechami morfologicznymi i plonotwórczymi u ryżu (Thomson, 2003), określenia lokalizacji QTL-a determinującego wysokość roślin sorgo (Pereria i Lee, 1995), czy badań nad heterozją rzodkiewnika (Mitchel-Olds, 1995).

Rozwinięciem metody IM jest regresyjne mapowanie interwałowe (RIM, ang. Regression Interval Mapping), opracowane w roku 1992 przez Haleya i Knotta (Haley i Knott, 1992). RIM zwiększa prawdopodobieństwo mapowania i ogranicza czas poświęcony na obliczenia, jednakże zdolność wykrywania QTL-i w porównaniu z tradycyjnym IM jest tu mniejsza. W 1993 roku Jansen (1993) i Zeng (1993, 1994) połączyli mapowanie interwałowe ze złożoną analizą regresyjną. Zeng (1993) nazwał to podejście złożonym mapowaniem interwałowym (CIM, ang. Composite Interval Mapping). CIM

pozwała na sprawdzanie efektu QTL-i w innych interwałach lub chromosomach dzięki czemu zwiększa się precyzja ich wykrywania. Innym rozwiązaniem jest opisane w 1999 (Kao i in., 1999) wielokrotne mapowanie interwałowe (MIM, ang. Multiple Interval Mapping). Metoda ta pozwala na zredukowanie oddziaływania QTL-i w tej samej grupie sprzężeń. Dzięki temu uzyskane informacje o lokalizacji QTL-a pozbawione są błędów wynikających z jego obecności w innych regionach chromosomu. Poza tym stosowanie tej metody pozwala wykryć epistazę i oszacować indywidualną wartość genotypową oraz zdolność dziedziczenia. Wielokrotne mapowanie interwałowe jest obecnie najbardziej precyzyjną metodą wykrywania QTL. Z powodzeniem zostało wykorzystane do mapowania licznych QTL u wielu gatunków roślin (Qi in., 2014; Zeng i in., 2014; Chankaew i in., 2014).

Entuzjazm związany z wykorzystaniem IM, a następnie wariantów CIM został stopniowo zastąpiony umiarkowanym optymizmem. Okazało się, że mimo selekcji określonych obszarów genomu kontrolowanej za pomocą markerów DNA, podejmowane działania nie dawały oczekiwanych rezultatów (Mauricio, 2001; Broman, 2001). Uznano, że przyczyną niepowodzeń mogło być niekorzystne usytuowanie obszarów kodujących cechę w genomie biorcy, czy też brak sekwencji genomowych odpowiedzialnych za jej ekspresję. Nawet w przypadku udanej selekcji (np. gen przywracania płodności pyłku u żyta z CMS pampa lokalizujący się na chromosomie 4R) pożądanego genu za pomocą silnie sprzężonych markerów wraz z cechą przenoszone były geny niekorzystne (przeciąganie genetyczne) (Hospital, 2005). Kolejnym problemem okazała się niska powtarzalność lokalizacji cechy na mapach genetycznych. Uznano, że może to być przyczyną niepowodzenia przy próbach selekcji cech ilościowych (Boopathi, 2013). Kolejnym ograniczeniem metody mapowania cech ilościowych okazało się ich "uzależnienie" od form rodzicielskich populacji mapujących (Soto-Cerda i Cloutier, 2012). Możliwym było identyfikowanie wyłącznie tych QTL, które były reprezentowane przez rodziców populacji. Selekcja nie mogła wykorzystywać całej dostępnej puli genetycznej cechy czy też licznych form allelicznych poszczególnych genów (Semagn i in., 2010). Początkowo rozwiązaniem było prowadzenie wielu populacji mapujących. Jednak takie podejście okazało się kosztowne i pracochłonne i nie dawało gwarancji powodzenia. Istotne znaczenie miało to dla cech kodowany przez wiele *loci*, które miały średni i niski wkład do wytłumaczonej zmienności fenotypowej. Zaletą metody było jednak opracowanie licznych map genetycznych gatunków roślin uprawnych. Stworzyło to podwaliny rozwoju zaawansowanych metod analitycznych.

Omawiane powyżej metody w zdecydowanej większości bazują na jednej populacji mapującej. Takie podejście jest mało wydajne, gdyż uwzględnia zmienność dostępną jedynie w obrębie badanych form rodzicielskich, a identyfikowane QTL niekoniecznie muszą być najistotniejszymi QTL-ami cechy (Boopathi, 2013). Ponadto, pojedyncze mapy dostarczają ograniczonej informacji o lokalizacji markerów DNA i przez to typowanie nowych markerów cechy jest utrudnione i wymaga wyprowadzania dodatkowych populacji mapujących. Rozwiązaniem jest opisane w 1995 roku tworzenie zintegrowanych map konsensusowych. Polega ono na poszukiwaniu wspólnych markerów dla wielu populacji mapujących i aproksymowaniu lokalizacji pozostałych markerów na wirtualnych mapach

gatunku, co daje pełniejsze pokrycie genomu (Yap i in., 2003). Dane z różnych populacji stają się wspólną pulą informacji o położeniu *loci* na chromosomach i ich wzajemnego uszeregowania. Jako pierwsze powstały mapy konsensusowe dla jęczmienia (Langridge i in., 1995) i pszenicy (Gale i in. 1995). Następnie opracowano mapy m.in. dla żyta (Börner i Korzun, 1998; Milczarski i in. 2011), ryżu (Kennard, 1999) czy pszenżyta (Alheit i in., 2011).

Populacje wielorodzicielskie

Zwiększenie zmienności genotypowej można osiągnąć poprzez zastosowanie populacji wielorodzicielskich NAM lub MAGIC. Do utworzenia NAM wykorzystuje się liczne dwurodzicielskie populacje mapujące (najlepiej RIL), które mają wspólnego rodzica (Yu i in., 2008). Otrzymane populacje wykorzystuje się zarówno do mapowania interwałowego jak i asocjacyjnego. W populacjach MAGIC wyjściową pulę, zwykle ośmiu wyrównanych linii rodzicielskich krzyżuje się ze sobą tak by uzyskać zestaw roślin pokolenia F₁ o ściśle zdefiniowanym dwurodzicielskim pochodzeniu reprezentowanym przez wszystkie możliwe kombinacje form rodzicielskich (Cavanagh i in., 2008). Rośliny pokolenia F₁ krzyżuje się następnie, aż do uzyskania mieszańców łączących całą pulę ośmiu wyjściowych genotypów i wyprowadza populacje RIL. Taki model pozwala na bardzo precyzyjną detekcję QTL-i z dokładnością do sub-centymorganów przy 1000 osobników. Ponadto 8-rodzicielskie populacje RIL charakteryzuje wysokie zróżnicowanie zmienności allelicznej i fenotypowej oraz barak silnej struktury (Cavanagh i in., 2008).

MODELE POPULACYJNE

Omawiane powyżej metody u swej podstawy były ukierunkowane na ściśle zdefiniowane obszary genomu odpowiedzialne za ekspresję badanej cechy. Okazało się jednak, że markery identyfikowane za pomocą opisanych metod wykorzystujących populacje dwu- oraz wielorodzicielskie, choć umożliwiały selekcję cechy, nie zawsze prowadziły do oczekiwanego rezultatu (Heffner, 2009). Taka sytuacja może mieć miejsce gdy cecha jest złożona, a geny odpowiedzialne za jej ekspresję warunkują niewielki stopień wytłumaczonej wariancji. Na ekspresję cechy może mieć również wpływ tło genetyczne (Hamblin, 2011). Tak więc selekcja tylko wybranych obszarów może być mało wydajna. Alternatywą do omawianych wyżej metod wykorzystujących do mapowania modele mendelowskie są stosowane od niedawna modele populacyjne, do których możemy zaliczyć mapowanie asocjacyjne (Pritchard, 2000) oraz rozwijaną obecnie selekcję genomową (Meuwissen i in., 2001).

Mapowanie asocjacyjne

Mapowanie asocjacyjne (AM, ang. Association Mapping), zakłada, że do badań brane są liczne materiały roślinne (zwykle niespokrewnione lub daleko spokrewnione o dużym stopniu wyrównania w obrębie poszczególnych linii) (Rafalski, 2010). Liczebność badanej populacji jest zależna od gatunku oraz cechy. W celu osiągnięcia pożądanego efektu wielkość populacji mapującej może się wahać od około 200 do 800 osobników (Hall i in., 2010). Zwiększanie rozmiaru populacji zwykle nie jest uzasadnione ekonomicznie (Semagn i in., 2010). Zastosowanie zróżnicowanych materiałów roślinnych powoduje, że

akty rekombinacji zachodzące we wcześniejszych etapach wyprowadzenia materiałów były częste, a zawarta pomiędzy nimi zmienność odzwierciedla zmienność danej puli genetycznej (Soto-Cerda i Cloutier, 2012). Tak więc identyfikowane markery cechy powinny być użyteczne na szerszym materiale roślinnym. Metoda zakłada, że wszystkie analizowane materiały roślinne są badane pod względem konkretnej cechy oraz są profilowane za pomocą odpowiednich markerów DNA. W przeciwieństwie do metod bazujących na populacjach biparentalnych w mapowaniu asocjacyjnym zamiast sprzężenia markera z cechą określa się jego asocjację. Im silniej marker jest asocjowany z cechą tym silniejsze niezrównoważenie sprzężeń (LD, ang. linkage disequilibrium). Oznacza to, że prawdopodobieństwo identyfikacji obszaru występowania cechy w genomie ulega zawężeniu do węższego obszaru genomu niż to ma miejsce w przypadku badań prowadzonych na populacjach dwurodzicielskich (Soto-Cerda i Cloutier, 2012). Wynika to z faktu, że podczas wielu aktów rekombinacji tylko markery występujące bardzo blisko poszukiwanych genów pozostały w ich sąsiedztwie podczas gdy pozostałe obszary genomu, w miarę kolejnych aktów rekombinacji, wykazują coraz niższe niezrównoważenie. Aby móc identyfikować markery cechy należy dysponować dużą liczbą markerów, które gęsto i równomiernie pokrywają genom. Gęstość takiego pokrycia jest zależna od wartości niezrównoważenia sprzężeń, które będzie zależęć od gatunku oraz cechy (Hall i in., 2010). W efekcie do takich badań konieczne są markery uzyskiwane metodami sekwencjonowania nowej generacji takie jak GBS (Elshire i in., 2011) czy DArTseq (Carling i in., 2015) oraz stosunkowo duże moce obliczeniowe (Hall i in., 2010).

W mapowaniu asocjacyjnym wyróżniono dwa podejścia: asocjacja genów kandydatów (ang. Candidate gene asociacion) oraz skanowanie całego genomu (GWAS, ang. Genome — Wide Association Study). W przypadku asocjacji genów kandydatów sprawdzana jest hipoteza: „czy jest jakaś korelacja pomiędzy polimorfizmem DNA w określonym genie a cechą”. Przy braku szczegółowej wiedzy biochemicznej związanej z poszukiwaną cechą, uzasadnionym jest podejście GWAS. GWAS poszukuje asocjacji cecha-marker w całym genomie i zakłada, że w obrębie genomu warunkującego ekspresję cechy znajdują się markery wykazujące niezrównoważenie sprzężeń (Rafalski, 2010). Początkowo mapowanie asocjacyjne wykonano u kukurydzy (Bar-Hen i in., 1995), ryżu (Virk i in., 1996) i owsa (Beer i in., 1997) stosując izozymy, RAPD i RFLP. Prace te nie uwzględniały struktury populacji, której istotność została opisana przez Pritcharda i in. (2000), a następnie po raz pierwszy zastosowana w badaniach u kukurydzy (Pritchard 2001). Wraz z rozwojem wydajnych metod markerowania oraz udostępnieniem programów statystycznych liczba badanych gatunków wzrosła, a identyfikowane tą metodą markery DNA są obecnie stosowane w praktyce hodowlanej (Soto-Cerda i Cloutier, 2012). Mapowanie asocjacyjne okazało się użyteczne do identyfikacji markerów cech których loci ilościowe tłumaczą znaczną zmienności cechy (Soto-Cerda i Cloutier, 2012). Metoda ta ma jednak ograniczone zastosowanie w przypadku cech złożonych o słabych efektach poszczególnych loci (Hall i in., 2010).

Selekcja genomowa (SG)

Selekcja oparta o wybrane regiony genomu nie uwzględniała tła genetycznego oraz roli "mało" istotnych genów czy też ich oddziaływań epistatycznych. Dopiero podejście

populacyjne ukierunkowane na analizę efektów porozrzucanych po całym genomie, a występujących w populacji wydawało się być właściwym rozwiązaniem. Przez długi czas realizacja takich badań była praktycznie niemożliwa ze względu na niewystarczające moce obliczeniowe komputerów oraz brak odpowiednich narzędzi statystycznych. Wraz z rozwojem mocy obliczeniowych oraz opracowaniem statystycznych podstaw SG możliwym stało się przetwarzanie danych, które można wykorzystać do prowadzenia selekcji genomowej. Metoda po raz pierwszy została opisana przez Meuwissen w 2001 roku (Meuwissen i in., 2001). Selekcja genomowa wykorzystuje nowe modele statystyczne i narzędzia bioinformatyczne połączone z coraz większą wiedzą o genomie. W SG markerom molekularnym przypisuje się pewne wartości powiązania z badaną cechą bazując na precyzyjnym fenotypowaniu zróżnicowanych materiałów roślinnych. Takie powiązanie z badaną cechą określa się jako genomową szacunkową wartość hodowlaną (GEBV — Genetic Estimated Breeding Value) (Heffner i in., 2009), a rośliny charakteryzujące się najwyższą wartością GEBV są wykorzystywane do krzyżowań (Jonas i Koning, 2013). Żaden z zastosowanych markerów nie jest pomijany w analizie dzięki czemu uwzględniony jest całokształt zmienności cechy oraz tła genetycznego. Zastosowania GEBV stało się wykonalne w połączeniu z wysokoprzepustowymi platformami genetycznymi (Hamblin i in., 2011).

Aby prowadzić SG formuje się populację treningową składającą się z linii odzwierciedlających materiał programu hodowlanego. Taka populacja treningowa powinna składać się z co najmniej 500 form. Następnie przeprowadza się jej genotypowanie i fenotypowanie. Dzięki temu uzyskuje się zbiór danych zawierający zapis fenotypowej i genotypowej charakterystyki osobników. Populacja treningowa jest wykorzystywana jedynie w celu oszacowania wartości GEBV. Precyzyjne określenie GEBV jest znaczącym czynnikiem w selekcji genomowej. Pozwala ono ustalić prognozowaną wartość danej linii dla której zostaje wyznaczone dzięki czemu hodowca może określić jakimi cechami będzie charakteryzować się roślina przed jej wysianiem. Aby uzyskać dokładne GEBV stosuje się odpowiednie metody statystyczne. Najczęściej są to: najbliższa liniowa nieobciążona predykcja (BLUP ang. Best Linear Unbiased Prediction) i jej rozszerzenia RR-BLUP (Ridge Regression BLUP) czy GBLUP (Genomic BLUP) oraz metoda Bayesian (Meuwissen i in., 2001, Habier i in., 2013). Wszystkie te modele statystyczne ulegają ciągłym przekształceniom, przez co powstają nowe platformy wykorzystywane w SG (Habier i in., 2013).

Po oszacowaniu GEBV do dalszego krzyżowania wykorzystywane są osobniki o najwyższej wartości GEBV dzięki czemu uzyskuje się populacje mieszańców wzbogaconych o pożądane allele. Typowanie kolejnych osobników do krzyżowań odbywa się na podstawie genotypowej wartości hodowlanej (z pominięciem etapu fenotypowania). Zwykle przeprowadza się kilka cykli selekcji w oparciu o GEBV z ewentualnym etapem korygowania modelu treningowego. Wyselekcjonowaną pulę form wzbogaconych o pożądane allele zwykle należy wyrównać pod względem genetycznym poprzez chów wsoalny lub metodami kultur *in vitro*. Zaletą SG jest zachowanie wyjściowej zmienności badanych materiałów z jednoczesnym wyprowadzeniem form wzbogaconych o pożądane allele. SG minimalizuje czasochłonną ocenę fenotypową cech wielokrotnie złożonych i

skraca długość cyklu hodowlanego powodując wzrost korzyści w jednostce czasu. Sprawdza się to zwłaszcza dla roślin wieloletnich, które wymagają kilkuletniej oceny fenotypowej (Jonas i Koning, 2013).

PODSUMOWANIE

Od dłuższego czasu można zauważyć wzrost nacisku na zwiększenie i zrównoważenie produkcji żywnościowej. W związku z tym opracowuje się coraz nowsze narzędzia gwarantujące większą dokładność selekcji. Obecnie wykorzystywane metody selekcji zostały wzbogacone o osiągnięcia biologii molekularnej oraz modele statystyczne umożliwiające zarówno identyfikację markerów poszczególnych cech będących wynikiem działania pojedynczych genów jak i takich które są uwarunkowane wieloma QTL-ami w różnym stopniu tłumaczącymi zmienność fenotypową cechy. Najnowsze metody selekcji umożliwiają realizację prac zarówno z wykorzystaniem modeli mendelowskich, jak i populacyjnych. Każdy z wymienionych modeli ma swój zdefiniowany obszar działania, które definiuje jego wykorzystanie w ramach konkretnych potrzeb eksperymentalnych.

LITERATURA

- Alheit K. V., Reif J. C., Maurer H. P., Hahn V., Weissmann E.A., Miedaner T., Würschum T. 2011. Detection of segregation distortion *loci* in triticale (*x Triticosecale* Wittmack) based on a high density DArT marker consensus genetic linkage map. *BMC Genomics* 12: 380.
- Bar-Hen A., Charcosset A., Bourgoin M., Guiard J. 1995. Relationship between genetic markers and morphological traits in a maize inbred line collection. *Euphytica* 84, 2: 145 — 154.
- Beer S., Siripoonwiwat W., O'Donoghue L., Souza E., Mathews D., Sorrells M. 1997. Associations between molecular markers and quantitative traits in oat germplasm pool: can we infer linkages? *J. Agric. Genom.* 3 <http://wheat.pw.usda.gov/jag/papers97/paper197/jqt>.
- Boopathi M. N., 2013. Genetic mapping and marker assisted selection: basics, practice and benefits. Published by Springer India: 23–37 pp., 117 — 163 pp.
- Börner A., Korzun V. 1998. A consensus linkage map of rye (*Secale cereale* L.) including 374 RFLPs, 24 isozymes and 15 gene loci. *Theor. Appl. Genet.* 97, 8: 1279 — 1288.
- Broman K. W. 2001. Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal.* 30, 7: 44 — 52.
- Carling J., Heller-Uszyńska K., Jaccoud D., Machado A., Hopper C., Xia L., Vippin C., Caig V., Uszyński G., Kilian A. 2015. DArTTM and DArTseqTM genome profiling for breeding, pre-breeding and population genetics applications. Contribution P0052, XXIII Plant and Animal Genome, San Diego, CA 10 — 14.
- Cavanagh C., Morell M., Mackay I., Powell W. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Cur. Opin. Plant Biol.* 11: 215 — 221.
- Chankaew S., Isemura T., Naito K., Ogiso-Tanaka E., Tomooka N., Somta P., Kaga A., Vaughan D. A., Srinives P. 2014. QTL mapping for salt tolerance and domestication-related traits in *Vigna marina* subsp. *oblonga*, a halophytic species. *Theor. Appl. Genet.* 127, 3: 691 — 702.
- Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., Buckler E. S., Mitchell S. E. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* 6, 5: e19379.
- Gale M. D., Atkinson M. D., Chinoy C. N., Harcourt R. L., Jia J., Li Q. Y., Devos K. M. 1995. Genetic maps of hexaploid wheat. In *Proc 8th Int. Wheat Genet Symp.* China Agricultural Sciencetech Press, Beijing (pp. 29 — 40).
- Habier D., Fernando R. L., Garrick D. J. 2013. Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194: 597 — 607.

- Haley C. S., Knot S. A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315 — 324.
- Hall D., Tegström C., Ingvarsson P. K. 2010. Using association mapping to dissect the genetic basis of complex traits in plants. *Brief Funct Genomics* 9, 2: 157 — 165.
- Hamblin M. T., Buckler E. S., Jannink J. L. 2011. Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27, 3: 98 — 106.
- Heffner E. L., Sorrells M. E., Jannink J.-L. 2009. Genomic selection for crop improvement. *Crop Sci.* 49: 1 — 12.
- Hospital F. 2005. Selection in backcross programmes. *Phil. Trans. R. Soc. B.* 360: 1503 — 1511.
- Jansen R. C. 1993. Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205 — 211.
- Jonas E., de Koning D.-J. 2013. Does genomic selection have a future in plant breeding?, *Trends Biotechnol.* 31, 9: 497 — 504.
- Kao C.-H., Zeng Z.-B., Teasdale R. D. 1999. Multiple Interval Mapping for Quantitative Trait Loci. *Genetics.* 152, 3: 1203 — 1216.
- Kearsey M. J., Farquhar A. G. L. 1998. QTL analysis; where are we now? *Heredity* 80, 2: 137 — 142.
- Kennard W., Phillips R., Porter R., Grombacher A., Phillips R. L. 1999. A comparative map of wild rice (*Zizania palustris* L. $2n=2x=30$). *Theor Appl Genet.* 99: 5: 793 — 799.
- Lander E. S., Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 1: 185 — 199.
- Langridge P., Karakousis A., Collins N., Kretschmer J., Manning S. 1995. A consensus linkage map of barley. *Mol Breed.* 1: 389 — 395.
- Mansur L. M., Lark K. G., Kross H., Oliveira A. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). *Theor. Appl. Genet.* 86, 8: 907 — 913.
- Mauricio R. 2001. Mapping Quantitative Trait loci in plants: uses and caveats for evolutionary biology. *Nature* 2: 370 — 380.
- McMullen M. D., Louie R. 1989. The linkage of molecular markers to a gene controlling the symptom response in maize to maize dwarf mosaic virus. *Mol Plant Microbe Interact.* 2: 309 — 314.
- Meuwissen T. H. E., Hayes B. J., Goddard M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819 — 1829.
- Milczarski P., Bolibok-Bragoszewska H., Myśków B., Stojalowski S., Heller-Uszyńska K., Góralaska M., Brągoszewski P., Uszyński G., Kilian A., Rakoczy-Trojanowska M. 2011. A high density consensus map of rye (*Secale cereale* L.) based on DArT markers. *PLoS ONE* 6:12: e28495.
- Mitchell-Olds T. 1995. Interval mapping of viability loci causing heterosis in Arabidopsis. *Genetics.* 140, 3: 1105 — 1109.
- Nakaya A., Isobe S. N. 2012. Will genomic selection be a practical method for plant breeding? *Ann Bot.* 110: 1303 — 1316.
- Pereira M. G., Lee M. 1995. Identification of genomic regions affecting plant height in sorghum and maize. *Theor Appl Genet.* 90, 3-4: 380 — 388.
- Pritchard J. 2001. Deconstructing maize population structure. *Nat Genet.* 28 (3): 203 — 204.
- Pritchard J. K., Stephens M., Rosenberg N.A., Donnelly P. 2000. Association mapping in structured populations. *Am J Hum Genet.* 67: 170 — 181.
- Qi Z., Han X., Hou M., Xin D., Wang Z., Zhu R., Hu Z., Jiang H., Li C., Liu C., Hu G., Chen Q. 2014. QTL analysis of soybean oil content under 17 environments. *Can. J. Plant Sci.* 94: 245 — 261.
- Rafalski J. A. 2010. Association genetics in crop improvement. *Current Opinion in Plant Biology* 13, 2: 174 — 180.
- Sax K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8: 552 — 560.
- Semagn K., Bjørnstad Å., Ndjiondjop M. N. 2006. An overview of molecular marker methods for plants. *Afr. J. Biotechnol.* 5, 25: 2540 — 2568.
- Semagn K., Bjørnstad A., Xu Y. 2010. The genetic dissection of quantitative traits in crops. *Electron J Biotechnol.* 13: 5, <http://dx.doi.org/10.2225/vol13-issue5-fulltext-14>.

- Soto-Cerda B. J., Cloutier S. 2012. Genetic Diversity in Plants. Association mapping in plant genomes. Red. Çalışkan M., <http://www.intechopen.com/books/genetic-diversity-in-plants/association-mapping-in-plant-genomes>, pp. 29 — 54.
- Thomson M. J., Tai T. H., McClung A. M., Lai X-H., Hinga M. E., Lobos K. B., Xu Y., Martinez C. P., McCouch S. R. 2003. Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theor. Appl. Genet.* 107, 3: 479 — 93.
- Virk P., Ford-Lloyd B., Jackson M., Pooni H., Clemeno T., Newbury H. 1996. Predicting quantitative variation within rice germplasm using molecular markers. *Heredity* 76, 3: 296 — 304.
- Xu Y., Crouch J. H. 2008. Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* 48: 391 — 407.
- Yap I. V., Schneider D., Kleinberg J., Matthews D., Cartinhour S., McCouch S. R. 2003. A graph-theoretic approach to comparing and integrating genetic, physical and sequence-based maps. *Genetics* 165: 2235 — 2247.
- Young N. D., Zamir D., Ganai M. W., Tanksley S. D. 1988. Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the Tm-2a gene in tomato. *Genetics* 120: 579 — 585.
- Yu J., Holland J. B., McMullen M. D., Buckler E. S. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178: 539 — 551.
- Zeng A., Chen P., Shi A., Wang D., Zhang B., Orazaly M., Florez-Palacios L., Brye K., Song Q., Cregan P. 2014. Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Sci.* 54, 2: 554 — 564.
- Zeng Z.-B. 1993. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proc. Natl. Acad. Sci.* 90: 10972 — 10976.
- Zeng Z.-B. 1994. Precision mapping of quantitative trait loci. *Genetics* 136: 1457 — 1468.