

SPRAWOZDANIE MERYTORYCZNE

z realizacji zadania na rzecz postępu biologicznego w produkcji roślinnej w 2015 roku

Tytuł zadania: **Fenotypowanie i genotypowanie lubinu wąskolistnego pod względem wybranych cech morfologicznych, plonotwórczych i parametrów technologicznych nasion (93)**

Zespół badawczy: Renata Galek r. hab. inż., Bartosz Kozak dr inż., Dariusz Zalewski dr inż.

Celem zadania było: scharakteryzowanie materiałów kolekcyjnych pod względem cech jakościowych i wybranych elementów struktury plonu, dobór form rodzicielskich zróżnicowanych pod względem grubości okrywy nasiennej i parametrów jakościowych nasion do programu krzyżowan i otrzymanie nasion mieszańcowych pokoleń F₁ oraz wytypowanie markerów SNP użytecznych do genotypowania lubinu.

Materiał badawczy stanowiły przekazane przez hodowców materiały hodowlane oraz materiały kolekcyjne zgromadzone w Katedrze Genetyki, Hodowli Roslin i Nasiennictwa. W roku 2015 poddano ocenie 50 genotypów w zakresie: parametrów technologicznych nasion (zawartości białka ogólnego, tłuszczu, alkaloidów, włókna – oznaczenia wykonano w COBORU zgodnie z obowiązującymi normami) oraz plonu i cech z nim związanych (masą tysiąca nasion, długością kwiatostanu pędu głównego i bocznego, liczbą okółków, liczbą zawiązanych strąków na kwiatostanie pędu głównego i bocznego oraz w przypadku form o zdeterminowanym typie wzrostu łączną liczbą strąków z pędu głównego i bocznego, wysokością do pierwszego produktywnego rozgałęzienia, liczbą rozgałęzień bocznych I-rzędu). Określono również wysokość roślin i pędu głównego. Doświadczenie polowe zostało założone na terenie stacji Hodowli Roślin Smolice sp. z o.o. Oddział Przebędowo, metodą pasów prostokątnych w trzech powtórzeniach, na poletkach o powierzchni 3,9 m². Dla zweryfikowania hipotezy zerowej o braku zróżnicowania badanych obiektów pod względem analizowanych cech, przeprowadzono analizę wariancji. W przypadku wykazania istotnych różnic obliczono wartość NIR, podano wartość minimalną i maksymalną oraz współczynnik zmienności. Dodatkowo obliczono współczynnik korelacji dla badanych cech oraz pogrupowano obiekty za pomocą analizy skupień, wykreślając dendrogram metodą Warda. Przeprowadzono również analizę składowych głównych (PCA), w celu określenia, które z cech odgrywają istotną rolę w kształtowaniu zmienności w ocenianej kolekcji.

Dla wybranych 7 genotypów (wytypowanych pod względem grubości okrywy nasiennej, zawartości białka ogólnego, alkaloidów) z 50, które wzięto wstępnie pod uwagę do programu krzyżowań przeprowadzono dodatkowe analizy. Obejmowały one określenie udziału frakcji białek (stosując 4 sposoby ich izolacji). Poszczególne frakcje białek zostały rozdzielane za pomocą elektroforezy SDS-PAGE prowadzonej według metody opisanej przez Laemmli (1970), stosując 12% (w/v) i 5% (w/v) żele poliakrylamidowe. Supernatant zawierający wszystkie białka rozpuszczalne był analizowany ilościowo za pomocą metody Lowry i in. (1951). Ponadto określono udział poszczególnych aminokwasów (g/16 g N) oraz węglowodanów rozpuszczalnych (mg/g nasion) (Piotrowicz-Cieślak 2005).

Dla wyprowadzenia pokolenia F₂, a w kolejnych etapach linii RILs przeprowadzono program krzyżowań wg poniższego schematu:

- a) linia australijska o cienkiej okrywie nasiennej (R004) x linia polska o grubej okrywie nasiennej (R001)
- b) linia polska o grubej okrywie nasiennej (R001) x linia australijska o cienkiej okrywie nasiennej (R004)
- c) linia o wysokiej zawartości białka (R009) jednocześnie niskiej alkaloidów x linia o niskiej zawartości białka (R005) i wysokiej zawartości alkaloidów

- d) linia o wysokiej zawartości alkaloidów (R005) x linia o niskiej zawartości alkaloidów (R009)

Dobór form rodzicielskich nastąpił w oparciu o własne dane (Clements i in. 2014, Sawicka – Sienkiewicz i in. 2006) dotyczące cechy cienkościemności oraz po konsultacji z hodowcą.

W pierwszym etapie badań nad genotypowaniem łubinu wąskolistnego należało wyłonić zestaw markerów SNP. Do analiz bioinformatycznych wykorzystano sekwencje 1152 markerów SNP opracowanych dla populacji mapującej ExL otrzymanych w analizach DArTseq. Analizowane sekwencje o długości 69 nukleotydów porównano z wykorzystaniem algorytmu BLAST (Madden T. The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK21097/>) z referencyjnym genomem dla łubinu wąskolistnego (odmiana Tanjil, Yang i in. 2013). Na potrzeby projektu napisano specjalistyczny program komputerowy. W programie wykorzystano biblioteki: BioPython (http://biopython.org/wiki/Main_Page), Pandas (<http://pandas.pydata.org/pandas-docs/stable/index.html>) oraz Numpy. Przed przystąpieniem do analiz BLAST utworzono lokalną bazę danych wykorzystując referencyjne sekwencje genomowe dostępne dla łubinu wąskolistnego (Yang i in. 2013) z publicznej internetowej bazy danych (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000338175.1_Lupin_genome_scaffold/GCA_000338175.1_Lupin_genome_scaffold_genomic.fna.gz). Do analiz BLAST wykorzystano sekwencje uzyskane dla odmiany Emir w analizach DArTseq.

Genomowe DNA do analiz wyizolowano metodą CTAB (Doyle i Doyle 1990) z liści dwutygodniowych siewek. Po izolacji ilość i jakość wyizolowanego DNA sprawdzono elektroforetycznie. Genotypowanie metodą KASP (ang. *Kompetitive Allele Specific PCR*) przeprowadzono z wykorzystaniem zmodyfikowanego protokołu PCR oraz sond typu FRET (ang. *Fluorescence Resonance Energy Transfer*) z barwnikami fluorescencyjnymi (FAM i HEX). Analizy przeprowadzono w urządzeniu RT-PCR Illumina Eco.

Analizując wyniki składu jakościowego nasion dla materiałów kolekcyjnych ocenianych w jednorocznym doświadczeniu, stwierdzono że średnia zawartość białka wynosi 33%, tłuszczu 6,5%, alkaloidów 0,134% a włókna 16%. Wśród ocenianych genotypów znalazły się takie, które zdecydowanie odbiegały od wartości średniej. Podwyższoną zawartością białka powyżej 35% charakteryzowały się takie obiekty jak: R009, R020, R021, R023, R031. Dwie linie charakteryzowały się bardzo niską zawartością białka - na poziomie 25 i 29%. Linia o bardzo niskiej zawartości białka R044 charakteryzowała się jednocześnie najwyższą zawartością tłuszczu, zawierała też sporo włókna. Linia R038 na podstawie jednorocznych wyników charakteryzowała się najniższą zawartością włókna 13,99% a najwyższą 17,69% linia R003. Zawartość sumy alkaloidów wyniosła dla linii R050 - 0,002 a dla czterech linii (R002, R003, R005 i R006) zawierała się w przedziale 1,740-1,440.

Po przeprowadzeniu szczegółowych analiz składu alkaloidów stwierdzono, że analizowane genotypy różnią się między sobą ich składem. Zasadniczo oznaczono 6 alkaloidów: isoangustifolinę, angustifolinę, izolupaninę, lupaninę, 13OH lupaninę, hydroxylupaninę. Pierwszy z nich był charakterystyczny tylko dla 8% linii (R002, R003, R005 i R006), drugi dla 54% badanych linii, trzeci 46%. Lupanina występowała u wszystkich genotypów, natomiast 13OH lupaniny nie stwierdzono tylko u linii R029. Hydroxylupanina była charakterystyczna tylko dla czternastu linii. Najwięcej średnio nasiona zawierały lupaniny (61%), 13OHlupaniny (29%), angustifoliny (10%), izolupaniny (6,8%), hydroxylupaniny (2,7%) oraz isoangustifoliny (0,3%).

Zastosowane sposoby izolacji białek pozwoliły na uzyskanie profili różnicujących badane 7 genotypów R001, R004, R005, R006, R007, R008, R009, zwłaszcza pod względem

albumin oraz białek rozpuszczalnych w wodzie niskocząsteczkowych. Odnotowano też zróżnicowany udział poszczególnych frakcji w obrębie genotypu.

Badane genotypy pod względem zawartości poszczególnych aminokwasów różniły się istotnie, z wyjątkiem metioniny, której średnia zawartość wyniosła 0,75 (g/16 g N). Największą zmienność obserwowano pod względem zawartości poszczególnych aminokwasów u analizowanych siedmiu genotypów, rzędu 7 do 11% stwierdzono dla lizyny i tyrozyny oraz treoniny, a dla pozostałych współczynnik zmienności nie przekroczył 4%.

Pod względem zawartości węglowodanów rozpuszczalnych w wodzie stwierdzono znaczne zakresy zmienności. Największą zmienność obserwowano dla D-*chiro*-inozytolu (21,6%), fruktozy (17,7%), D-pinitolu, *myo*-inozytolu i galaktinolu (13-14,7%). Najmniej zróżnicowane genotypy były pod względem zawartości rafinozy i stachiozy (V% poniżej 5%) oraz werbaskozy i sacharozy – współczynnik zmienności na poziomie 7,6-8,5%. Dwa genotypy z 7 zawierały glukozę Uzyskane wyniki fenotypowania na podstawie elementów struktury plonu wskazują na istotne zróżnicowanie analizowanych obiektów, a wartości współczynników zmienności (od 13,7% dla plonu z poletka do 69% w przypadku masy tysiąca nasion) uwidaczniają duże zróżnicowanie w obrębie poszczególnych cech. Obliczone współczynniki korelacji wskazały na istotną ujemną zależność między zawartością tłuszczu a białka, liczbą produktywnych rozgałęzień a plonem, masą 1000 nasion a liczbą produktywnych rozgałęzień, liczbą zawiązanych strąków z pędu głównego a liczbą produktywnych rozgałęzień oraz wysokością do I produktywnego rozgałęzienia a długością kwiatostanu pędu głównego. Dodatkowo korelacje stwierdzono między takimi cechami jak: plonem a wysokością do I produktywnego rozgałęzienia i masą 1000 nasion oraz liczbą strąków zawiązanych na kwiatostanie pędu głównego; masą 1000 nasion a wysokością pędu głównego i całej rośliny, wysokością pędu głównego i całej rośliny oraz długością kwiatostanu pędu głównego i bocznego oraz liczbą okółków kwiatostanu pędu głównego. Analiza skupień pozwoliła na porównanie analizowanych obiektów pod względem ich podobieństwa fenotypowego - zostało utworzone 7 skupień. Do dwóch najliczniejszych zakwalifikowano po 13 genotypów, a samodzielny klaster utworzył genotyp R005. Warto zauważyć, iż genotypy przeznaczone do programu krzyżowań znalazły się w odrębnych klastrach i uplasowały się w znacznej odległości od siebie. Wykorzystując analizę składowych głównych stwierdzono, że pierwsze sześć wyjaśniało odpowiednio: 20,81%, 16,7%, 13,3%, 9,5%, 8,8%, 7,6%, co łącznie odpowiadało 76,71% obserwowanej zmienności. Pierwsza składowa główna była związana z wysokością pędu głównego, długością kwiatostanu pędu głównego i bocznego, liczbą zawiązanych strąków na pędzie bocznym; druga z plonem z poletka, liczbą rozgałęzień produktywnych; trzecia z liczbą okółków i zawiązanych strąków na pędzie głównym, czwarta z liczbą zawiązanych strąków na kwiatostanie pędu głównego, piąta z zawartością białka i tłuszczu a szósta z zawartością sumy alkaloidów i włókna. Stwierdzone zróżnicowanie badanych genotypów już w pierwszym roku badań zarówno pod względem cech jakościowych nasion oraz elementów plonowania daje podstawę do określenia w przyszłości, po kolejnych cyklach doświadczeń, roli interakcji genotypowo-środowiskowej a poziomem cech. Szczegółowa ocena potencjalnych form rodzicielskich nie tylko pod względem cech związanych z elementami plonowania, ale też składu jakościowego, dadzą podstawę do genotypowania *L. angustifolius*. Interesujący jest fakt znacznego zróżnicowania w zawartości frakcji białek, co daje podstawę do zintensyfikowania prac z wykorzystaniem elektroforezy 2D dla ich precyzyjnej identyfikacji.

Doświadczenia własne wskazują, że proces krzyżowania u łubinu wąskolistnego jest stosunkowo efektywny, aczkolwiek obserwuje się znaczne różnice w krzyżowaniach form w obu kierunkach, jak i między różnymi układami krzyżowania, bazującymi na różnych komponentach rodzicielskich. Nie bez znaczenia na pewno ma przebieg pogody. Przykładowo wykonywane programy krzyżowań w latach 2001, 2002 czy 2008 wskazują, że

w zależności od kierunku krzyżowania efektywność przekracza 70%. Uzyskane wyniki w bieżącym roku 2015, zważywszy na bardzo wysokie temperatury panujące w okresie kastrowania roślin i ich przepylania, pokazały gorsze zawiązywanie strąków, zwłaszcza, gdzie jednym z komponentów rodzicielskich była linia R004, przy której obserwowano znaczące opadanie wykastrowanych kwiatów i zawiązków małych strąków. W wyniku przeprowadzonego programu krzyżowań na podstawie efektywności zawiązanych strąków w stosunku do przepylonych kwiatów (od 38 – 68%) można stwierdzić, że nie wszystkie genotypy sokały się dobrymi formami matecznymi, a kierunek krzyżowania odegrał znaczącą rolę. Przeprowadzony program krzyżowań daje podstawę do wyprowadzenia pokolenia F₂ a następnie linii RILs, niezbędnych do genotypowania *L. angustifolius*.

Opracowany program do analiz bioinformatycznych prawidłowo konwertuje sekwencje DNA pomiędzy formatami zapisu danych powszechnie stosowanymi w bioinformatyce (csv, fasta). Opracowane narzędzie pozwala na przeprowadzenie analizy porównawczej BLAST dowolnej liczby sekwencji w zapisanych w formacie fasta lub csv z sekwencją lokalnej bazy danych. Pozwala także na selekcję wyników BLAST według zadanych kryteriów. Opracowany program pozwala również na wybranie z referencyjnego genomu (bazy danych) określonych przez użytkownika sekwencji lub ich fragmentów. Jedną z najbardziej użytecznych funkcji programu jest możliwość wstawienia do sekwencji DNA oznaczeń SNP w określonym przez użytkownika miejscu. Program umożliwia wykonanie wszystkich tych czynności (konwersja formatu, analiza BLAST, selekcja wyników BLAST, wybór sekwencji lub fragmentu sekwencji z bazy danych, wstawienie oznaczenia SNP w określonym przez użytkownika miejscu) dla dowolnej liczby sekwencji w jednej analizie. Analiza BLAST pozwoliła na wybranie z 1152 sekwencji markerów SNP 400 sekwencji z referencyjnego genomu łubinu wąskolistnego. Wybrane z genomu referencyjnego sekwencje miały długość od 1 000 do ponad 35 000 nukleotydów. Do zaprojektowania starterów do analiz KASP potrzeba minimum 101 nukleotydów, w związku z tym, wykorzystując opracowany program komputerowy dla każdej z 400 wybranych sekwencji wycięto fragment długości 269 nukleotydów. W każdym z tak wyselekcjonowanych fragmentów wybranych sekwencji miejsce SNP znajduje się pomiędzy 101 a 169 nukleotydem. Taki wybór pozwolił na zaprojektowanie starterów do analiz KASP, wykorzystanych w pierwszym etapie do genotypowania 15 form kolekcyjnych. Pozytywne wyniki genotypowania (wyraźny sygnał fluorescencji umożliwiający podział analizowanego materiału na dwa klastry: Homozygota A oraz Homozygota B) uzyskano dla 289 (72,3 %) z 400 analizowanych markerów. Do oceny informatywności markerów wykorzystano współczynnik PIC. Wartość współczynnika PIC dla wszystkich badanych markerów przedstawiono na histogramie. Maksymalną wartość współczynnika PIC - 0.374 uzyskano dla 33 markerów. Markery te dzieliły genotypowane linie na dwa klastry liczące 7 i 8 obiektów. Minimalnie współczynnik PIC dla badanych markerów wynosiła - 0.117 i została stwierdzona w przypadku 19 markerów. Markery te pozwalały na identyfikację dwóch klastrów w obrębie genotypowanych linii, liczących odpowiednio 14 i 1 linię. Z analizowanych markerów 167 miało wartość współczynnika PIC ≤ 0.3 . Markery te jako najbardziej informatywne zostaną wykorzystane do genotypowania kolekcji łubinu wąskolistnego. Wybrane markery w ten sposób reprezentują 19 z 20 grup sprzężeń wyznaczonych na genetycznej mapie sprzężeń opracowanej dla populacji ExL. Spośród 289 markerów dających pozytywny wynik genotypowania, 160 wykazywało polimorfizm pomiędzy genotypami R001 i R004, które wykorzystano do opracowania populacji mapującej. Markery te reprezentują wszystkie 20 grup sprzężeń opracowanych dla populacji mapującej ExL. Grupy 2, 3 i 11 reprezentowane są przez pojedyncze markery. Pozostałe grupy reprezentowane są przez 2 do 11 markerów. Wszystkie 160 markerów polimorficznych pomiędzy liniami R001 i R004 wykorzystana zostanie do genotypowania populacji F₂ w kolejnych etapach projektu.