

Sprawozdanie z realizacji zadania badawczego nr 11 za 2022 rok

Tytuł zadania **Identyfikacja mechanizmu molekularnego odporności żyta ozimego na rdzę brunatną.**

Kierownik zadania: **dr hab. inż. Kamila Nowosad , prof. Uczelni**

Celem prac realizowanych w 2022 roku było:

- 1) Ocena podatności genotypów żyta ozimego na porażenie przez rdzę brunatną w warunkach sztucznej inokulacji.
- 2) Wytworzenie populacji mapującej żyta ozimego.
- 3) Identyfikacja markerów użytecznych w genotypowaniu populacji mapującej.

W roku 2022 wykonano następujące prace:

Ad 1)

Materiał badawczy stanowiły przekazane przez hodowców materiały wyjściowe do hodowli nowych odmian żyta ozimego, linie wsobne z kolekcji oraz zarejestrowane odmiany, które zostały użyte jako wzorce.

1. Formy żyta ozimego przekazane przez:

1.1. DANKO Hodowla Roślin Sp. z o. o.

S 43 IV, S 46 IV, S 50 IV, S 53 IV, S 55 IV, S 57 IV, S 60 IV, S 65 IV, S 68 IV, S 71 IV, S72 IV, S 73 IV, S 75 IV, S 86 IV, S 88 IV, S 89 IV, S 93 IV, S 97 IV, S 99 IV, S 129 IV, SO 17 R/20, SO 22 R/20, SO 28 R/20, SO 56 R/20, SO 62 /20, SO 118 R/20, SO 119 R/20, SO 133 R/20, SO 179 R/20, SO 180 R/20, SO 200 R/20, SO 204 R/20, SO 229 R/20, SO 231 R/20, SO 235 R/20, SO 251 R/20, SO 260 R/20, SO 280 R/20, SO 285 R/20, SO 288 R/20, LS 532 N22, LS 533 N22, LS 534 N22, LS 540 N22, LS 545 N22, LS 546 N22, LS 548 N22, LS 549 N22, LS 550 N22, LS 551 N22, LS 554 N22, LS 557 N22, LS 559 N22, LS 568 N22, LS 569 N22, LS 571 N22, LS 573 N22, LS 574 N22, LS 575 N22 ,LS 576 N22

1.2. Hodowla Roślin Smolice Sp. z o. o. Grupa IHAR

HRSM 599 - 4R, HRSM 601 - 4R, HRSM 602 - 4R, HRSM 604 - 4R, HRSM 606 - 4R, HRSM 609 - 4R, HRSM 611 - 4R, HRSM 613 - 4R, HRSM 614 - 4R, HRSM 615 - 4R, HRSM 619 - 4R, HRSM 620 - 4R, HRSM 621 - 4R, HRSM 622 - 4R, HRSM 623 - 4R, HRSM 624 - 4R, HRSM 625 - 4R, HRSM 627 - 4R, HRSM 633 - 4R, HRSM 652 - 3R, HRSM 656 - 3R, HRSM 659 - 3R, HRSM 662 - 3R, HRSM 663 - 3R, HRSM 681 - 2R, HRSM 682 - 2R, HRSM 685 - 2R, HRSM 688 - 2R, HRSM 689 - 2R, HRSM 690 - 2R, HRSM 691 - 2R, HRSM 692 - 2R, HRSM 696 - 2R, HRSM 697 - 2R, HRSM 699 - 2R, HRSM 705 - 2R, HRSM 710 - R, HRSM 712 - R, HRSM 716 - R, HRSM 722 - R, HRSM 727 - R, HRSM 737 - R, HRSM 739 - R, HRSM 741 - R, HRSM 746 - R, HRSM 757 - R, HRSM 758 - R, HRSM 761 - R, HRSM 762 - R, HRSM 764 - R, HRSM 767 - R, HRSM 769 - R, HRSM 770 - R, HRSM 771 - R, HRSM 772 - R, HRSM 773 - R, HRSM 774 - R, HRSM 778 - R, HRSM 782 - R, HRSM 783 - R

1.3 Linie Upwr

UPWR 101-UPWR 150

W roku 2022 testowaniu w kierunku poszukiwania odporności na rdzę brunatną poddano łącznie 170 genotypów żyta ozimego w doświadczeniach infekcyjnych. Były to dostarczone przez hodowców linie wsobne, populacje oraz odmiany wzorcowe.

Ocenę podatności genotypów żyta na porażenie przez rdzę brunatną (*Puccinia recondita*) wykonano metodą laboratoryjną. Na szlakach Petriego przygotowano pożywkę składającą się z agaru z dodatkiem benzomidazolu. Liście dziesięciodniowych siewek układano na pożywce, a następnie infekowano przygotowanym wcześniej inokulatem rdzy brunatnej. Materiał do przygotowania zawiesiny inokulacyjnej zebrano w różnych miejscowościach (m.in: Wrocław, Smolice, Kondratowice, Radzików), aby otrzymać szeroką genetycznie populację zarodników patogena. Oceny porażenia wykonano po dziesięciu dniach od inokulacji, w skali czterostopniowej, gdzie 1 – brak porażenia, a 4 - silne objawy porażenia rdzą brunatną na liściu.

Wyniki ocen zostały poddane analizom statystycznym. W celu prawidłowego statystycznego opracowania otrzymanych danych eksperymentalnych przeprowadzono ich transformację, w celu spełnienia warunków koniecznych do analiz, czyli uzyskania ciągłości i normalności rozkładu empirycznego. Wyniki surowe przekształcono wykorzystując wzór podany przez Węgrzyna i in. (1996), a następnie przeprowadzono jednoczynnikową analizę wariancji zgodnie z modelem kompletnej randomizacji (ANOVA), co umożliwiło sprawdzenie istotności zróżnicowania badanych genotypów żyta ozimego pod względem odporności na rdzę brunatną. Do grupowania genotypów o tej samej wartości średnich, czyli do ich podziału na grupy jednorodne wykorzystano wielokrotny test Duncana.

W doświadczeniach infekcyjnych wykonano oceny podatności posiadanych form żyta na porażenie przez *Puccinia recondita*. Oceną objęto łącznie 170 genotypów żyta.

Wyniki ocen dla badanych genotypów żyta zostały poddane analizie statystycznej. Do analizy wariancji oraz podziału genotypów na grupy jednorodne wykorzystano dane po transformacji, które po pogrupowaniu zostały z powrotem przeliczone na skalę oceny.

Analiza wariancji wykazała istotne zróżnicowanie badanych genotypów żyta pod względem ich odporności na rdzę brunatną, to znaczy że stopień porażenia liści przez rdzę brunatną w warunkach sztucznej inokulacji był zróżnicowany w zależności od genotypu. Wśród 170 badanych genotypów żyta, znaleziono 18 obiektów które charakteryzowały się pożądaną, wysoką odpornością na tego patogena. W zestawieniu wiele z badanych genotypów było podatnych na porażenie i zostały one ocenione na 3 lub 4 w stosowanej czterostopniowej skali oceny porażenia. Wystąpiło jednak 18 obiektów o znikomym porażeniu lub bez śladów zarodników na liściach (SO 204 R/20, UPWR121, UPWR150, HRSM 623-4R, HRSM 624-4R, HRSM 652 - 3R, SO 229 R/20, S 53 IV, SO 231 R/20, SO 280 R/20, SO 285 R/20, LS 545 N2, LS 549 N22, HRSM 773 – R, S72 IV, S 89 IV, S 93 IV, S 97 IV), które można uznać za odporne na rdzę brunatną. Genotypy te utworzyły pierwszą grupę jednorodną. Kolejne grupy jednorodne tworzyły genotypy, które w teście laboratoryjnym uzyskały oceny świadczące o większej wrażliwości na porażenie przez rdzę brunatną. Z informacji literaturowych wynika, że istnieje wysoka korelacja pomiędzy odpornością określoną metodą laboratoryjną, a połową odpornością na rdzę brunatną. Wymienione genotypy można będzie wykorzystać w programach hodowli nowych, tolerancyjnych na rdzę brunatną odmian żyta.

Uzyskane w bieżącym roku wyniki ocen porażenia genotypów żyta przez rdzę brunatną nie odbiegają znacząco do rezultatów z innych badań [Sawinska 2011; Pietrusińska i in 2018]. Wśród analizowanych genotypów żyta ozimego stwierdzono nieliczne formy charakteryzujące się pełną odpornością rdzę brunatną. Zakres zmienności porażenia badanych linii przez patogena wynosił od 1,0 do 4,0. Analiza statystyczna otrzymanych wyników wykazała istotne zróżnicowanie podatności na porażenie rdzę brunatną badanych genotypów. Wśród badanych genotypów wystąpiły formy odporne na *Puccinia recondita*. Podobne wyniki uzyskano w badaniach linii wsobnych wyprowadzonych w Katedrze Genetyki, Hodowli Roślin i Nasiennictwa Uniwersytetu Przyrodniczego we Wrocławiu. Linie wsobne żyta ozimego wchodzące w skład całej kolekcji charakteryzowały się również istotnym zróżnicowaniem podatności na rdzę brunatną [Weber i in. 2015;Bujak i in. 2019].

Ad 2)

Materiał badawczy stanowiły linie wsobne (homozygotyczne) o potwierdzonej odporności na rdzę brunatną, jak i linie silnie podatne na porażenie. Linie podatne o sprawdzonej odporności zostały wyprowadzone z różnych populacji wyjściowych i w wyniku wieloletniego chowu wsobnego stanowią homozygoty. Linie odporne zostały przetestowane w warunkach polowych w latach poprzedzających wykonanie doświadczenia. Spośród linii wytypowanych przez hodowców jako odporne, pełną odporność w warunkach polowych udało się potwierdzić tylko dla jednej formy RB1/ TN1874 pochodząca z kolekcji Instytutu Hodowli i Aklimatyzacji Roślin w Radzikowie. Linia ta została wybrana jako komponent do krzyżowania i wytworzenia pokolenia F1

Po przezimowaniu uzyskanego pokolenia F1 wybrane zostały po 25 osobników, które zostały zaizolowane w celu przeprowadzenia kontrolowanego samozapylenia.

Efekty krzyżowania były zadowalające pod względem otrzymanej liczby ziarniaków mieszańcowych F2 dla wszystkich osobników F1

Ziarniaki mieszańców zostały wysiane w celu wyprowadzenia populacji mapującej. Nasiona wysiewano punktowo w rozstawie 20 x 5 na poletkach czterorzędowych. Na każdym z poletek wysiano po 40 ziarniaków. Kombinacje mieszańcowe po przezimowaniu, zostaną ocenione i poddane dalszym analizom.

Zastosowanie różnych kombinacji linii stanowiło dobre rozwiązanie aby uniknąć efektu samoniezdgodności u roślin żyta ozimego. Wykonanie krzyżowań linii wsobnych w układzie A x B i BxA, z 10 liniami pozwoliło na uzyskanie odpowiedniej liczby ziarniaków do wysiewu i wytworzenia populacji mapującej. Pokolenie F1 po przezimowaniu poddane zostało samozapyleniu na drodze którego wytworzona została zadowalająca liczba ziarniaków pokolenia F2

Ad 3)

Materiał do analiz WGS stanowiło całkowite DNA wyizolowane z młodych liści genotypów żyta. Sekwencjonowanie genomu wykonano w technologii Illumina 150 PE. Analiza bioinformatyczna wykonana została w oparciu o standardowy protokół analityczny z wykorzystaniem narzędzi: bowtie2, samtools, Picard oraz haplotypcaller

Izolacja DNA, sekwencjonowanie

Markery SNP w roku 2022 opracowane zostały na podstawie danych sekwencjonowania całego genomu (WGS) techniką sekwencjonowania przez syntezę (SDS). Sekwencjonowanie dwóch form rodzicielskich, wytypowanych na podstawie wyników z 2021 roku, przeprowadzone zostało usługowo na urządzeniu Illumina NovaSeq 6000. Wyizolowane genomowe DNA z analizowanych genotypów wykorzystane zostało do przygotowania biblioteki sekwencyjnych. Biblioteki utworzone zostały według protokołu TruSeq DNA PCR Free (350). W wyniku sekwencjonowania wygenerowane zostały odczyty długości 150 o sparowanych końcach (PE). Sekwencjonowanie wykonane zostało z pokryciem minimum 20x. Otrzymane odczyty w postaci plików fastq, po wstępnej filtracji wykorzystane zostały w dalszych analizach bioinformatycznych do predykcji markerów pojedynczego polimorfizmu (SNP).

Identyfikacja markerów SNP

Predykcja markerów SNP wykonana została przy użyciu potoku analitycznego GATK (Genome Analysis Toolkit) [McKenna i in. 2010]. Analizy przeprowadzono zgodnie z zaleceniami najlepszej praktyki dostarczonymi na stronie projektu GATK (<https://gatkforums.broadinstitute.org/gatk/discussion/3892/the-gatk-best-practices-for-variant-calling-on-rnaseq-in-full-detail>). Oczyszczone dane zostały przyrównane do dwóch dostępnych obecnie wersji genomu referencyjnego żyta [Li i in 2021; Rabanus-Wallace i in 2021]. Do mapowania odczytów użyty został program bowtie2 [Langmead i Salzberg 2012], zgodnie z zaleceniami programu GATK. W kolejnym etapie uzyskane w mapowaniu pliki sam przekształcone zostały w pliki binarne

BAM, posortowane, zindeksowane i zaznaczono zduplikowane odczyty. Modyfikacje te wykonano za pomocą programów samtools [Li i in. 2009] oraz Picard (<http://broadinstitute.github.io/picard/>). Na tak przygotowanych danych uruchomiona została funkcja „HaplotypeCaller” z pakietu GATK (zgodnie z zaleceniami najlepszej praktyki). Program został uruchomiony dla każdego genotypu osobno. Otrzymane pliki VCF zostały następnie ze sobą połączone za pomocą programu CombineGVCFs z pakietu GATK. W ten sposób wyłoniłone zostały warianty polimorficzne pomiędzy analizowanymi formami rodzicielskimi. Po wstępnej identyfikacji markerów SNP wykonana została ich filtracja zgodnie z sugestiami podanymi w zaleceniach programu GATK. Do filtracji wykorzystane zostaną następujące parametry: filter clusters minimum 3 SNPs w oknie 35 zasad między nimi, przy użyciu wartości Fisher Strand ($FS > 30.0$) i wartości Qual By Depth ($QD < 2.0$).

Przeprowadzone zostały ponownie analizy bioinformatyczne danych uzyskanych w sekwencjonowaniu transkryptomu (dane z roku 2021), ale tym razem analizy polimorfizmów zostały przeprowadzone dla genomu referencyjnego. Zastosowano potok analityczny, w oparciu o GATK i program haplotypeCaller. Analizy wykonano analogicznie jak dla danych WGS.

Seqwencjonowanie genomowego DNA pozwoliło na uzyskanie ponad 170 mln odczytów oraz ponad 20 miliardów odczytanych zasad (PI - 70,537,620 odczytów 10,651,180,620 odczytanych zasad; PII - 70,537,620 odczytów 10,618,230,306 odczytanych zasad). Jakość uzyskanych wyników była bardzo wysoka (ponad 90% odczytów Q30, ponad 95% odczytów Q20. Zawartość par CG w uzyskanych odczytach wynosiła 36,62% u genotypu PI oraz 36,95% u genotypu PII.

W wyniku sekwencjonowania transkryptomu z 6 bibliotek uzyskano łącznie ponad 26 mln odczytów oraz ponad 246 mln odczytanych par zasad. Jakość odczytów była bardzo wysoka parametr Q30 wyniósł ponad 93% a Q20 ponad 97% we wszystkich 6 analizowanych bibliotekach. Średnia zawartość par CG w uzyskanych odczytach wynosiła 43,2,% u genotypu PI oraz 44,6% u genotypu PII.

Wyniki mapowania odczytów zostały zapisane w plikach SAM, które następnie zostały przekształcone w pliki BAM. Średnio ok 82% odczytów mapowało się unikalnie do sekwencji genomu referencyjnego. Uzyskane pliki BAM zostały następnie poddane dalszym modyfikacjom (sortowanie, indeksowanie, usunięcie duplikatów PCR, splitNeigarstring). Modyfikacje te zostały wykonane za pomocą programów samtools oraz Picard. Zmodyfikowane w ten sposób pliki BAM zostały wykorzystane do identyfikacji polimorfizmów SNP za pomocą programów GATK (haplotypecaller). W wyniku działania tych programów uzyskano pliki VCF zawierające informacje o lokalizacji wykrytych polimorfizmów SNP. W kolejnym kroku przeprowadzono filtrację uzyskanych polimorfizmów tak aby usunąć warianty o niskiej wiarygodności lub monomorficzne pomiędzy analizowanymi genotypami (PI oraz PII). Po filtracji łącznie (wyniki DNA oraz RNA, dwa potoki analityczne) uzyskano ponad 802 tyś. unikalnych polimorfizmów SNP (745 760 dane WGS oraz 56 989 dane RNAseq).

Na podstawie rozmieszczenia w genomie z uzyskanego zestawu wybrano (częściowo losowo) zestaw 50 000 loci SNP dla których wykonano projektowanie starterów do analiz targetGBS. Startery udało się zaprojektować dla 42 826 locii. Z tej puli wybrany został zestaw 2500 potencjalnych markerów SNP. Zestaw ten został wybrany tak aby reprezentować każdą z możliwych kombinacji: typ danych (DNA/RNA).

Na podstawie wyników z roku 2021 wybrany został zestaw 126 sekwencji mikrosatelitarnych. Markery wybrane zostały na podstawie lokalizacji w genomie referencyjnym. Wybrany zestaw przekraczał założone w projekcie 100 markerów do walidacji, ponieważ udało się zaprojektować większą liczbę markerów z uzyskanych danych bez zwiększania kosztów uzyskania. Chcieliśmy przez to zwiększyć liczbę pozytywnie zwalidowanych markerów, które na kolejnych etapach będą mogły być wykorzystane do mapowania. Wybrane markery są zlokalizowane w wszystkich 7 chromosomach genomu żyta. Dla otrzymanego zestawu zaprojektowano startery do reakcji PCR. Do projektowania starterów wykorzystany został algorytm primer3plus (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) z domyślnymi ustawieniami. W kolejnym kroku zaprojektowane

markery zostały zwalidowane w reakcji PCR. Jako matryce wykorzystane zostało genomowe DNA form rodzicielskich populacji mapujących. PCR przeprowadzony został w standardowej reakcji z wykorzystaniem polimerazy Taq. Produkty amplifikacji rozdzielono z użyciem zestawu do elektroforezy kapilarnej na urządzeniu QIAxcel Advance. Produkty oceniono pod kątem specyficzności produktu, jakości sygnału (ilości amplikonu), a także różnic w długości produktu pomiędzy formami rodzicielskimi. Na tej podstawie wybrany został zestaw co 48 markerów, który w kolejnych etapach prowadzonych prac będzie wykorzystany do genotypowania analizowanej populacji mapującej.

Przyjęty w pracy potok analityczny oparty o mapowanie odczytów NGS do genomu referencyjnego w oparciu o program bowtie2, a następnie identyfikacja markerów SNP za pomocą programu haplotypcaller z pakietu GATK okazała się być skuteczna w identyfikacji markerów SNP pomiędzy formami rodzicielskimi populacji mapującej. W literaturze dostępne są liczne przykłady wykorzystania analogicznego potoku analityczny do idnetyfikacji markerów SNP, a następnie wykorzystania ich z sukcesem do genotypowania populacji mapującej i konstrukcji mapy genetycznej [Ray i Satya 2014; Duitama i in. 2015; Kalemkeridou i in. 2022]

Dla opracowania użytecznej mapy genetycznej konieczne jest genotypowanie populacji mapującej za pomocą zestawu wysoce polimorficznych markerów DNA równomiernie rozmieszczonych w genomie referencyjnym [Rafalski 2002]. Do konstrukcji mapy genetycznej dla genomu wielkości porównywalnej z żytem konieczne jest użycie conajmniej kilku tysięcy unikalnych markerów rozmieszczonych równomiernie w genomie [Li i in. 2021; Rabanus-Wallace 2021]. Opracowany w ramach realizacji zadania badawczego 3 „Identyfikacja markerów użytecznych w genotypowaniu populacji mapującej” Zestaw markerów SNP liczy ponad 800 tysięcy markerów SNP. Markery te zostały opracowane na podstawie sekwencjonowania form rodzicielskich populacji mapującej. Dzięki temu gwarantowany jest (w przypadku większości markerów) polimorfizm w populacji mapującej. Na podstawie wyników uzyskanych w innych pracach realizowanych w KGHRiN, a także danych literaturowych [Gardner i in. 2016; Kim i in. 2016; Chen i in. 2019] można oczekiwać iż co najmniej 80% opracowanych markerów wykaże polimorfizm w populacji mapującej, a rozkład markerów będzie zgodny z rozkładem teoretycznym 1:2:1 ($pval < 0.0001$). Z 800 tys opracowanych markerów możliwy będzie wybór ostatecznego zestawu markerów SNP, który w kolejnych etapach prowadzonych prac zostanie wykorzystany do genotypowania populacji mapującej. Ponieważ markery w opracowanym zestawie pokrywają cały genom żyta możliwe będzie tak dobranie ostatecznego zestawu markerów SNP, aby możliwe było stworzenie gęstej mapy genetycznej (odpowiednie zagęszczenie markerów) reprezentującej cały genom żyta. Dzięki temu identyfikacja obszarów genomu, a w szczególności sekwencji genów związanych z odpornością na rdzę burunatną powinna być możliwa w dalszych etapach prowadzonych badań [Young 1996].

Otrzymany zestaw markerów SSR będzie możliwy do wykorzystania w czasie genotypowania populacji mapującej w dalszych etapach prowadzonych prac.

Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-genome sequencing strategy to increase yield potential of rice, *Oryza sativa* L. *Rice*, 9(1), 1-17.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357-359.

Li, G., Wang, L., Yang, J., He, H., Jin, H., Li, X., ... & Wang, D. (2021). A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nature genetics*, 53(4), 574-584.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297-1303.

Rabanus-Wallace, M. T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., ... & Stein, N. (2021). Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nature genetics*, 53(4), 564-573.

Rafalski, J. A. (2002). Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant science*, 162(3), 329-333.

Ray, S., & Satya, P. (2014). Next generation sequencing technologies for next generation plant breeding. *Frontiers in plant science*, 5, 367.

Young, N. D. (1996). QTL mapping and quantitative disease resistance in plants. *Annual review of phytopathology*, 34(1), 479-501.