

Szkoła Główna Gospodarstwa Wiejskiego
w Warszawie
Instytut Biologii

Ewa Borzęcka

Wykorzystanie biblioteki BAC do integracji
map fizycznych i genetycznych oraz
identyfikacji genów związanych z
udomowieniem i odpornością na choroby
żyta

BAC library application in integration of physical and genetic
maps and identification of genes related to domestication and
disease resistance of rye

Praca doktorska
Doctoral thesis

Praca wykonana pod kierunkiem:
dr hab. Hanny Bolibok-Brągoszewskiej,
profesor uczelni
Katedra Genetyki, Hodowli i Biotechnologii Roślin

Recenzenci:
prof. dr hab. Dariusz Grzebelus
Katedra Biologii Roślin i Biotechnologii
Wydział Biotechnologii i Ogrodnictwa
Uniwersytet Rolniczy im. Hugona Kołłątaja w Krakowie

prof. dr hab. Henryk Bujak
Katedra Genetyki, Hodowli Roślin i Nasiennictwa
Wydział Przyrodniczo-Technologiczny
Uniwersytet Przyrodniczy we Wrocławiu


prof. dr hab. Edyta Paczos-Grzęda, profesor uczelni
Instytut Genetyki, Hodowli i Biotechnologii Roślin
Uniwersytet Przyrodniczy w Lublinie

Warszawa, 2022

Oświadczenie promotora pracy

Oświadczam, że niniejsza praca została przygotowana pod moim kierunkiem i stwierdzam, że spełnia ona warunki do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data 5-07-2022 r.

Podpis promotora pracy 

Oświadczenie autora pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przez mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data 5-07-2022 r.

Podpis autora pracy 

*Składam podziękowania dla **Pani Profesor Hanny Bolibok-Brągoszewskiej** za nieocenioną pomoc i wsparcie przy realizacji badań oraz pisaniu niniejszej pracy doktorskiej.*

Pragnę podziękować również wszystkim pracownikom, koleżankom i kolegom z Katedry Genetyki, Biotechnologii i Hodowli Roślin za owocną współpracę, wsparcie naukowe i miłą atmosferę.

Specjalne podziękowania dla moich rodziców, w szczególności mamy której pomoc na „ostatniej prostej” była bezcenna.

Pracę dedykuję córkom i mężowi, który zawsze wierzył w ukończenie tego projektu

Badania zostały wykonane w ramach projektu badawczego pt. "Wpływ selekcji na genom rośliny uprawnej - identyfikacja i charakterystyka sekwencji, na które nakierowana była presja selekcyjna w trakcie udomowienia i hodowli żyta (*Secale cereale* L.)." finansowanego przez Narodowe Centrum Nauki, nr. UMO 2014/14/E/NZ9/00285.

Część badań dotyczących próby identyfikacji genów związanych z odpornością na choroby u żyta współfinansowano w ramach grantu Dziekana WOBiAK na prowadzenie badań naukowych służących rozwojowi młodych naukowców oraz uczestników studiów doktoranckich SGGW. Tytuł projektu: "Wstępna charakterystyka strukturalna nowych genów potencjalnie związanych z odpornością na choroby u żyta" (numer ewidencji księgowej: 505-10-041000-P00286-99).

Streszczenie

Wykorzystanie biblioteki BAC do integracji map fizycznych i genetycznych oraz identyfikacji genów związanych z udomowieniem i odpornością na choroby żyta

Żyto (*Secale cereale* L.) jest ważną rośliną uprawną w Europie Środkowej i Wschodniej, o dużej tolerancji na różne stresy abiotyczne i biotyczne. Celem pracy była integracja map fizycznych i genetycznych oraz identyfikacja genów związanych z udomowieniem i odpornością na choroby przy wykorzystaniu biblioteki BAC linii wsobnej żyta L318. Biblioteki BAC są ważnym narzędziem w genomice strukturalnej, a przypisanie klonów BAC do określonych pozycji mapy genetycznej jest niezbędne do integracji map genetycznych i fizycznych. Wykorzystując technologię DArTseq zgenotypowano 16 pul klonów BAC reprezentujących 41% całej biblioteki BAC i 65% genomu żyta (~ 5,2 Gbp). Uzyskano 221 122 adresów marker DArTseq-klon BAC. Wykazano, że technologia DArTSeq w połączeniu z trójwymiarową strategią pulowania jest efektywnym sposobem na zakotwiczenie klonów BAC na mapie genetycznej. Proponowane podejście można zastosować także w przypadku innych gatunków, zwłaszcza tych o złożonych genomach. Zastosowanie wyników adnotacji funkcjonalnej sekwencji markerów DArT żyta i wdrożenie technologii Oxford Nanopore do sekwencjonowania pojedynczych klonów BAC umożliwiło identyfikację 28 genów przypuszczalnie związanych z odpornością na choroby. Większość z nich (82%) należała do największej grupy genów odporności - NB-LRR. Dla jednego ze zidentyfikowanych genów (*Sck33*) przy pomocy techniki qRT-PCR wykazano prawdopodobną rolę w reakcji na zakażenie patogenem *Puccinia recondita f. sp. secalis*, powodującym rdzę brunatną. W sześciu klonach BAC odnaleziono sekwencje homologiczne do pięciu genów związanych z udomowieniem: *Q*, *GIF1*, *BTR2*, *VRN1* i *VRN2*. Dodatkowo w pełnych sekwencjach klonów BAC zidentyfikowano 128 przypuszczalnych genów o różnych funkcjach biologicznych. Uzyskane wyniki stanowią punkt wyjścia do dalszych badań nad funkcją zidentyfikowanych genów i ich różnicowaniem allelicznym, co przyczyniłoby się do lepszego poznania genetycznej kontroli istotnych cech żyta.

Słowa kluczowe: *Secale cereale*, biblioteka BAC, kotwiczenie klonów BAC, DArTseq, geny odporności, udomowienie

Summary

BAC library application in integration of physical and genetic maps and identification of genes related to domestication and disease resistance of rye

Rye (*Secale cereale* L.) is an important crop in Central and Eastern Europe, with great tolerance against different biotic and abiotic stresses. The aim of this research was to integrate physical and genetic maps and to identify genes related to domestication and disease resistance using a BAC library of rye inbred line L318. BAC libraries are an important tool in structural genomics and the assignment of BAC clones to positions on a genetic map is indispensable for integration of genetic and physical maps. Using DArTSeq technology we genotyped 16 BAC clone pools comprising over 43 thousand clones that represented 41% of the whole BAC library and 65% of the rye genome (~ 5.2 GpZ). We obtained 221 122 x DArTseq to BAC addresses. These results indicate that DArTseq technology combined with the three-dimensional pooling strategy provides an efficient method of BAC clone anchoring. Proposed approach can be used in other species as well, especially those with complex genomes. The use of functional annotation results of the rye DArT markers sequences and the implementation of Oxford Nanopore approach to sequence individual BAC clones allowed identification of 28 genes with functions likely related to pathogen response. Most of them (82%) belonged to the biggest group of resistance genes called NB-LRR. A possible role of one of these genes - *Sck33* - in response to infection by the pathogen causing brown rust (*Puccinia recondita* f. sp. *secalis*) was demonstrated using the qRT-PCR method. Six BAC clones containing homologs of five domestication related genes: *Q*, *GIF1*, *BTR2*, *VRN1* and *VRN2* were identified. In addition, we identified 128 possible genes with different biological functions in full sequences of six BAC clones. The obtained results constitute a starting point for further research on the function of the identified genes and their allelic variation, which would contribute to a better understanding of the genetic control of important features of rye.

Keywords: *Secale cereale*, BAC library, BAC clones anchoring, DArTseq, disease resistance genes, domestication

SPIS TREŚCI

1.	WYKAZ SKRÓTÓW	11
2.	WSTĘP, CEL PRACY I HIPOTEZY BADAWCZE	13
3.	PRZEGLĄD LITERATURY	15
3.1	Charakterystyka żyta	15
3.1.1	Systematyka, właściwości i zastosowanie	15
3.1.2	Genomika żyta	15
3.1.3	Geny warunkujące ważne użytkowo cechy żyta.....	16
3.2	Biblioteki BAC i ich zastosowania.....	18
3.2.1	Zastosowanie bibliotek BAC w identyfikowaniu genów	18
3.2.2	Zastosowanie bibliotek BAC w cytogenetyce	20
3.2.3	Zastosowanie kotwiczenie klonów BAC w integrowaniu map fizycznych i genetycznych i sekwencjonowaniu genomów	21
3.3	Sekwencjonowanie nowej generacji w genomice	24
3.3.1	Metody sekwencjonowania nowej generacji	24
3.3.2	Zastosowania technik sekwencjonowania nowej generacji	27
3.4	Udomowienie roślin i jego wpływ na genom.....	32
3.4.1	Geny związane z udomowieniem roślin	35
3.4.2	Geny związane z odpornością na choroby	39
4.	MATERIAŁY I METODY	43
4.1	Biblioteka BAC	43
4.2	Konstrukcja pul klonów BAC	43
4.3	Izolacja BAC DNA z pul zbiorczych	44
4.4	Izolacja DNA pojedynczych klonów BAC	44
4.5	Genotypowanie pul klonów BAC metodą Genotyping by Sequencing (DArTseq).....	45
4.6	Rozkodowanie wyników genotypowania klonów BAC - ustalenie adresów marker DArTSeq-klon BAC.....	45
4.7	Kotwiczenie klonów BAC do map genetycznych.....	46
4.8	Walidacja adresów marker DArTseq-klon BAC.....	46
4.9	Sekwencjonowanie indywidualnych klonów BAC	47
4.10	Elektroforeza pulsacyjna (PFGE).....	48
4.11	Identyfikacja potencjalnych genów związanych z odpornością na choroby....	49
4.12	Identyfikacja przypuszczalnych żytnich homologów genów związanych z udomowieniem	50
4.13	Analiza bioinformatyczna zidentyfikowanych potencjalnych genów.....	51

4.14 Ilościowy RT-PCR	51
5. WYNIKI	53
5.1 Genotypowanie DNA pul klonów BAC metodą DArTseq i ustalenie adresów marker DArTseq-klon BAC	53
5.2 Kotwiczenie klonów BAC do map genetycznych.....	53
5.3 Weryfikacja adresów marker DArTseq-klon BAC	56
5.3.1 Weryfikacja sekwencyjna	56
5.3.2 Weryfikacja na podstawie pozycji na mapie.....	57
5.4 Wdrożenie sekwencjonowania nowej generacji Oxford Nanopore	58
5.5 Identyfikacja potencjalnych genów związanych z odpornością na choroby u żyta	62
5.6 Identyfikacja genów kandydujących przypuszczalnie związanych z udomowieniem u żyta.....	69
6. DYSKUSJA	78
6.1 Genotypowanie DArTseq oraz kotwiczenie klonów BAC na mapie genetycznej żyta	78
6.2 Sekwencjonowanie klonów BAC.....	84
6.3 Identyfikacja z wykorzystaniem biblioteki BAC sekwencji genów przypuszczalnie związanych z odpornością na choroby oraz z udomowieniem.....	88
7. WNIOSKI	97
8. ANEKS	98
9. LITERATURA	109

1. WYKAZ SKRÓTÓW

BAC- ang.: *bacterial artificial chromosome*; sztuczny chromosom bakteryjny

DArT- ang.: *Diversity Arrays Technology*; technologia genotypowania wykorzystująca mikromacierze

DArTSeq- jedna z metod genotypowania poprzez sekwencjonowanie

dNTP- ang.: *deoxyNucleotide TriPhosphate*; mieszanina czterech dezoksynukleotydów

EDTA- ang.: *EthyleneDiamineTetraacetic Acid*; kwas etylenodiaminotetraoctowy

FPC- ang.: *fingerprinted contigs*; program służący do składania kontigów na podstawie wyników trawienia restrykcyjnego klonów

GBS- ang.: *genotyping by sequencing*; genotypowanie przez sekwencjonowanie

gDNA- DNA genomowe

HC- ang.: *high confidence genes*; geny o wysokim stopniu ufności, dla których znane są ich produkty białkowe

HICF- ang.: *high information content fingerprinting technology*; technika charakteryzowania klonów poprzez trawienie restrykcyjne wykorzystująca znakowanie fluorescencyjnym

kb- ang.: *kilo base pair*; kilo par zasad (1000 par zasad)

LB- pożywka mikrobiologiczna Lurii-Bertaniego

MTP- ang.: *minimal tiling path* – najmniejszy możliwy zbiór klonów BAC tworzących mapę fizyczną docelowego genomu i nakładających się na siebie

FMTP- ang.: *Fingerprint-based minimal tiling path*; oprogramowanie do obliczania ścieżki MTP mapy fizycznej na podstawie wzorów trawienia restrykcyjnego klonów BAC

N50- długość najkrótszego kontigu w zbiorze kontigów tworzących 50% długości całego złożenia

NGS- ang.: *next-generation sequencing*; sekwencjonowanie nowej generacji

PCR- ang.: *Polymerase Chain Reaction*; reakcja łańcuchowa polimerazy

PFGE- ang.: *Pulsed Field Gel Electrophoresis*; elektroforeza żelowa pulsacyjna w zmiennym polu elektrycznym

pz- para zasad

RFLP- ang.: *restriction fragment length polymorphism*; polimorfizm długości fragmentów restrykcyjnych

RIL- rekombinacyjna linia wsobna

rpm- ang.: *revolutions per minute*; obroty na minutę

SDS- siarczan dodecyłu sodu

SLS- N-lauroilosarkozyna sodu

SP- superpula

2. WSTĘP, CEL PRACY I HIPOTEZY BADAWCZE

Żyto (*Secale cereale* L.) jest jedną z najważniejszych roślin zbożowych uprawianych w Polsce ze względu na jej dobre przystosowanie do panujących tu warunków glebowo - klimatycznych, a powierzchnia jego upraw wynosi ok. 903 800 hektarów (<https://fao.org>). Posiada unikalne cechy, które czynią je atrakcyjnym zarówno dla producentów jak i dla konsumentów. Odznacza się wysoką odpornością na stresy biotyczne i abiotyczne oraz zdolnością do bardzo efektywnego wykorzystania składników pokarmowych z podłoża, przez co nie wymaga intensywnego nawożenia i ochrony chemicznej. Udowodniono również jego prozdrowotne właściwości.

Żyto ma jeden z największych genomów wśród roślin zbożowych (ok. 8 Gpz), wysoce zmetylowany i z dużym udziałem sekwencji powtarzalnych - ok. 92%. Zasoby genomiczne żyta były przez długi okres czasu bardzo ograniczone (w porównaniu do zbóż takich jak jęczmień i pszenica), co bardzo utrudniało prowadzenie badań z zakresu genetyki i genomiki tego gatunku. Intensywny rozwój metod biologii molekularnej, w tym wysokoprzepustowych technik genotypowania i narzędzi bioinformatycznych przyspieszył prace nad poznaniem genomu żyta i zaowocował opublikowaniem jego sekwencji (Rabanus-Wallace i in., 2021; Li i in., 2021). Zmieniło to sposób prowadzenia badań nad żytem, w tym również realizacji niniejszej pracy doktorskiej.

W przedstawionej pracy, w celu integracji mapy fizycznej i genetycznej oraz zidentyfikowania i scharakteryzowania genów związanych z udomowieniem i odpornością na choroby u żyta, zastosowano kompleksowe i ukierunkowane podejście. Obejmowało ono nowoczesne metody badawcze oraz efektywne wykorzystanie istniejących zasobów genomicznych (takich jak biblioteka BAC). Pozwoliło to na przypisanie klonów BAC do określonych pozycji na mapie genetycznej oraz zidentyfikowanie i charakterystykę sekwencji poszukiwanych genów związanych z odpornością na choroby i udomowieniem.

Cel ogólny:

Integracja mapy fizycznej i genetycznej żyta oraz uzyskanie pełnej sekwencji wybranych genów związanych z odpornością na choroby i udomowieniem żyta przy wykorzystaniu biblioteki BAC.

Cele szczegółowe:

1. Identyfikacja na podstawie wyników geotypowania DArTseq, klonów BAC zawierających poszczególne markery DArTseq – ustalenie adresów klon BAC-marker DArTseq.
2. Weryfikacja wiarygodności uzyskanych adresów klon BAC-marker DArTseq.
3. Integracja mapy fizycznej i genetycznej przy wykorzystaniu adresów klon BAC-marker DArTseq i mapy genetycznej markerów DArTseq.
4. Opracowanie metodyki sekwencjonowania indywidualnych klonów BAC z wykorzystaniem technologii Nanopore.
5. Zidentyfikowanie klonów BAC zawierających geny związane z odpornością na choroby oraz geny przypuszczalnie związane z udomowieniem oraz ich charakterystyka.

Hipotezy badawcze:

1. Genotypowanie DArTseq w połączeniu z zastosowaniem trójwymiarowej strategii tworzenia pul zbiorczych klonów BAC jest efektywnym sposobem na ustalanie adresów marker DArTseq-klon BAC i w konsekwencji kotwiczenie klonów BAC na mapie genetycznej gatunku o złożonym genomie jakim jest żyto.
2. Wykorzystanie wyników adnotacji funkcjonalnej żytnich markerów DArT i biblioteki BAC linii wsobnej żyta L318 pozwala na poznanie pełnej sekwencji przypuszczalnych genów związanych z odpornością na choroby, wykazujących zróżnicowanie alleliczne w zasobach genowych żyta.
3. Technologia Nanopore pozwala na uzyskanie wysokiej jakości sekwencji klonów BAC.
4. Genom żyta zawiera homologi genów udomowienia zidentyfikowanych u innych zbóż.

3. PRZEGLĄD LITERATURY

3.1 Charakterystyka żyta

3.1.1 Systematyka, właściwości i zastosowanie

Żyto uprawne (*Secale cereale* L.) należy do rodziny wiechlinowatych (*Poaceae*), podrodziny wiechlinowych (*Pooideae*) oraz plemienia *Triticeae*. Jest blisko spokrewnione z pszenicą (*Triticum* L.) i jęczmieniem (*Hordeum* L.) jednak w odróżnieniu od tych gatunków żyto uprawne jest rośliną obcopylną, występuje u niego samoniezgodność i depresja wsobna, co utrudnia prowadzenie badań i hodowlę tego gatunku.

Żyto jest ważną rośliną uprawną w Środkowej, Wschodniej i Północnej Europie. Polska zalicza się do wiodących producentów żyta, dzięki dobremu przystosowaniu tego gatunku do warunków glebowo-klimatycznych panujących w kraju (<http://faostat.fao.org>). Żyto używane jest głównie do wyrobu chleba, alkoholu i paszy dla zwierząt hodowlanych.

Żyto jest interesującą rośliną uprawną ze względu na posiadanie cech atrakcyjnych zarówno dla producentów jak i konsumentów. Wykazuje dużą tolerancję na niskie temperatury oraz inne rodzaje stresów abiotycznych i biotycznych. Może służyć jako model odporności dla innych roślin zbożowych o większym znaczeniu gospodarczym (pszenica, jęczmień) (Martis i in., 2013). Charakteryzuje się także walorami prozdrowotnymi, między innymi ze względu na wysoką zawartość błonnika pokarmowego. Jest cennym źródłem minerałów takich jak cynk, żelazo, fosfor. Zawiera również szereg związków bioaktywnych między innymi kwas foliowy, sterole i foliany (Jonsson i in., 2018). Udowodniono, że spożywanie chleba żytniego ma pozytywny wpływ na zdrowie ludzkie, poprawia funkcjonowanie jelit i przyczynia się do obniżenia ryzyka zachorowania na raka jelita grubego (Gråsten i in., 2000; Holma i in., 2010).

3.1.2 Genomika żyta

Żyto jest diploidem i jego materiał genetyczny jest zorganizowany w 7 par chromosomów homologicznych. Wyróżnia je jeden z największych genomów wśród roślin zbożowych, wysoce zmetylowany i z dużym udziałem sekwencji powtarzalnych - ok. 92% (Flavell i in., 1974), co stanowi wyzwanie dla genetyków i hodowców. Według najnowszych badań całkowita długość genomu dla linii wsobnej L318 wynosi 8,01 Gpz (Rabanus-Wallace i in., 2021).

W ostatnich latach nastąpił znaczny postęp w genomice żyta. Pierwsza wysokoprzepustowa technologia genotypowania została wdrożona w 2009 roku wraz z utworzeniem panelu do genotypowania żyta zawierającego 1520 klonów DArT (*Diversity Arrays Technology*) (Bolibok-Brągoszewska i in., 2009). Poźniejszy postęp manifestował się, między innymi, pojawieniem się map genetycznych opartych na tych markerach (DArT) (Milczarski i in., 2011) oraz na markerach opracowanych na podstawie sekwencji transkryptów (Haseneyer i in., 2011; Martis i in., 2013). Jednak, w momencie rozpoczynania badań składających się na tę rozprawę doktorską, wiedza na temat genomu żyta była ciągle jeszcze niewielka. W kolejnych latach zaczęto upowszechniać wysokoprzepustowe technologie genotypowania - Milczarski i in. (2016) zaprezentowali wykorzystanie metody DArTseq - jednego z wariantów GBS (genotypowanie przez sekwencjonowanie) do skonstruowania wysokorozdzielczej mapy genetycznej. Opublikowano również częściową sekwencję rejonów niskokopijnych genomu żyta linii wsobnej Lo7, wykorzystywanej w niemieckiej hodowli heterozyjnej (Bauer i in., 2017), jednakże jakość tego zasobu (N50=1708 pz, min. dł. kontigu 300 pz) ograniczała jego praktyczne zastosowanie. Nowa wersja sekwencji genomu wspomnianej wyżej linii jest dostępna od grudnia 2019 roku (Rabanus-Wallace i in., 2019; Rabanus-Wallace i in., 2021). Składa się ona z ponad 109 tysięcy skafoldów o łącznej długości 6,74 Gpz. Skafoldy zostały uporządkowane i ustawione w odpowiedniej orientacji dzięki użyciu szeregu niezależnych źródeł danych takich jak.: i) odczyty specyficzne dla chromosomów (ang. *chromosome-specific shotgun (CSS) reads*), (ii) powiązane odczyty 10X Chromium (ang. *10X Chromium linked reads*), (iii) markery mapy genetycznej, (iv) dane o interakcji chromatyny uzyskane metodą Hi-C (ang. *3D chromosome conformation capture sequencing (Hi-C)*) i (v) optyczna mapa genomu w technologii Bionano. Ostatecznie 92% ze złożonej sekwencji (tj. ~ 78% całkowitej wielkości genomu) zostało ułożone najpierw w super skafoldy (N50> 29 Mpz), a następnie w pseudomolekuły. Zastosowanie różnych metod umożliwiło oszacowanie liczby genów na 34 441 (geny o wysokim poziomie ufności), które najprawdopodobniej stanowią 97,9% wszystkich genów (Rabanus-Wallace i in., 2021).

3.1.3 Geny warunkujące ważne użytkowo cechy żyta

Dotychczas scharakteryzowano niewiele genów warunkujących ważne użytkowo cechy żyta. Przy wykorzystaniu klasycznych metod analizy genetycznej i mapowania QTL ustalono, m. in., że krótkie ramię chromosomu 1R (1RS) zawiera liczne geny warunkujące

odporność na patogeny i stresy abiotyczne (Bartos i in., 2008), natomiast na chromosomach 6R i 4R zlokalizowano po dwa geny związane z odpornością żyta na rdzę zbożową (Schlegel i Korzun 2016).

Gen *ScMATE1*, ortolog *HvMATE1* i *TaMATE1* (z jęczmienia i pszenicy), jest zaangażowany w odporność na aluminium (Al) żyta (Silva-Navas i in., 2012), które znane jest z wysokiej tolerancji na niskie pH. Aluminium jest głównym czynnikiem ograniczającym produkcję roślinną na glebach kwaśnych (Kochian i in., 2004). Wkład *ScMATE1* w tolerancję Al jest złożony. Podobnie jak w przypadku innych gatunków, u żyta występuje kilka wersji allelicznych tego genu, o różnym wpływie na tolerancję podwyższonej zawartości Al w badanych odmianach i gatunkach żyta (Santos i in., 2018).

W celu poszerzenia wiedzy na temat genetycznego podłoża biosyntezy benzoksazynoidów (BX) u żyta zespół profesor Rakoczy-Trojanowskiej uzyskał sekwencje 5 genów (*ScBx1*, *ScBx2*, *ScBx3*, *ScBx4* i *ScBx5*) kontrolujących syntezę tych głównych wtórnych metabolitów roślin należących do rodziny *Poaceae*. Dokonano tego wykorzystując sekwencję genów ze spokrewnionych gatunków i przy zastosowaniu biblioteki BAC linii wsobnej L318. Wykonano również analizę strukturalną i filogenetyczną genów *ScBx3*, *ScBx4*, *ScBx6* i *ScBx7* oraz analizę ekspresji genów *ScBx1* ÷ *ScBx6* w wybranych wsobnych liniach żyta w warunkach kontrolnych i po ekspozycji rośliny na stres zasolenia (Bakera i in., 2015).

Myśków i in. (2012) określili pozycje loci cech ilościowych (QTL) dla aktywności α -amylazy na mapie genetycznej populacji rekombinacyjnych linii wsobnych S120 x S76 i porównali je ze znanymi już wcześniej QTL dla porostania (ang. *preharvest sprouting*-PHS) i wczesnego kwitnienia (ang. *heading earliness* - HE). Porostanie zbóż to przedwczesne kiełkowanie ziarniaków w kłosach, które przynosi znaczne straty gospodarcze spowodowane obniżeniem wysokości plonów i pogorszeniem wartości technologicznej ziarna (Piasecka i in. 2017). Porostanie jest główną przyczyną nadmiernej aktywności α -amylazy (AA), co prowadzi do obniżenia jakości ziarna, ponieważ enzymatyczna hydroliza skrobi podczas produkcji żywności może prowadzić do problemów z przetwarzaniem i niezadowalającej jakości finalnego produktu (Tjin Wong Joe i in., 2004; Mares i Mrva, 2014). Znalezione 14 QTL zlokalizowanych na wszystkich siedmiu chromosomach żyta, odpowiedzialnych za 6,09–23,32% zmienności cechy aktywności α -amylazy. Stwierdzono kolokalizację sześciu QTL dla cech AA i PHS (na 1R, 3R, 4R, 6R, 7R) i również sześciu QTL dla PHS i HE (na 1R, 2R, 6R, 7R). Na długim

ramieniu chromosomu 1R znaleziono QTL mający wspólną część dla wszystkich trzech cech (AA, PHS i HE).

QTL zaangażowane w kontrolę cech o dużym znaczeniu agronomicznym takich jak wysokość rośliny, długość kłosa i liczba ziaren w kłosie zidentyfikowano dzięki wykorzystaniu map genetycznych o wysokiej gęstości dwóch populacji rekombinacyjnych linii wsobnych linii (RIL-S i RIL-M). Na mapie genetycznej populacji RIL-S zlokalizowano 35 QTLi związanych z badanymi cechami, a na mapie populacji RIL-M-52. Rejony te były rozmieszczone na wszystkich chromosomach żyta (Myśków i in., 2014).

3.2 Biblioteki BAC i ich zastosowania

Sztuczne chromosomy bakteryjne (BAC) to zmodyfikowane bakteryjne czynniki F (Peterson i in., 2000), bardzo stabilne, wolne od chimeryzmu i rearanżacji insertów (Venter i in., 1998). Pojawiły się jako alternatywa dla sztucznych chromosomów drożdżowych (YACs) (Marek i Shoemaker, 1997) i stały się cennym narzędziem genomiki strukturalnej. Są relatywnie łatwe w manipulacji i późniejszej aplikacji (Wu i in., 2015). Posiadają one markery selekcyjne, takie jak gen oporności na antybiotyki, i miejsce wielokrotnego klonowania (tzw. polilinker, ang.: *Multiple Cloning Site*, MCS), zawierające sekwencje rozpoznawane przez różne enzymy restrykcyjne (Priyadarshan, 2019). Zwykle średnia wielkość insertu waha się od 80 do 200 Kb, aczkolwiek może też sięgać nawet 500 kb (Ariyadasa i Stein, 2012).

Biblioteki BAC są przydatne w badaniach genomicznych, zwłaszcza u gatunków z niepoznaną jeszcze sekwencją genomu. Mogą służyć do izolacji genów, w wysokoprzepustowym fizycznym mapowaniu i sekwencjonowaniu genomów, rozwoju nowych markerów czy w cytogenetyce (Zhang i Wu, 2001; Ariyadasa i Stein, 2012). Efektywność wykorzystania bibliotek BAC została potwierdzona dla genomów roślin z wysoką zawartością sekwencji powtarzalnych (Visendi i in., 2016). Wielkość insertów w klonach BAC jest wystarczająca dla zastosowania zaawansowanej genetyki odwróconej. Możliwe jest wstawienie do klonu BAC całego fragmentu genomu, w którym zlokalizowany został QTL (Wu i in., 2015).

3.2.1 Zastosowanie bibliotek BAC w identyfikowaniu genów

Biblioteki BAC są cennym narzędziem genomiki strukturalnej umożliwiającym zidentyfikowanie rejonu genomu zawierającego wybraną sekwencję, poznanie sekwencji

sąsiadujących z interesujących nas fragmentem i w efekcie zidentyfikowanie pełnej sekwencji genu.

Wykorzystanie biblioteki BAC linii wsobnej AI237 kukurydzy, odpornej na glin pozwoliło na zbadanie struktury genomowej locus *MATE1*, związanego z tolerancją nadmiaru tego pierwiastka. Po przeszukiwaniu biblioteki BAC, przy pomocy starerów specyficznych do genu *MATE1* zidentyfikowano klon pozytywny. Sekwencjonowanie tego klonu metodą pojedynczej cząsteczki DNA w technologii PacBio (SMRT) ujawniło, że w genomie linii wsobnej AI237 występują trzy kopie genu *MATE1*, ułożone bezpośrednio jedna po drugiej. Każda powtarzająca się jednostka jest wielkości prawie 30 kbp i wykazuje duże podobieństwo sekwencji z pozostałymi kopiami (Maron i in., 2013). Zwiększenie liczby kopii genu *MATE1* związane jest z jego zwiększoną ekspresją, co z kolei skutkuje lepszą tolerancją nadmiaru glinu. Ustalono również, że linie kukurydzy posiadające trzy kopie allelu *MATE1* pochodzą z regionu geograficznego, w którym tolerancja podwyższonego poziomu glinu w glebie jest ważną cechą adaptacyjną i gdzie prowadzona była intensywne hodowla pod kątem tolerancji na niskie pH gleby (Maron i in., 2013).

Zidentyfikowanie genów *btr1* i *btr2*, odpowiadających za niełamliwość osadki u jęczmienia było możliwe dzięki skonstruowaniu kontigu BAC o długości 403 kbp, opartego na sekwencjach odmian Morex (Yu i in., 2002), Haruna Nijo (Saisho i in., 2007) oraz dzikiej linii jęczmienia - OUH602 (Pourkheirandish i in., 2015). Biblioteki BAC przeszukano z wykorzystaniem strategii pulowania (Stein i in., 2005), a następnie wykorzystano technikę „chromosome walking” rozpoczynającą się od dwóch zakotwiczonych markerów DNA. Sprzężenie tych markerów z genami *btr1* i *btr2* zostało potwierdzone przez genotypowanie zestawu rekombinacyjnych linii wsobnych (Pourkheirandish i in., 2015).

Dwie biblioteki BAC proso (*Panicum virgatum* L.), na które składało się 96 000 klonów, zostały wykorzystane do poszukiwania genów wpływających na biomasa i genów związanych z odpowiedzią na stres (Sharma i in., 2016). Posiłowano się sekwencjami ortologów tych genów u ryżu, sorgo, kukurydzy i *Brachypodium*. Użycie pełnych sekwencji 311 klonów BAC odpowiadających 51,7 Mbp genomu proso pozwoliło na zidentyfikowanie prawie 4000 unikalnych genów (Sharma i in., 2016). Wśród nich 695 należało do poszukiwanych rodzin genów (*GT*, *GH*, kinaz i *ERF*), pełniących role związane ze składem ściany komórkowej, wydajnością scukrzania i tolerancją na stres.

Biblioteka BAC DNA pojedynczej rośliny szpinaku, która była homozygotyczna w locus *RPF1* umożliwiła jego dokładne zbadanie (Feng i in., 2015). Gen *Rpfl* bierze udział w reakcji obronnej na patogena z gromady Oomycetes *Peronospora farinosa* f. sp. *spinaciae* (*Pfs*) przyczyniającego się do powstania mączniaka rzekomego, który jest uciążliwą chorobą szpinaku, powodującą duże straty ekonomiczne (She i in., 2018; Bhattarai i in., 2020). Wykorzystana biblioteka BAC zawierała 73 728 klonów, ze średnią wielkością insertu 183 kpz, co zapewniało 13-krotnie pokrycie badanego genomu (989 Mpz). W celu weryfikacji jakości skonstruowanej biblioteki BAC oraz wygenerowania markerów, zsekwencjonowano końce 1920 klonów BAC. Uzyskano 3405 sekwencje (BES) o długości od 100 do 935 pz. Pośród nich, 38,59 % zawierało powtarzające się DNA. Zidentyfikowano w nich geny należące do różnych kategorii, w tym czynniki transkrypcji, geny biorące udział w reakcji na stres oraz związane z reprodukcją. Znalaziono również domniemane geny odporności na choroby zawierające domeny wiążące nukleotydy i domeny bogate w powtórzenia leucynowe (NBS-LRR), z których jeden był ściśle powiązany z locus *RPF1* (Feng i in., 2015).

3.2.2 Zastosowanie bibliotek BAC w cytogenetyce

Użycie bibliotek BAC w połączeniu z fluorescencyjną hybrydyzacją in-situ (FISH) jest efektywnym sposobem fizycznego mapowania specyficznych sekwencji DNA oraz identyfikacji chromosomów u różnych organizmów, w tym również u roślin, takich jak ryż (Jiang i in., 1995; Cheng i in., 2002), bawełna (Hanson i in., 1995), sorgo (Gomez i in., 1997), rzodkiewnik (Fransz i in., 2000; Jackson i in., 2000), ziemniak (Dong i in., 2000), cebula (Suzuki i in., 2001), rośliny strączkowe (Pedrosa i in., 2002) i pszenica (Zhang i in., 2004). W zależności od wybranego celu, duża część sukcesu metody BAC-FISH zależy od proporcji powtarzających się sekwencji i unikalnych sekwencji docelowych w wybranych klonach BAC.

Wyniki badań zespołu Zang i in. (2004) potwierdziły, że BAC-FISH jest skuteczną techniką identyfikacji molekularnych markerów cytogenetycznych w dużych roślinnych genomach. Na podstawie wyników przeszukiwania trzech różnych bibliotek BAC pszenicy z wykorzystaniem genetycznie zmapowanych markerów RFLP wybrano 56 klonów BAC przyporządkowanych do poszczególnych loci. Dla 40 klonów BAC, zaobserwowano hybrydyzację do wszystkich chromosomów, na całej ich długości, spowodowaną prawdopodobnie obecnością w tych klonach elementów powtórzonych powszechnie

występujących w genomie pszenicy. Dla trzech klonów BAC nie zaobserwowano żadnego sygnału hybrydizacyjnego. W pozostałych przypadkach, w zależności od wykorzystanego klonu, obserwowano różne wzory dystrybucji sygnałów FISH, odzwierciedlające obecność w genomie różnych typów elementów powtórzonych. Zidentyfikowano również kilka nowych klas elementów powtórzonych i wskazano 5 klonów BAC, które zawierają powtórzenia przydatne w identyfikacji chromosomów. Wyniki sekwencjonowania ujawniły, że wszystkie te elementy powtórzone to transpozony, sugerując tym samym ich ważną rolę w ewolucji genomu pszenicy.

Przykładem wykorzystania strategii BAC-FISH u marchwi (*Daucus carota* subsp. *sativus*) jest praca Iovene i in. (2011). Przedstawili oni badanie kariotypu tego gatunku i przyporządkowanie odpowiednich grup sprzężeń do chromosomów na podstawie mapowania FISH 17 zakotwiczonych na mapie klonów BAC. Większość klonów BAC wykorzystanych do mapowania FISH została zidentyfikowanych po przeszukaniu dwóch bibliotek BAC marchwi (Cavagnaro i in., 2009) z wykorzystaniem zmapowanych sekwencji i strategii opartej na PCR. Klony specyficzne dla poszczególnych chromosomów zostały użyte do fizycznej lokalizacji genetycznie niezakotwiczonych sekwencji DNA, ustanowienia relacji pomiędzy dystansem genetycznym i fizycznym oraz poszerzenia wiedzy na temat ewolucji chromosomów w rodzaju *Daucus*.

W pracy Iorizzo i in., (2016) w eksperymentach FISH mających na celu ocenę spójności i pokrycia sekwencji genomu marchwi w regionach telomerowych, wykorzystano klony BAC specyficzne dla subtelerowych rejonów krótkiego i długiego ramienia poszczególnych chromosomów (1, 2, 4, 5, 6, 8, 9). Użyto 4 717 sekwencji końców BAC i nowo opracowanej mapy sprzężeń z 2 075 markerami. Wyniki przeprowadzonych badań potwierdziły, że uzyskana w pracy sekwencja nukleotydowa zapewnia dobre pokrycie genomu marchwi.

3.2.3 Zastosowanie kotwiczenie klonów BAC w integrowaniu map fizycznych i genetycznych i sekwencjonowaniu genomów

Kotwiczenie pojedynczych klonów BAC i kontigów na mapach genetycznych o wysokiej rozdzielczości ma kluczowe znaczenie przy konstrukcji i eksploatacji map fizycznych. Pozwala na określenie orientacji kontigów i uporządkowanie ich wzdłuż odpowiedniego chromosomu (Meksem i in., 2005).

Biblioteki BAC wykorzystywano przy sekwencjonowaniu genomów w strategii klon po klonie (ang.: *BAC by BAC*), co zastosowano, m. in., w przypadku rzodkiewnika, ryżu oraz kukurydzy (Arabidopsis Genome Initiative, 2001; International Rice Genome Sequencing Project, 2005; Schnable i in., 2009). Schnable (2015) przedstawił zsekwencjonowanie genomu kukurydzy przy użyciu MTP (ang.: *minimal tiling path*) czyli minimalnego zestawu zachodzących na siebie klonów BAC pochodzących ze zintegrowanej mapy fizycznej i genetycznej.

Do budowy superskafoldów wykorzystanych w zintegrowanej mapie marchwi użyto sekwencji końców klonów BAC wraz z sekwencjami uzyskanymi technologią Illumina (Iorizzo i in., 2016).

Yim i in. (2002) wykorzystali trzy biblioteki BAC w pracach nad konstrukcją mapy fizycznej kukurydzy. Każda z bibliotek była stworzona przy użyciu innego enzymu restrykcyjnego: *HindIII*, *EcoRI*, i *MboI*. Zastosowanie takiej strategii miało zminimalizować problem zbyt małej reprezentacji niektórych rejonów genomu, który może wystąpić przy użyciu jednego enzymu restrykcyjnego do konstrukcji biblioteki (Frijters i in., 1997). Te trzy biblioteki BAC zapewniały 26-krotne pokrycie genomu. Łącznie przetestowano 331 776 klonów BAC, ze średnią długością insertu 136 kpz. Do złożenia klonów BAC w kontigi użyto programu FPC (Soderlund i in., 2000). Zestaw markerów RFLP hybrydyzowano do sześciu filtrów o wysokiej gęstości z klonami BAC z każdej z trzech bibliotek BAC, aby uzyskać informacje o organizacji chromosomów i zawartości organellarnego DNA. Wyniki hybrydyzacji pokazały, że klony BAC odpowiadają kluczowym elementom struktury chromosomów takim jak centromery, telomery i rDNA. Dodatkowo, aby stworzyć szkielet do integracji mapy genetycznej i fizycznej, wykorzystano drugi zestaw sond, składający się z 90 zmapowanych genetycznie markerów RFLP kukurydzy, którym przeszukano bibliotekę *HindIII*. Dla każdego z markerów RFLP zidentyfikowano przynajmniej trzy klony BAC.

W pracy Munoz-Amatriain i in. (2015) zaprezentowano zastosowanie i optymalizację zestawu MTP obejmującego prawie 16 000 klonów BAC. Wykorzystano bibliotekę BAC jęczmienia (cv. Morex) skonstruowaną przez Yu i in. (2000), składającą się z 313 344 klonów BAC z pokryciem genomu sięgającym 6,3 x. Ze względu na duży rozmiar genomu jęczmienia i wysoką zawartość elementów repetytywnych, skupiono się na klonach BAC bogatych w geny. Po przeglądaniu biblioteki zidentyfikowano 83 832 takich klonów i poddano analizie metodą HICF (ang.: *high information content*

fingerprinting technology). Następnie przy zastosowaniu metody FMTP wytypowano 15 711 unikalnych klonów BAC stanowiących zestaw MTP. Klony te zostały następnie połączone w osiem zestawów, które poddano sekwencjonowaniu na urządzeniu Illumina HiSeq2000, po czym odczyty o najlepszej jakości zostały przyporządkowane do pojedynczych klonów BAC i złożone „klon po klonie” przy pomocy oprogramowania Velvet v. 1.2.09 (Zerbino i Birney, 2008). Ostatecznie uzyskano sekwencje 15 622 klonów BAC, a łączna długość tych sekwencji wyniosła około 1.7 Gpz, co stanowi 33% genomu jęczmienia (~5,1 Gpz; IBSC, 2012). Liczba klonów BAC z fragmentem genomu jęczmienia zawierającym gen wyniosła 83 831. Na jeden klon BAC średnio przypadało 2,38 modeli genów wysokiej ufności (ang.: *high confidence genes*, HC). Uzyskano również ciekawe informacje na temat struktury genomu jęczmienia - potwierdzono, że dystalne części chromosomów, charakteryzujące się wysokim współczynnikiem rekombinacji, zawierają więcej genów niż rejony bliższe centromerom. Jednak wykryto też rejony bogate w geny, gdzie nie ma rekombinacji (Munoz-Amatriain i in., 2015).

Mimo postępu w technologiach sekwencjonowania i składaniu genomu, mapy fizyczne chromosomów pszenicy zawierające klony BAC pozostają cennym zasobem, umożliwiającym szybki dostęp i szczegółową analizę rejonu będącego przedmiotem zainteresowania (Tulpova i in., 2019). W przypadku pszenicy płaskurki, sekwencja referencyjna pokrywa 87,5% wielkości genomu i brakujące części są uważane za kombinacje sekwencji powtarzalnych, jak i rejonów trudnych do sekwencjonowania (Avni i in., 2017). Ponadto część skafoldów (4,1% u tego gatunku) pozostaje nieprzypisana do określonej pozycji genomu. Do skutecznego klonowania genów niezbędna jest kompletna i dokładna sekwencja rejonu będącego przedmiotem badań. Dostępność klonów BAC ze znaną pozycją w genomie umożliwia ukierunkowane i niedrogi ponowne sekwencjonowanie poszczególnych rejonów bardziej zaawansowanymi technologiami, tj. PacBio i Oxford Nanopore, ułatwiając identyfikację genów (Shneider i in., 2016; Roseli i in., 2017).

Przykładem wykorzystania biblioteki BAC w mapowaniu fizycznym jest skonstruowanie mapy fizycznej krótkiego ramienia chromosomu 3D (3DS) pszenicy (Holušová i in., 2017). Badania te przeprowadzono w ramach projektu sekwencjonowania genomu pszenicy, aby ułatwić składanie sekwencji genomu referencyjnego, a także klonowanie pozycyjne genów. Wykorzystana biblioteka BAC składała się z 36 864 klonów o średniej długości insertu 110 kpz i reprezentowała 11-krotne pokrycie ramienia

chromosomu 3DS (Šafář i in., 2007). W pracy zastosowano metodę HICF oraz oprogramowanie FPC. Ostateczna mapa fizyczna zawierała 869 kontigów zbudowanych z 18 927 klonów o łącznej długości 274,5 Mpz, stanowiących 85,5% oczekiwanej wielkości ramienia chromosomu. Najdłuższy kontig (2 311 kpz) zawierał 232 klony BAC a 28 kontigów było dłuższych niż 1 Mpz. W pracy tej wykonano również sekwencjonowanie nowej generacji klonów BAC z mapy fizycznej (połączonych w pule). Odczyty przypisano *in silico* do kontigów skonstruowanej mapy fizycznej. Długość sekwencji przyporządkowanych do pojedynczych kontigów wynosiła od 200 do 917 922 pz (Holusova i in., 2017).

Tulpova i in. (2019) zbudowali opartą na klonach BAC mapę fizyczną krótkiego ramienia chromosomu 7D pszenicy (7DS) (*Triticum aestivum* L.), na którym zlokalizowanych jest szereg ważnych użytkowo genów i QTLi (Roder i in., 2008). Zaprezentowana w tej pracy wersja mapy fizycznej 7DS łączyła 1 713 markerów z mapy radiacyjnej i trzech map genetycznych (jedna z nich to mapa genomu D *Ae. Tauschii* - przodka pszenicy zwyczajnej) z uporządkowanymi i zsekwencjonowanymi klonami BAC z MTP. Do ułożenia kontigów mapy fizycznej chromosomu 7DS wykorzystano kilka metod. Pierwszą z nich było przeszukanie pul klonów BAC z użyciem 35 genetycznie zmapowanych markerów z bazy danych GrainGenes. Drugie podejście wykorzystywało integrację mapy fizycznej 7DS z mapą fizyczną *Ae. tauschii*. Połączenie klonów BAC z obu tych źródeł w jedną mapę fizyczną pozwoliło na wykrycie 635 markerów SNP na chromosomie 7DS, które zostały zmapowane do 285 różnych pozycji. Wykorzystanie zasobów genomu *Ae. tauschii* ułatwiło uporządkowanie kontigów mapy fizycznej nawet w nie ulegającym rekombinacji regionie centromeru, co zwykle stanowi główne wyzwanie w pracach nad składaniem sekwencji całego genomu. Kontigi BAC z obu wyżej wymienionych gatunków wykorzystano do porównania zmienności ich genomów. Zlokalizowano rejon w centromerze, wykazujący strukturalną zmienność między *Ae. tauschii* i pszenicą (Tulpova i in., 2019).

3.3 Sekwencjonowanie nowej generacji w genomice

3.3.1 Metody sekwencjonowania nowej generacji

Pierwszą techniką sekwencjonowania, która znalazła szerokie zastosowanie w praktyce, jest metoda dideoksy, wykorzystująca reakcję enzymatyczną, wynalaziona przez Friedricha Sangera w latach 70-tych XX wieku (Sanger i in., 1977). Sekwencjonowanie

metodą Sangera polega na stopniowym wydłużaniu nici DNA z użyciem mieszaniny deoksynukleotydów oraz dideoksynukleotydów (współcześnie znakowanych fluorescencyjnie). Dodanie jednego z czterech inaczej oznakowanych dideoksynukleotydów (ddNTP) powoduje wykrywalne zakończenie łańcucha (Sanger i in., 1977). Znakowane fragmenty DNA są następnie rozdzielane zgodnie z ich długością przez zastosowanie elektroforezy w matrycy żelowej. Długość odczytu sekwencji może wynosić od 250 do 1 100 nukleotydów. Pomimo uzyskiwanej dużej dokładności odczytów i wysokiej wiarygodności sekwencji ta metoda sekwencjonowania okazała się zbyt mało wydajna dla szybkiego odczytu sekwencji dużych genomów. Dlatego z upływem czasu rosnące zainteresowanie sekwencjonowaniem kwasów nukleinowych przyczyniło się do powstania i rozwoju technik sekwencjonowania wysokoprzepustowego. Ich celem było umożliwienie jednoczesnej analizy wielu próbek w jednej reakcji. Sekwencjonowanie nowej generacji – NGS (ang.: *next generation sequencing*), pozwala uzyskiwać miliony odczytów w czasie nieporównywalnie krótszym niż w klasycznym sekwencjonowaniu metodą Sangera.

Pierwsze technologie sekwencjonowania drugiej generacji wprowadzono w 2005 roku (Ansorge 2009). Z platform sekwencjonowania drugiej generacji najbardziej popularne to Ion Torrent, Roche/454 i Illumina.

System Roche/454 (www.454.com) był pierwszym systemem sekwencjonowania drugiej generacji wdrożonym na rynek (Morozova i Marra, 2008). Próbki były amplifikowane za pomocą emulsyjnego PCR i następnie poddawane iteracyjnemu pirosekwencjonowaniu (Margulies i in., 2005). Długość uzyskiwanych odczytów sięgała 700 pz z dokładnością sekwencjonowania wynoszącą 99,9% (Liu i in., 2011). Pirosekwencjonowanie (Ronaghi 2001) to metoda sekwencjonowania przez syntezę, która opiera się na kaskadzie enzymatycznej, w której uwalnianie pirofosforanu podczas wprowadzania nukleotydu powoduje reakcję chemiluminescencyjną, którą można wykryć za pomocą kamery o wysokiej rozdzielczości ze sprzężonym urządzeniem ładującym (CCD) (Bayes i in., 2012). Światło rejestrowane jest w postaci pików, zwanych pirogramem.

Metoda Ion Torrent bazuje na mierzeniu z zastosowaniem półprzewodników zmian pH powstałych w wyniku uwolnienia jonów H⁺ po włączeniu nukleotydu, co wyróżnia ją spośród innych technologii NGS (Rothberg i in., 2011).

Firma Illumina w 2006 roku rozwinęła strategię pozwalającą na otrzymanie z jednego sekwencjonowania około 13 GB danych w 45 milionach sparowanych odczytów. Długość odczytów generowanych za pomocą tej technologii wynosi od 50 do 300 pz (Priyadarshan, 2019). W tej technologii próbki amplikowane są w bliskim sąsiedztwie poprzez tzw. „wzmocnienie mostka na powierzchni stałej” za pomocą mikroprzepływowej stacji klastrowej (Bentley i in., 2008). Sekwencjonowanie przebiega na tej samej powierzchni przy użyciu czterech nukleotydów zawierających terminatory, które są odpowiednio wyznakowane specyficznymi dla nich barwnikami fluorescencyjnymi. Tożsamość każdego włączonego nukleotydu określana jest przez rejestrację fluorescencji kamerą CCD. Po detekcji terminator jest odcinany, co umożliwia dołączenie następnej zasady (Bayes i in., 2012).

Do metod sekwencjonowania trzeciej generacji zaliczamy Pacific Bioscience RS (PacBio) i Oxford Nanopore Technologies (ONT). Platformę PacBio (sekwencjonowanie pojedynczych cząsteczek - SMRT (ang.: *Single Molecule, Real-Time*) wprowadzono w 2011 roku (www.pacificbiosciences.com) i jest ona uważana za pierwszą metodę sekwencjonowania trzeciej generacji, ponieważ daje możliwość sekwencjonowania pojedynczych fragmentów DNA bez potrzeby jego amplifikacji. Prowadzi to do ujednolicenia pokrycia we wszystkich rejonach DNA: bogatych w AT, długich homopolimerach lub sekwencjach palindromowych (Noceń i in., 2018). Pojedyncze cząsteczki polimerazy DNA są unieruchomione na dole nanometrowych komór aperturowych zwanych falowodami w trybie zerowym (Korlach i in., 2010). PacBio oferuje wysoką przepustowość, generując od 50 000 do 70 000 odczytów na reakcję (Zhou i in., 2016). Umożliwia wygenerowanie długich odczytów, sięgających 15 000 pz. Dłuższe odczyty ujawniają złożone zmiany strukturalne obecne w sekwencji, np. informują gdzie wystąpiły różnice w ilości kopii w stosunku do sekwencji odniesienia (Noceń i in., 2018).

Firma Oxford Nanopore Technologies wprowadziła platformę sekwencjonowania opartą na nanoporach (www.nanoporetech.com) prezentując w 2014 r. urządzenie MinION. W przeciwieństwie do poprzednich metod, sekwencjonowanie odbywa się nie poprzez syntezę, ale poprzez kontrolowane odcinanie nukleotydu po nukleotydzie z analizowanej nici DNA (Lu i in., 2016). Zmiana potencjału membrany (rzędu pikoamperów) w nanoporze służy do identyfikacji przechodzących przez nią nukleotydów (Schadt i in., 2010). Długości odczytów wynoszą maksymalnie do kilkuset tysięcy pz. Jednak odczyty uzyskane w technologii Nanopore mają wyższy poziom błędów, z

dokładnością w zakresie 85% - 90% (Nicholls i in., 2019; Loit i in., 2019). Technologia Nanopore umożliwia uzyskanie bardzo długich sekwencji nawet z rejonów genomu z wysoką zawartością elementów repetetywnych (Noceń i in., 2018). Unikatowa, kompaktowa struktura urządzenia MinION daje możliwość szybkiego sekwencjonowania w mniej korzystnych warunkach czy lokalizacjach (Lu i in., 2016). Dodatkowym atutem technologii Nanopore jest uzyskiwanie danych w czasie rzeczywistym.

Wymienione powyżej techniki sekwencjonowania uzupełniają się, zapewniając różne długości i jakości odczytów.

3.3.2 Zastosowania technik sekwencjonowania nowej generacji

Do najważniejszych zastosowań technik NGS należy sekwencjonowanie genomów, transkryptomów i genotypowanie.

Postępy w sekwencjonowaniu DNA zmieniły sposób prowadzenia badań z zakresu genomiki ze względu na możliwość uzyskiwania obszernych danych dotyczących sekwencji genomu i transkryptomu w znacznie krótszym czasie i przy dużo niższych kosztach (Kircher i in., 2011). Przyczyniło się to do rozwoju metody sekwencjonowania genomów, ang.: *whole-genome shotgun DNA sequencing* - WGS. Metoda została zastosowana 25 lat temu do sekwencjonowania pierwszego bakteryjnego genomu i 20 lat temu do genomu ludzkiego (Venter i in., 2001; Lander i in., 2001; Fleischmann i in., 1995). Opierała się ona na idei sekwencjonowania przypadkowych fragmentów DNA i późniejszemu ich uporządkowaniu poprzez dopasowanie nachodzących na siebie fragmentów odczytów (Messing i in., 1981). Jednym z możliwych udoskonaleń metodycznych strategii WGS jest zastosowanie sekwencjonowania dającego sparowane odczyty (ang.: *paired-end sequencing*). Pozwala ono sekwencjonować oba końce fragmentu i generować wysokiej jakości sekwencje. Technika WGS ułatwia wykrywanie rearanżacji genomu i powtarzalnych elementów sekwencji, a także fuzji genów i nowych transkryptów (Vieira i Messing, 1982).

W projektach sekwencjonowania genomów wykorzystywane są różne platformy NGS. Podczas sekwencjonowania genomów gatunków takich jak: kakao (*Theobroma cacao*) czy miskant (*Miscanthus sp.*) wykorzystano platformę Roche/454 (Sheffler i in., 2009; Swaminathan i in., 2010). Większość publikacji dotyczących wykorzystania technologii Ion Torrent koncentruje się na sekwencjonowaniu genomów bakteryjnych (Howden i in. 2011; Rothberg i in. 2011). Platforma Illumina została użyta do

sekwencjonowania genomów wielu gatunków i ze względu na uzyskiwanie sekwencji o wysokiej jakości służy do zidentyfikowania wewnątrzgatunkowej zmienności genetycznej (Bentley i in., 2008). Większość dużych genomów roślinnych zawiera wiele powtarzających lub duplikujących się sekwencji (Pellicer i in., 2018). Dlatego platformy sekwencjonowania trzeciej generacji (PacBio i Nanopore) generujące dobrej jakości, długie odczyty dają lepszą ciągłość w pokryciu genomu przy dużo niższych kosztach.

Metodę PacBio wykorzystano do sekwencjonowania ludzkiego genomu (Clark i in., 2012) oraz w analizach gatunków roślinnych, np. ogórka (*Cucumis sativus* L.) (Osipowski i in., 2020). W tych badaniach uzyskano sekwencję linii wsobnej B10 ogórka, o długości 343,58 Mbp, zawierającą 8 096 kontigów o maksymalnej długości 12,66 Mbp i parametrze N50 wynoszącym 842 kbp. Osipowski i in. (2020) potwierdzili, że technologia PacBio generując długie, dobrej jakości odczyty zapewnia wysoką ciągłość sekwencji genomu. Technika SMRT może służyć do udoskonalania genomów referencyjnych dla roślin użytkowych, jak np. kukurydzy (Jiao i in., 2017). Berlin i in. (2015) zaprezentowali wykorzystanie takich odczytów do złożenia *de novo* genomów człowieka oraz 4 organizmów modelowych (*Escherichia coli*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Drosophila melanogaster*). Wyniki tych badań pokazują, że technika SMRT może doprowadzić do prawie kompletnego zakończenia sekwencjonowania genomów eukariotycznych przy niewielkich kosztach. W projektach sekwencjonowania dużych genomów roślinnych znajduje zastosowanie również technologia Nanopore. Dużym atutem tej technologii jest jej duża skuteczność w prawidłowym odczytywaniu sekwencji rejonów genomu zawierających sekwencje powtarzające się (Schmidt i in., 2017).

W sekwencjonowaniu genomów stosowane są również strategie „hybrydowe” łączące różne technologie sekwencjonowania, wykorzystujące zalety każdej z nich. To podejście zwiększa dokładność długich odczytów poprzez mapowanie na nich krótkich odczytów w celu wygenerowania wspólnej sekwencji (Jung i in., 2019).

NGS umożliwia bezwzględną kwantyfikację transkryptów (Mortazavi i in., 2008). Pomaga to przy wykrywaniu zmian poziomów ekspresji genów w różnych rodzajach komórek czy tkanek, poddanych różnym traktowaniom (Cullum i in., 2011; Zhou i in., 2010). Zaletą NGS jest umożliwianie analizy wpływu różnych czynników oddziałujących na ekspresję genów w skali całego genomu (Mutz i in., 2013).

Do badania ekspresji genów stosuje się metody oparte na krótkich fragmentach cDNA reprezentujących geny (ang.: *tag-based*), m. in. Tag-Seq oraz technologię sekwencjonowanie całego transkryptomu - RNA-Seq (Cullum i in., 2011). Obie metody wykrywają 96,2% tych samych transkryptów z wysokim współczynnikiem korelacji sięgającym 0,967. Sekwencjonowanie całego transkryptomu daje bardziej kompletny obraz. RNA-Seq pozwala na zidentyfikowanie nowych i alternatywnych transkryptów. Jest to możliwe dzięki odtwarzaniu pełnych sekwencji całych genów na podstawie pominiętych eksonów, miejsc startowych niedostępnych w adnotacji, zachowanych intronów, genów fuzyjnych i innych zmian w transkryptach (Mortazavi i in., 2008; Cullum i in., 2011).

Przykładem wykorzystania analizy transkryptomu RNA-Seq w połączeniu z funkcjonalną adnotacją do odkrywaniu nowych genów jest, m in., praca Furtado i in. (2015), którzy zidentyfikowali w ten sposób gen produkcji chleba pszennego (ang. *wheat bread making*, wbm) oraz geny kontrolujące plon mąki. Ciekawym przykładem badań w tej dziedzinie jest również praca zespołu Rangan i in. (2017) wykorzystająca do sekwencjonowania platformę Illumina. Przeanalizowano transkryptomy 35 odmian pszenicy metodą RNA-Seq w dwóch momentach rozwojowych: podczas wczesnego (14 dni po kwitnięciu - 14 dpa) oraz późnego (30 dni po kwitnięciu- 30 dpa) etapu procesu wypełniania ziarna u pszenicy. W etapie 14 dpa jako dominujące wśród procesów biologicznych zidentyfikowano procesy metabolizmu węglowodanów i fotosyntezę, co sugeruje duży wpływ fotosyntezy kłosów na plon ziarna. W 14 dpa większość transkryptów była związana z syntezą głównych składników ziarniaków, w tym białek zapasowych i skrobi. W 30 dpa wiele różnorodnych genów wykazywało niską ekspresję. Wśród nich przeważały geny związane z rozwojem nasion i ich jakością oraz tolerancją na stres. Wyróżniono także kluczowe transkrypty różnych szlaków metabolicznych, które pokazywały zróżnicowaną ekspresję na poszczególnych etapach procesu rozwoju ziarna pszenicy. Wykazano, że fotosynteza związana z owocnością była bardzo aktywna w 14 dpa, ale została wstrzymana w 30 dpa. Stwierdzono też, że geny dotyczące plonu mąki i jakości chleba miały wpływ na jakość ziarna pszenicy, głównie ze względu na ich wzorce ekspresji obserwowane na wcześniejszym etapie rozwoju nasion.

NGS znalazło również zastosowanie w identyfikacji i ilościowym oznaczeniu transkryptomów małego RNA (Zhou i in., 2010). Jest to nie-kodujące RNA (ncRNA) i pełni ono funkcje regulatorów ekspresji genów, a także jest związane z degradacją mRNA i modyfikacją chromatyny (Czech i Hannon, 2011). Kilka z tych rodzajów ncRNA, takich jak

mikro RNA (miRNA) lub małe interferujące RNA (siRNA), jest przedmiotem dużego zainteresowania w badaniach, ponieważ są zaangażowane w potranskrypcyjną regulację licznych procesów biologicznych (Morozova i Marra, 2008). Dostępność technologii NGS umożliwiła odkrycie kilku nowych klas ncRNA. Badania z wykorzystaniem technologii sekwencjonowania 454 przyczyniły się do odkrycia nowej klasy małych RNA - RNA oddziałujących z Piwi (ang.: *Piwi-interacting RNAs*, piRNA) (Ozata i in., 2019).

Technologie NGS są przydatne także w poznawaniu różnych typów polimorfizmów, tj. pojedynczych nukleotydów (SNP) czy indeli (Noceń i in., 2018). Jednym z zastosowań metod NGS z tym związanych jest genotypowanie poprzez sekwencjonowanie (ang.: *genotyping by sequencing*, GBS). Spośród innych metod genotypowania wyróżnia ją możliwość identyfikacji markerów i genotypowania w tym samym czasie (Poland i Rife, 2012). Technologia GBS umożliwia identyfikację nawet kilkuset tysięcy kodominujących markerów DNA (Niedziela i in., 2015).

Trzy podstawowe kroki w technologii GBS to konstrukcja biblioteki, sekwencjonowanie i analiza danych. Przygotowanie biblioteki rozpoczyna się ilościową weryfikacją genomowego DNA za pomocą metody fluorescencyjnej. Następnie wymagana jest normalizacja potrzebna do zapewnienia równej reprezentacji wszystkich próbek oraz jednakowej molarności genomowego DNA i adapterów, potrzebnych w kolejnych etapach procedury. Kolejnym etapem jest trawienie genomowego DNA enzymami restrykcyjnymi np. *ApeKI*, *PstI*, *MspI* (Baird i in., 2008). Dobór odpowiednich enzymów restrykcyjnych, wrażliwych na metylację, pozwala na uniknięcie elementów powtarzalnych genomu, takich jak np. elementy mobilne. Analizowane są rejony niskopijne (z 2- lub 3-krotnie wyższą wydajnością). Zmniejsza to trudność przyrównywania odczytów podczas bioinformatycznej analizy danych, zwłaszcza u gatunków o wysokim poziomie różnorodności genetycznej (Gore i in., 2009). Fragmenty DNA pocięte przez enzymy restrykcyjne poddawane są ligacji z adaptorami. Ponieważ te ostatnie zawierają w sobie identyfikatory (ang.: *barcode*), pochodzenie każdej próby jest ściśle zdefiniowane (Poland i in., 2012). Kolejnym etapem jest PCR, a uzyskane dzięki niemu produkty są analizowane pod względem rozmiaru i stanowią bibliotekę genomową, która jest następnie sekwencjonowana za pomocą NGS (Pachota i in., 2016). Zaletami GBS są szybkość, niezwykła specyficzność, wysoka powtarzalność i umożliwienie dotarcia do ważnych regionów genomu niedostępnych przy zastosowaniu innych strategii (Orban i in., 2011).

Dzięki stosowaniu technik, które redukują złożoność genomu, metoda ta obniża koszty analiz genetycznych (Zhou i in., 2016).

NGS zastosowano do genotypowania dużych u różnych gatunków roślin uprawnych, takich jak kukurydza, ryż, jęczmień, pszenica i soja (Huang i in., 2009; Poland i in., 2012; Li i in., 2014; Schreiber i in., 2019; Cao i in., 2022).

Markery GBS są bardzo użyteczne w mapowaniu genetycznym i umożliwiają konstrukcję map genetycznych o wysokiej gęstości. Przykładem takiego zastosowania są mapy stworzone dla kukurydzy z 5 296 oraz 6 533 markerami (Chen i in., 2014; Li i in., 2015). Fragmenty DNA z bibliotek zostały zsekwencjonowane przy użyciu platformy Illumina (HiSeq2000). Obie te mapy zostały wykorzystane do zidentyfikowania genetycznych czynników warunkujących czas kwitnienia (Chen i in., 2014; Li i in., 2015). Zhou i in. (2016) skonstruowali mapę genetyczną kukurydzy na podstawie genotypowania GBS. Platformą NGS wykorzystaną do uzyskania krótkich sekwencji była również Illumina. Wykazano wysoki stopień korelacji pomiędzy tą mapą a genomem referencyjnym linii kukurydzy B73. Jakość i dokładność mapy została zweryfikowana przez mapowanie znanego genu warunkującego barwę kolby (*pericarp color 1- P1*). Wykryto 35 QTLi wpływających na architekturę roślin, w tym 14 dla wysokości rośliny, 14 dla wysokości kłosa i siedem dla liczby międzywęźli (Zhou i in., 2016).

Jedną z wersji GBS jest metoda DArTseq. Zaletami techniki DArTseq są wysoka przepustowość, atrakcyjna cena, uzyskiwanie dużej liczby markerów oraz krótki czas oczekiwania na wygenerowanie danych (Li i in., 2015). Technologia ta jest optymalizowana dla każdego gatunku i dla żyta wybrano metodę wykorzystującą enzymy *PstI* i *HpaII*. Al-Beyroutiowa i in. (2016) wykorzystali ją w badaniach nad filogenezą rodzaju *Secale*. Przeprowadzone analizy, wykorzystujące ponad 13 000 polimorficznych markerów DArTseq, pozwoliły na poszerzenie wiedzy dotyczącej taksonomii żyta. Ostatecznie potwierdzono występowanie trzech głównych gatunków w rodzaju *Secale*: *S. sylvestre*, *S. strictum* i *S. cereale*.

Technologię DArTSeq zastosowano również do oceny różnorodności genetycznej 150 linii wsobnych żyta o różnym pochodzeniu (Targońska-Karasek i in., 2017). Do analizy użyto 14 892 markerów silicoDArT i 5 031 SNP. Uzyskane wyniki wskazywały na stosunkowo wysoki stopień różnorodności genetycznej w obrębie badanych linii wsobnych

ze średnim podobieństwem genetycznym 0,15 i heterozygotycznością od 0,01 do 0,39 (Targońska-Karasek i in., 2017).

Li i in. (2015) zastosowali genotypowanie poprzez sekwencjonowanie typu DArTseq dla trzech populacji rekombinacyjnych linii wsobnych pszenicy dzielących jednego rodzica (PBW343). Uzyskane dane pozwoliły na stworzenie mapy konsensusowej, na którą składało się 28 644 markerów w 3 757 unikalnych pozycjach. Całkowita długość mapy wyniosła 3 302,45 cM, odległość między markerami oscylowała pomiędzy 0 a 28,3 cM, przy średniej odległości 0,88 cM. Mapę tą wykorzystano do lokalizacji genów związanych z odpornością na choroby. Zidentyfikowano trzy geny związane z odpornością na rdzę (*Sr58/Lr46/Yr29*, *Sr2/Yr30/Lr27* i *Sr57/Lr34/Yr18*) na chromosomie 1BL, 3BS i 7DS odpowiednio. Zweryfikowano również 15 opublikowanych QTL związanych z rdzą u pszenicy. Zidentyfikowano cztery nowe QTL łączone z odpornością na rdzę żdźbłową i żółtą: *QSr.cim-3BS1*, *QSr.cim-3BS2*, *QSr.cim-4AS* i *QSr/Yr.cim-6BL*. Długość zidentyfikowanych rejonów genomu związanych z odpornością na rdzę żdźbłową wahała się od 0,1 do 15,8 cM.

Praca zespołu Milczarskiego (Milczarski i in., 2016) opisuje pierwsze wykorzystanie markerów DArTseq do mapowania genetycznego u żyta. Na skonstruowanej przez nich mapie znalazło się 19 081 markerów, w tym 18 132 uzyskanych z platformy DArTseq. Całkowita długość mapy wynosiła 1 593 cM, średnia odległość pomiędzy markerami wynosiła 0,47 cM. Zastosowanie platformy DArTseq do stworzenia mapy konsensusowej wysokiej gęstości umożliwiło zlokalizowanie genu *Rfc1*, który przywraca męską płodność żyta. Gen *Rfc1* został umiejscowiony na długim ramieniu chromosomu 4R.

3.4 Udomowienie roślin i jego wpływ na genom

Udomowienie (domestykacja) jest wynikiem procesów przekształcania się właściwości fizjologicznych, morfologicznych i rozwojowych dzikiego przodka udomowionego gatunku na skutek działalności człowieka (Diamond, 2002; Zeder, 2006). Hammer (1984) postulował istnienie szeregu charakterystyk wpisujących się w „syndrom domestykacji”, odróżniających rośliny uprawne od ich przodków. Te cechy to utrzymanie nasion do zbioru poprzez utratę mechanizmów rozsiewania nasion, redukcja wzrostu pędów bocznych, zwiększenie rozmiaru owoców i nasion, a także modyfikacja czasu kwitnienia (Doebley i in., 2006). Zmiany te ułatwiają zbiory, podnoszą plony oraz wpływają na długość okresu wegetacyjnego, dopasowując go do nowych warunków czy

środowisk (Kantar i in., 2017). Porównując rośliny udomowione do ich przodków również obserwuje się różnice na poziomie fizjologii. Obejmują one zmiany związane ze zjawiskiem spoczynku nasion, zmiany we wrażliwości fotoperiodycznej oraz zsynchronizowanie kwitnienia (Doebly i in., 2006). Bardzo wczesne praktyki rolne pozostawiły swoje ślady w różnorodności genetycznej roślin uprawnych. Powodem było używanie przez człowieka przy zakładaniu upraw nasion niewielkiej liczby osobników reprezentujących dany dziki gatunek, co wpływało na niewykorzystanie całego potencjału zmienności takiego „przodka”. Nieświadomie dokonywano także selekcji, dotyczącej na przykład wyboru nasion tylko z najlepszych osobników do powstawania kolejnych pokoleń (Doebly i in., 2006). Z dzikiego przodka żyta rolnicy wybierali osobniki o niełamliwych osadkach i wydłużonych nasionach (Sidhu i in., 2019), kształtując w ten sposób nowożytnę *Secale cereale*. Takie praktyki spowodowały efekt szyjki butelki (ang.: *genetic bottleneck*) (Doebly i in., 2006). Jest to sytuacja występująca w okresach bardzo małych liczebności populacji, kiedy następują zmiany frekwencji poszczególnych alleli, co prowadzi do zmniejszenia różnorodności genetycznej w obrębie populacji (Doebly, 1989). Skala tego zjawiska zależy od wielkości populacji podczas udomowienia oraz długości trwania procesu (Eyre-Walker i in., 1998). Spadek zmienności nie obejmował w takim samym zakresie wszystkich genów w genomie. Geny odpowiedzialne za pożądane cechy były narażone na najbardziej drastyczny spadek zmienności, ponieważ rośliny posiadające pożądane allele były wykorzystywane w największym stopniu do tworzenia kolejnych pokoleń i z tego względu pozostałe allele tych genów były eliminowane z populacji (Wright i in., 2005). Liczba genów związanych z udomowieniem i późniejszym powstawaniem odmian lokalnych jest dość duża - u kukurydzy ślady presji selekcyjnej zaobserwowano u 2 - 4% genów (Purugganan i Fuller, 2009).

Udomowienie traw tysiące lat temu było momentem przełomowym w historii cywilizacji. Zboża zapewniły naszej populacji bezpieczeństwo żywnościowe (Sang, 2009). Trzy rośliny zbożowe o największej światowej produkcji to kukurydza (*Zea mays*), ryż (*Oryza sativa*) i pszenica (*Triticum* spp.) (<http://faostat.fao.org>). Zostały one udomowione w różnych regionach geograficznych świata od 10 000 do 7 000 lat temu (Sang 2009). Jęczmień, plasujący się na czwartej pozycji, również należy do jednych z najwcześniej udomowionych zbóż. Następne pod względem rocznej produkcji jest sorgo, a za nim kolejno: proso, owies (*Avena sativa*) i żyto (<http://faostat.fao.org/>).

Ciekawym przykładem udomowionego gatunku jest kukurydza, która powstała w wyniku przemiany dzikiej trawy *Balsas teosinte* około 6 - 8,5 tysięcy lat temu (Piperno i in., 2009). Przeszła ona bardzo drastyczne zmiany morfologiczne podczas procesu udomowienia w porównaniu do pozostałych wymienionych wyżej gatunków traw. Sprowokowało to liczne badania poświęcone próbom rozszyfrowania podstaw molekularnych tych procesów (Doebley, 2004; Doebley i in., 2006). Wiedza na temat poszczególnych genów lub rejonów w genomie, które zostały poddane presji selekcyjnej podczas procesu udomowienia oraz hodowli jest wciąż niewielka wśród innych roślin zbożowych. U żyta nie przeprowadzono dotychczas wielu analiz dedykowanych temu zagadnieniu. Zrozumienie podstaw genetycznych procesu udomowienia oraz rozwoju roślin może przyczynić się do ulepszenia odmian roślin uprawnych (Doebley i in., 2006). Przykładem wykorzystania takiej wiedzy jest zastosowanie technologii edycji genomu CRISPR-Cas9 do połączenia pożądaných agronomicznych cech z ważnymi użytkowo cechami obecnymi w liniach dzikich. Jest to ciekawa strategia do ponownego łączenia utraconych, korzystnych cech, takich jak tolerancja na stres czy odporność na choroby, z wysokim plonowaniem. Edycja sześciu loci, które są ważne dla plonu i produktywności współczesnych linii uprawnych pomidorów, umożliwiła udomowienie *de novo* dzikiego *Solanum pimpinellifolium*. Linie zmodyfikowane tą metodą w porównaniu do swojego dzikiego przodka mają trzykrotnie większe owoce oraz 10-krotnie zwiększoną ich liczbę (Zsogon i in., 2018).

Postępy w technologii sekwencjonowania oraz rozwój metod analizy bioinformatycznej DNA dają coraz większe możliwości zrozumienia mechanizmów procesu udomowienia na poziomie genów i całego genomu organizmu (Edwards i Henry, 2011). W celu identyfikacji takich genów lub obszarów w genomie, w których mogłyby się one znajdować, przeprowadzono analizy QTL w populacjach uzyskanych z krzyżowań pomiędzy dzikim przodkiem a jego udomowionym potomkiem. Klonowanie QTL jest powolne i pracochłonne i z tego względu wprowadzono alternatywę w postaci metod genetyki populacyjnej (Doebley i in., 2006). Smith i Haigh (1974) postulowali, że geny poddane presji selekcyjnej podczas procesu udomowienia mogą charakteryzować się obniżoną zmiennością nukleotydową lub zwiększoną nierównowagą sprzężeń (ang.: *linkage disequilibrium*, LD). Jako „selective sweep” określali oni ewolucyjny proces, w którym pojawia się silnie korzystna mutacja i rozprzestrzenia się w populacji. Do wykrywania tego zjawiska stworzono wiele algorytmów. Analizując dane obejmujące cały

genom, uzyskane przy wykorzystaniu sekwencjonowania nowej generacji, stosuje się, m. in., takie narzędzia jak OmegaPlus, SweeD lub SweepFinder (Alachiotis i in., 2012; Pavlidis i in., 2013; Nielsen i in., 2005). Program OmegaPlus opiera się o charakterystykę poziomu LD i dlatego może być używany tylko wtedy, gdy znane są haplotypy. W przeciwieństwie do niego, SweeD może analizować próbki, dla których ustalona jest częstość alleli. Dzięki temu może przetwarzać zbiory danych lub nawet próby zbiorcze. Aby polepszyć skuteczność wykrywania śladów presji selekcyjnej można stosować różne narzędzia jednocześnie, np. OmegaPlus i SweeD (Pavlidis i in., 2010). Pavlidis i Alachiotis (2017) zaprezentowali również algorytm RAiSD (ang.: *Raised Accuracy in Sweep Detection*) łączący różne strategie identyfikacji regionów genomu, na które wywierana była presja selekcyjna. RAiSD umożliwia do 1 000 razy szybsze przetwarzanie danych niż inne powszechnie stosowane narzędzia.

3.4.1 Geny związane z udomowieniem roślin

W wyniku badań prowadzonych nad wpływem udomowienia na genomy roślin uprawnych zidentyfikowano dotychczas szereg genów związanych z tym procesem. Geny biorące udział w ważnych przemianach związanych z procesem domestykacji są genami regulatorowymi. Mutacje w nich zachodzące mogą prowadzić do szeregu zmian fenotypowych, które mogą stać się odpowiednim celem do działania mocnej presji selekcyjnej (Doebly i Lukens, 1998; Doebly i in., 2006). Regulatory transkrypcji to najliczniejsza grupa pośród genów udomowienia co sugeruje, że odegrały one kluczową rolę w tym procesie. Należą do nich geny takie jak: *teosinited branched1 (tb1)* (Doebly i in., 1997); *Fruitweight2.2 (fw2)* (Frary i in., 2000); *teosinte glume architecture (tga1)* (Wang i in., 2008); *Q; qSH1; Shattering4 (sh4)* (Sang i in., 2009); *reduced height-1 (Rht)* (Evans 1998); *vernalization (vrn)* (Muterko i in., 2015). Oprócz regulatorów transkrypcji wśród głównych genów udomowienia również zidentyfikowano kinazy i oksydazy cytokininowe, np. *Grain number1 (gn1)* czy *OsGA20oxidase-2 (OsGA20ox-2)* u ryżu (Doebly i in., 2006).

Próby zwiększenia plonów spowodowały nacisk na zwiększenie wielkości i liczby jadalnych części roślin w efekcie udomowienia. Do zidentyfikowanych dotychczas genów udomowienia związanych ze zwiększeniem wielkości jadalnych części roślin należy gen *fw2.2*. Został on zidentyfikowany jako QTL o dużym efekcie fenotypowym, kontrolujący

30% różnicy w masie owocu między dzikim i uprawnym pomidorem. Gen *fw2.2* działa jako negatywny regulator podziałów komórkowych (Frary i in., 2000).

Cechy związane z jakością nasion również były istotne przy udomowieniu wielu gatunków. Przykładami genów udomowienia związanych z jakością nasion są *GRAIN INCOMPLETE FILLING (GIF1)* i *qSW5*. Gen *GIF1* u ryżu koduje inwertazę ściany komórkowej potrzebną do asymilacji węgla podczas wczesnego etapu wypełniania ziarna (Krishnan i Dayanandan, 2003). Wypełnienie ziarna (ang.: *grain filling*) jest ważną cechą, która znacznie przyczynia się do wzrostu jego masy (Wang i in., 2008). Wang i in. (2008) udowodnili, że *GIF1* reguluje dystrybucję sacharozy w tkance naczyniowej załączni i znamieniu słupka do syntezy skrobi w bielmie podczas procesu wypełniania ziarna.

Gen *qSW5* u ryżu został zidentyfikowany jako główny QTL kontrolujący szerokość ziarna u ryżu, wyjaśniający 38,55 % naturalnej zmienności w populacji F₂ (Shomura i in., 2008). Koduje białko wiążące kalmodulinę i działa jako negatywny regulator zarówno szerokości ziaren, jak ich masy w zależności od ścieżki sygnałowej brasinosteroidów (Liu i in., 2017).

Gen *grain number1 (gn1)* wpływa na liczbę ziaren u różnych odmian ryżu. Koduje oksydazę/dehydrogenazę degradującą hormon roślinny - cytokininę (Ashikari i in., 2005).

Gen *teosinte glume architecture1 (tga1)* został zidentyfikowany jako QTL kontrolujący formowanie osłonki ziarniaków u przodka kukurydzy (Wang i in., 2005). U kukurydzy ziarna są nieosłonięte. Należy do grupy czynników transkrypcji *SBP* (ang.: *squamosa-promoter binding protein family*), których przedstawiciele bezpośrednio regulują geny z grupy *MADS-box* (Cardon i in., 1999). *Tga1* ma fenotypowy wpływ na różnorodne cechy, w tym lignifikację komórek, odkładanie krzemionki w komórkach, trójwymiarowy wzrost narządów i ich wielkość (Dorweiler i Doebley, 1997).

Eliminacja opadania nasion była jednym z krytycznych etapów procesu udomowienia u zbóż (Sang i in., 2009), ułatwiającym człowiekowi zbiory. Geny warunkujące tę cechę u różnych gatunków to *shattering4 (sh4)* (Li i in., 2006), *qSH1* (Konishi i in., 2006), *Brittle rachis (Btr1 i Btr2)* (Komatsuda i in., 2004), *Q i Tg* (Simonetti i in., 1999, Simons i in., 2006).

Gen *shattering4 (sh4)*, zidentyfikowany u ryżu, jest głównym QTL kontrolującym czy nasiona opadają z rośliny, tak jak u dzikiego ryżu, czy pozostają na niej jak w odmianach uprawnych (Li i in., 2006). Udowodniono, że białko *sh4* jest kodowane przez

gen, który jest homologiem czynnika transkrypcji *Myb3*. Wykazano, że za zanik cechy osypywania się nasion odpowiada mutacja powodująca zmianę pojedynczego aminokwasu (Lys do Asn) w domenie wiążącej DNA (Li i in., 2006).

Gen *qSH1* jest kolejnym głównym QTL kontrolującym opadanie nasion u ryżu (Konishi i in., 2006). Stosując różne strategie udowodniono, że pojedyncza zmiana nukleotydów w elemencie regulatorowym *cis* wywołuje zanik tej cechy (Konishi i in., 2006).

Odpowiednikiem omawianej cechy u jęczmienia jest łamliwość osadki, kontrolowana przez ściśle powiązane loci: *Btr1* i *Btr2* znajdujące się na krótkim ramieniu chromosomu 3H (Komatsuda i in., 2004). Podczas udomowienia delekcje w obu tych loci wpłynęły na pojawienie się niełamliwych osadek (Pourkheirandish i in., 2015). Wystąpienie cechy niełamliwej osadki uwarunkowane jest obecnością homozygotycznego recesywnego genotypu w jednym z loci, *btr1btr1/Btr2Btr2* lub *Btr1Btr1/btr2btr2*. Równoczesna obecność zarówno allelu *Btr1*, jak i *Btr2* jest konieczna do wystąpienia cechy łamliwej osadki w jęczmieniu (Takahashi i Hayashi, 1964).

U pszenicy zidentyfikowano dwa loci odpowiedzialne za osypywanie się nasion: *Q* - zlokalizowany na długim ramieniu chromosomu 5A oraz *Tg* na krótkim ramieniu chromosomu 2B (Simonetti i in., 1999). Gen *Q* określany jest jako super gen udomowienia, oddziałujący na wiele cech związanych z udomowieniem, w tym morfologię kłosa, który u odmian uprawnych jest kompaktowy (Simons i in., 2006). Gen ten należy do rodziny *AP2* regulatorów transkrypcji. Ta grupa genów reguluje szereg charakterystyk wpływających na rozwój roślin, a szczególnie związanych z kwitnieniem i morfologią kwiatostanów (Simons i in., 2006).

Wskazano wiele genów oddziałujących na inne cechy morfologiczne. Na architekturę i pokrój rośliny wpływa gen *teosinite branched1 (tb1)*. Koduje on czynnik transkrypcyjny z rodziny *TCP*, odpowiedzialny głównie za redukcję wzrostu pędów bocznych (Doebley i in., 1997). Kontrola wzrostu roślin jest ważnym aspektem udomowienia. Większość dzikich gatunków wykazuje tendencję do intensywnego wzrostu, co czyni je bardziej podatnymi na uszkodzenie przez wiatr czy deszcz (Lenser i in., 2013). W związku z tym wiele roślin uprawnych wybrano ze względu na „karłowaty” rozmiar, który zwiększa plony kosztem ogólnej biomasy. Jest to przydatna cecha również pod względem możliwości mechanizacji zbiorów. Wysokość roślin jest cechą poligeniczną

(Lenser i in., 2013). Wpływają na nią geny, takie jak *Ghd7* u ryżu lub *Q* u pszenicy. Opisano dwie strategie powodujące karłowatość u różnych gatunków: ingerencja w metabolizm hormonów lub sygnalizację (Salamini, 2003) oraz zmiany tożsamości merystemu powodujące określony wzrost rośliny (McGarry i in., 2013). Mutacja utraty funkcji w dwóch różnych genach (*OsKO2* i *OsGA20ox-2*), wpływających na szlak metaboliczny giberelin (GA), powoduje karłowatość u ryżu. Podobny mechanizm zaobserwowano u jęczmienia, gdzie zidentyfikowano gen *HvGA20ox-2*, który jest ortologiem genu *GA20ox-2* (Jia i in., 2009).

U pszenicy półkarłowaty fenotyp jest determinowany obecnością odpowiedniego allelu genu *reduced height-1 (Rht1)*. Geny *Rht-B1* i *Rht-D1* są obecne w większości nowoczesnych odmian pszenicy (Evans 1998). Są ortologami genu *Arabidopsis Gibberellin Insensitive (GAI)*, genu karłowatości u kukurydzy *dwarf-8 (d8)* (Chandler i in., 2002; Peng i in., 1999) oraz jęczmienia *Slender1 (HvSln1)* (Wen i in., 2013). Przeprowadzone badania wykazały, że allele *Rht* kodują białka DELLA (Peng i in. 1999), które są regulatorami transkrypcji zaburzającymi sygnalizację giberelin (Pearce i in. 2011). Zmutowane allele zawierają pojedyncze substytucje nukleotydów w rejonie kodującym N-końcową część peptydu, które powodują pojawienie się przedwczesnych kodonów stop (Peng i in., 1999). Bezpośrednim efektem mutacji jest redukcja wysokości roślin, ale również obserwuje się ważny efekt plejotropowy powodujący zwiększone przekazywanie asymilatów do rozwijających się kłosów i zwiększoną liczbę ziaren w kłosie (Flintham i in., 1997). Dodatkowo, niektóre allele genu *Rht* odgrywają rolę w odporności na choroby (Saville i in. 2012).

Czas kwitnienia jest powiązany z wysokością rośliny, ponieważ wczesne kwitnienie automatycznie redukuje czas dla wzrostu wegetatywnego, prowadząc do otrzymania mniejszych roślin (Lenser 2013). Stwierdzono, że niektóre geny związane z udomowieniem takie jak *Q* u pszenicy i *Ghd7* u ryżu wpływają na obie te cechy (Gross i in., 2010; Xue i in., 2008). Odpowiedni czas kwitnienia jest istotny dla roślin uprawnych, ponieważ uzależniona od niego może być ich ekspansja geograficzna, często związana z koniecznością dostosowania się danego gatunku do innych warunków klimatycznych (Weller i in., 2012). Zauważono, że zdolność kwitnienia u gatunków udomowionych jest mniej zależna od czynników środowiskowych, takich jak długość dnia lub temperatura (Murphy i in., 2011). Wernalizacja to procesy biochemiczne zachodzące pod wpływem niskiej temperatury w roślinach, potrzebne do przejścia z fazy wegetatywnej do

reprodukcyjnej. Odpowiedź pszenicy na wernalizację jest kontrolowana przez allele genów *Vrn*. Locus *Vrn2* jest kluczowym elementem represji kwitnienia pszenicy ozimej przed wernalizacją, podczas gdy *Vrn1* określa głównie czas kwitnienia odmian jarych (Muterko i in., 2015). Przeprowadzone badania wykazały, że odporność na niskie temperatury i wernalizacja są ze sobą powiązane (Kosova i in., 2008; Galiba i in., 2009).

3.4.2 Geny związane z odpornością na choroby

Jedną z najważniejszych cech roślin użytkowych jest odporność na choroby. Rośliny wykształciły efektywne mechanizmy do rozpoznawania i odpowiedzi na infekcje spowodowane patogenami (Sekhwal i in., 2015). Genami *R* określa się geny odpowiedzialne za przekazywanie dziedzicznej odporności na patogeny. Roślinne geny *R* determinują odporność na chorobotwórcze wirusy, bakterie, grzyby, lęgniowce, a także pasożytnicze rośliny, owady oraz nicienie (Krzymowska, 2019). Zawierają one charakterystyczne, konserwatywne domeny oraz motywy odgrywające specyficzne role w odporności na patogeny (Sekhwal i in., 2015). Większość produktów tych genów posiada zachowaną ewolucyjnie sekwencję wiążącą nukleotydy (ang.: *nucleotide binding* - NB) oraz domenę zawierającą wielokrotne powtórzenia sekwencji aminokwasowych bogate w leucynę (ang.: *leucine rich region* - LRR), stąd nazywane są białkami typu NB-LRR (Akira i in., 2001). Domena LRR daje białkom NB-LRR możliwość specyficznego rozpoznania obecności patogenu, zachowując się w niektórych przypadkach jak receptor (Meyers i in., 1999). Wykazano, że oprócz wiązania różnego typu ligandów domeny LRR mogą brać udział w tworzeniu oddziaływań wewnątrzcząsteczkowych, odpowiedzialnych za modulację sygnału (Hwang i Williamson, 2003). Domena odgrywająca rolę w interakcjach białko-białko jest umiejscowiona na aminowym końcu łańcucha peptydowego. Na podstawie rodzaju tej domeny wyróżnia się dwa typy białek R (Krzymowska 2019). Pierwsza z nich obejmuje białka z domeną zwiniętego kłębka (ang.: *coiled coil*, CC), która zlokalizowana jest w odległości 25 do 50 reszt aminokwasowych (aa) od N-końca (Meyers i in., 1999). Do tej grupy należą, mi in. białka RPS2 i RPM1 z *Arabidopsis thaliana*, zapewniające odporność na *Pseudomonas syringae*. Charakterystyczną cechą białek drugiej klasy jest posiadanie domeny homologicznej do cytoplazmatycznej części białka Toll oraz do receptora TIR-interleukiny I systemu immunologicznego ssaków (TIR) (Meyers i in., 1999). Przedstawicielem tej klasy jest białko N tytoniu warunkujące odporność na wirus mozaiki tytoniu TMV (Krzymowska 2019). TIR jest wysoce konserwatywnym motywem, zidentyfikowanym pierwotnie u owadów, a następnie u

ssaków i roślin, który bierze udział w procesach odpowiedzi wrodzonej na atak patogenów (Cannon i in., 2002).

Do genów związanych z reakcją roślin na patogeny należą również receptory kinazowe (ang.: *receptor like kinase*, RLK) oraz proteinowe (ang.: *receptor like protein*, RLP). Są one głównymi składnikami pierwszej linii odpowiedzi immunologicznej rośliny wyzwalanej przez elicytory mikrobiologiczne, co ma zwykle miejsce w ścianach komórkowych (Sekhwal i in., 2015). Struktura białek kodowanych przez geny z obu tych grup jest podobna. Zawierają one pojedynczy peptyd na końcu N, domenę zewnątrzkomórkową do rozpoznawania wzorców pochodzenia bakteryjnego poprzez powtórzenia bogate w leucynę oraz transmembranową domenę helix umożliwiającą kotwiczenie RLP i RLK w błonie komórkowej (Becraft, 2002). RLP różnią się od RLK brakiem wewnątrzkomórkowej domeny kinazowej. Obie te rodziny białek są zaliczane do receptorów rozpoznających wzorce (ang.: *pattern recognition receptors*, PRRs), rozpoznających elicytory takie jak lipidy, białka, kwasy nukleinowe i węglowodany (Monaghan i Zipfel, 2012).

Wskazanie przypuszczalnej funkcji homologów genów *R* u spokrewnionych gatunków jest utrudnione, ze względu na ich zdolność do szybkiego ewoluowania (Michelmore i Meyers, 1998). Ciekawy przykład ewolucji różnorodności allelicznej stanowią geny *Mla* jęczmienia biorące udział w w odporności na patogen wywołujący mączniaka (*Blumeria graminis f. sp. hordei*). Analiza paleontologiczna retrotranspozonów z długimi powtórzeniami końcowymi wskazuje, że obecny region *Mla* ewoluował przez 7 milionów lat w wyniku różnych zdarzeń takich jak duplikacje, inwersje i insercje transpozonowe (Wei i in., 2002). Różne warianty locus *Mla* występują u odmian uprawnych na całym świecie, stanowiąc unikalny system eksperymentalny, w ramach którego bada się ewolucję różnorodności allelicznej genów *R* u gatunków jednoliściennych (Jørgensen, 1994; Wise, 2000). Homologi *Mla* odkryto także u innych gatunków traw. Periyannan i in. (2013) udowodnili, że gen *Sr33*, związany z odpornością na rdzę żdźbłową i występujący u *Aegilops tauschii*, jest homologiem genów *Mla*. Locus *Mla* u jęczmienia zawiera co najmniej 6 funkcjonalnych genów odporności (Gu i in., 2015). Jest zorganizowany jako trzy wyspy bogate w geny, oddzielone dwoma zagnieżdżonymi kompleksami zawierającymi transpozony i regionem ubogim w geny o długości 45 kb (Wei i in., 2002). Większość genów odporności na mączniaka w jęczmieniu znajduje się blisko telomerowego końca chromosomu 5S (1HS) (Wei i in., 1999). U pszenicy gen kinazy

receptorowej *Lrk10* i locus odporności na mączniaka prawdziwego - *Pm3* znajdują się w odległości zaledwie 0,7 cM od homologu *Mla* - *TaMla*. Świadczy to tym, że najbardziej polimorficzne locus odporności na mączniaka prawdziwego u jęczmienia i pszenicy mogły ewoluować równolegle w dwóch blisko powiązanych homeoloci (Zhou i in., 2001).

W ostatnich dekadach sklonowano i scharakteryzowano ponad 100 genów *R* u różnych gatunków roślin (Liu i in., 2007; Yang i in., 2013). Jednak dotychczas scharakteryzowano niewiele genów związanych z odpornością na choroby u żyta. U innych gatunków zbożowych naszej strefy klimatycznej blisko spokrewnionych z żytem, takich jak pszenica czy jęczmień, geny *R* są szerzej zbadane i opisane (Gu i in., 2015). Jednym z przykładów badań nad charakterystyką genów *R* u żyta jest praca Hurni i in., (2013). Sklonowali oni gen *Pm8*, który jest żytnim ortologiem genu *Pm3* pszenicy, biorącym udział w odporności na mączniaka. Wykazano, że podobieństwo sekwencji białek PM8 i PM3B sięga 81% i różnice na poziomie nukleotydów zlokalizowane są głównie w domenie LRR. W domenie CC, która jest najbardziej zachowana ewolucyjnie, zaobserwowano różnice jedynie w 10 na 158 aminokwasów. Analizy wirulencji przeprowadzone na różnorodnych izolatach potwierdziły również podobieństwo funkcjonalne pomiędzy *Pm8* i *Pm3*, co prawdopodobnie opiera się na rozpoznawaniu podobnych białek efektorowych (Hurni i in., 2013).

Zidentyfikowano i zmapowano także pięć dominujących genów odporności na rdzę brunatną u żyta, która jest jedną z najważniejszych chorób żyta, wywoływaną przez grzyba *Puccinia recondita* f. sp. *secalis* (*Prs*): *Pr1*, *Pr2*, *Pr3*, *Pr4* i *Pr5* (Wehling i in., 2003; Roux i in., 2004). Geny *Pr1* i *Pr2* znajdują się na chromosomach 6R i 7R (Wehling i in., 2003). Gen *Pr3* zlokalizowano w centromerowym regionie krótkiego ramienia chromosomu żyta 1R (1RS), natomiast *Pr4* i *Pr5* przypisano do regionu centromerowego długiego ramienia chromosomu 1R (1RL) (Roux i in., 2004).

Rakoczy-Trojanowska i in. (2017) zasugerowali możliwy związek między zawartością benzoksazynoidów i odpornością na rdzę brunatną (BR). Ustalono, że polimorfizm pojedynczego nukleotydu (*ScBx4_1583*) w genie *ScBx4* (kodującym monooksygenazę cytochromu P450) jest związany z odpornością na BR. Wcześniej pokazano korelację dwóch SNP: *ScBx4_1583* i *ScBx4_1627* z zawartością BX w korzeniach roślin (Rakoczy-Trojanowska i in., 2016). W dalszych badaniach zespołu, postanowiono wyjaśnić wpływ BR na syntezę sześciu BX (HBOA, GDIBOA, DIBOA, GDIMBOA, DIMBOA i MBOA) oraz ekspresję siedmiu genów (*ScBx1-5*, *ScIgl* i *Scglu*).

Na podstawie przeprowadzonych obserwacji postawiono hipotezę, że trzy BX: GDIBOA, GDIMBOA i MBOA są kluczowymi składnikami reakcji obronnych żyta, przy czym szczególnie dwa pierwsze (zwłaszcza GDIMBOA) zapewniają ochronę roślin przed BR. Odnotowano znaczący wzrost ich zawartości w decydujących momentach dla rozwoju choroby, czyli kiedy zaczynają się tworzyć komórki macierzyste haustoriów (ang.: *haustoria mother cells*), wzrost patogenu jest najszybszy i rozwijają się zmiany nekrotyczne. Kluczowymi komponentami reakcji obronnej wszystkich analizowanych linii wsobnych żyta na BR wśród analizowanych genów były: *ScBx1*, *ScBx2*, *ScBx4* i *Scglu*. Zmiany w poziomach ekspresji genów były zwykle dodatnio związane z opornością na chorobę (Święcicka i in., 2020).

4. MATERIAŁY I METODY

4.1 Biblioteka BAC

W badaniach wykorzystano bibliotekę BAC linii wsobnej żyta L318, ze średnią długością insertu 121 kpb i z około 1,5-krotnym pokryciem genomu. Biblioteka składa się z 105 216 pojedynczych klonów, uporządkowanych w 273 płytkach, każda z 384 studzienkami (Bakera i in., 2015). Do konstrukcji biblioteki wykorzystano wektor pIndigoBAC-5 (Epicenter, Madison, USA). DNA poddano trawieniu enzymem *HindIII*. Udział pustych klonów został oszacowany na 0,7 % (Šimková i Rakoczy-Trojanowska, 2010).

4.2 Konstrukcja pul klonów BAC

Do tworzenia pul zbiorczych zastosowano technikę trójwymiarową wg Liu i in., (2011). W tej metodzie bibliotekę dzieli się na superpule (SP) zawierające siedem 384-dołkowych płytek. Przed rozpoczęciem przygotowania pul wykonano przy pomocy replikatora 384-igłowego repliki płytek biblioteki w pożywce płynnej LB zawierającej 25 mg chloramfenikolu w ml, po 80 µl pożywki w każdym dołku płytki. Bakterie namnażały się przez 16 godzin w temperaturze 37 °C. Następnie zebrano równe objętości kultur bakteryjnych z odpowiednich dołków. W celu stworzenia pul płytkowych połączono 35µl kultury z każdego dołka danej płytki przy pomocy elektronicznych pipet wielokanałowych. Przy przygotowywaniu pul rzędowych połączono 35 µl kultury z każdej studzienki z danego rzędu, np. rzędu A, z każdej z 7 płytek wchodzących w skład SP. Podobny schemat zastosowano przy konstrukcji pul kolumnowych. W ten sposób, dla każdej SP po pierwszym etapie pulowania uzyskano siedem pul płytkowych (od 1 do 7), 16 rzędowych (od A do P) i 24 kolumnowych (od 1 do 24).

Zgodnie ze schematem opisanym przez Liu i in. (2011) i przedstawionym w Tabeli 1 następnie łączono równe porcje pul danego typu stworzonych w pierwszym etapie w celu stworzenia tzw. pul matrix. Uzyskano w ten sposób pięć pul płytkowych nazwanych matrix (W1 do W5), jak również 8 pul rzędowych (R1 do R8) i 10 kolumnowych (T1 do T10). Każda z pul stworzonych w pierwszym etapie była reprezentowana w 2 pulach typu matrix.

Tab.1. Schemat tworzenia pul typu matrix z pul rzędowych, kolumnowych i płytkowych w danej superpuli

Pule rzędowe matrix	Pule rzędowe wchodzące w skład puli rzędowej matrix	Pule kolumnowe matrix	Pule kolumnowe wchodzące w skład puli kolumnowej matrix	Pule płytkowe matrix	Pule płytkowe wchodzące w skład puli kolumnowej matrix
R1	A, B, C, D	T1	1,2,3,4,5,6	W1	1, 2, 3
R2	E, F, G, H	T2	7,8,9,10,11,12	W2	4, 5, 6
R3	I, J, K, L	T3	13,14,15,16,17,18	W3	7, 1, 4
R4	M, N, O, P	T4	19,20,21,22,23,24	W4	2, 5, 7
R5	A, E, I, M	T5	1, 7, 13, 19	W5	3, 6
R6	B, F, J, N	T6	2, 8, 14, 20		
R7	C, G, K, O	T7	3, 9, 15, 21		
R8	D, H, L, P	T8	4, 10, 16, 22		
		T9	5, 11, 17, 23		
		T10	6, 12, 18, 24		

To pozwoliło na kontrolę wiarygodności wyników przeglądania biblioteki: sygnał dla danego klonu BAC powinien wystąpić w określonych 2 pulach płytkowych matrix, 2 pulach kolumnowych matrix i 2 pulach rzędowych matrix z danej superpuli.

4.3 Izolacja BAC DNA z pul zbiorczych

Do izolacji BAC DNA z pul zbiorczych wykorzystano zmodyfikowaną metodę lizy alkalicznej (Sambrook 2001), którą wyróżnia zastąpienie siarczanu dodecylu sodu (SDS) N-lauroilosarkozyną sodu (SLS) w buforze przeznaczonym do lizy (Sinnert i in., 1998). Po wytrąceniu etanolem o stężeniu 96% i wypłukaniu etanolem o stężeniu 70%, BAC DNA zostało zawieszony w 50 µl buforu TE (pH 8.0, 10mM Tris-HCl, 1mM EDTA) i zostawione w lodówce w temperaturze 4°C na noc w celu rozpuszczenia. Stężenie i jakość wyizolowanego DNA oceniono przeprowadzając elektroforezę w 0,8% żelu agarozowym oraz wykonując pomiary spektrofotometryczne za pomocą NanoDrop 2000 (Thermo Scientific). Roztwory DNA BAC były przechowywane w temperaturze -20° C.

4.4 Izolacja DNA pojedynczych klonów BAC

Do izolacji DNA pojedynczych klonów BAC (BAC DNA) zastosowano zestaw PhasePrep BAC DNA (SIGMA Aldrich). Procedurę przeprowadzono w skali mini. Ze stocków glicerolowych wykonano posiew redukcyjny na pożywkę stałej LB z chloramfenikolem [25 mg/ml]. Po całonocnej inkubacji w cieplarni w temp. 37°C z

pojedynczej kolonii założono pre-kulturę płynną w objętości 15 ml, z której po 24-godzinnym wytrząsaniu (280-300 rpm) w temperaturze 37°C przeniesiono 150 µl do 50 ml pożywki LB. Następnie kulturę w pożywce płynnej LB inkubowano w temperaturze 37°C na wytrząsarce przez 20 godzin. Kolejne kroki wykonano zgodnie z instrukcją producenta dołączoną do zestawu. Ostatecznie DNA zostało zawieszono w 25 µl sterylnej dejonizowanej wody a jego stężenie i jakość została sprawdzona jak wyżej.

4.5 Genotypowanie pul klonów BAC metodą Genotyping by Sequencing (DArTseq)

BAC DNA pul zbiorczych zostało wysłane do genotypowania do firmy Diversity Arrays Technology Pty Ltd (Canberra, Australia). Metodą DArTseq wykonano genotypowanie prób zbiorczych (matrix) z 16 SP reprezentujących 43 008 klonów. Przy uwzględnieniu średniej wielkości insertu i udziału pustych klonów dla badanej biblioteki BAC przeszukano zatem około 5,17 Gbp DNA żyta.

Zastosowano procedurę opisaną przez Al-Beyrouthova i in. (2015), wykorzystującą kombinację enzymów restrykcyjnych *PstI-HpaII*. Jedynymi modyfikacjami zastosowanymi w niniejszych badaniach było przeprowadzenie analizy sekwencyjnej na urządzeniu HiSeq2500 (Illumina, USA) oraz użycie tylko markerów Silico-DArT, informujących tylko o obecności lub braku danej sekwencji w badanej próbce DNA (Li i in., 2015; Milczarski i in., 2016).

4.6 Rozkodowanie wyników genotypowania klonów BAC - ustalenie adresów marker DArTSeq-klon BAC

Wyniki genotypowania otrzymano w postaci plików excel. Obecność markera DArTseq w danej puli matrix oznaczona była jako „1”, a jej brak „0”. Dane z każdej SP analizowano oddzielnie. Najpierw wybrano markery z dokładnie sześcioma pozytywnymi wynikami w 23 pulach matrix: 2 w pulach płytkowych, 2 w rzędowych i 2 w kolumnowych. W ten sposób zidentyfikowano markery DArTseq występujące tylko raz w danej SP. Następnie użyto formuł w programie Microsoft Excel do identyfikacji płytki i studzienki (informacje o rzędzie i kolumnie) z klonem BAC zawierającym odpowiedni marker.

Wyróżniono 3 grupy markerów DArTseq ze względu na liczbę ich wystąpień w klonach BAC: TS1 – markery DArTseq z dowolną liczbą wystąpień, TS2 – markery, które wystąpiły raz lub dwa razy i TS3 – markery, które wystąpiły tylko raz. Ten podział zastosowano by umożliwić wskazanie markerów pochodzących z niskokopijnych rejonów genomu.

4.7 Kotwiczenie klonów BAC do map genetycznych

Kotwiczenie klonów BAC za pomocą markerów DArTseq odbyło się z wykorzystaniem zintegrowanej mapy genetycznej markerów DArTseq wygenerowanej przy użyciu oprogramowania Biomercator V4.257 (Sosnowski i in., 2012) z indywidualnych map czterech populacji posiadających wspólnego rodzica: 541 × 2020 LM (92 RIL), 541 × DD1 (91 osobników F₂), 541 × S44 (89 osobników F₂) i 541 × EM1 (88 osobników F₂). Długość mapy wynosiła 1590 cM, a średnia odległość między markerami to 0,47 cM. Zlokalizowano na niej 32 699 markerów DArTseq w 11 393 pozycjach (Borzęcka i in., 2018). Jedyna opublikowana wcześniej mapa żyta oparta na markerach DArTseq – mapa populacji 541 × 2020LM (Milczarski i in., 2016) – była wykorzystana jako mapa referencyjna w procesie kompilacji wyżej wymienionych map. Do ilustracji dystrybucji zakotwiczonych klonów wzdłuż mapy genetycznej wykorzystano program Circos (Krzywiński i in., 2009).

4.8 Walidacja adresów marker DArTseq-klon BAC

Aby sprawdzić wiarygodność przyporządkowania markerów DArTseq do klonów BAC zastosowano dwie metody: pierwszą z użyciem algorytmu BLAST (Altschul i in., 1990), a drugą wykorzystującą informacje o położeniu markerów na mapie genetycznej.

Pierwsza metoda polegała na poszukiwaniu algorytmem BLAST sekwencji homologicznych do sekwencji oczekiwanych markerów DArTseq w sekwencjach wybranych klonów BAC z genotypowanych SP. Jeśli sekwencja homologiczna spełniała następujące kryteria: wartość $e < 0,00001$, podobieństwo $\geq 75\%$, długość dopasowania $\geq 75\%$ długości sekwencji markera, przyjmowano, że została potwierdzona obecność danego markera DArTseq w danym klonie BAC.

W przypadkach, gdy wiele genetycznie zmapowanych markerów DArTseq było adresowanych do tego samego klonu BAC, porównano lokalizacje tych markerów na mapie genetycznej. Przeanalizowano (i) czy wiele markerów przypisanych do jednego klonu BAC znajdowało się na tym samym chromosomie na zintegrowanej mapie i (ii) jaka była odległość pomiędzy takimi markerami na mapie genetycznej.

4.9 Sekwencjonowanie indywidualnych klonów BAC

Do sekwencjonowania indywidualnych klonów BAC wykorzystano dwie metody sekwencjonowania nowej generacji: Illumina oraz Oxford Nanopore. Jeden klon BAC zsekwencjonowano obiema metodami.

Siedem klonów BAC dostarczono do sekwencjonowania w Instytucie Biochemii i Biofizyki (Warszawa, Polska). Wykonano je z użyciem platformy MiSeq, Illumina. Otrzymano krótkie (ok. 150 nukleotydów), zachodzące na siebie odczyty, które zostały złożone w dłuższe kontigi przy pomocy narzędzia Newbler v. 3.0 (Roche Applied Science).

Wybrano również 12 klonów BAC do sekwencjonowania technologią Nanopore na urządzeniu MinION (Oxford Nanopore Technologies) w KGHIBR we współpracy z dr. Piotrem Gawrońskim. Izolację DNA z klonów BAC wykonano przy użyciu zestawu The PhasePrep™ BAC DNA Kit, jak opisano powyżej. Koncentrację i czystość wstępnie oszacowano metodą spektrofotometryczną z użyciem urządzenia NanoDrop 2000 (Thermo Scientific), a następnie dokonano precyzyjnych pomiarów koncentracji DNA BAC za pomocą fluometru Qubit (Qubit dsDNA Assay, Thermo Fisher). Do sekwencjonowania użyto 1 µg BAC DNA. Kolejnym etapem była fragmentacja BAC DNA. Do tego procesu użyto kolumn gTUBE (Covaris). Próbkówki g-TUBE z roztworem BAC DNA wirowano przez 1 minutę w temperaturze pokojowej z prędkością 5 000 rpm. Następnie próbkówki odwracano i ponownie wirowano przez 1 minutę w celu zebrania pofragmentowanego BAC DNA.

Po fragmentacji nastąpiła naprawa końców fragmentów DNA BAC z zastosowaniem modułu NEBNext End Repair Module (NEB). Następnie do każdej próbki BAC DNA z osobna przyłączono fragmenty DNA pozwalające na późniejszą identyfikację próbki, tzw. barkody. Wykorzystano w tym celu zestaw Native Barcoding Kit 1D (EXP-NBD103). Ostatnim etapem było przyłączenie adaptorów (Ligation Sequencing Kit 1D SQK-LSK108). Na każdym etapie sprawdzano ilość uzyskanego DNA za pomocą Qubit (Qubit dsDNA Assay).

Pierwszym etapem sekwencjonowania z wykorzystaniem MinION (Oxford Nanopore) była konfiguracja urządzenia i płytki sekwencyjnej SpotON (ang.: *flow cell*) (Oxford Nanopore) z komputerem oraz wprowadzenie danych konfiguracyjnych. Przy użyciu dedykowanego oprogramowania (MinKNOW) sprawdzono poprawność funkcjonowania układu, oraz ustalno liczbę aktywnych porów sekwencyjnych.

Przed przystąpieniem do nakładania biblioteki DNA na płytkę sekwencyjną nakładany był tzw. bufor startowy (priming buffer), w którego skład wchodzi 576 μ l buforu RBF (Oxford Nanopore) oraz 624 μ l sterylnej wody. Następnie do 12 μ l biblioteki DNA dodawano 35 μ l RBF, 25,5 μ l buforu LLB (Oxford Nanopore) oraz 2,5 μ l sterylnej wody, całość delikatnie mieszano przy użyciu pipety. Na płytkę sekwencyjną nałożono 75 μ l tak przygotowanej mieszaniny, w postaci pojedynczych kropli. Sekwencjonowanie było prowadzone przez 48 h. Postępy sekwencjonowania, liczba i średnia długość odczytów oraz liczba aktywnych porów były monitorowane na bieżąco. Po zakończeniu sekwencjonowania, płytki sekwencyjne, które zachowały dużą liczbę aktywnych porów były płukane z wykorzystaniem zestawu Wash Kit (Oxford Nanopore) według protokołu producenta, a następnie przechowywane w lodówce w temperaturze 2 - 8°C.

Bioinformatyczną analizę rozpoczęto od procedury zwanej „basecalling” czyli identyfikacji nukleotydów na podstawie zmian w przepływie prądu (program Albacore, Sahoo 2017). Kolejnym etapem było filtrowanie otrzymanych odczytów za pomocą programu Filtrlong (<https://github.com/rrwick/Filtrlong>). Po przefiltrowaniu odczytów, składano kontigi za pomocą Canu (Koren i in., 2017). Po złożeniu kontigów niezbędne było uwzględnienie kolistej struktury BACów do czego zastosowano program Circlator (Hunt i in., 2015).

4.10 Elektroforeza pulsacyjna (PFGE)

W celu określenia długości insertów w wybranych do sekwencjonowania klonach BAC wykorzystano technikę elektroforezy pulsacyjnej (PFGE). Najpierw strawiono DNA poszczególnych klonów BAC enzymem FastDigest *NotI* (ThermoFisher) przez godzinę w temperaturze 37°C. Mieszanina reakcyjna składała się z 3 μ l DNA (0.5-1 μ g/ μ L), 1 μ l enzymu *NotI* (#ER0591 300 U), 14 μ l wody MiliQ oraz 2 μ l buforu (10X Buffer O). Elektroforezę przeprowadzono za pomocą urządzenia CHEF-DR III System, w 0,8% żelu agarozowym (Agarose Plus, Prona) z buforem 0,5xTBE. Wykorzystano marker wielkości MidRange (50 μ g/ml, New England BioLabs) oraz 10 μ l DNA po trawieniu *NotI*. Zastosowano następujące parametry: napięcie 6V/cm, początkowy czas pulsu 5 s, końcowy czas pulsu 15 s, kąt 120°, temperatura 12°C, całkowity czas elektroforezy 16 godzin (Peterson i in., 2000). Następnie żel wybarwiano w wodnym roztworze bromku etydyny (1mg/ml, Merck) przez 40 minut. Po odpłukaniu z roztworu bromku żel oglądano na

transiluminatorze, oraz wykonywano fotografię przy użyciu systemu do dokumentacji żeli (The Gel Doc XR+, Biorad).

4.11 Identyfikacja potencjalnych genów związanych z odpornością na choroby

W badaniach wykorzystano wyniki analizy funkcjonalnej sekwencji klonów DArT i genotypowania pul klonów BAC metodą DArT (Gawroński i in., 2016; Borzecka i in., 2018). Do analiz wytypowano klony BAC, w których spodziewano się obecności markerów DArT, którym w analizie funkcjonalnej przyporządkowano funkcje związane z odpornością na choroby (Gawroński i in., 2016; sekwencje zdeponowane w bazie GSS banku genów GenBank: numery akcesyjne sekwencji KS366081-KS372098, numer akcesyjny biblioteki LIBGSS_039286).

Obecność spodziewanych markerów DArT w klonach BAC potwierdzano zarówno bioinformatycznie jak i eksperymentalnie.

Do weryfikacji eksperymentalnej wykorzystano technikę PCR. Startery do powielenia fragmentów genów związanych z odpornością na choroby zaprojektowano na podstawie sekwencji markerów DArT, wykorzystując narzędzie Primer-BLAST (Ye i in., 2012). Do przeprowadzenia reakcji przygotowano mieszaninę o objętości 15 μ l zawierającą 5 μ l DNA [10ng/ μ l]; 0,6 μ l starterów F i R [5 μ M], 0,3 μ l [10 mM] dNTPs, 0,15 μ l [0,5 mM] MgCl₂; 1,5 μ l buforu (DreamTaq™ Green Buffer 10X, ThermoFisher) i 0,3 μ l [5 U/ μ l], polimerazy DreamTaq (Thermo). Matrycą w reakcji było DNA klonów BAC oraz DNA genomowe linii wsobnej L318. W zależności od użytego startera zastosowano różne profile termiczne (temperatury przyłączenia starterów w zakresie od 50 do 55°C), a podstawowy program do amplifikacji składał się z następujących etapów; (1) 10 cykli: 94 °C przez 1 min, 64°C przez 30 s (temperatura obniżana o 1°C co cykl), 72°C przez 1 min i później (2) 30 cykli: 94 °C przez 1 min, 55 °C (lub inna temperatura w zależności od startera) przez 1 min, 72 °C przez 1 min; (3) 72 °C przez 5 min. Produkty amplifikacji były wizualizowane przy wykorzystaniu elektroforezy w 0,8-1,5% żelu agarozowym (im mniejszy produkt tym wyższe stężenie agarozy pozwalające na lepszy rozdział krótszych fragmentów kwasów nukleinowych) z bromkiem etydyny (0,5 μ g/ml, Merck). Produkty PCR poddano sekwencjonowaniu metodą Sanger (GENOMED S.A w Warszawie). Analizę bioinformatyczną wykonano z użyciem algorytmu BLAST (Altschul i in., 1998), posilując się różnymi bazami danych z NCBI (nucleotide collection, reference genomic sequences, reference RNA sequences).

Po uzyskaniu sekwencji insertów klonów BAC wykorzystano algorytm BLASTn (z ustawieniami domyślnymi) do poszukiwania sekwencji homologicznych do sekwencji oczekiwanych markerów DArT w poszczególnych klonach BAC.

4.12 Identyfikacja przypuszczalnych żytnich homologów genów związanych z udomowieniem

Ze względu na niewielką wiedzę na temat genów związanych z udomowieniem u żyta, zgromadzono informację na ich temat u różnych gatunków roślin zbożowych: *T. aestivum*, *T. monococcum*, *O. sativa*, *H. vulgare*. Pobrano ich sekwencje z bazy danych NCBI i zaprojektowano startery do amplifikacji fragmentów tych genów. Do projektowania starterów użyto narzędzia Primer-Blast (Ye i in., 2012)

Po pojawieniu się sekwencji genomu linii wsobnej żyta Lo7 przeszukano ją w celu znalezienia homologów wyżej wymienionych genów (Rabanus-Wallace i in., 2019; Rabanus-Wallace i in., 2021) algorytmem BLAST (Altschul i in., 1998). Na podstawie sekwencji identyfikowanych fragmentów homologicznych projektowano startery do powielenia z genomu żyta fragmentów przypuszczalnych homologów genów związanych z udomowieniem. Startery do genu *VRN1* uzyskano z publikacji Li i in. 2011. Skład mieszaniny reakcyjnej i program amplifikacji był taki sam jak opisany powyżej - przy identyfikacji potencjalnych genów związanych z odpornością na choroby. Ostatnim krokiem było przeszukanie biblioteki BAC linii wsobnej L318 przy pomocy starterów zaprojektowanych do amplifikacji fragmentów przypuszczalnych genów związanych z udomowieniem.

Kolejnym krokiem było przeprowadzenie PCR z BAC DNA (w koncentracji 100 ng/1µl) 39 superpul. Do przeprowadzenia reakcji przygotowano mieszaninę o objętości 15 µl zawierającą 1,5 µl DNA; 0,6 µl starterów F i R [5 µM], 0,3 µl [10 mM] dNTPs, 0,15 µl [0,5 mM] MgCl₂; 1,5 µl buforu (DreamTaq™ Green Buffer 10X, ThermoFisher) i 0,3 µl [5 U/µl], polimerazy DreamTaq (Thermo). Produkty amplifikacji były wizualizowane przy wykorzystaniu elektroforezy w 0,8-1,5% żelu agarozowym z bromkiem etydyny (0,5µg/ml, Merck).

Kolejnym etapem było przeprowadzenie PCR z pulami matrix płytkowymi (W), rzędownymi (R) i kolumnowymi (T) z superpul wyselekcjonowanych w pierwszym etapie. Produkty amplifikacji były wizualizowane przy wykorzystaniu elektroforezy w 0,8-1,5% żelu agarozowym z bromkiem etydyny (0,5µg/ml, Merck). Dzięki tej procedurze

identyfikowano klony, które przypuszczalnie zawierały sekwencje poszukiwanych genów (Tab. 2).

Tab. 2. Lista lokalizacji w bibliotece klonów BAC prawdopodobnie zawierających poszukiwane geny.

	Gen	SP	Płytką	Rząd	Kolumna	Klon
1	<i>Q</i>	18	1	K	5	S18p1K5
2	<i>GIF1</i>	12	7	K	11	S12p7K11
		30	5	H	22	S30p5H22
3	<i>BTR2</i>	30	2	A	6	S30p2A6
4	<i>VRN1</i>	16	7	H	5	S16p7H5
5	<i>VRN2</i>	19	5	N	17	S19p4N17

Ostatnim etapem przeszukiwania biblioteki BAC była weryfikacja obecności poszukiwanej sekwencji w wytypowanych klonach BAC. Przeprowadzono analizę PCR według schematu opisanego powyżej. Jako matrycę wykorzystano DNA poszczególnych klonów BAC (izolacja opisana w rozdziale 4.4 - „Izolacja DNA pojedynczych klonów BAC”), DNA genomowe linii wsobnej L318 (kontrola pozytywna) oraz wodę (kontrola negatywna).

4.13 Analiza bioinformatyczna zidentyfikowanych potencjalnych genów

Do wstępnego określenia sekwencji i struktury nowo zidentyfikowanych genów potencjalnie związanych z odpornością na choroby u żyta oraz związanych z udomowieniem użyto programu FGESH (Ustawienia: Monocots plants generic, Corn, Rice, Wheat, Barley- <http://www.softberry.com/>) (Salamov i Solovyev, 2000). Do analizy i porównywania sekwencji zastosowano oprogramowanie Sequencher Software ver. 5.4.1 (Applied Biosystems). Program Bioedit służył do wizualizacji sekwencji i ich edycji (<http://www.mbio.ncsu.edu/BioEdit/>).

Analizie bioinformatycznej poddano sekwencje klonów BAC, sekwencje genomowe wybranych genów z różnych gatunków oraz sekwencje białkowe przypuszczalnych genów związanych z odpornością na choroby oraz związanych z udomowieniem żyta, wygenerowane w programie FGESH.

4.14 Ilościowy RT-PCR

Materiał roślinny, udostępniony przez prof. dr hab. Monikę Rakoczy-Trojanowską, stanowiły części nadziemne dwóch linii wsobnych żyta L318 i D33, które zebrano po 8,

17, 24 i 48 h od inokulacji zarodnikami rdzy brunatnej *Puccinia recondita f. sp. secalis*. Rośliny stanowiące kontrole inokulowano wodą z dodatkiem Tween 20. RNA wyizolowano ze 100 mg utartej tkanki roślinnej stosując GeneMATRIX Universal RNA Purification Kit (version 1.2) (Eurx, Gdańsk, Polska). RNA rozpuszczano w 40 μ l wody wolnej od RNaz. Przy użyciu spektrofotometru NanoDrop 2000 określono koncentrację i czystość wyizolowanego RNA. Następnie syntetyzowano cDNA stosując High-Capacity cDNA Reverse Transcription kit (Thermo Fisher Scientific).

Reakcje qRT-PCR wykonano w aparacie LightCycler 96 (Roche, Polska, Application Software ver. 1.1), w 3 powtórzeniach biologicznych i 3 powtórzeniach technicznych. Reakcje przeprowadzono w objętości 20 μ l, a w skład mieszaniny reakcyjnej wchodziło 4 μ l cDNA (10 ng/ μ l), 1 μ l każdego ze starterów [5 mM], 4 μ l wody wolnej od Rnazy i 10 μ l FastStart Essential DNA Green Master (Roche). Reakcję przeprowadzono w następujących warunkach: 10 min wstępnej preinkubacji w 95°C; 35 cykli - 95°C przez 10 s, 55°C przez 10 s, 72°C przez 15 s; następnie 95°C przez 10 s, 65°C przez 10 s. i 97°C przez 1s.), Jako gen referencyjny wykorzystano gen aktyny *HvAct* (AY 145451) (Starter Fw: CCCCTTTGAACCCAAAAGCC, Starter Rv: GAAAGCACGGCCTGAATAGC) zgodnie z metodyką opracowaną przez Świącicką i in. (2020). Poziom ekspresji badanych genów został określony przy użyciu metody $\Delta\Delta C_t$, w której wyznacza się różnicę pomiędzy poziomem ekspresji badanego genu, a poziomem ekspresji genu referencyjnego (Livak i Schmittgen, 2001).

5. WYNIKI

5.1 Genotypowanie DNA pul klonów BAC metodą DArTseq i ustalenie adresów marker DArTseq-klon BAC

W wyniku genotypowania 16 superpul klonów BAC metodą DArTseq zidentyfikowano 221 122 potencjalnych adresów marker DArTseq-klon BAC. Z tych adresów wybrano 53 370 zawierających 41 438 żytnich markerów DArTseq (Tab. 3). Łącznie 16 superpul obejmuje około 43 000 klonów BAC, co stanowi 41% biblioteki BAC linii L318. Reprezentuje to 65% genomu żyta (~ 5,2 Gpz). Długość żytnich markerów DArTseq, które znalazły się w adresach, wynosiła pomiędzy 20 a 69 nukleotydów (średnio 56). Zostały one przyporządkowane do 19 695 klonów BAC. Na jeden klon BAC średnio przypadało 2,1 markera DArTseq (Tab. 3). Większość (78%) markerów DArTseq w zidentyfikowanych adresach występowało tylko raz.

Przy zastosowaniu ustawień TS1, przy których brano pod uwagę markery DArTseq z dowolną liczbą wystąpień, uzyskano 53 370 adresów marker DArTseq-klon BAC. Liczba adresów obejmujących markery, które wystąpiły tylko raz lub dwa razy (ustawienie TS2) lub tylko raz (ustawienie TS3) wynosiła odpowiednio 46 499 i 32 217 (Tab. 3).

Tab. 3. Podsumowanie wyników genotypowania pul klonów BAC metodą DArTseq uzyskanych przy różnych ustawieniach.

Kryteria wyboru markerów	Wszystkie typy markerów DArTseq	Markery DArTseq występujące raz lub dwa razy	Markery DArTseq występujące tylko raz
Ustawienia adresowania DArTseq-BAC	TS1	TS2	TS3
Liczba adresów DArTseq-BAC	53 370	46 499	32 217
Liczba markerów DArTseq w adresach	41 438	39 357	32 217
Liczba klonów BAC w adresach	19 695	18 482	14 845
Liczba markerów DArTseq/ klon BAC	2,1	2,1	2,2

W 40-50% klonów BAC zidentyfikowano jeden marker DArTseq, w zależności od ustawień, w 20% – dwa markery a w 10% – trzy. Klony BAC, do których przypisano więcej niż 1 marker, wykorzystano do weryfikacji prawidłowości adresowania.

5.2 Kotwiczenie klonów BAC do map genetycznych

Pozycja na mapie genetycznej wykorzystanej do kotwiczenia była dostępna dla 5 436 markerów DArTseq z adresów DArTseq-BAC. Umożliwiło to przyporządkowanie 4 525 klonów BAC (z całkowitej liczby 19 695 klonów występujących w 53 370 adresach) do 3

324 pozycji na mapie genetycznej, ze średnią odległością wynoszącą 0,47 cM (ustawienia AS1) (Tab. 4). Przy najbardziej restrykcyjnych ustawieniach (AS3), 2 783 klonów BAC przyporządkowano do 2 544 pozycji na mapie ze średnią odległością 0,62 cM.

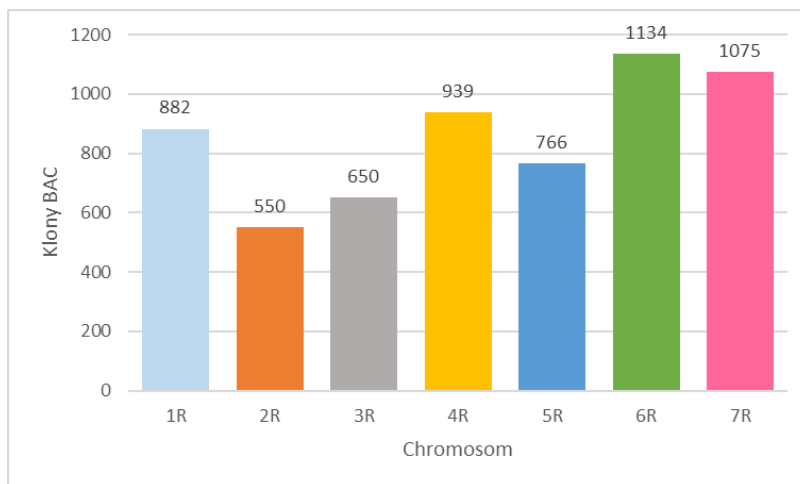
Tab. 4. Podsumowanie wyników kotwiczenia klonów BAC uzyskanych przy różnych ustawieniach.

Kryteria wyboru markerów	Wszystkie typy markerów DArTseq	Markery DArTseq występujące raz lub dwa razy	Markery DArTseq występujące tylko raz
Ustawienia kotwiczenia BAC poprzez DArTseq:	AS1	AS2	AS3
Liczba zakotwiczonych adresów DArTseq-BAC	6 847	6 208	4 044
Liczba kotwiczących markerów DArTseq	5 436	5 127	4 044
Liczba zakotwiczonych klonów BAC	4 525	4 022	2 783
Liczba klonów BAC zakotwiczonych przez przynajmniej dwa markery DArTseq	1 602	1 346	799
Liczba klonów BAC z konfliktami kotwiczenia	213	169	91
Procent klonów BAC z konfliktami kotwiczenia	13,3	12,6	11,4
Liczba klonów BAC z przynajmniej 2 markerami DArTseq z tego samego chromosomu	1 507	1 249	740
Liczba klonów BAC z odległością pomiędzy markerami DArTseq poniżej 5 cM	713	598	356
% klonów BAC z odległością pomiędzy markerami DArTseq poniżej 5 cM	47,3	47,9	48,1
Liczba klonów BAC z odległością pomiędzy markerami DArTseq powyżej 20 cM	336	282	159
% klonów BAC z odległością pomiędzy markerami DArTseq powyżej 20 cM	22,3	22,6	21,5

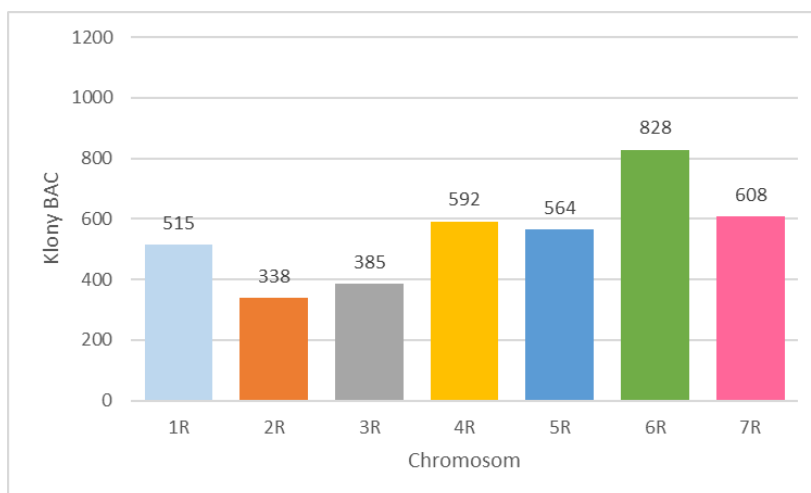
Liczba zakotwiczonych klonów BAC wahała się od 259 na chromosomie 2R do 528 na chromosomie 7R (Rys. 1 C) przy najbardziej surowych ustawieniach (AS3). Przy najłagodniejszych ustawieniach (AS1) ta liczba wyniosła 550 na chromosomie 2R i 1075 na chromosomie 6R (Rys. 1 A).

Rozmieszczenie zakotwiczonych klonów BAC wzdłuż mapy genetycznej żyta nie było równomierne (Rys. 2). Najwięcej klonów BAC znalazło się na chromosomach 6R i 7R, a najmniej na chromosomie 2R.

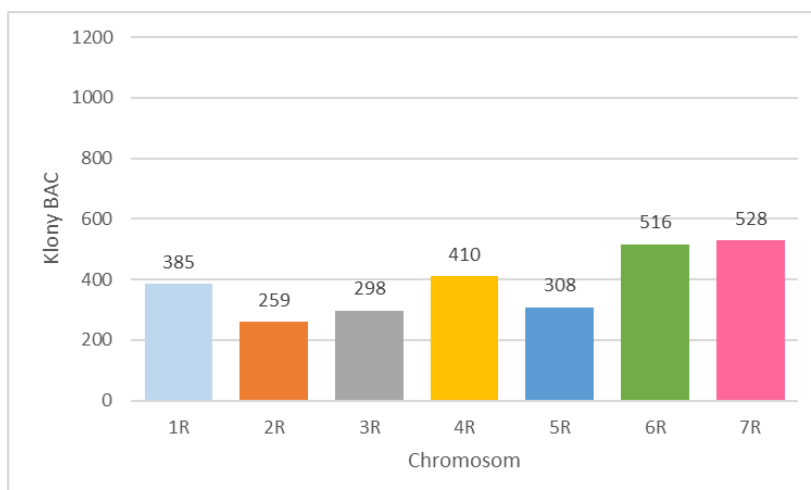
A



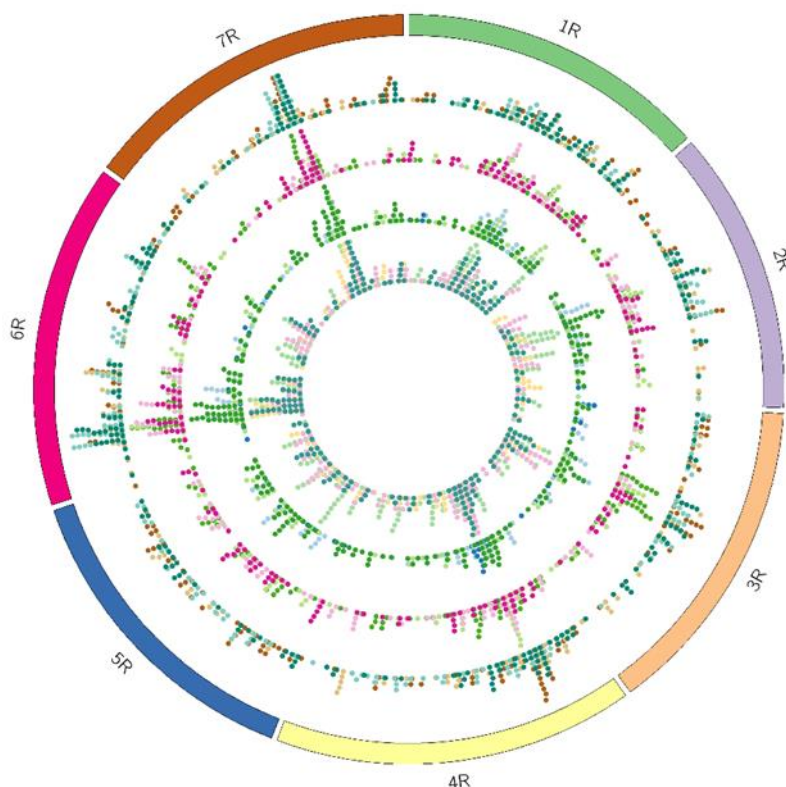
B



C



Rys. 1. Liczba klonów BAC zakotwiczonej na mapach genetycznych poszczególnych chromosomów według 3 różnych ustawień (A - AS1, B - AS2, C - AS3).



Rys. 2. Dystrybucja zakotwiczonych klonów BAC (według ustawień A1) na siedmiu chromosomach żyta. Paski na obwodzie koła reprezentują mapy genetyczne chromosomów, kropki przedstawiają klony BAC. Różne kolory reprezentują klony BAC z różnych superpul. Każdy wewnętrzny krąg przedstawia wyniki kotwiczenia dla 4 superpul genotypowanych równocześnie w jednym eksperymencie.

Wyniki genotypowania pul BAC metodą DArTseq i kotwiczenia klonów BAC do mapy genetycznej zostały opublikowane w pracy Borzęcka i in., 2018.

5.3 Weryfikacja adresów marker DArTseq-klon BAC

5.3.1 Weryfikacja sekwencyjna

Pierwsza metoda zastosowana do sprawdzenia wiarygodności adresowania markerów DArTseq do klonów BAC opierała się na wykorzystaniu sekwencji markerów DArTseq, sekwencji klonów BAC i algorytmu BLAST. Z wykorzystaniem platformy Illumina uzyskano sekwencje siedmiu klonów BAC, których długość wynosiła pomiędzy 94 983 a 145 571 pz (średnio 122 016 pz) (Tab. 5).

Tab. 5. Lista klonów BAC poddanych sekwencjonowaniu przy użyciu platformy Illumina.

L.p.	Klon BAC	L. skafoldów	Długości skafoldów	Łączna długość (pz)
1	S23p4E24	5	42 957 - 2 356	99 987
2	S31p7D16	7	60 235 - 2 383	119 159
3	S34p1J5	3	116 272 - 9 049	145 725
4	S34p2K21	3	103 517 - 6 593	123 881
5	S34p4M5	15	31 385 - 2 145	124 436
6	S34p7K24	6	108 323 - 2 146	132 611
7	S34p7P12	10	81 886 - 2 884	133 891

Zostały one zdeponowane w bazie danych GenBank (NCBI) z numerami akcesyjnymi MG646011-MG646017. Są to sekwencje składające się z wielu (od 5 do 15) skafoldów. Na podstawie wyników genotypowania oczekiwano w nich 91 markerów DArTseq przy najmniej restrykcyjnych ustawieniach (TS1). Istotnie statystycznie dopasowania do odpowiednich sekwencji w klonach BAC odnotowano dla 98,8% do 100% spodziewanych markerów DArTseq (w zależności od ustawień adresowania), ze średnią długością dopasowania wynoszącą ponad 99% i przy średniej identyczności sekwencji również ponad 99% (Tab. 6).

Tab. 6. Wyniki weryfikacji sekwencyjnej adresowania markerów DArTseq do klonów BAC.

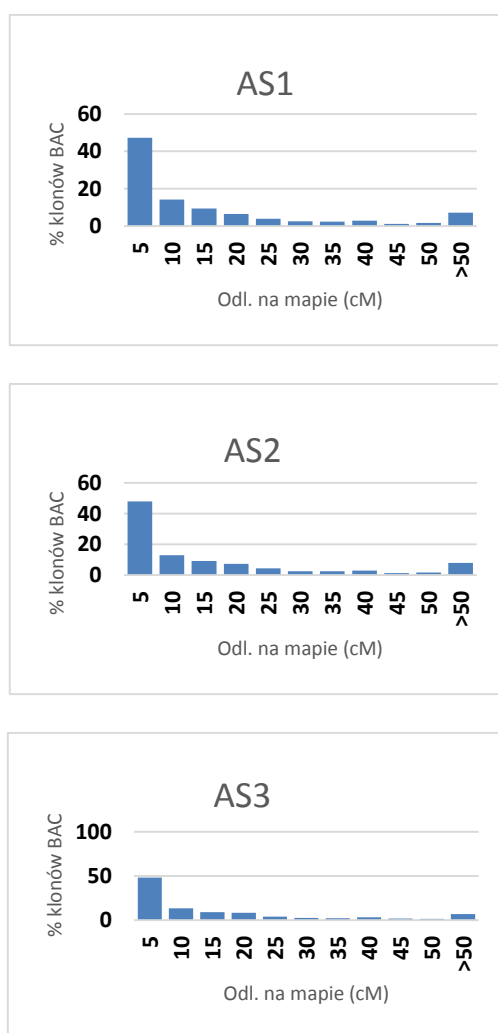
Klon BAC	Liczba przypisanych markerów DArTseq	Średnie podobieństwo ^a	Średnia długość porównania (%)
S23p4E24	14	100,00	100,00
S31p7D16	4	100,00	100,00
S34p1J5	26	99,85	99,88
S34p2K21	19	100,00	100,00
S34p4M5	15	93,33	93,33
S34p7K24	6	100,00	100,00
S34p7P12	7	100,00	100,00

a - średnie wartości dla wszystkich markerów DArTseq występujących w danym klonie BAC

5.3.2 Weryfikacja na podstawie pozycji na mapie

Druga metoda weryfikacji prawidłowości adresowania opierała się na użyciu informacji o położeniu markerów DArTseq na mapie genetycznej. Analizom tym poddano 1 602 klonów BAC, do których przypisano po przynajmniej dwa markery DArTseq o znanej pozycji na mapie genetycznej. W większości przypadków, gdy wiele markerów DArTseq zostało przyporządkowanych do tego samego klonu BAC, markery te były zlokalizowane blisko siebie na mapie genetycznej. Odsetek klonów BAC zakotwiczonych do mapy genetycznej przez przynajmniej 2 markery DArTseq z tego samego chromosomu,

które znajdowały się w odległości mniejszej niż 5cM od siebie wynosił 47%. Klony BAC zawierające markery DArTseq oddalone od siebie o więcej niż 20 cM na mapie genetycznej stanowiły 22%, niezależnie od zastosowanych kryteriów (Rys. 3). Odsetek klonów BAC, w których stwierdzono obecność markerów DArTseq pochodzących z różnych chromosomów wynosił od 11.4% do 13.3%.



Rys. 3. Weryfikacja wiarygodności adresowania wykorzystująca informację o położeniu markerów na mapie genetycznej: odsetek klonów BAC zawierających markery rozproszone na danym dystansie na mapach genetycznych, w zależności od ustawień (AS1-AS3).

5.4 Wdrożenie sekwencjonowania nowej generacji Oxford Nanopore

Do sekwencjonowania wybranych klonów BAC zastosowano technologię Nanopore z użyciem urządzenia MinION. Wykorzystano wysokiej jakości DNA indywidualnych klonów BAC w ilości 1 µg. Zsekwencjonowano 13 klonów BAC, w tym 6 klonów wytypowanych podczas prac związanych z poszukiwaniem genów związanych z odpornością na choroby (Tab. 7), 6 klonów wytypowanych podczas prac związanych z

poszukiwaniem genów związanych z udomowieniem (Tab. 7) oraz (pilotażowo) klon S34p7K24. Dodatkowo wyselekcjonowany klon - S34p4M5 był zsekwencjonowany wcześniej metodą Illumina, ale uzyskano dla niego najgorszą jakość złożenia spośród wszystkich zsekwencjonowanych klonów BAC - 15 skafoldów o łącznej długości 124 436 pz.

W wyniku sekwencjonowania techniką Nanopore klonów BAC uzyskano od 34 836 nukleotydów przy sekwencjonowaniu klonu S34p4M5 do 393 307 przy klonie S16p7H5. Łączna długość tych odczytów wynosiła od 281 270 031 do 3 637 831 244 nt. Do składania sekwencji klonów BAC wykorzystano tylko przefiltrowane odczyty. Odrzucono wszystkie odczyty krótsze niż 5000 nt (--min_length 5000) oraz 10% o najgorszej jakości (--keep_percent 90). Ostatecznie wykorzystano odczyty o najlepszej jakości, które dawały łącznie ~12Mpz (--target_bases 12000000).

Tab. 7. Wyniki sekwencjonowania klonów BAC techniką Nanopore.

Klon BAC	Wszystkie uzyskane odczyty			Odczyty użyte do złożenia sekwencji klonów BAC	
	Liczba odczytów	Łączna długość odczytów (nt)	Zakres długości odczytów	Wszystkie odczyty	Łączna długość odczytów (nt)
S3p5G8	451	12 001 193	10 049 - 81 174	44 512	348 268 367
S4p4E5	538	12 006 661	9 974 - 61 308	45 582	358 399 077
S21p7A1	524	12 008 452	10 162 - 73 886	44 520	334 250 459
S24p1H6	547	12 019 076	8 793 - 64 391	44 325	314 689 241
S34p4M5	462	12 009 377	8 714 - 69 590	34 836	281 270 031
S37p4L8	593	12 017 997	9 644 - 60 390	61 468	439 311 588
S30p2A6	382	12 016 021	16 386 - 72 317	171 353	1 504 862 706
S16p7H5	495	12 015 471	13 044 - 58 667	393 307	2 632 769 167
S30p5H22	383	12 002 891	13 910 - 79 149	241 868	1 650 555 853
S12p7K11	553	12 017 549	11 405 - 64 011	270 025	1 826 785 409
S19p4N17	769	12 005 148	9 533 - 39 097	138 142	784 770 507
S18p1K5	356	12 033 118	14 771 - 85 675	385 883	3 637 831 244

Dla każdego z klonów BAC sekwencjonowanych metoda Nanopore uzyskano pełną ciągłą sekwencję zawierającą wektor wraz z insertem.

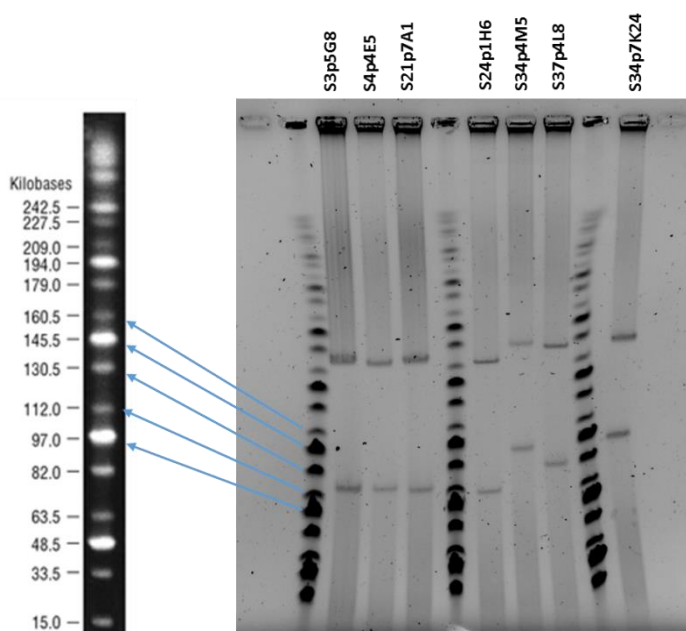
Długości insertów w wybranych do sekwencjonowania klonach BAC, ustalone z wykorzystaniem techniki elektroforezy pulsacyjnej, wynosiły od 100 do 138 kpz (Tab. 8, Rys. 4). Długości insertów ustalone na podstawie wyników sekwencjonowania Nanopore wahały się między 105 a 138 kpz.

Tab. 8. Lista zsekwencjonowanych klonów BAC wraz z ich długościami, ustalonymi na podstawie PFGE i odczytów z dwóch metod sekwencjonowania

Klon BAC	Długość PFGE (kpz)	Długość Illumina ([kpz]+l. skafoldów)	Długość Nanopore [kpz]
S3p5G8	105	-	105 865
S4p4E5	105	-	115 666
S21p7A1	105	-	112 573
S24p1H6	100	-	108 648
S34p4M5	132	124 436 (15)	138 081
S37p4L8	123	-	125 528
S34p7K24	138	132 611 (6)	141 447

Długości insertów w klonach BAC ustalone w wyniku sekwencjonowania Nanopore i przy wykorzystaniu elektroforezy pulsacyjnej były zbliżone.

Rys. 4. Określenie długości klonów BAC metodą PFGE. Po prawej stronie marker wielkości DNA Midrange PFG Marker (New England, BioLabs, źródło zdjęcia: www.international.neb.com; długości fragmentów podane w kpz)



W celu określenia dokładności sekwencjonowania uzyskanej na urządzeniu MinION porównano sekwencję wektora pIndigoBAC-5 (o długości 7 506 pz) z każdego zsekwencjonowanego metodą Nanopore klonu BAC z sekwencją zdeponowaną w bazie danych GenBank (nr akcesyjny: EU140754). Identyczność uzyskana po porównaniu tych sekwencji wynosiła średnio 99,26% (Tab. 9). Porównanie sekwencji wykazało obecność substytucji pojedynczych nukleotydów (SNP) oraz insercji/delecji (InDel) w sekwencjach uzyskanych na urządzeniu MinION metodą Nanopore. W poszczególnych klonach liczba

polimorfizmów typu SNP wynosiła od 6 do 19, a liczba polimorfizmów typu InDel - od 18 do 81 (Tab. 9). Najwięcej różnic wystąpiło w klonach S30p2A6 i S3p5G8.

Tab. 9. Wynik porównania sekwencji wektora pIndigoBAC-5 (EPICENTRE Biotechnologies) z poszczególnych klonów BAC uzyskanych metodą Nanopore z sekwencją wektora z bazy danych.

Klon BAC	Q ^a	I ^b	Liczba SNP	Typy SNP oraz frekwencja danego typu (%)	Liczba InDel
S3p5G8	99%	99,85%	19	A/G (21), T/C (5), A/T (5), G/A (16), C/G (11), CT/AC (42)	65
S4p4E5	100%	98,88%	13	A/G (8), T/C (8), A/T (15), G/A (8), C/G (8), TG/AC (8), AG/GT (8), CT/AC (5)	54
S21p7A1	100%	98,71%	14	A/G (1), T/C (8,5), A/T (2), G/A (2), G/C (2), TG/CT (7), AG/GT (36)	49
S24p1H6	100%	98,91%	10	A/G (10), T/C (40), A/T (10), G/A (10), AG/GT (20), CT/AC (10)	56
S34p4M5	100%	98,85%	10	A/G (10), T/C (10), A/T (10), G/A (10), C/G (10), G/T(20), AG/GT (10), CT/AC (10), TG/AC (10)	54
S37p4L8	100%	98,88%	9	A/G (11), T/C (22), A/T (11), G/A (11), G/C (11), AG/GT (11), CA/GT (11), CT/AC(11)	57
S30p2A6	100%	99,48%	9	A/G (11), T/C (22), A/T (11), G/A (11), AG/GT (22), CT/AC (22)	81
S16p7H5	100%	99,41%	12	A/G (17), T/C (17), A/T (8), G/A (8), G/C (8), AG/GT (17), CT/AC (17), CT/TC (8)	27
S30p5H22	100%	99,52%	7	A/G (29), T/C (14), A/T (14), G/A (14), AG/GT (14), CT/AC (14)	26
S12p7K11	100%	99,47%	10	A/G (10), T/C (30), A/T (10), G/A (10), CT/AC (20), AG/GT (20)	22
S19p4N17	100%	99,59%	6	A/G (16,6), T/C (16,6), A/T (16,6), G/A (16,6), CT/AC (16,6), AG/GT (16,6)	24
S18p1K5	100%	99,52%	11	A/G (9), T/C (18), G/C (9), A/T (9), G/A (9), AG/GT (36,4), CAG/TGT (9)	18

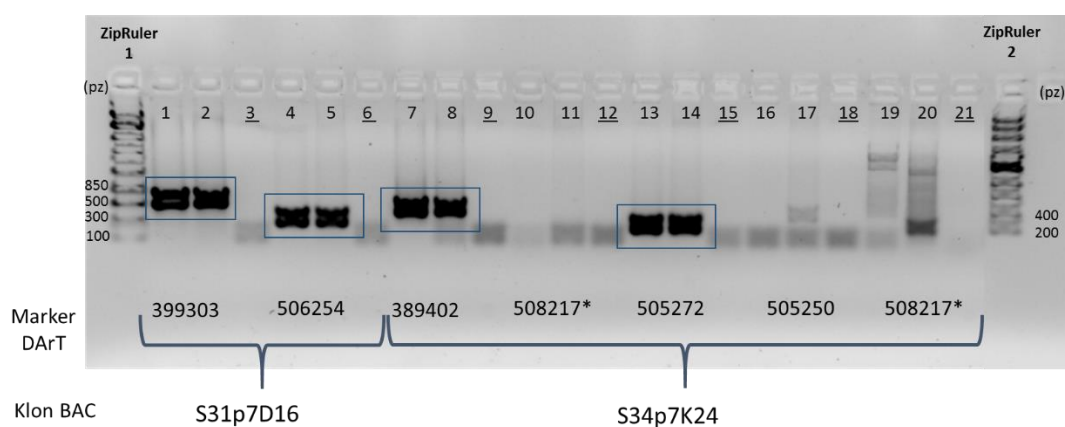
a - pokrycie [%], b - poziom identyczności [%]

Dwa z wymienionych klonów - S34p7K24 i S34p4M5, zostały zsekwencjonowane obiema wymienionymi metodami, aby porównać jakość uzyskanych sekwencji (Tab. 8). Dopasowanie sekwencji wskazało podobieństwo sięgające 99% przy pokryciu wynoszącym ponad 98%. Długość sumaryczna skafoldów klonu S34p7K24 przy sekwencjonowaniu Illumina wynosiła 132 611 kbpz a długość ciągłej sekwencji klonu uzyskanej z odczytów Nanopore - 141 447 kbpz. W przypadku klonu S34p4M5 długości te wynosiły odpowiednio: 124 436 i 138 081 kbpz (Tab. 8). Różnice w długości wynosiły 8 836 kbpz pomiędzy sekwencjami uzyskanymi metodą Illumina i Nanopore w klonie S34p7K24. Natomiast w klonie S34p4M5 różnica długości wynosiła 13 645 kbpz.

5.5 Identyfikacja potencjalnych genów związanych z odpornością na choroby u żyta

W badaniach przeanalizowano wszystkie (126) sekwencje markerów DArT, skojarzone na podstawie analiz bioinformatycznych z funkcjami związanymi z odpornością na choroby (ang.: *disease resistance genes*, DRG) (Gawroński i in, 2016). Na podstawie wyników genotypowania pul klonów BAC metodą DArT (Borzęcka i in., 2018) zidentyfikowano 11 klonów BAC przypuszczalnie zawierających przynajmniej dwa markery DArT powiązane z odpornością na choroby. W tych klonach BAC łącznie oczekiwano obecności 27 markerów DArT, w tym 21 o funkcjach związanych z odpornością. Do 20 z nich zaprojektowano startery (Tab. A1- Aneks).

W celu sprawdzenia wiarygodności wyników genotypowania i wyselekcjonowania najbardziej obiecujących klonów BAC do sekwencjonowania, zastosowano technikę PCR. Eksperymentalnie potwierdzono obecność 11 markerów DArT w 7 klonach BAC (Rys. 5, Tab. 12).



Rys. 5. Przykład wyniku elektroforezy produktów PCR uzyskanych przy pomocy starterów zaprojektowanych do powielenia fragmentów sekwencji markerów DArT. W każdym przypadku przeprowadzono trzy reakcje: w pierwszej matrycą był DNA BAC (ścieżki: 1, 4, 7, 10, 13, 16, 19), w drugiej matrycą był DNA linii wsobnej L318 (ścieżki: 2, 5, 8, 11, 14, 17, 20), w trzeciej nie dodano DNA (kontrola negatywna, ścieżki: 3, 6, 9, 12, 15, 18, 21). W pierwszej i ostatniej studziencie na żelu znajdują się drabinki wielkości DNA: ZipRuler 1 oraz ZipRuler 2 (SM1373, Thermo Scientific). Niebieskimi ramkami zaznaczone są markery DArT, których obecność udało się potwierdzić w poszczególnych klonach BAC.

Z uzyskanych produktów PCR 8 oczyszczono i wysłano do sekwencjonowania metodą Sanger przez firmę zewnętrzną. Następnie porównano algorytmem BLAST uzyskane sekwencje z sekwencjami markerów DArT zdeponowanymi w bazie danych

GenBank (Gawroński i in., 2016). Podobieństwo analizowanych fragmentów oscylowało między 72 a 99% (Tab. 10).

Tab. 10. Wyniki porównania algorytmem BLAST fragmentów sekwencji uzyskanych po sekwencjonowaniu produktów PCR z sekwencjami markerów DArT z bazy GenBank.

Marker DArT	Klon BAC ^a	Dł. prod PCR ^b (pz)	Q [%] ^c	E ^d	I [%] ^e
505520	S21p7A1	306	89	3,00E-38	72,62
506540	S24p1H6	330	95	2,00E-111	88,85
401485		193	89	4,00E-53	88,59
399682	S37p4L8	315	95	4,00E-88	84,21
505760		374	96	8,00E-174	98,82
389394	S3p5G8	412	96	4,00E-120	83,76
398573		212	68	4,00E-53	90,91
506219	S4p4E5	156	93	1,00E-22	76,00

a - klon BAC, w którym spodziewano się obecności markera DArT; b - długość produktu uzyskanego z wykorzystaniem starterów projektowanych na podstawie sekwencji markerów DArT z bazy GenBank; c - pokrycie; d - E-wartość; e - poziom identyczności

Do dalszych analiz (sekwencjonowania) wybrano 5 klonów BAC (S21p7A1, S24p1H6, S37p4L8, S3p5G8, S4p4E5), dla których udało się uzyskać metodą PCR potwierdzenie występowania w nich oczekiwanych klonów DArT. Dodatkowo wyselekcjonowano klon S34p4M5. Klony BAC poddano sekwencjonowaniu metodą Nanopore. Wyniki sekwencjonowania opisano w poprzednim rozdziale.

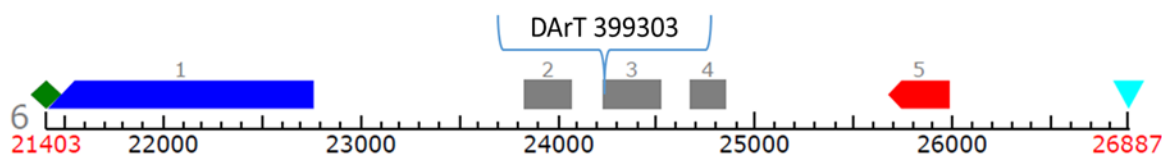
Uzyskane sekwencje klonów BAC poddano analizie bioinformatycznej. W tym etapie badań wykorzystano również uprzednio uzyskane sekwencje 2 klonów BAC (S31p7D16 i S34p7K24), zsekwencjonowane metodą Illumina, w których też na podstawie genotypowania oczekiwano obecności markerów DArT, w tym 8 przypuszczalnie związanych z odpornością na choroby (Tab. 12). Analiza bioinformatyczna algorytmem BLAST pozwoliła na potwierdzenie obecności w 7 wyżej wymienionych sekwencjach klonów BAC sekwencji wszystkich oczekiwanych 21 markerów DArT powiązanych z odpornością na choroby, wymienionych w tabeli 12. Podobieństwo sekwencji markerów DArT zdeponowanych w bazie GenBank, a uzyskanych po sekwencjonowaniu insertów z klonów macierzy genotypowania DArT, do sekwencji do nich homologicznych, znalezionych w sekwencjach klonów BAC wynosiło od 85 do 99%. Średnie pokrycie porównywanych fragmentów sięgało prawie 90% (Tab. 11).

Tabela 11. Wyniki porównania algorytmem BLAST sekwencji konów DArT zdeponowanych w bazie GenBank i sekwencji do nich homologicznych zidentyfikowanych w zsekwencjonowanych klonach BAC.

Marker DArT	Dł. sekw. DArT ^a	Klon BAC ^b	Q [%] ^c	E ^d	I [%] ^e
505520	659 pz	S21p7A1	93	5,00E-131	99,23
506540	636 pz	S24p1H6	87	3,00E-90	87,7
401485	575 pz		90	2,00E-69	98,01
399682	650 pz	S37p4L8	93	5,00E-139	98,92
505760	718 pz		96	1,00E-171	98,82
389394	524 pz	S3p5G8	97	0.0	98,47
398573	540 pz		65	2,00E-58	96,85
506219	310 pz	S4p4E5	94	2,00E-36	85,48

a - długość sekwencji markera DArT z bazy GenBank; b - klon BAC, w którym spodziewano się obecności markera DArT; c - pokrycie; d - E-wartość; e - poziom identyczności

Analiza programem SoftBerry/FGENESH (Salamov A., Solovyev V., 2000) sekwencji klonów BAC pozwoliła na zidentyfikowanie przypuszczalnej pełnej sekwencji i struktury obecnych w nich genów, w tym 28 genów potencjalnie związanych z odpornością na choroby (DRG) u żyta (Tab. 12, Rys. 6).



Rys. 6. Przykład wyniku przewidywania struktury potencjalnych genów związanych z odpornością zidentyfikowanych w sekwencjach klonów BAC: przypuszczalna struktura genu S31p7D16.6. Zielonym kolorem oznaczono fragment poliadenylowy (poliA), ciemnoniebieskim - sekwencję kodującą zakończoną kodonem STOP, szarym - wewnętrzne eksony, czerwonym - sekwencję kodującą rozpoczynającą się kodonem START, jasnoniebieskim - start transkrypcji (sekwencja TATA). Liczby pod osią oznaczają pozycje odnoszące się do całej sekwencji klonu BAC, w której wyszukano dany gen. Niebieską klamrą zaznaczona jest lokalizacja markera DArT 399303 w klonie.

Przy pomocy Fgenesh w sekwencjach klonów BAC zidentyfikowano łącznie 130 genów. Znalazionym genom nadawano numery kolejne w każdym z klonów. W poszczególnych klonach znaleziono od 2 do 8 genów, o długości od 1 192 do 17 272 pz, z czego 21,5 % stanowiły geny z możliwymi funkcjami związanymi z odpornością na choroby. Większość z nich (ponad 82%) należała do grupy genów NB-LRR (np. *RPM1*, *RPP13-like*, *RGAs*). Pozostałe geny (18%) wykazywały podobieństwo do kinazy *Lr10-like* występującej w *Aegilops tauschii* i *Oryza sativa* (Tab. 12).

Wszystkie przypuszczalne geny zawierające markery DArT wykazywały podobieństwo do genów z grupy NB-LRR.

Tab. 12. Podsumowanie wyników analizy programem SoftBerry/FGENESH sekwencji klonów BAC z listą genów przypuszczalnie związanych z odpornością na choroby.

Klon BAC	Marker DArT	L.genów ^a	L. DRG ^b	DRG ^c	Dł (pz) ^d	Pozycje genów ^e	blastp baza "ref_seq" ^f
S3p5G8	389394 398573 389569	16	3	S3p5G8.3 S3p5G8.6 S3p5G8.12	4682 10499 1192	9141..13823 28762..39261 82514..83706	<i>A. tauschii</i> RPP13-like protein 3, RPP8-like <i>A. tauschii</i> RPM1-like <i>A. tauschii</i> RPP13-like protein 3
S4p4E5	506219 505974 509502	13	3	S4p4E5.1 S4p4E5.2 S4p4E5.13	17272 12136 10620	165..17437 20666..32802 102105..112725	<i>B. distachyon</i> RGA2 isoform X1 <i>A. tauschii</i> RGA2-like isoform X2 i X1 <i>A. tauschii</i> RPP13-like protein 3, RGA2-like
S21p7A1	505520 506219	13	3	S21p7A1.1 S21p7A1.2 S21p7A1.4	6311 10867 13313	111..6422 8896..19763 27306..40619	<i>A. tauschii</i> RGA2-like <i>A. tauschii</i> RPP13-like protein 4, RGA4-like <i>A. tauschii</i> RPP13-like protein 4, RGA2-like
S24p1H6	506540 401485	17	2	S24p1H6.4 S24p1H6.16	2990 2660	18630..21620 92776..95436	<i>A. auschii</i> RPP13-like protein 3, <i>Triticum urartu</i> RPM1 <i>A. tauschii</i> RPP13-like protein 3
S34p4M5		21	8	S34p4M5.1 S34p4M5.3 S34p4M5.6 S34p4M5.11 S34p4M5.12 S34p4M5.13 S34p4M5.19 S34p4M5.20	4827 5280 1955 3581 3384 3103 4054 3314	2242..7069 10037..15317 42000..43955 80616..84197 84666..88050 90526..93629 128254..132308 132478..135792	<i>A. tauschii</i> rust resistance kinase Lr10-like isoform X2 <i>A. tauschii</i> RPM1-like <i>A. tauschii</i> RPM1-like <i>A. tauschii</i> rust resistance kinase Lr10-like <i>O. sativa</i> Japonica Group rust resistance kinase Lr10 <i>A. tauschii</i> RPM1-like <i>A. tauschii</i> , <i>O. sativa</i> rust resistance kinase Lr10-like <i>A. tauschii</i> rust resistance kinase Lr10-like isoform X2
S37p4L8	399682 505760 348128	15	2	S37p4L8.13 S37p4L8.15	4846 8540	93344..98190 116474..125014	<i>A. tauschii</i> RPM1-like <i>A. tauschii</i> RPP13-like protein 3
S34p7K24	505272 389402 508217 505250 507036 390286	19	5	S34p7K24.10 S34p7K24.11 S34p7K24.12 S34p7K24.13 S34p7K24.16	7383 4969 13291 5401 11193	39094..46477 46723..51692 55613..68904 73089..78490 101774..112967	<i>A. tauschii</i> RGA3 <i>A. tauschii</i> RPP13-like protein 4, protein ALP1-like <i>A. tauschii</i> RGA3, RPP13-like protein 4 <i>B. distachyon</i> receptor-like protein EIX2 <i>A. tauschii</i> RGA3, RPP13-like protein 4
S31p7D16	399303 506254	16	2	S31p7D16.5 S31p7D16.6	3093 2382	17796..19990 21180..22530	<i>A. tauschii</i> RGA4 <i>B. distachyon</i> resistance protein At3g14460

a - liczba genów według programu Fgenesh; b - liczba genów DRG; c - nazwy genów DRG; d - długości genów; e - pozycje (początek i koniec) genów w odpowiednich klonach BAC; f - wynik przeszukiwania sekwencji genów (wskazanych przez program Fgenesh) w bazie danych RefSeq z NCBI.

Pogrubioną czcionką zaznaczone są markery DArT, których obecność w klonach BAC potwierdzono eksperymentalnie poprzez PCR.

Wyniki analizy funkcjonalnej przypuszczalnych genów odporności z klonów BAC pokrywają się z wynikami analizy funkcjonalnej sekwencji klonów DArT (Tab. 13).

Tabela 13. Zestawienie wyników analizy funkcjonalnej sekwencji markerów DArT oraz analizy funkcjonalnej przypuszczalnych genów, w których te markery są zlokalizowane, zidentyfikowanych bioinformatycznie w sekwencjach klonów BAC.

Klon BAC	Marker DArT	Wynik analizy funkcjonalnej fragmentu sekwencji przypuszczalnego genu, uzyskanej z klonu biblioteki BAC	Wynik analizy funkcjonalnej sekwencji klonu DArT (Gawroński i in., 2016)
S21p7A1	505520	<i>A. tauschii</i> RGA2-like	disease resistance rpp13-like protein 1
	506219	<i>A. tauschii</i> RGA3	disease resistance rpp13-like protein 1
S24p1H6	506540	<i>A. tauschii</i> RPP13-like protein 3	disease resistance protein rpm1
	401485	<i>A. tauschii</i> RPP13-like protein 3	disease resistance protein rpm1
S37p4L8	399682	<i>A. tauschii</i> RPP13-like protein 4	disease resistance protein rpm1
	505760	<i>A. tauschii</i> RPM1-like	disease resistance protein rpm1
	348128	<i>A. tauschii</i> RPM1-like	disease resistance protein rpm1
S3p5G8	389394	<i>A. tauschii</i> RPP13-like protein 3, RPP8-like	disease resistance protein rpm1
	398573	<i>A. tauschii</i> RPM1-like	disease resistance protein rpm1
	389569	<i>A. tauschii</i> RPM1-like	disease resistance protein rpm1
S4p4E5	506219	<i>A. tauschii</i> RPP13-like protein 4	disease resistance rpp13-like protein 1
	505974	<i>A. tauschii</i> RGA4-like	disease resistance RGA4 isoform X1
	509502	<i>A. tauschii</i> RPP13-like protein 4	disease resistance rpp13-like protein 1
S31p7D16	399303	<i>A. tauschii</i> RGA2-like	RGA2-like
	506254	<i>B. distachyon</i> putative disease resistance protein At3g14461	Disease resistance protein RGA4
S34p7K24	505272	RGA2-like	Disease resistance protein RGA2
	389402	<i>A. tauschii</i> RGA3	disease resistance protein rga4
	508217	disease resistance rpp13-like protein 1	disease resistance protein rga4
	505250	<i>A. tauschii</i> RGA3	disease resistance protein rga3
	507036	<i>A. tauschii</i> RGA2-like	disease resistance protein rga2
	390286	<i>A. tauschii</i> RGA3	disease resistance protein rga2

Pogrubiona czcionka oznacza markery DArT których obecność w danym klonie BAC potwierdzono eksperymentalnie za pomocą PCR.

Analiza BLAST przy wykorzystaniu bazy danych RefSeq z NCBI wskazała, że geny występujące w klonie S34p4M5 (Tab. 12) są podobne do genów, które mogą brać udział w odpowiedzi na zakażenie rośliny patogenem powodującym rdzę brunatną. Do bardziej szczegółowej analizy został wybrany gen S34p4M5.1. Spośród wytypowanych genów (S34p4M5.11, S34p4M5.12 S34p4M5.19 i S34p4M5.20) wykazywał on najwyższe podobieństwo do genów związanych z odpornością na rdzę brunatną. W wyniku przeszukania nukleotydowej bazy danych, zaobserwowano najwyższe podobieństwo sekwencji genu S34p4M5.1, do sekwencji klonu biblioteki kosmidowej genomu pszenicy

Triticum aestivum (nr. akcesyjny: AF325197.1), zawierającego geny *Lrk 33*, *Lrk19* oraz *Tak33*. Podobieństwo to wyniosło ponad 93%, ze średnią długością dopasowania porównywanych fragmentów przewyższającą 80% (Tab. 14). Bliższa analiza wykazała najwyższe podobieństwo S34p4M5.1 do sekwencji genu *Tak33*. Z tego względu przyjęto dla niego nazwę *Sck33*.

Na podstawie fragmentu sekwencji klonu S34p4M5 - przypuszczalnego genu *Sck33*- zaprojektowano dwie pary starterów (2Fw/Rv: CAAACATGGGTTTCGTCAAGC/TCTTTGGCATGCTCAACCTT i 5Fw/Rv: TGGGCGAGGTCATAGAGATT/TTCTTTGGCATGCTCAACCT) i wykonano PCR. Amplifikacja przy wykorzystaniu starterów 2Fw/Rv zaskutkowało uzyskaniem 2 produktów, z tego względu nie wykorzystano tej pary starterów w dalszych analizach. Przy zastosowaniu pary starterów 5Fw/Rv uzyskano jeden produkt o oczekiwanej długości (Rys. 7). Starter Forward (5Fw) był umiejscowiony na początku egzonu 4, a starter Reverse (5Rv) w intronie poprzedzającym egzon 5.



Rys. 7. Wynik elektroforezy produktów PCR uzyskanych przy pomocy 2 par starterów (2Fw/Rv i 5Fw/Rv) zaprojektowanych do powielenia fragmentu sekwencji genu *Sck33*. W pierwszej studzience na żelu znajduje się drabinka wielkości DNA: 100bp DNA Ladder (100bp Plus DNA Ladder, Thermo Scientific). Czerwoną kółkiem zaznaczony jest uzyskany produkt o oczekiwanej długości.

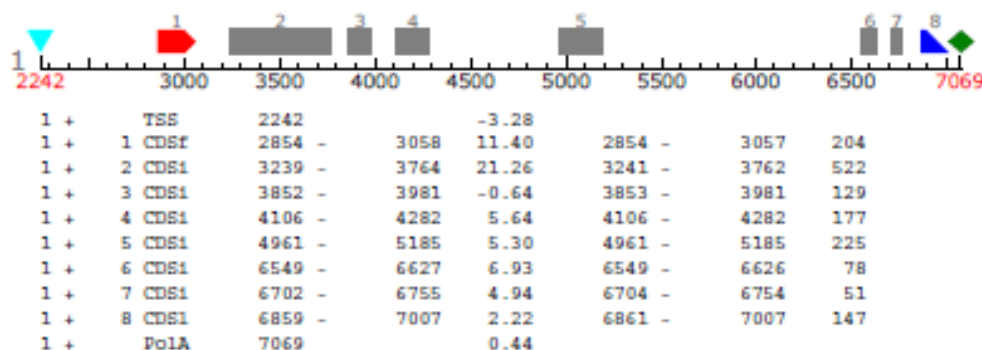
Sekwencję uzyskanego amplikonu o długości 711 pz obejmującego fragment genu *Sck33* z klonu S34p4M5 poddano analizie BLAST (Tab. 14). Analiza wykazała najwyższe podobieństwo (95.5%) genu *Sck33* do sekwencji genu *Tak33*. Na trzecim miejscu pojawiał się tak jak poprzednio, gen *Lr10-like* z dopasowaniem 94.87% ale pokrycie fragmentu wynosiło jedynie 48%. Gen *Sck33* wykazywał wysokie podobieństwo do tej sekwencji genu *Lr10-like* w rejonie od 1 101 do 2 408 pz co odpowiada sekwencji kodującej.

Tab. 14. Trzy najlepsze wyniki porównania algorytmem BLAST sekwencji przypuszczalnego genu *Sck33* znalezione w klonie BAC S34p4M5 z nukleotydową bazą danych z NCBI oraz najlepszy wynik porównania sekwencji fragmentu genu *Sck33* uzyskanej po sekwencjonowaniu produktów PCR.

Klon BAC	Gen	Podobieństwo do sekwencji w bazie NCBI	Q ^a	E ^b	I ^c
S34p4M5	S34p4M5.1	<i>T. aestivum Lrk33</i> and <i>Tak33</i> genes, complete cds (AF325197.1)	81	0.0	93.43
		<i>T. aestivum Pst19, Lrr19, Tak19-1,</i> and <i>Lrk19</i> gene(AF325196.1)	80	0.0	93.35
		PREDICTED: <i>A. tauschii</i> subsp. <i>tauschii</i> LEAF RUST 10 DISEASE-RESISTANCE LOCUS (XM_020324814.1)	30	0.0	93.65
	amplikon	<i>T. aestivum Lrk33</i> and <i>Tak33</i> genes, complete cds (AF325197.1)	95	0.0	95.51

a - pokrycie [%]; b - E-wartość; c - poziom identyczności [%]

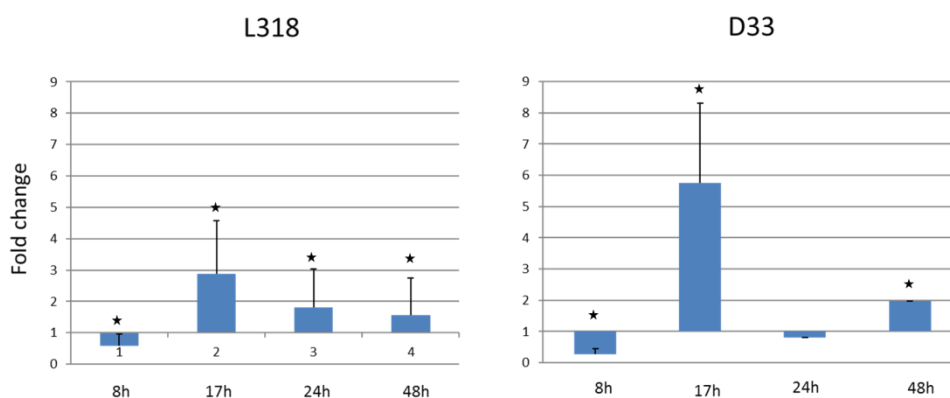
Na podstawie analizy bioinformatycznej, uwzględniającej predykcje algorytmów programu Fgenesh, przyjęto, że gen *Sck33* składa się z ośmiu egzonów (Rys. 8), a jego długość całkowita to 4 908 pz. Gen *Tak 33* ma długość 2 352 pz i składa się z 3 egzonów (Feuillet i in., 2001).



Rys. 8. Wynik przewidywania struktury genu *Sck33* według Fgenesh. Zielonym kolorem oznaczono PolA, ciemnoniebieskim - sekwencję kodującą zakończoną kodonem STOP, szarym - wewnętrzne eksony, czerwonym - sekwencję kodującą rozpoczynającą się kodonem START, jasnoniebieskim - start transkrypcji (sekwencja TATA).

Porównano również sekwencje genu *Sck33* do sekwencji genomu linii Lo7 żyta (Rabanus-Wallace i in., 2021). Został on przyporządkowany do chromosomu 1R z podobieństwem sięgającym 91,4% przy pokryciu wynoszącym 85%. Jego pozycja w genomie nie odpowiada żadnemu genowi opisanemu w dostępnej adnotacji (Rabanus-Wallace i in., 2019).

Zbadano poziom ekspresji genu *Sck33* w dwóch liniach wsobnych żyta po inokulacji patogenem *Puccinia recondita* f. sp. *secalis* w 4 punktach czasowych (Rys. 9). Amplifikowano fragment pierwszego egzonu genu *Sck33*, obejmujący około 100 pz (startery: Fw- GCCTTCGAGGCCGTTTTTCT, Rv- ACCACACTCAATTCCTCCTCTG). Linia wsobna żyta L318 oraz linia D33 wykazują różny stopień odporności na rdzę bruntaną powodowaną infekcją wspomnianym patogenem (Święcicka i in., 2020). Profile ekspresji tego genu w obu liniach są podobne. W stosunku do roślin kontrolnych, obserwowano statystycznie istotny spadek ekspresji u roślin porażonych grzybem w 8h po inokulacji, a następnie statystycznie istotny wzrost ekspresji po 17h od inokulacji. Może to sugerować, że ten gen jest zaangażowany w odpowiedź na patogen jednak potwierdzenie tej hipotezy wymaga dalszych badań.



Rys. 9. Profil ekspresji genu *Sck33* w częściach nadziemnych w dwóch liniach wsobnych żyta: L318 i D33 po 8, 17, 24 i 48 h od inokulacji *Puccinia recondita* f. sp. *secalis*. * zmiany istotne statystycznie (program REST, $P < 0,05$).

5.6 Identyfikacja genów kandydujących przypuszczalnie związanych z udomowieniem u żyta

Do identyfikacji genów potencjalnie związanych z procesem udomowienia u żyta wykorzystano informację o genach związanych z udomowieniem u innych, blisko spokrewnionych gatunków, takich jak jęczmień, pszenica i ryż. Na podstawie danych literaturowych wytypowano 13 genów o długościach od 486 do 4 496 pz (Tab. 15). Na podstawie sekwencji tych genów z innych gatunków zaprojektowano startery (Tab. A2-aneks) Jedynym uzyskanym przy pomocy tych starterów produktem był fragment genu *VRN2*.

Tab. 15. Lista 13 genów związanych z udomowieniem wykorzystanych do projektowania starterów, wraz z ich numerami akcesyjnymi z bazyGenBank oraz funkcjami.

	Gen	Genebank ID	Funkcja	Źródło	Dł. genu (pz)
1	<i>qSH-1</i>	AB071333.1	opadanie nasion	Konishi i in., 2005	4 496
2	<i>Q</i>	KX580304.2	opadanie nasion	Simons i in., 2006	3 729
3	<i>sh4</i>	GU220964.1	opadanie nasion	Li i in., 2006	3 553
4	<i>GIF1</i>	GU797937.1	jakość ziarna	Wang i in., 2008	1 797
5	<i>BTR1</i>	LT727746.1	niełamliwość osadki	Pourkheirandish i in., 2015	987
6	<i>BTR2</i>	LT727769.1	niełamliwość osadki	Pourkheirandish i in., 2015	724
7	<i>Rht-B1</i>	JX993615.1	półkarłowatość	Peng i in., 1999	4 138
8	<i>Rht-D1</i>	HE585643.1	półkarłowatość	Peng i in., 1999	2 253
9	<i>VRN1</i>	ACI24357.2	wernalizacja	Yan i in., 2003	732
10	<i>VRN2</i>	AAS60237.1	wernalizacja	Yan i in., 2003	642
11	<i>GA20ox2</i>	KX611234.1	biosynteza giberelin	Xu i in., 2017	486
12	<i>qSW5</i>	HQ658633.1	szerokość ziarna	Shomura i in., 2008	1 021
13	<i>Ghd7</i>	EU286801.1	represor kwitnienia	Xue i in., 2008	2 720

Po pojawieniu się sekwencji genomu referencyjnego żyta linii wsobnej Lo7 (Rabanus-Wallace i in., 2019) i jego przeszukaniu algorytmem BLAST w poszukiwaniu sekwencji homologicznych do wyżej wymienionych 13 genów, zidentyfikowano 11 genów przypuszczalnie związanych z udomowieniem (Tab. 16). Nie znaleziono sekwencji homologicznych do genów *qSW5* i *Ghd7*.

Podobieństwo sekwencji genów związanych z udomowieniem z blisko spokrewnionych gatunków z homologicznymi sekwencjami w genomie linii wsobnej Lo7 wynosiło średnio 87% (od 77 do 98%). Przypuszczalne homologe genów udomowienia zostały zlokalizowane na chromosomach 2R, 3R, 4R i 5R. Pozycje 5 z tych 11 genów pokrywały się z pozycjami genów w dostępnej adnotacji (Rabanus-Wallace i in. 2019). Tak było w przypadku przypuszczalnych homologów genów: *qSH-1*, *sh4*, *BTR1*, *VRN1* oraz *GA20ox2* (Tab. 16). Trzem z tych genów (wykazującym homologię do genów *sh4*, *BTR1* oraz *GA20ox2*) w adnotacji przypisano funkcje odpowiadające literaturowym genom udomowienia. Znaleziona w genomie Lo7 sekwencja homologiczna do genu *qSH1* stanowi część przypuszczalnego genu, który w adnotacji funkcjonalnej został określony jako kodujący białko zawierające homeobox typu BEL1 (Yoon i in., 2014). Z kolei sekwencja homologiczna dla genu *VRN1* zidentyfikowana w genomie Lo7 w adnotacji opisana została jako czynnik MADS box, który należy do jednej z największej grupy czynników

transkrypcyjnych. Są one kluczowymi regulatorami wielu procesów rozwojowych roślin (Yan i in., 2003).

Na podstawie zidentyfikowanych 11 sekwencji homologicznych w genomie Lo7 (Tab. 16) zaprojektowano startery do amplifikacji fragmentów tych genów z genomu żyta. Startery do amplifikacji genu *VRN1* uzyskano z publikacji Li i in., 2011. Dla *VRN2* wykorzystano startery wspomniane powyżej, zaprojektowane na podstawie sekwencji genów ze spokrewnionych gatunków (Tab. A2 - aneks). Ostatecznie uzyskano 6 produktów o oczekiwanej długości (Tab. 17).

Tab. 16. Informacja o położeniu w genomie referencyjnym żyta sekwencji homologicznych do znanych genów związanych z udomowieniem wraz z wynikiem PCR w linii wsobnej L318 oraz długością otrzymanego produktu.

	Gen	Lok.^a	I [%]^b	Dł. fragm. (pz)^c	E [%]^d	Pozycja start^e	Pozycja end^f	Adnotacja start^g	Adnotacja end^h	Opisⁱ	Dł. prod.^j
1	<i>qSH-1</i>	chr3R	77,62	706	5.51e-96	831702835	831703502	831699250	831703595	BEL1-like	-
2	<i>Q</i>	chr5R	89,2	2732	0.0	797898391	797901044	-	-	-	368
3	<i>sh4</i>	chr2R	80,68	735	5.23e-145	907651838	907651115	907650952	907653008	Sh4	-
4	<i>GIF1</i>	chr2R	80,27	365	2.88e-80	570570903	570571251	-	-	-	127
5	<i>BTR1</i>	chr3R	82,07	619	1.85e-139	104436985	104436380	104436397	104436993	BTR1-like	-
						104708192	104708797	104708184	104708780	BTR1-like	-
6	<i>BTR2</i>	chr3R	82,79	651	1.69e-158	104435353	104436004	-	-	-	372
7	<i>Rht-B1</i>	chr4R	90,19	2753	0.0	28385221	28387901	-	-	-	600
8	<i>Rht-D1</i>	chr4R	92,55	2188	0.0	28385466	28387594	-	-	-	-
9	<i>VRN1</i>	chr5R	98,39	185	3.15e-86	715823746	715823930	715810716	715823930	MADS box factor	158
10	<i>VRN2</i>	chr5R	89,56	316	5.75e-108	861612995	861613308	-	-		156
11	<i>GA20ox2</i>	chr3R	93,02	487	0.0	891310867	891311349	891308630	891311260	Gibberellin 20 oxidase	-

a -lokalizacja chromosomalna w genomie Lo7; b - poziom identyczności; c - długość znalezionej sekwencji homologicznej w genomie żyta; d - E-wartość; e,f - pozycje w genomie żyta; g,h - pozycje genów według adnotacji; i - opis genów według dostępnej adnotacji; j - długość otrzymanego produktu PCR, „-” brak produktu

Sekwencje zamplifikowanych fragmentów genów uzyskanych po sekwencjonowaniu produktów PCR wprowadzone do bazy danych NCBI wykazały najwyższe podobieństwo do genów udowodnienia ze spokrewnionych gatunków, które zostały wykorzystane jako referencyjne w tej pracy. Wynosiło ono od 82 do 100 % (Tab. 17).

Tab. 17. Wyniki porównania sekwencji 6 fragmentów genów uzyskanych po sekwencjonowaniu produktów PCR do sekwencji zawartych w bazie nukleotydowej NCBI.

	Gen ^a	Dł. prod. ^b	Sekwencja wykazująca najwyższe podobieństwo	Q ^c	E ^d	I [%] ^e
1	<i>Q</i>	368	<i>T. spelta</i> floral homeotic protein (Q), MK450626.1	97	1E-179	93,48
2	<i>GIF1</i>	127	<i>Oryza sativa</i> Japonica isolate 27762 GIF1 gene, GU797949.1	66	2E-15	89,80
3	<i>BTR2</i>	372	<i>H. vulgare</i> <i>BTR2</i> , LT727748.1	99	7E-106	81,79
4	<i>VRN1</i>	158	<i>S. cereale</i> MADS-box protein vrn1, EU525891.1	100	2E-72	100
5	<i>VRN2</i>	156	<i>A tauschii</i> cultivar AS75 ZCCT2 (<i>VRN2</i>) gene, KM503043.1	85	3E-75	93,97
6	<i>Rht-B1</i>	600	<i>T. aestivum</i> Rht-B1a_10 DELLA protein, JX993617.1	97	0.0	94,86

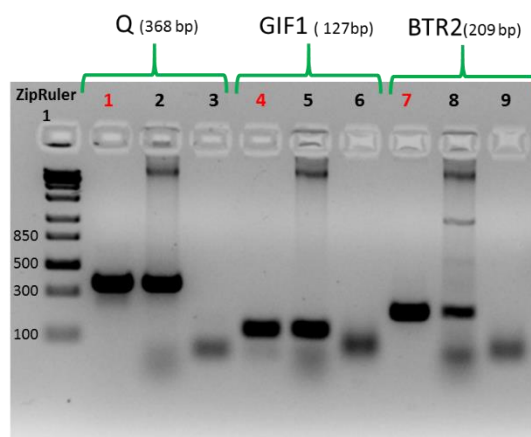
a -fragment genu; b - długość oczekiwanego i uzyskanego produktu PCR; c - pokrycie [%]; d - E-wartość; e - poziom identyczności [%]; d - wynik PCR z DNA linii L318

Następnie startery do amplifikacji fragmentów przypuszczalnych genów udowodnienia z genomu żyta wykorzystano do przeszukiwania biblioteki BAC linii wsobnej L318. Odnaleziono fragmenty 5 z tych genów w 6 klonach BAC (Tab. 18, Rys. 10). Nie udało się odnaleźć klonu zawierającego fragment genu *Rht-B1*. Klony, w których zidentyfikowano poszukiwane geny zsekwencjonowano metodą Nanopore. Porównanie sekwencji fragmentów tych genów uzyskanych poprzez PCR z odpowiadającymi im rejonami w klonach BAC wykazało podobieństwo sięgające od 91 do 100 % (Tab. 18).

Tab. 18. Wyniki poszukiwania klonów BAC zawierających geny udowodnienia i porównania algorytmem BLAST sekwencji amplikonów uzyskanych z DNA genomowego linii L318 z homologicznym obszarem w zidentyfikowanych klonach BAC oraz sekwencje starterów wykorzystanych do amplifikacji.

	Poszukiwany Gen	Klon ^a	Q ^b	E ^c	I ^d	Fw ^e	Rv ^f
1	<i>Q</i>	S18p1K5	97	0.0	98,90	AAGAGGGACAACATCATCGC	ACCTGTACGTTCCGGGTAGAA
2	<i>GIF1</i>	S12p7K11	90	4,00E-39	97,09	GAACGTCTTGGAGGCGTAGA	CCGGTACGACTACTACACCAT
		S30p5H22	82-88	9,00E-31	91,76		
3	<i>BTR2</i>	S30p2A6	99	0	98,92	CCTGAGCATGACGTTGAAGAC	ACGAATGAGATAGCGGAGGC
4	<i>VRN1</i>	S16p7H5	100	4,00E-79	100	GGAGATTCGCACGTACGAT	ATGACTCGGTGGAGAACTCG
5	<i>VRN2</i>	S19p4N1	100	0	100	GGCGCTATGACAAGCAAATC	GAACCATCCGAGGTGAAGTT

a - wykryty klon; b - pokrycie [%]; c - E - wartość; d - poziom identyczności [%], e - starter forward, f - starter reverse



Rys. 10. Przykład wyniku elektroforezy produktów PCR uzyskanych z DNA klonów BAC wytypowanych po przeglądaniu biblioteki przy pomocy starterów zaprojektowanych do powielenia fragmentów przypuszczalnych genów udomowienia. W każdym przypadku przeprowadzono trzy reakcje dla każdego z genów oznaczonych ponad zdjęciem zielonymi klamrami: w pierwszej matrycą był DNA BAC (ścieżki: 1, 4, 7), w drugiej matrycą był DNA linii wsobnej L318 (ścieżki: 2, 5, 8), w trzeciej nie dodano DNA (kontrola negatywna, ścieżki: 3, 6, 9). W pierwszej studzience na żelu znajduje się drabinka wielkości DNA: ZipRuler 1 (SM1373, Thermo Scientific).

Porównano również sekwencje genów udomowienia z genomu żyta Lo7 do sekwencji do nich homologicznych występujących w klonach BAC zidentyfikowanych w wyniku przeglądania biblioteki BAC (L318) i wykazano wysokie podobieństwo wynoszące od 83,5 do 100% (Tab. 5). Podobieństwo genów udomowienia z blisko spokrewnionych gatunków (oznaczonych „*ref*”) z sekwencjami homologicznymi z klonów BAC (L318) wynosiło od 67,7% w przypadku genu *GIF1* (klon S12p7K11) do 98% dla genu *VRNI*. Poziom identyczności genów udomowienia z blisko spokrewnionych gatunków z odpowiadającymi im sekwencjami w genomie żyta Lo7 wynosił od 79% (*GIF1*) do 98% (*VRNI*) (Tab. 19).

Tab. 19. Wyniki porównania sekwencji 5 genów związanych z udomowieniem z sekwencjami homologicznymi z genomu linii wsobnej żyta Lo7 i z linii wsobnej L318 (ustalonymi na podstawie sekwencjonowania klonów BAC).

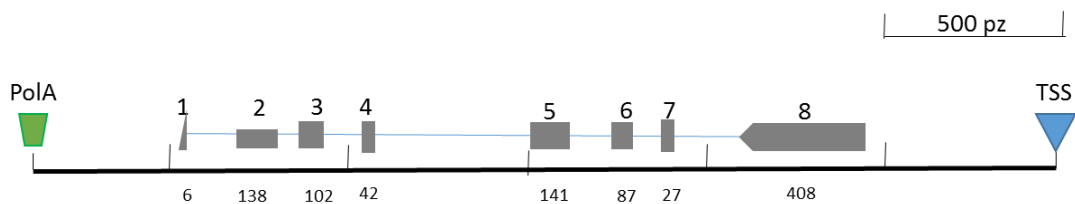
Porównywane sekwencje		Q ^a	E ^b	I ^c
<i>Q</i> _Lo7	<i>Q</i> _L318 (S18p1K5)	100	0.0	96.12
<i>Q</i> _ref	<i>Q</i> _L318 (S18p1K5)	91	0.0	89.60
<i>Q</i> _ref	<i>Q</i> _Lo7	72	0.0	88.82
<i>GIF1</i> _Lo7	<i>GIF1</i> _L318 (S12p7K11)	99	1,00E-91	83.53
<i>GIF1</i> _ref	<i>GIF1</i> _L318 (S12p7K11)	70	3,00E-81	67.78
<i>GIF1</i> _ref	<i>GIF1</i> _Lo7	100	3,00E-95	79.12
<i>GIF1</i> _Lo7	<i>GIF1</i> _L318 (S30p5H22)	99	0.0	100.00
<i>GIF1</i> _ref	<i>GIF1</i> _L318 (S30p5H22)	86	1,00E-119	72.14
<i>BTR2</i> _Lo7	<i>BTR2</i> _L318 (S30p2A6)	100	0.0	98.61
<i>BTR2</i> _ref	<i>BTR2</i> _L318 (S30p2A6)	88	9,00E-139	81.52
<i>BTR2</i> _ref	<i>BTR2</i> _Lo7	88	0.0	81.34
<i>VRN1</i> _Lo7	<i>VRN1</i> _L318 (S16p7H5)	100	1,00E-100	100.00
<i>VRN1</i> _ref	<i>VRN1</i> _L318 (S16p7H5)	99	8,00E-95	98.38
<i>VRN1</i> _ref	<i>VRN1</i> _Lo7	25	4,00E-92	98.38
<i>VRN2</i> _Lo7	<i>VRN2</i> _L318 (S19p4N17)	100	2,00E-173	99.68
<i>VRN2</i> _ref	<i>VRN2</i> _L318 (S19p4N17)	84	6,00E-114	90.20
<i>VRN2</i> _ref	<i>VRN2</i> _Lo7	48	4,00E-120	89.17

a - pokrycie [%]; b- E-wartość; c- poziom identyczności [%]

Analiza programem SoftBerry/FGENESH (Salamov A., Solovyev V., 2000) sekwencji 6 wyżej wspomnianych klonów BAC pozwoliła na zidentyfikowanie przypuszczalnej pełnej sekwencji i struktury obecnych w nich genów, w tym potencjalnie związanych z udomowieniem u żyta (Tab. A3 - aneks).

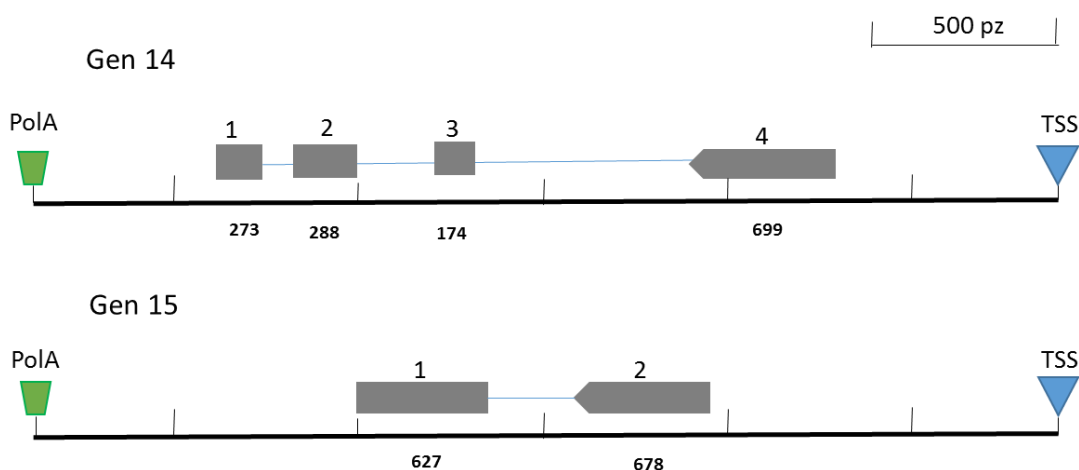
W poszczególnych klonach zidentyfikowano od 16 do 29 przypuszczalnych genów, łącznie 128 genów o długościach od 165 do 20 820 pz, wykazujących różne funkcje (Tab. A3 - aneks). Dokonano poszukiwania sekwencji homologicznych w nukleotydowej bazie danych „nr/nt” oraz białkowej „RefSeq” NCBI i podobieństwo dopasowanych sekwencji wynosiło od 41 do 100%.

W klonie S30p2A6 znaleziono przypuszczalny żytni homolog genu *BTR2*, w S16p7H5 genu *VRN1* oraz w klonie S18p1K5 genu *Q*. Według algorytmów programu Fgenesh przypuszczalny żytni homolog genu *Q* składa się z 8 egzonów (gen 15 w klonie S18p1K5) oraz 7 intronów o długościach: 176, 85, 95, 518, 134, 98 oraz 225 pz (Rys. 11). Według danych literaturowych gen *Q* pszenicy o długości 3229 pz ma 10 egzonów. Jego sekwencja kodująca ma długość 1344 pz (Simons i in., 2006) a jego żytniego ortologa - 951 pz. Oba te geny wykazywały poziom identyczności sekwencji transkryptu sięgający 93.73% przy pokryciu 92%. Przypuszczalna sekwencja białka wykazała podobieństwo do APETALA2-like protein.



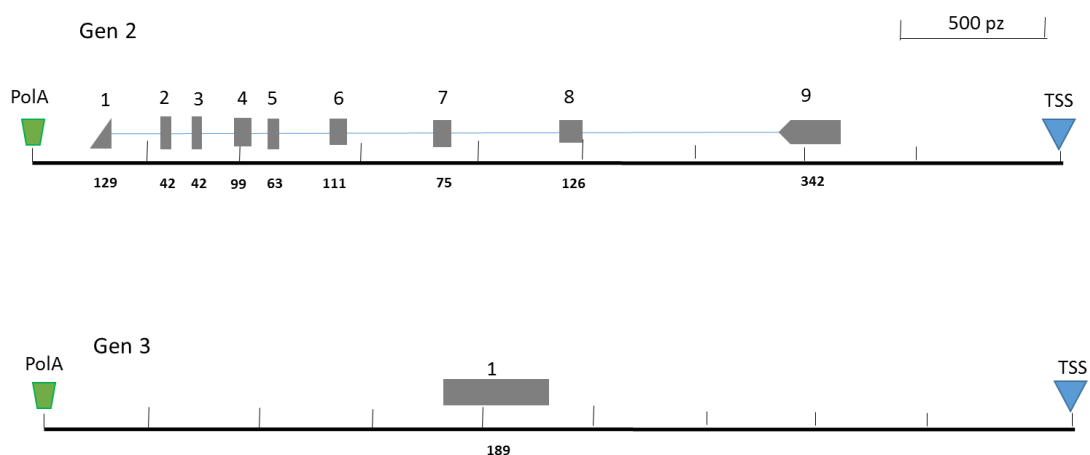
Rys. 11. Wynik predykcji bioinformatycznej struktury genu zlokalizowanego w klonie S18p1K5, wykazującego homologię do genu *Q*. Szare figury przedstawiają egzony a niebieskie linie między nimi introny.

W sekwencji klonu BAC S30p2A6 zidentyfikowano dwa rejony wykazujące podobieństwo do genu *BTR2*. Pierwszy z nich, o długości sekwencji kodującej 1 437 pz i z 2 egzonami, został oznaczony jako gen 14. Drugi rejon, oznaczony jako gen 15, oddalony od genu 14 o 181 pz, również mógłby być homologiem *BTR2* posiadającym 4 egzony (1 308 pz) (Rys.12). Gen *BTR2* jęczmienia składa się z jednego o długości 609 pz (Zeng i in., 2020). Gen zlokalizowany jest na chromosomie 3H, a u żyta najprawdopodobniej na trzecim chromosomie 3R. (Tab. 16). Przypuszczalne homologi genu *BTR2* (gen 14 i 15 zidentyfikowany w klonie S30p2A6) na odcinku o długości 563 pz wykazują podobieństwo między sobą sięgające 91%. W tym samym rejonie wykazują podobieństwo do genu *BTR2* jęczmienia, wynoszące 82%. Przypuszczalna sekwencja białka genu 14 przypominała w 61% BTR2-like-2 protein [*Triticum monococcum*], natomiast genu 15 wykazywała podobieństwo do BTR2-like-1 protein [*Triticum monococcum subsp. aegilopoides*].



Rys. 12. Wynik predykcji bioinformatycznej struktury genów zlokalizowanych w klonie S30p2A6, wykazujących homologię do genu *BTR2*. Szare figury przedstawiają egzony a niebieskie linie między nimi introny.

W sekwencji klonu BAC S16p7H5 znaleziono dwa rejony wykazujące podobieństwo do genu *VRNI*. Pierwszy z nich (gen 2) według predykcji programu Fgenesh może posiadać 9 egzonów, a przypuszczalna długość sekwencji kodującej to 1 038 pz. Drugi rejon homologiczny (gen 3) jest oddalony od pierwszego o 6 395 pz. Według predykcji Fgenesh posiada tylko 1 egzon o długości 189 pz (Rys. 13). Znajduje się na chromosomie 5R. Gen *VRNI* u pszenicy zlokalizowany jest na chromosomie 5A. Sekwencja zidentyfikowanego genu 2 porównana do sekwencji genu *VRNI* w pszenicy wykazuje poziom identyczności sięgający 96% na 42% swojej długości. W bazie danych nie ma nukleotydowej sekwencji żytniego *Vrn1* tylko częściowa sekwencja białkowa (ACD74577.1). Przypuszczalna sekwencja białka genu 2 wykazała podobieństwo do MADS-box protein *vrn-1* [*Secale cereale*] w 71%, natomiast genu 3 do *Vrn-A1a* [*Triticum aestivum*] w 100%.



Rys. 13. Wynik predykcji bioinformatycznej struktury genów zlokalizowanych w klonie S16p7H5, wykazujących homologię do genu *VRNI*. Szare figury przedstawiają egzony a niebieskie linie między nimi introny.

Ostatnim etapem było porównanie sekwencji 128 genów wytypowanych w programie Fgenesh z klonów BAC z sekwencją genomu żyta Lo7. Dopasowania w znaleziono dla 107 z tych przypuszczalnych genów (83%). Dla 65 z nich nie uzyskano dopasowań pokrywających przynajmniej 75% ich sekwencji.

6. DYSKUSJA

W tej pracy zaprezentowano wykorzystanie biblioteki BAC żyta do integracji map fizycznych i genetycznych. Po raz pierwszy do kotwiczenia klonów BAC na mapie genetycznej zostały wykorzystane markery typu GBS - DArTseq. Bibliotekę BAC wykorzystano również do identyfikacji genów związanych z udomowieniem i odpornością na choroby żyta. W momencie rozpoczęcia badań do niniejszej pracy zasoby genomiczne żyta były ograniczone. W trakcie realizacji pracy doktorskiej rozpoczął się projekt sekwencjonowania genomu żyta i w roku 2021 opublikowano sekwencję genomu referencyjną żyta linii Lo7, reprezentowaną przez skafoldy o łącznej długości 6,74 Gpz (Rabanus-Wallace i in., 2021). Nie zmienia to jednak faktu, że wyniki zaprezentowane w niniejszej dysertacji dostarczają nowych informacji dotyczących genomu żyta. W opisanych badaniach wykorzystano bibliotekę BAC linii wsobnej żyta L318, posiadającej inne pochodzenie niż linia Lo7. Dodatkowo, pierwsze wersje sekwencji genomowej, zwłaszcza u gatunków o dużych i złożonych genomach często posiadają różne niedoskonałości, które mogą ograniczać praktyczne wykorzystanie takiego zasobu (Jung i in., 2019; Giani i in., 2020). W pracy przedstawiono również implementację i optymalizację rozwiązań metodycznych, takich jak sekwencjonowanie pojedynczych klonów BAC metodą Nanopore - ułatwiających wykorzystanie bibliotek BAC w badaniach z zakresu genetyki i genomiki.

6.1 Genotypowanie DArTseq oraz kotwiczenie klonów BAC na mapie genetycznej żyta

Kotwiczenie pojedynczych klonów BAC i kontigów na mapach genetycznych o wysokiej rozdzielczości ma kluczowe znaczenie przy konstrukcji i eksploatacji map fizycznych, a także przy klonowaniu genów. W niniejszej pracy przetestowano skuteczność i efektywność metody przeszukiwania biblioteki BAC z wykorzystaniem technologii DArTseq.

Upowszechnienie wysokoprzepustowych technologii genotypowania przyspieszyło badania nad złożonym genomem żyta. Milczarski i in. (2016) jako pierwsi zaprezentowali wykorzystanie markerów DArTseq do zagęszczenia mapy genetycznej żyta. Całkowita długość mapy, na której zlokalizowano 19 081 markerów (w tym 18 132 DArTseq) wynosiła 1 593 cM. W porównaniu z wcześniej dostępnymi mapami, opartymi na

markerach DArT (Milczarski i in., 2011), gdzie udało się zlokalizować 4 048 markerów, był to duży postęp.

Ze względu na to, że markery DArTseq okazały się tak skuteczne w analizach genomu żyta, podjęto próbę wykorzystania ich do kotwiczenia klonów BAC na mapie genetycznej i tworzenia szkieletu mapy fizycznej. Dzięki genotypowaniu ok. 43 000 klonów BAC przeanalizowano ponad 65% genomu linii wsobnej żyta L318 (~5,2 Gpz), co pozwala przypuszczać, że uzyskane wyniki są miarodajne a niniejsza praca jest cennym wkładem w badania nad genomem żyta.

W tej pracy dla zwiększenia efektywności adresowania DArTseq-BAC zastosowano kilka strategii: wykorzystano trójwymiarową metodę konstrukcji pul zbiorczych, zawężono liczbę markerów DArTseq uwzględnionych w adresowaniu, spełniających określone kryteria oraz zadbano o równą reprezentację poszczególnych klonów BAC w pulach.

Aby zoptymalizować efektywność przeszukiwania bibliotek BAC w celu wykrycia poszukiwanych sekwencji zostały opracowane różne techniki tworzenia pul zbiorczych: sześciowymiarowa (Klein i in., 2000), pięciowymiarowa (Luo i in., 2009) oraz zastosowana w tej pracy - trójwymiarowa (3-D) (Liu i in., 2011). W porównaniu do innych technik pulowania, metoda 3-D ma wiele zalet: (a) pulowanie jest proste do wykonania, co zmniejsza w ten sposób ryzyko popełnienia błędów, (b) klony BAC z danej SP są reprezentowane tylko przez 23 próbki, co wpływa na wzrost przepustowości przeszukiwania, (c) każdy klon BAC z SP jest obecny w dwóch pulach matrix każdego typu, zapewniając dodatkową kontrolę wewnętrzną zarówno dla jakości przygotowania próbki, jak i genotypowania, (d) stosunkowo niewielka część genomu zawarta jest w jednej SP (ok. 1/25 genomu żyta w tym badaniu), co obniża prawdopodobieństwo, że dany region genomu o pojedynczej kopii będzie obecny w więcej niż jednym klonie BAC z danej SP i w efekcie skutkuje większą liczbą identyfikacji jednoznacznych adresów marker-klon BAC (Luo i in., 2009; You i in., 2010).

Wybór markerów DArTseq wykorzystywanych do adresowania polegał na ograniczeniu się do tych sekwencji markerowych, które zostały wcześniej odkryte podczas standardowego genotypowania całkowitego genomowego DNA żyta i zostały zdefiniowane jako markery DArTseq dla żyta. Taka strategia została zastosowana w celu wyeliminowania markerów pochodzących z powtarzających się rejonów genomu. W pracy podane są osobno dane dotyczące tylko tych markerów, które wystąpiły raz lub dwa razy

lub tylko raz, w wynikach przeszukiwania biblioteki. Wdrożenie rygorystycznej selekcji markerów z jednokopijnych regionów genomu miało na celu uzyskanie jeszcze bardziej wiarygodnych i jednoznacznych przyporządkowań DArTseq-klon BAC.

Kolejną strategią zastosowaną dla uzyskania jak największej liczby jednoznacznych adresów marker DArTseq-klon BAC było zapewnienie równej reprezentacji klonów BAC w pulach BAC i izolatach DNA BAC. Aby to osiągnąć wykorzystano różne rozwiązania. Zastosowano 384-dołkowe płytki podczas namnażania klonów BAC do tworzenia pul zbiorczych, zapewniające oddzielny wzrost każdego klonu z puli. Takie rozwiązanie pozwala uniknąć błędu systematycznego wynikającego z różnego tempa wzrostu poszczególnych klonów, który może pojawić się, jeżeli wiele klonów namnażanych jest równocześnie w jednym naczyniu z pożywką. Ważnym krokiem było przeprowadzenie izolacji DNA BAC zoptymalizowaną w naszym laboratorium metodą, co skutkowało uzyskaniem stosunkowo dużej ilości DNA BAC o wysokiej jakości. Pozwoliło to na zrezygnowanie z amplifikacji BAC DNA przed genotypowaniem - rozwiązania stosowanego czasem w projektach kotwiczenia klonów BAC na mapach genetycznych (Pinard i in., 2006). Wykorzystanie amplifikacji BAC DNA może powodować nierównomierność w reprezentacji sekwencji poszczególnych klonów BAC w próbie zbiorczej, ze względu na niekompletną amplifikację lub zanieczyszczenie (Pinard i in., 2006; Blainey i in., 2013). Zanieczyszczenia mogą pochodzić z trzech źródeł: (1) z samej próbki, (2) środowiska laboratoryjnego oraz (3) odczynników i przyrządów używanych do przygotowania próbki (Blainey i in., 2013). Często przy amplifikacji dochodzi do kontaminacji fragmentami pochodzącymi z genomowego DNA bakterii (Blainey i in., 2013). Błędy związane z PCR w produktach amplifikowanych ze złożonych mieszanin mogą być związane z zawartością zasad GC w sekwencji, etapem wydłużania produktu czy różnicami w energii wiązania starterów (Benita i in., 2003). Co więcej, nie wszystkie fragmenty genomu są podatne na amplifikację, powodując uzyskanie niekompletnych sekwencji (Jordan i in., 2002). Może również dojść do tworzenia homodupleksów podczas przejścia z etapu denaturacji do wydłużania w trakcie PCR (Kurata i in., 2004). Wszystkie te czynniki i błędy mogą skutkować wytworzeniem aż do 70% niespecyficznego artefaktów amplifikacji prowadząc do niepełnego pokrycia powielanych sekwencji (Cheung i in., 1996).

Zastosowany układ eksperymentalny zaowocował preferencyjną identyfikacją markerów jednokopijnych - 95% markerów zaangażowanych w adresy DArTseq-BAC

wystąpiło raz lub dwa razy w całym zbiorze danych, co jest zgodne z pokryciem genomu linii L318 poprzez bibliotekę BAC. Ponieważ całkowite pokrycie genomu przez użytą bibliotekę było ok. 1,5-krotne, można oczekiwać, że dany region genomu żyta mógł być reprezentowany co najwyżej dwukrotnie w analizowanych SP. Oznacza to, że markery genetyczne, które wystąpiły raz lub dwa razy w przeszukiwanej frakcji biblioteki, z dużym prawdopodobieństwem będą markerami jednokopijnymi - występującymi tylko w jednej kopii w genomie żyta.

Zaobserwowano, że dystrybucja chromosomalna zakotwiczonych klonów BAC nie była równomierna. Niezależnie od ustawień najwięcej klonów BAC zakotwiczono na chromosomach 6R i 7R, a najmniej na 2R. Nie wykryto korelacji między liczbą zakotwiczonych w nieniejszej pracy klonów BAC a fizyczną długością poszczególnych chromosomów, dlatego inne faktory związane z organizacją i polimorfizmem genomu wpływają na ten aspekt, np. czynniki związane z wykorzystaną do tego procesu mapą genetyczną, takie jak poziom zróżnicowania form rodzicielskich. Wyższy poziom polimorfizmu w danym regionie genomu powoduje większą liczbę segregujących markerów, a to umożliwia zakotwiczenie większej liczby klonów BAC.

Do potwierdzenia prawidłowości przyporządkowania markerów DArTseq do odpowiednich klonów BAC zastosowano strategię opartą na sekwencji markerów DArTseq i klonów BAC oraz weryfikację opartą na pozycji markerów DArTseq na mapie genetycznej.

Stosując te sposoby potwierdzono wysoką wiarygodność adresowania markerów DArTseq do klonów BAC: (a) potwierdziliśmy obecność 98,8% oczekiwanych markerów DArTseq w sekwencjach odpowiednich klonów BAC (91% przy ustawieniach TS1), ze średnią długością dopasowania wynoszącą ponad 99,95% i przy średniej identyczności sekwencji również ponad 99,95%, (b) wiele genetycznie zmapowanych markerów zaadresowanych do tego samego klonu BAC miało sąsiadujące pozycje na mapach genetycznych - tak było w przypadku 1 602 klonów BAC, które były zakotwiczone do mapy genetycznej przez przynajmniej 2 markery DArTseq.

W zsekwencjonowanych klonach BAC nie udało się potwierdzić obecności wszystkich markerów, co można wytłumaczyć niewystarczającą jakością użytych sekwencji klonów BAC. W wykorzystanych przy weryfikacji sekwencjach klonów BAC, używanych na podstawie sekwencjonowania metodą Illumina, generującą krótkie odczyty,

wystąpiły rejony o gorszej jakości, zawierające powtarzające się fragmenty. Odnotowano też obecność fragmentów zawierających niemożliwe do zidentyfikowania zasady oznaczone jako „N”. Liczba skafoldów uzyskanych przy sekwencjonowaniu siedmiu klonów BAC wynosiła od 5 do 15. Podobne wyniki uzyskali Liu i in. (2011). Przeprowadzali oni badania nad jęczmieniem, który podobnie jak żyto posiada duży genom (5,3 Gpz), z wysoką zawartością powtarzalnego DNA. Autorzy tej pracy do kotwiczenia klonów BAC wykorzystali genotypowanie pul zbiorczych DNA BAC przy pomocy macierzy Agilent, reprezentującej 42 302 geny. Do weryfikacji wiarygodności adresowania wykorzystano wyniki sekwencjonowania metodą Illumina DNA BAC wyizolowanego z poszczególnych pul matrix. Liu i in., (2011) potwierdzili w ten sposób obecność 82,1% zaadresowanych genów w dostępnych sekwencjach, uzyskanych z całkowitego DNA BAC odpowiedniej SP. Autorzy zapostulowali, że brak potwierdzenia obecności pozostałych genów w odczytach z sekwencjonowania BAC DNA może wynikać ze słabej jakości odczytów lub występowania SNP w miejscach wiązania starterów.

W niniejszej pracy w części klonów BAC stwierdzono obecność markerów DArTseq pochodzących z różnych chromosomów. W zależności od zastosowanych ustawień było to od 11,4 do 13,3% klonów BAC zawierających przynajmniej dwa markery z odmienną lokalizacją chromosomalną na mapie genetycznej. Niższy poziom występowania konfliktów kotwiczenia (8%) został pokazany przez Cviková i in. (2015), którzy wykorzystali mapę fizyczną i sekwencjonowanie NGS pul klonów BAC z biblioteki chromosomu 3DS pszenicy oraz zasób zastępujący mapę genetyczną, oparty na syntenii genomów traw - GenomeZipper.

Istnieje kilka możliwych przyczyn takiego zjawiska. Może to być spowodowane, m. in., przez nieprawidłowości przy konstruowaniu biblioteki - umieszczenie więcej niż jednego klonu BAC w jednej studziencie przy przenoszeniu klonów po transformacji do płytek wielodołkowych lub powstanie podczas ligacji klonów chimerycznych. Innym powodem mogą być kwestie związane z mapą genetyczną - błędne przypisanie markerów do chromosomów (Khan i in., 2012; Gan i in., 2017). Przyczyną występowania konfliktów kotwiczenia może być również strukturalne zróżnicowanie między liniami rodzicielskimi w populacji mapującej. W przedstawionej pracy mapa genetyczna do kotwiczenia składała się z indywidualnych map czterech populacji posiadających wspólnego rodzica: 541 × 2020 LM; 541 × DD1; 541 × S44 i 541 × EM1. Linie rodzicielskie były odległe genetycznie od linii wykorzystanej do konstrukcji biblioteki BAC (L318). Chociaż nie ma jeszcze

wyczerpujących badań o strukturalnej zmienności żyta, pojawiające się dane dotyczące innych gatunków roślin pokazują, że zakres tego typu zmienności może być istotny (Zhang i in., 2016; Golicz i in., 2016). Zhang i in. (2016) porównali dwie linie ryżu *indica*: Zhenshan 97 i Minghui 63, które reprezentują dwie główne grupy odmianowe tego podgatunku. Analizy porównawcze tych dwóch genomów ujawniły zaskakujące różnice strukturalne, zwłaszcza w odniesieniu do inwersji, translokacji i duplikacji segmentowych. Tylko około 42% genów związanych z elementami niepodlegającymi transpozycji było identycznych w obu genomach. Golicz i in. (2016) zbadali skalę zmienności strukturalnej u *B. oleracea*. Do tego celu wykorzystano dziewięć zróżnicowanych morfologicznie odmian *B. oleracea* i jej dzikiego krewnego- *Brassica macrocarpa*. Zaobserwali, że prawie 18,7% pangenomu składa się z genów, które posiadają strukturalne warianty. Rabanus-Wallace i in. (2021) podjęli pierwszą próbę zbadania częstości występowania zmienności strukturalnej u żyta. Porównano dwie linie wsobne: Lo7 i Lo225, reprezentujące dwie pule genetyczne wykorzystywane w hodowli mieszańcowej żyta. Zaobserwowano duże inwersje w czterech z siedmiu chromosomów żyta. Duże inwersje były obserwowane wcześniej w jęczmieniu i są one powszechną cechą zmienności strukturalnej w genomie tego gatunku (Himmelbach i in., 2018).

Uzyskana w niniejszej pracy efektywność adresowania markerów GBS (DArTseq) do klonów BAC pozostaje w zgodzie z doniesieniami literaturowymi wskazującymi na potencjał metod GBS do dostarczania dziesiątek tysięcy markerów w jednym eksperymencie (Li i in., 2015; Al-Beyroutiová i in., 2016). Przeszukiwanie biblioteki oparte na technologii DArTseq dało około siedmiokrotnie więcej markerów genetycznych przypisanych do poszczególnych klonów BAC w porównaniu do procedury opartej na markerach DArT, przeprowadzonej w takim samym układzie eksperymentalnym (Borzęcka i in., 2018). Liczba klonów BAC, które zostały zakotwiczone na mapie genetycznej przez markery DArTseq była około 3,6 razy większa niż liczba klonów BAC zakotwiczonych przez markery DArT.

W przedstawionej pracy zakotwiczone 4 525 klonów BAC za pomocą 5 436 markerów DArTSeq. Za pomocą genotypowania mikromacierzami (Li i in., 2011) udało się przypisać zaledwie 1 364 klony BAC do 3 089 sond. Zastosowanie mikromacierzy umożliwiło zidentyfikowanie ponad pięciokrotnie mniejszej liczby klonów BAC zakotwiczonych poprzez markery występujące raz lub dwa razy niż przy zastosowaniu genotypowanie DArTSeq (738 vs 4 022). W niniejszej pracy zakotwiczone ponad

czterokrotnie więcej klonów BAC poprzez jeden marker w porównaniu z wynikami uzyskanymi przez Li i in. (2011) (2 783 vs 626).

Ocena efektywności adresowania w porównaniu do innych prac wypadła zatem korzystnie. Ten wyższy wskaźnik sukcesu w porównaniu do innych układów eksperymentalnych jest najprawdopodobniej wynikiem zarówno zastosowania bardziej przepustowej metody genotypowania jak i ulepszonej metodologii tworzenia pul zbiorczych i analizy danych, przedyskutowanej szczegółowo powyżej.

Przedstawione badania pokazały, że metoda DArTseq w połączeniu z trójwymiarową strategią tworzenia pul zbiorczych w zakotwiczeniu klonów BAC na mapie genetycznej żyta jest efektywna, wysokoprzepustowa i wiarygodna. Zastosowanie jej u żyta, które wyróżnia się posiadaniem dużego i złożonego genomu, wskazuje na możliwość wykorzystania podobnego podejścia dla większości roślin, zwłaszcza o słabo poznanych i złożonych genomach.

6.2 Sekwencjonowanie klonów BAC

Technologia sekwencjonowania Nanopore jest coraz częściej wykorzystywana do sekwencjonowania dużych, złożonych genomów zwierząt i roślin (Li i in., 2017). Wyzwaniem przy sekwencjonowaniu skomplikowanych genomów jest częsta obecność fragmentów powtarzających się. Uzyskanie prawidłowego złożenia takich fragmentów genomów na podstawie krótkich odczytów uzyskiwanych przy zastosowaniu metod sekwencjonowania drugiej generacji jest niezwykle trudne, a wręcz niemożliwe (Jung i in., 2019). Metody sekwencjonowania generujące długie, ciągłe sekwencje, takie jak sekwencjonowanie Nanopore, mogą być pomocne w takich przypadkach, ponieważ długie odczyty mogą obejmować fragmenty trudne do sekwencjonowania innymi metodami, łącząc dwa kontigi w skafold o odpowiedniej długości (Li i in., 2017). Fragmenty zsekwencjonowane dzięki technologii Nanopore są źródłem cennych informacji biologicznych, ponieważ mogą zawierać całe geny, elementy regulacyjne, powtarzające się elementy transpozycyjne (TE), centromery i telomery (Li i in., 2017).

Jednym z przykładów wykorzystania technologii Nanopore do sekwencjonowania klonów BAC jest praca Mitsuhashi i in., (2017). Jest to nie tylko przykład wykorzystania tej technologii do sekwencjonowania klonów BAC ale również pokazuje skuteczność w sekwencjonowaniu fragmentów genomu zawierających sekwencje powtórzone. Badania Mitsuhashi i in., (2017) dotyczyły wysoce powtarzalnego rejonu subtelomerycznego,

zwanego regionem pLAM, który jest charakterystycznym elementem dla osób cierpiących na dystrofię twarzowo-łopatkowo-ramieniową (FSHD). Ze względu na powtarzalny charakter uzyskanie prawidłowej sekwencji tego regionu było niezwykle trudne ale technologia Nanopore okazała się bardzo skuteczna do tego celu. Używając urządzenia MinION, Mitsuhashi i in. (2017) zsekwencjonowali cały region pLAM z dokładnością sięgającą 100%.

W niniejszej pracy wdrożono technologię Nanopore do sekwencjonowania pojedynczych klonów BAC. Jest to pierwszy przykład zastosowania tej technologii dla żyta. W trakcie wcześniejszych badań pozyskano sekwencję siedmiu klonów BAC przy zastosowaniu platformy Illumina. Dzięki tej technice produkowane są krótkie (ok. 150 nukleotydów), zachodzące na siebie odczyty, które następnie podczas analiz bioinformatycznych składane są w dłuższe kontigi. Niestety w przypadku wyżej wymienionych klonów BAC nie udało się złożyć odczytów Illumina w jeden ciąg, odpowiadający oczekiwanej długości klonu BAC (ok. 130 kbp). Jest to typowy wynik uzyskiwany przy sekwencjonowaniu fragmentów złożonych genomów, takich jak genom żyta, bogatych w sekwencje powtórzone. Ze względu na obecność licznych powtórzeń składanie krótkich odczytów, typowych dla metody Illumina w jedną ciągłą, całkowitą sekwencję całego insertu jest szczególnie skomplikowane (Bakera i in., 2015). Dlatego zdecydowano się na zastosowanie do sekwencjonowania indywidualnych klonów BAC technologii Nanopore, umożliwiającej uzyskanie długich odczytów - na tyle długich, że potencjalnie mogą objąć cały problematyczny fragment klonu BAC zawierający powtórzenia. Technologią Nanopore zsekwencjonowano 13 klonów BAC i dla każdego z nich uzyskano ich pełną, ciągłą sekwencję.

Pierwszym krokiem analizy sygnałów sekwencjonowania nanoporów była identyfikacja nukleotydów (ang.: *basecalling*) za pomocą algorytmów oprogramowania Albacore (Sahoo 2017), rekomendowanego przez Oxford Nanopore Technologies. Następnie przetestowano czy można zredukować ilość wykorzystanych odczytów bez utraty jakości ostatecznego złożeń. Okazało się, że do uzyskania wysokiej jakości sekwencji potrzeba około 12 mln nukleotydów na każdy klon BAC. Oceniano to za pomocą sprawdzenia ilości mutacji w rejonie wektora BAC. W związku z tym przefiltrowano otrzymane odczyty według poszukiwanej długości i jakości za pomocą narzędzia Filtrlong. Wykorzystuje ono zarówno długość odczytu (im dłuższe tym lepsze) jak i jego poziom identyczności (im wyższa tym lepsza) przy wyborze, które odczyty przechodzą przez filtr.

Uzyskane odczyty składano w kontigi za pomocą programu Canu (Koren i in., 2022) stworzonego z myślą o usprawnieniu i jak najlepszym wykorzystaniu danych z sekwencjonowania Nanopore oraz PacBio. Zastosowana procedura składa się z trzech etapów: korekty, przycinania i składania - z których każdy może działać niezależnie lub szeregowo (kolejno po sobie) (Koren i in., 2022). Canu jest obecnie najbardziej wydajnym assemblerem dostępnym dla dużych genomów. Wymaga około 20 000 godzin pracy procesora na złożenie ludzkiego genomu w porównaniu z około 60 000 godzin wymaganych dla FALCON (Chin i in., 2016) oraz ponad 250 000 niezbędnych dla Celera Assembler v8.2 (Berlin i in., 2015). Oprócz ulepszeń umożliwiających dużą redukcję czasu pracy, w Canu zaimplementowano również rozwiązania pozwalających na uzyskanie większej ciągłości złożeń niż przy pomocy innych assemblerów (Koren i in., 2022).

Po złożeniu kontigów zastosowano Circlator, pierwsze narzędzie do automatyzacji cyrkulacji złożenia (Hunt i in., 2015). Zastosowane w niniejszej pracy programy do składania odczytów uzyskanych dzięki technologii sekwencjonowania Nanopore okazały się bardzo skuteczne. Ze względu na uzyskanie sekwencji o satysfakcjonujących nas parametrach (ciągłość złożenia, odpowiednia dokładność określona przez porównanie z sekwencją referencyjną wektora i długość - zweryfikowana eksperymentalnie przy wykorzystaniu metody PFGE) opracowana procedura sekwencjonowania i obróbki bioinformatycznej odczytów może zostać wdrożona do kolejnych tego typu analiz.

Początkowo istotną wadą technologii Nanopore był wysoki wskaźnik błędów w porównaniu z innymi platformami NGS - uzyskiwano odczyty charakteryzujące się współczynnikiem błędów wynoszącym 5 - 15% (Rang i in., 2018). Problem zminimalizowano stosując techniki konsensusu czyli wykorzystanie różnych sposobów sekwencjonowania umożliwiających uzyskanie odczytów o różnych długościach. Scheunert i in. (2020) zaproponowali hybrydową strategię polegającą na użyciu technologii Nanopore i Illumina. Na podstawie wyników swoich badań dowiedli, że długie odczyty Nanopore są odpowiednią alternatywą dla krótkich odczytów Illumina w filogenomicie plastomowej. Podobną taktykę obrano do sekwencjonowania genomu grzyba *Saccharomyces cerevisiae*. Dzięki użyciu odczytów Nanopore i Illumina uzyskano bardzo dobre rezultaty. Długość kontigu N50 była dziesięć razy dłuższa od złożenia uzyskanego dzięki technologii Illumina (678 Kpz w porównaniu do 59,9 Kpz). Ten kontig zawierał znacznie pełniejszą reprezentację elementów powtórzonych i struktur, nieobecnych w

odczytach Illumina (Goodwin i in., 2015). Poziom błędów przy zastosowaniu technologii Nanopore w tym przypadku był na niskim poziomie 0,12%.

Ciągły rozwój nowych narzędzi bioinformatycznych do korekcji błędów odczytu (polerowania, ang.: *polishing*) oraz generowania konsensusu znacznie pomógł w poprawie dokładności sekwencjonowania Nanopore (Ciuffreda i in., 2020). Dzięki tym ulepszeniom uzyskuje się niższy poziom błędów (w naszych analizach ok. 0,5%) i uzyskana dokładność sekwencjonowania Nanopore jest wystarczająca dla wielu zastosowań, na przykład w projektach dotyczących klasyfikacji taksonomicznej (Wick i in., 2019).

Dla oceny dokładności sekwencjonowania Nanopore uzyskanej w niniejszej pracy porównano sekwencje wektora (pIndigoBAC-5) z klonów BAC zsekwencjonowanych techniką Nanopore z sekwencją wektora zdeponowaną w bazie NCBI. Wykazano wysokie podobieństwo sięgające 99%. Jest to dokładność w zupełności wystarczająca do bioinformatycznego przewidywania struktury genu (Li i in., 2017).

Uzyskana przez nas dokładność jest porównywalna do wartości otrzymanych w innych pracach, w których zastosowano technologię Nanopore. Loman i in. (2015) złożyli *de novo* chromosom *Escherichia coli* K-12 MG1655 w pojedynczy kontig o wielkości 4,6 Mbp z dokładnością sekwencjonowania sięgającą 99,5%. Scheunert i in. (2020) w badaniach poświęconych sekwencjonowaniu genomu chloroplastowego u dwóch gatunków należących do rodziny astrowatych (*Leucanthemum* Mill.) stosując podobną strategię do Loman i in. (2015) osiągnęli porównywalną skalę podobieństwa sekwencji wynoszącą 99,59%. Zastosowane w obu tych pracach metody to korekta odczytu z wykresami POA - zależny algorytm konsensusu zaimplementowany w CANU - wykorzystany również w naszej pracy oraz polerowanie końcowego złożenia za pomocą oprogramowania NANOPOLISH.

Dokładniejsza analiza uzyskanych w niniejszej pracy metodą Nanopore sekwencji wektora wykazała obecność substytucji pojedynczych nukleotydów (SNP) oraz insercji/delecji (InDel). Dane literaturowe wskazują, że w przeciwieństwie do losowego profilu błędów obserwowanego przy zastosowaniu technologii PacBio, niektóre błędy w odczytach Nanopore pojawiają się częściej niż inne. Laehnemann i in. (2016) sugerowali, że najczęściej występujące typy SNP to substytucje i delecje. Postulowali, że substytucje A/T są najmniej prawdopodobne spośród wszystkich rodzajów substytucji. W naszej pracy substytucje A/T stanowiły od 2 do 17% wszystkich typów SNP zależnie od klonu. Rang i

in. (2018) i Scheunert i in. (2020), przy zastosowaniu technologii Nanopore odnotowali, że większość wykrytych błędów to InDele (81,5%), a znaczna większość z nich jest częścią regionów homopolimerowych. Wyniki uzyskane w niniejszej pracy pozostają w zgodzie z tymi doniesieniami - w poszczególnych klonach BAC polimorfizmy typu InDel stanowiły od 69 do 90% wszystkich wykrytych błędów.

Uzyskane w niniejszej pracy wyniki pokazały, że technologia sekwencjonowania Nanopore jest efektywną i wiarygodną metodą uzyskiwania pełnych, ciągłych sekwencji insertów biblioteki BAC żyta. Podobnie jak w innych pracach (Scheunert i in., 2020) zauważono, że przy porównywalnych kosztach uzyskania sekwencji jednego klonu sekwencjonowanie Nanopore pozwala na uzyskanie lepszej jakości sekwencji klonów BAC niż zastosowanie metody Illumina. Przy zastosowaniu szeregu ulepszeń jest możliwe uzyskanie poziomu błędów na niskim poziomie wynoszącym 0,5 - 1%.

6.3 Identyfikacja z wykorzystaniem biblioteki BAC sekwencji genów przypuszczalnie związanych z odpornością na choroby oraz z udomowieniem

W momencie rozpoczęcia badań, które przedstawiono w niniejszej rozprawie doktorskiej sekwencja genomu żyta pozostawała nieznana, dlatego skorzystano z biblioteki BAC jako cennego narzędzia genomiki strukturalnej. Klony BAC zawierają długie fragmenty genomu (rzędu 130 kbp), które można zsekwencjonować uzyskując wysoką jakość odczytu. Jest to dużo łatwiejsze do osiągnięcia niż dla całego genomu o długości 8Gbp. Dostępność wysokiej jakości sekwencji insertów klonów BAC znacznie ułatwia zidentyfikowanie kompletnych sekwencji poszukiwanych genów. Dodatkowo, dostępność sekwencji genomu referencyjnego dla danego gatunku nie rozwiązuje wszystkich problemów związanych z procesem identyfikacji genów - w pierwszych wersjach genomów, składanych z krótkich odczytów często adnotacje zawierały dużo błędów, a liczba i struktura genów nie była przewidziana prawidłowo z powodu niskiej jakości sekwencji (Giani i in., 2020). Jakość złożenia genomu bezpośrednio wpływa na jakość adnotacji. Zostało udowodnione, że technologie sekwencjonowania umożliwiające uzyskanie długich odczytów nie tylko ulepszają jakość złożenia, ale również pozwalają na skuteczne kotwiczenie i prawidłową orientację skafoldów, co jest kluczowe do identyfikacji genów za pomocą precyzyjnego mapowania i analizy syntenii genomów (Hulse-Kemp i in., 2021).

Wykorzystanie klonów biblioteki BAC zawierających interesujący nas fragment genomu umożliwia poznanie jego dokładniejszej sekwencji. Takie podejście zostało

zastosowane w pracach nad izolacją genu *MATE1*, związanego z tolerancją wysokiej zawartości glinu w glebie u kukurydzy (Maron i in., 2013) oraz genów *btr-1* i *btr-2* odpowiadających za niełamliwość osadki u jęczmienia (Pourkheirandish i in., 2015). Innym przykładem takiego zastosowania sekwencji klonów BAC była praca zespołu Tucker i in. (2017). Wykorzystali oni biblioteki BAC dwóch odmian pszenicy do identyfikacji i charakterystyki genu *Ms1*, odpowiedzialnego za przywracanie płodności. Koduje on białko przenoszące lipidy, niezbędne do rozwoju eksyny pyłkowej (Tucker i in., 2017).

Opisane są różne strategie umożliwiające identyfikację genów odporności, tj. klasyczna analiza genetyczna (mapowanie genetyczne) (Dyda i in., 2022), mapowanie asocjacyjne (Li i in., 2011; Vendelbo i in., 2021), poszukiwanie ortologów (Bakera i in., 2015) czy wykorzystanie do tego celu danych z RNAseq (Rabanus-Wallace i in., 2021; Li i in., 2021). Bioinformatycznie można stosunkowo łatwo znaleźć potencjalne geny DRG w sekwencjach genomu, ale nie wiadomo, które z nich faktycznie wykazują zróżnicowanie alleliczne u danego gatunku i mają potencjalne praktyczne znaczenie w hodowli odpornościowej. Również nie wszystkie strategie są możliwe do stosowania w wielu gatunkach ze względu na takie czynniki, jak wielkość genomu, poliploidalność i związane z tym koszty sekwencjonowania i opracowania procedur (Hawliczek i in., 2020).

Żyto wykazuje dużą tolerancję na niskie temperatury i inne rodzaje stresów abiotycznych i biotycznych. Może służyć jako model odporności dla innych roślin zbożowych o większym znaczeniu gospodarczym (pszenica, jęczmień) (Martis i in., 2013). Coraz więcej wiadomo na temat genów związanych z odpornością na choroby u żyta. Analogi genów odporności roślin (RGA), jako potencjalne geny odporności (R), mają konserwatywne domeny i motywy, które odgrywają określoną rolę w odporności na patogeny. Miejsca wiążące nukleotydy bogate w leucynę, receptory kinazy i białka receptorowe są charakterystycznymi elementami dla genów RGA (Sekhwal i in., 2015). Opisano kilka genów R w życie (geny Pr): *Pr1-5*, *Pr-d-f*, *Pr-i-l*, *Pr-n*, i *Pr-p-t*, związanych z odpornością na rdzę brunatną spowodowaną zakażeniem patogenem *P. recondita f. sp. secalis* (*Prs*) (Roux i in., 2007). W sumie zidentyfikowano dziesięć głównych genów R związanych z odpornością na rdzę brunatną, na chromosomach pszenicy syntenicznych z ramieniem chromosomu żyta 7RS, w tym dwa na 4AL (*Lr28*, *Lr30*), trzy na 5BS (*LrK1*, *Lr52*), dwa na 5BL (*Lr18*, *Lr*), trzy na 5DS (*Lr57*, *Lr70*, *Lr76*) i dwa na chromosomie 5DL (*Lr1*, *LrSyn137*) (McIntosh i in., 2017). W pracy dotyczącej genomu referencyjnego żyta

(Rabanus-Wallace i in., 2021) zidentyfikowano bioinformatycznie 792 genów NB-LRR pełnej długości u żyta, znajdując je w dystalnych regionach chromosomów. Podobna dystrybucja genów NB-LRR występuje w w genomie pszenicy (Steuernagel i in., 2020). Według pracy Li i in. (2021) największa liczba DRG znajduje się u żyta na chromosomach 2R-4R (296-301), a najmniej na chromosomie 7R (227). Łącznie zidentyfikowali oni 1 989 DRG.

W różnych pracach zastosowano metodę DArT wykorzystującą mikromacierze i znaleziono powiązania markerów DArT do genów warunkujących ważne cechy agronomiczne, np. dwuletność u marchwi (Grzebelus i in., 2014). Potwierdzono również efektywność metody DArT w analizach genomu żyta (Bolibok-Brągoszewska i in., 2009; Bolibok-Brągoszewska i in., 2014; Milczarski i in., 2011). W badaniach nad genomem eukaliptusa wykazano, że niemal 70 % markerów DArT zlokalizowanych było w przewidywanych genach (Petroli i in., 2012). Zaproponowana w tej pracy doktorskiej strategia, obejmująca wykorzystanie wyników adnotacji funkcjonalnej sekwencji żytnich markerów DArT (Gawroński i in., 2016) i biblioteki BAC genomu linii wsobnej żyta L318 pozwoliła na poznanie sekwencji przypuszczalnych genów związanych z odpornością na choroby.

Badania opisane w niniejszej pracy, dotyczące identyfikacji genów związanych z odpornością na choroby u żyta, bazują na wynikach zespołu naukowców z Katedry Genetyki, Hodowli i Biotechnologii Roślin. Prace te obejmowały analizę bioinformatyczną 6 177 sekwencji żytnich markerów DArT, o całkowitej długości 3 140 801 pz i średniej długości 508 pz (Gawroński i in., 2016). Do analizy i porównań zostały wykorzystane publicznie dostępne zasoby sekwencyjne (referencyjne sekwencje genomowe, sekwencje cDNA) gatunków *Brachypodium*, pszenica i jęczmień. Podczas analizy funkcjonalnej sekwencje żytnich markerów DArT zostały przeanalizowane poprzez przeszukiwanie białkowej bazy danych GenBank z użyciem algorytmu BLASTX. Przypisano potencjalne funkcje dla 1 101 sekwencji, w tym funkcje związane z odpornością na choroby zostały przyporządkowane do 126 sekwencji. Największą liczbę markerów DArT związanych z odpornością na choroby zidentyfikowano w chromosomach 4R i 6R (Gawroński i in., 2016). Ze względu na to, że przeciętna długość markerów DArT wynosi kilkaset nukleotydów (w przypadku żyta 508) zidentyfikowane w powyższej pracy sekwencje potencjalnie powiązane z odpornością na patogeny nie reprezentowały całych genów lecz jedynie ich fragmenty. Potencjalną zaletą zastosowanego w niniejszej pracy podejścia do

identyfikacji genów związanych odpornością na patogeny (poprzez sekwencje markerów DArT), w porównaniu do analizy czysto bioinformatycznej jednego genotypu jest zwiększona szansa znalezienia genów wykazujących zróżnicowanie alleliczne w zasobach genowych żyta. Związane jest to z tym, że do sekwencjonowania wybierano markery DArT, które wykrywały polimorfizm w różnego typu analizach zasobów genowych żyta (Gawroński i in., 2016).

Na podstawie wyników genotypowania pul klonów BAC metodą DArT (Borzęcka i in., 2018) zidentyfikowano 11 klonów przypuszczalnie zawierających 21 markerów DArT o możliwych funkcjach związanych z odpornością wskazanych w omawianych wyżej badaniach Gawrońskiego i in., (2016). Za pomocą PCR potwierdzono obecność 11 markerów DArT w 7 klonach BAC. Analiza bioinformatyczna sekwencji wytypowanych klonów BAC pozwoliła na zidentyfikowanie 28 genów potencjalnie związanych z odpornością na choroby. Należą one do największej grupy genów odporności na choroby NB-LRR (np. *RPM1*, *RPP13-like*, *RGAs*).

Podobieństwo sekwencyjne nie jest jednak wystarczającym potwierdzeniem przypuszczalnej funkcji danego genu. Jest szereg metod pozwalających na zbadanie funkcji genu. Między innymi można wykorzystać analizy poziomu ekspresji z wykorzystaniem łańcuchowej reakcji polimerazy w czasie rzeczywistym (ilościowy RT-PCR). Jest to czuła i precyzyjna technika umożliwiająca ilościową analizę mRNA w komórkach lub tkankach. W jednym z zespołów w KGHIBR prowadzone są badania związane z identyfikacją genów odporności na rdzę brunatną. Jest to groźna choroba żyta, powodująca straty plonu rzędu 30-40% (Wehling i in., 2003), powodowana infekcją patogenem *Puccinia recondita f. sp. secalis*. Badania nad identyfikacją żytnich ortologów pszenicznych genów *Lr* związanych z odpornością na tą chorobę są prowadzone przez zespół prof. Rakoczy-Trojanowskiej (Święcicka i in., 2019). W trakcie tych badań uzyskano unikalny materiał badawczy, umożliwiający sprawdzenie ekspresji genów na różnych etapach infekcji patogenem (po 8, 17, 24 i 48 h). Materiał ten został użyczony przez prof. Rakoczy-Trojanowską w celu zbadania przez nas zmian ekspresji jednego z genów zidentyfikowanych w analizowanym klonie BAC (S34p4M5) – *Sck33*. Wybór genu dokonany został na podstawie wyników analizy BLAST. Przy pomocy tego algorytmu znaleziono w tym klonie sekwencje wykazujące podobieństwo do genów, które mogą brać udział w odpowiedzi na zakażenie rośliny patogenem powodującym rdzę brunatną. Poziom ekspresji *Sck33* zbadano w tkankach dwóch linii wsobnych żyta zebranych w różnych punktach czasowych po

inokulacji patogenem *Puccinia recondita f. sp. secalis*. Poziom ekspresji genu *Sck33* po upływie 8 h od inokulacji był niższy w porównaniu do roślin kontrolnych. W kolejnych godzinach doświadczenia (17, 24 i 48 h) poziom ekspresji genu *Sck33* był statystycznie wyższy u roślin porażonych. Podwyższone i obniżone poziomy ekspresji były najczęściej wykrywane odpowiednio po 8 i 17 h. Przeprowadzona analiza ekspresyjna wykazała statystycznie istotny wzrost poziomu ekspresji w czasie rozwoju choroby dla genu *Sck33*, podobnie jak dla genu *Lr1* badanego przez Świąćicką i in. (2019), co sugeruje ich możliwą rolę w odpowiedzi na porażenie rdzą brunatną u żyta. Jednak, aby to potwierdzić, wskazane są dalsze badania. Uzyskane rezultaty stanowią punkt wyjścia do dalszych badań nad strukturą i funkcją wykrytych przypuszczalnych genów, na przykład przy wykorzystaniu metod genetyki odwróconej.

W niniejszej pracy podjęto również próbę poszukiwania genów związanych z procesem udomowienia. Geny związane z tym procesem kodują głównie regulatory transkrypcji oraz enzymy. Większość poznanych genów udomowienia należy do tych grup (Doebly i in., 2006).

Podczas udomowienia, a następnie hodowli, rośliny uprawne zostały podane intensywnej presji selekcyjnej nakierowanej na pewne pożądane cechy, takie jak wysoki i stabilny plon (Purugganan i Fuller 2009). Połączone skutki procesu udomowiania, ulepszanie zaawansowanych materiałów hodowlanych przy zastosowaniu krzyżowań wypierających oraz unikanie uwzględniania w programach hodowlanych krzyżowań z prymitywnymi formami o niskiej wartości użytkowej spowodowały spadek różnorodności genetycznej wśród zawansowanych materiałów hodowlanych roślin uprawnych (Kantar i in., 2017). Zwążająca się zmienność genetyczna w programach hodowlanych jest zjawiskiem niepożądanym. Jej następstwa to, m.in., zmniejszenie postępu hodowlanego i wzrost podatności na stesy biotyczne i abiotyczne, co może w krańcowej sytuacji skutkować zagrożeniem bezpieczeństwa żywnościowego (Fisher i in., 2010). Z tego punktu widzenia jest zatem niezwykle istotne, aby zrozumieć wpływ hodowli na genom rośliny uprawnej i zidentyfikować regiony genomu, na które wywierana była szczególnie silna presja i w efekcie charakteryzują się bardzo niewielką zmiennością we współczesnych odmianach hodowlanych. Podjęte prace, mające na celu poznanie sekwencji wybranych żytnich ortologów genów udomowienia są początkowym etapem większego projektu. Pozyskana wiedza ułatwi zaprojektowanie skutecznych starterów do amplifikacji tych genów podczas analiz zróżnicowania allelicznego w formach żyta o różnym stopniu

ulepszenia (na przykład przy wykorzystaniu metodyki opisanej przez Hawliczek i in., 2020) oraz, w dalszej perspektywie, wskazanie wariantów genów o przypuszczalnie zmienionej funkcji.

W porównaniu do innych gatunków, takich jak kukurydza czy soja, wiedza na temat poszczególnych genów lub rejonów w genomie, które zostały poddane presji selekcyjnej podczas procesu udomowienia oraz hodowli jest wciąż niewielka dla żyta. Dlatego na początku przedstawionych badań posiłkowano się sekwencjami genów udomowienia z innych roślin zbożowych. Projektowanie starterów na podstawie sekwencji innych gatunków jest powszechnym podejściem, opisywanym już tutaj wcześniej (Bakera i in., 2015; Świącicka i in., 2020). Hackauf i in. (2012) wskazali ortologiczne zestawy genów w ryżu, *Brachypodium* i sorgo i wykorzystali te modele genów jako szablony do opracowania zachowanych ewolucyjnie zestawów markerów ortologów (ang.: *conserved ortholog set*, COS) dla genu przywracającego męską płodność *Rfp1*, zlokalizowanego na długim ramieniu chromosomu żyta 4R. Markery COS umożliwiły integrację ulepszonej mapy genetycznej chromosomu 4R żyta z mapami pszenicy i jęczmienia. Pozwoliły również na identyfikację regionów ortologicznych względem *Rfp1* w pszenicy i jęczmieniu na krótkich ramionach chromosomów 6D i 6H. W niniejszej pracy strategia polegająca na projektowaniu starterów do genów związanych z udomowieniem na podstawie sekwencji innych gatunków okazała się nieskuteczna. Uzyskano oczekiwany produkt tylko dla genu *VRN2*. Może to być spowodowane tym, że mimo filogenetycznego pokrewieństwa żyta z pszenicą zwyczajną (*Triticum aestivum*) i jęczmieniem (*Hordeum vulgare*), poziom podobieństwa sekwencji nie był wystarczający.

Kolejnym krokiem było przeszukanie genomu referencyjnego żyta linii wsobnej Lo7 (Rabanus-Wallace i in., 2019) w celu znalezienia sekwencji homologicznych do genów udomowienia z blisko spokrewnionych gatunków. Poziom identyczności genów referencyjnych z ich homologami z genomu Lo7 był wysoki i wynosił prawie 90%. Nie udało się wskazać sekwencji homologicznych dla genu *qSW5*, kontrolującego szerokość ziarna u ryżu (Liu i in., 2017) oraz *Ghd7*, związanego z czasem kwitnienia oraz determinacją wysokości rośliny (Xue i in., 2008). Dla 5 z 11 sekwencji homologicznych znalezionych w genomie Lo7 pozycje pokrywały się z tymi ustalonymi w dostępnej adnotacji (Rabanus-Wallace i in., 2019). Różnice w identyfikacji genów wynikają z użytych przez autorów danych i narzędzi. Dane z RNAseq wykorzystane przez Rabanus-Wallace i in. (2019) mogły nie pozwolić na wykrycie w sekwencji genomu genów, które

nie ulegały ekspresji w badanych tkankach. Opierając się na sekwencji referencyjnej żyta (Rabanus-Wallace i in., 2019) zaprojektowano nowy zestaw starterów do amplifikacji fragmentów 11 genów. Udało się zamplifikować fragmenty 6 genów, przy wykorzystaniu genomowego DNA linii wsobnej L318. Uzyskane amplikony wykazały wysoki poziom identyczności z sekwencjami referencyjnymi genów udomowienia. Najwyższe podobieństwo zanotowano w przypadku genu *VRNI* (99%), a najniższe dla genu *BTR2* (82%). Wyniki te wskazują na efektywność zastosowanego podejścia i znalezienie interesujących nas fragmentów genomu oraz na stosunkowo silnie zachowane ewolucyjnie u zbóż fragmentów kilku z badanych genów udomowienia.

Udało się również zidentyfikować fragmenty 5 z 6 poszukiwanych genów w sekwencjach klonów BAC. Zaobserwowano wysokie podobieństwo sekwencji amplikonów uzyskanych z DNA genomowego linii L318 z homologicznym obszarem w identyfikowanych klonach BAC. Mimo wielu prób, przetestowania czterech par różnych starterów, nie udało się znaleźć sekwencji homologicznych do genu *Rht-B1*. Przyczyną nieudanej amplifikacji może być występowanie SNP/InDeli w miejscu przyłączenia starterów. Być może gen *Rht-B1* jest mniej zachowany ewolucyjnie od innych lub nie jest obecny w tej linii żyta. Poziom podobieństwa pomiędzy genami referencyjnymi a odpowiadającymi im sekwencjami homologicznymi w obu genomach żyta: L318 i Lo7 jest zbliżony. Poziom identyczności między sekwencjami nukleotydowymi odpowiadającymi przypuszczalnemu genom udomowienia z dwóch genomów linii wsobnych żyta: Lo7 i L318 wyniósł od 96% dla genu *Q* do 100% w przypadku *VRNI* i *GIF1* (z klonu S30p5H22). Jedynie dla genu *GIF1* z klonu BAC S12p7K11 ta wartość była niższa (83,5%). Wykazano zatem występowanie zróżnicowania sekwencyjnego pomiędzy liniami wsobnymi żyta - różnice w porównywanych sekwencjach obu genomów wynoszą maksymalnie do 4%.

Przewidziana bioinformatycznie przez program FGENESH (Salamov i Solovyev, 2000) struktura przypuszczalnego genu *Q* żyta była podobna do struktury jego ortologa w pszenicy. Żytni homolog genu *Q*, prawdopodobnie trzykrotnie krótszy, składa się z 8 egzonów, natomiast gen *Q* pszenicy ma ich 10. Zidentyfikowany przez nas homolog genu *BTR2* zawiera 2 (gen 14) lub 4 egzony (gen 15), tymczasem wśród dzikich gatunków Triticeae większość sekwencji podobnych do *Btr2* była pozbawiona intronów, tak jak w przypadku jęczmienia. Wyjątkiem jest *Ae. tauschii* gdzie według predykcji prognoz programu FGENESH występują 2 egzony (Zeng i in., 2020).

Warto mieć na uwadze, że czasem predykcja bioinformatyczna struktury przypuszczalnych genów może być błędna. Jest możliwe, że przewidywana sekwencja nie jest prawdziwym genem, albo przewidywany egzon faktycznie należy do jednego z sąsiednich genów zamiast być oddzielnym pojedynczym genem egzonowym. Tak mogło być w przypadku genu *VRN1*, ponieważ w jednym z klonów BAC znaleziono dwa rejony wykazującego do niego podobieństwo. Jeden z nich (gen 3 z klonu S16p7H5) według programu Fgenesh posiadał tylko 1 egzon. Zarówno w *Ae. tauschii* jak i w *T. aestivum*, locus *Vrn-D1* posiada 8 egzonów i 7 intronów (Takumi i in., 2011; Shcherban i in., 2013). Jego przewidziany homolog w życie może posiadać 9 egzonów i 8 intronów.

Fgenesh jest jednym z lepszych narzędzi służących do bioinformatycznej identyfikacji i przewidywania struktury genów. Yao i in., (2005) przetestowali pięć różnych programów (FGENESH, GeneMark.hmm, GENSCAN, GlimmerR and Grail) do predykcji genów kukurydzy. FGENESH okazał się najbardziej efektywnym i skutecznym narzędziem, natomiast na drugim miejscu uplasował się GeneMark.hmm. FGENESH był również dokładniejszy niż inne testowane programy do wykrywania genów ryżu i u ssaków (Solovyev, 2001; Yu i in., 2002).

Dane literaturowe, dotyczące kolejności genów u żyta, jęczmienia i pszenicy wskazują dobrze zachowaną kolinearność genomu między tymi trzema gatunkami. Jest to widoczne w dużych blokach syntenicznych, które są przerywane fragmentami odpowiadającymi rearanżacjom chromosomów (Naranjo i Fernández-Rueda, 1991; Devos i in., 1993; Martis i in., 2013). Lokalizacja chromosomalna wykrytych przypuszczalnych genów udomowienia w genomach żyta i pszenicy pokrywa się. W obu gatunkach gen *Q* i gen *VRN1* znajdują się na chromosomie 5, a gen *BTR2* na chromosomie 3. Gen *BTR2* u jęczmienia znajduje się na chromosomie 3H (Cross i in., 2022). Prawdopodobne funkcje tych przypuszczalnych genów w życie są jeszcze nieznanne. W pszenicy gen *Q* wykazuje efekt plejotropowy, warunkuje wiele cech związanych z udomowieniem (Simons i in., 2006). *BTR2* odpowiada za niełamliwość osadki jęczmienia, oddziałując na właściwości ściany komórkowej (Cross i in., 2022). Geny *VRN1* i *VRN2* odpowiadają za wernalizację. Locus *VRN-1* koduje czynnik transkrypcyjny MADS-box, który kontroluje przejście merystemu wierzchołkowego pędu wegetatywnego do fazy rozrodczej (Shcherban i in., 2013).

Na podstawie porównania sekwencji genów z klonów BAC z genomem żyta Lo7 zaobserwowano dosyć znaczne różnice między genomami tych linii wsobnych. Mimo, że uzyskano dopasowania dla 83% przypuszczalnych genów, były to ich krótkie fragmenty. Dla ponad 60% z nich długość pokrywających się sekwencji nie sięgała 75% ich całej długości. Znaczne zróżnicowanie między genomami różnych linii czy odmian tego samego gatunku zaobserwowano wcześniej w badaniach dotyczących innych roślin, np ryżu (Zhao i in., 2018). W tym momencie trudno jest ocenić na ile to jest odzwierciedlenie faktycznych różnic w sekwencji DNA na ile efekt niedoskonałości pierwszej wersji genomu żyta.

Wyniki uzyskane w niniejszej dysertacji poszerzyły dotychczasową wiedzę na temat genów związanych z odpornością na choroby oraz udomowieniem u żyta. Dalsze badania mogłyby przyczynić się do zrozumienia ich funkcji biologicznych u tego gatunku. Zaprezentowane wyniki dają możliwość analizy rejonów genomu, na które wywierana była presja selekcyjna podczas udomowienia i hodowli żyta. Jest to znaczący krok w kierunku wykorzystania dostępnych zasobów genowych w badaniach naukowych i programach hodowlanych.

7. WNIOSKI

1. Genotypowanie DArTseq w połączeniu z trójwymiarową strategią tworzenia pul zbiorczych umożliwia efektywne i wiarygodne ustalanie adresów marker-klon BAC i w konsekwencji kotwiczenie klonów BAC na mapie genetycznej żyta.
2. Technologia sekwencjonowania Nanopore jest efektywną i wiarygodną metodą uzyskiwania pełnych sekwencji klonów biblioteki BAC żyta.
3. Zaproponowana strategia, łącząca wykorzystanie wyników adnotacji funkcjonalnej sekwencji żytnich markerów DArT i biblioteki BAC genomu żyta pozwala na poznanie pełnej sekwencji przypuszczalnych genów związanych z odpornością na choroby.
4. Zidentyfikowane przypuszczalne geny związane z odpornością na choroby należą do największej grupy genów odporności na choroby (nucleotide binding site-leucine-rich repeat) NB-LRR .
5. Gen *Sck33* prawdopodobnie odgrywa rolę w odpowiedzi na porażenie rdzą brunatną u żyta.
6. W genomie żyta występują ortologi kilku znanych genów udomowienia innych gatunków zbóż, jednak poziom zachowania ewolucyjnego nie jest wysoki i nie pozwala na bezpośrednie wykorzystanie ich sekwencji do zaprojektowania starterów do amplifikacji ich żytnich ortologów, jednak wykorzystanie sekwencji genomu referencyjnego częściowo rozwiązuje ten problem.

8. ANEKS

Tab. A1. Lista starterów do amplifikacji markerów DArT potencjalnie związanych z odpornością na choroby zidentyfikowanych w siedmiu klonach BAC. Jako „-”, oznaczono brak produktu PCR.

	Klon BAC	Marker DArT	Starter Fw	Starter Rv	Dł.produktu (pz)	PCR
1	S31p7D16	399303	AAGGACGACCATGTGCTTTT	TGGGTTGGCTATTAGGACCA	476	+
		506254	TGTTGAGCTACCTTGACAGG	GCAACGCATCCAAGAGAAGA	231	+
2	S34p7K24	505272	GGCAGAATCAGAGTGCTCAA	GTGGAGCATGTCAAGGCATA	177	+
		389402	TGCAGATTGGGGAGATTGGT	GCAGTTCCTACAGGTCTCTC	360	+
		508217	GCGATCAATCCCAAGCTTTT	TTGAGGTGTATCCCTCCTGG	186	-
		505250	TCAGGTTGCTACCATTTCAGC	AATGCTATGGAAGGACGAGC	325	-
		507036	CAGCCGTTCTTTTCATGTCCA	AGTGGATGACGGACTAAAGC	397	-
		390286	GACGTTTCACTTCAGTCGCT	TGCTCAATCTCCCAGGTACT	345	-
3	S21p7A1	505520	TCTGTTGGTGTGAACGAGTC	TGCTGGGTTTCATGCTATCTG	306	+
		506219	CCGTAGGATCCTCTCTTGGT	GGCCGTGAATCTTGTCAAAC	156	-
4	S24p1H6	506540	TATACCGAGGGTGCCATCAT	CAGCTATTTCTCCTCCACGG	330	+
		401485	TATACCGAGGGTGCCATCAT	GCTTGTGAAACTCATGCTGC	193	-
5	S37p4L8	399682	TCGGTGGAGAAGTGTCCAAA	TAGGCGAGAAACGAGTGGAA	315	+
		505760	TAACGAATCGGCGTGTGAAT	CTCGGCCTGACTTACTGAAC	374	+
		348128	-	-	-	-
6	S3p5G8	389394	GCTACAAGCATTGGAAGCAC	GAAACCACCCCGTGAGAAAT	412	+
		398573	AAGTTGTTGAGGGGCAGAAC	CACCAGCCGACTCAGATTAG	212	+
		389569	CAAACCTGCGAAGGCTAGTT	ACAGCTATGCCATGATACGC	479	-
7	S4p4E5	506219	CCGTAGGATCCTCTCTTGGT	GGCCGTGAATCTTGTCAAAC	156	+
		505974	ATGACAGCTTGCATCAACCA	GGTGTGAACGAGTCAACAGT	239	-
		509502	ATCGCATCACGTTTGGATCA	TCCACGTGTCTTTCAACCTG	351	-

Tab. A2. Startery do amplifikacji potencjalnych żytnich homologów genów udomowienia, zaprojektowane na podstawie genów zidentyfikowanych u spokrewnionych gatunków (oznaczone „,*”) oraz genomu referencyjnego żyta linii wsobnej Lo7. Na czerwono zaznaczono startery umożliwiające znalezienie sekwencji homologicznych do poszukiwanych genów związanych z udomowieniem w linii wsobnej żyta L318.

Nazwa genu	Starter Fw	Starter Rv	Oczekiwana dl. produktu	Dł. amplikonu
<i>qSH1</i>	GAGGACGCGTAGGAGTAGAA	GCGAGAAGCTCCGGTTC	133	brak amplifikacji
	CCTTTGCAGAAAGAAAGGCG*	TGAAAAAGATGGTCGCCACA*	1225	brak amplifikacji
<i>Q</i>	AAGAGGGACAACATCATCGC	ACCTGTACGTTTCGGGTAGAA	368	368
	TTGAATTCCTTTTGCCCCGA*	GCGCGTGTGATGATAATGTG*	964	brak amplifikacji
	CTGATGATGACGACTGCTCC*	AGGCCAGCGATTGATTAGTG*	375	brak amplifikacji
<i>sh4</i>	AGAACCAGTGCAACGACAAG	GAGCATTACCGGAAACGGA	492	brak amplifikacji
	CGCTCGGTTGATTAGGAGAG*	CACTGCACGCAGCTTTATTC*	1355	brak amplifikacji
<i>GIF1</i>	GAACGTCTTGGAGGCGTAGA	CCGGTACGACTACTACACCAT	127	127
	CTCCGGACTGGGTATCACTT*	CGTAGAAGTTGCCGTAGTCG*	826	brak amplifikacji
<i>BTR1</i>	GAAGGCGATGTACCTGGATG	ATGAGAAGCATGAACCGGAC	417	brak amplifikacji
	CAATGGAAGGCGATGTACCA*	ACAAACGTACCAACCAACGA*	905	brak amplifikacji
<i>BTR2</i>	CCTGAGCATGACGTTGAAGAC	ACGAATGAGATAGCGGAGGC	372	372
	CTTGACCTACATCAACGCGA*	TCTACGCCGTA CTACAGAT*	638	brak amplifikacji
	CGCTATCGACCGGATCAAG*	TTCGTCTACGCCGTA CTCA*	480	brak amplifikacji
	GAGCATGACGTTGAAGACGC	ACGTCAGCGAGACGAATGAG	380	brak amplifikacji
	GAGAAGACGTTGCCCATGAC	AGCGAGACGAATGAGATAGCG	209	brak amplifikacji
	ACGGTTGTGTA CTTACAGCG	GTCAACTCCGTCTTCGAGAT	600	600
<i>Rht-B1</i>	CCTATGGCTTGCCCTGTATC*	CCCTACCTCGCTCTCTATC*	675	brak amplifikacji
	CCAGCAGGGGTATCTGCTTC	TGGATCCAAATCCCAACCCG	861	brak amplifikacji
	TCCTGGTACTCCCGCTTCAT	TCCATCGATCCAACGCTGTG	323	brak amplifikacji
	AGGTACACCTCCGACATCAC	CGTCAACTCCGTCTTCGAG	294	brak amplifikacji
<i>Rht-D1</i>	AGATAGAGAGGCGAGGTAGC*	CTCGAAGACTGAGTTGACGG*	1355	brak amplifikacji
	GGCGCTATGACAAGCAAATC	GAACCATCCGAGGTGAAGTT	156	156
<i>Vrn1</i>	GGAGATTCGCACGTACGAT	ATGACTCGGTGGAGAACTCG	158	158
	AGTTCTCCACCGAGTCATGT*	GCTCAGAAATGGATTCTGTGC*	291	brak amplifikacji
<i>GA20ox2</i>	ACACCTGGAAGAACCCGTG	GCAAGAACTATCCCTGACGC	371	brak amplifikacji
	AGCTCACACACCTCTCATCT*	GACACCTGGAAGAACCCGT*	413	413

Tab. A3. Lista wszystkich genów przewidzianych przez FGENESH w klonach BAC zsekwencjonowanych metodą Nanopore wraz z wynikami porównania ich sekwencji do sekwencji zawartych w bazie nukleotydowej NCBI ("nucleotide nr") oraz w bazie białkowej NCBI ("protein RefSeq").

Klon BAC ^a	L. genów ^b	Gen Fgenesh ^c	Dł. mRNA ^d	L. egzonów ^e	Pozycja w klonie BAC ^f	Baza NCBI "nucleotide nr" ^g	Baza NCBI "protein RefSeq" ^h	Q ⁱ	E ^j	I ^k
S30p2A6	20	gen 1	780	1	481..2474	brak podobieństwa	brak podobieństwa			
		gen 2	1536	4	19888...22763	S. cereale clone HK-26 repeat region		63	0.0	94.80
							Disease resistance protein RPM1 [H. vulgare]	63	2,00E-85	42.77
		gen 3	165	2	24016..22763	S. cereale clone BAC956-D7, complete sequence	brak podobieństwa	98	3,00E-34	96.81
		gen 4	3852	9	27404..35812	T. aestivum gamma gliadin-B1		77	0.0	56.79
							putative Opie4 pol protein [Zea mays]	91	0.0	81.77
		gen 5	186	1	35971..38975	brak podobieństwa	brak podobieństwa			
		gen 6	129	2	43596..45238	brak podobieństwa	brak podobieństwa			
		gen 7	3936	6	45449..53219	brak podobieństwa	polyprotein [H. vulgare]	95	0.0	78.96
		gen 8	2190	1	63431..66888	S.cereale clone BAC956-D7, complete sequence	brak podobieństwa	100	0.0	97.99
		gen 9	1527	2	67533..70513	S.cereale clone BAC956-D7, complete sequence	brak podobieństwa	99	0.0	94.39
		gen 10	2025	5	70725..76513	brak podobieństwa	plasmid-partitioning protein SopA [E.coli]	47	0.0	89.80
		gen 11	3183	8	76820..83675	brak podobieństwa	brak podobieństwa			
		gen 12	303	1	84344..85622	T. aestivum gamma gliadin-A1	brak podobieństwa	99	8,00E-109	88.29
		gen 13	1740	5	85965..89355	H. vulgare retrotransposon Sandra5; splicing factor SF2-like protein (SF2) and universal stress protein (USP1) genes		95	0.0	82.25

						putative receptor protein kinase ZmPK1 [H. vulgare]	99	0.0	80.39	
		gen 14	1437	4	98370..103136	T. aestivum clone BAC WCS1891K06 nonfunctional BTR1 (Btr1) gene; BTR2 (Btr2), BTR2-like protein	44	0.0	89.60	
						BTR2-A-like protein [T. aestivum]	40	1,00E-48	61.46	
		gen 15	1308	2	103317..108352	T. aestivum clone BAC WCS1891K06 nonfunctional BTR1 (Btr1) gene; BTR2 (Btr2), BTR2-like protein	99	0.0	88.04	
						BTR2-A [T. aestivum]	44	2,00E-53	61.11	
		gen 16	426	4	111304..112929	brak podobieństwa	histone H3.3-like [A. tauschii subsp. tauschii]	97	2,00E-79	85.51
		gen 17	390	2	118239..120681	brak podobieństwa	brak podobieństwa			
		gen 18	903	4	123346..125803	H. vulgare clone HV_Mba442-A15, complete sequence	92	1,00E-131	79.89	
						vacuolar protein sorting-associated protein 35B-like [H. vulgare]	72	1,00E-72	56.19	
		gen 19	231	1	125971..126972	brak podobieństwa	brak podobieństwa			
		gen 20	3423	9	131924..142168	S. cereale clone BAC956-D7, complete sequence	99	0.0	94.16	
S16p 7H5	16	gen 1	3369	7	86..5553	T. aestivum omega gliadin-D1	brak podobieństwa	95	0.0	88.11
		gen 2	1038	9	7316..13518	S. cereale MADS-box protein vrn-1 mRNA, partial cds	40	8,00E-121	99.21	
						MADS-box protein vrn-1 [S. cereale]	35	2,00E-30	71.20	
		gen 3	20820	1	19913..21760	T. aestivum vernalization 1 (Vrn1) gene, Vrn1-a allele, exon 1 and partial cds	98	1,00E-34	100.00	
						MADS-box transcription factor 14-like [Aegilops tauschii subsp. tauschii]	100	4,00E-89	98.94	

		gen 3	20820	1	19913..21760	brak podobieństwa	Vrn-A1a [T. aestivum]	100	9,00E-37	100.00
		gen 4	999	1	25378..26662	T. aestivum omega gliadin-D1	brak podobieństwa	100	0.0	92.61
		gen 5	2952	9	27523..31622	brak podobieństwa	microtubule cross-linking factor 1-like isoform X1 [A. tauschii subsp. tauschii]	83	0.0	58.88
		gen 5	2952	9	27523..31622	brak podobieństwa	serine/arginine repetitive matrix protein 2-like isoform X2 [A. tauschii]	78	2,00E-160	57.09
						T. aestivum gamma gliadin-A1		80	0.0	92.34
		gen 6	6480	15	31655..43943	T. monococcum actin (ACT-1) gene, putative resistance protein (RGA-2)	brak podobieństwa	99	0.0	95.17
		gen 7	321	2	45616..47595	brak podobieństwa	brak podobieństwa			
		gen 8	990	2	50321..53800	T.aestivum omega gliadin-D1	brak podobieństwa	100	0.0	89.31
		gen 9	444	3	60417..62706	T. aestivum omega gliadin-D1	brak podobieństwa	100	7,00E-130	97.15
		gen 10	252	2	65218..67816	S.cereale DNA for dispersed repeat sequence (R173-3)	brak podobieństwa	56	4,00E-65	99.31
		gen 11	2670	10	75274..86051	brak podobieństwa	brak podobieństwa			
		gen 12	571	5	87501..93407	T. aestivum omega gliadin-D1		96	0.0	90.80
		gen 13	2217	6	96188..101558	brak podobieństwa	brak podobieństwa			
		gen 14	4287	8	112536..118969	brak podobieństwa	brak podobieństwa			
		gen 15	1557	3	120330..123587	T. aestivum clone BAC 1354M21 cytosolic acetyl-CoA carboxylase (Acc-2)		71	0.0	94.04
		gen 15	1557	3	120330..123587		titin-like [A. tauschii]	52	5,00E-57	48.70
		gen 16	2190	1	124644..127715	S. cereale clone BAC956-D7, complete sequence		100	0.0	98.68
S30p5H22	29	gen 1	525	1	8965..10949	brak podobieństwa	brak podobieństwa			
		gen 2	3720	10	12079 - 20368	A. tauschii chromosome 1Ds prolamin	brak podobieństwa	71	0.0	92.25

		gen 3	1602	2		brak podobieństwa	brak podobieństwa			
		gen 4	876	1	26661 - 27536	T. aestivum isolate Dctg3 alpha-gliadin (Gli-D2) gene	aspartyl protease family protein, partial [Lactobacillus salivarius]	97	0.0	89.30
							uncharacterized protein	74	4,00E-116	77.17
		gen 5	339	1	35680 - 36018	S. cereale clone HK15-21 repeat region	brak podobieństwa	100	7,00E-104	86.13
		gen 6	387	1	37580 - 38513	brak podobieństwa	brak podobieństwa			
		gen 7	1656	5	41288 - 46037	brak podobieństwa	brak podobieństwa			
		gen 8	648	1	46177 - 46824	T. aestivum clone 4BS Ms1 gene	brak podobieństwa	90	4,00E-159	84.24
		gen 9	3804	14	48618 - 58288	S. cereale putative cell wall invertase c mRNA	brak podobieństwa	43	0.0	97.83
		gen 10	1134	5	64760 - 67161	A. tauschii chromosome 1Ds prolamin	brak podobieństwa	90	0.0	88.49
		gen 11	417	3	73427 - 74236	T. aestivum omega gliadin-D1	brak podobieństwa	86	1,00E-97	94.92
		gen 12	690	3	74710 - 77311	S. cereale clone HK-26 repeat region	brak podobieństwa	76	0.0	96.98
		gen 13	3651	13	79309 - 93498	brak podobieństwa	brak podobieństwa			
		gen 14	597	3	96508 - 98277	S.cereale DNA for dispersed repeat sequence (R173-1)	brak podobieństwa	61	2,00E-76	95.24
		gen 15	1065	3	100122 - 101582	H. vulgare vrs1 locus and Hox1 gene	brak podobieństwa	46	1,00E-151	86.94
		gen 16	576	1	103301 - 103876	S. cereale clone HK-26 repeat region	brak podobieństwa	100	0.0	93.23
		gen 17	387	2	107252 - 108026	S. cereale clone HK-52 repeat region	brak podobieństwa			
		gen 18	1458	5	118128 - 120991	brak podobieństwa	brak podobieństwa			
		gen 19	339	3	123436 - 126549	brak podobieństwa	brak podobieństwa			
		gen 20	1770	9	126557 - 134634	brak podobieństwa	brak podobieństwa			
		gen 21	573	2	135911 - 136821	T. aestivum gamma gliadin-B1	brak podobieństwa	99	1,00E-153	96.18
		gen 22	2955	9	138998 - 143989	T. aestivum gamma gliadin-B2	brak podobieństwa	91	0.0	96.95
		gen 23	525	1	151327 - 153310	brak podobieństwa	brak podobieństwa			

		gen 24	3594	10	155544 - 162854	A. tauschii chromosome 1Ds prolamin gene locus	brak podobieństwa	62	0.0	92.38
		gen 25	1479	2	163509 - 166861	brak podobieństwa	brak podobieństwa			
		gen 26	528	1	169023 - 169550	T. aestivum gamma gliadin-D1	brak podobieństwa	100	2,00E-161	87.33
		gen 27	522	1	180144 - 180665	brak podobieństwa	40S ribosomal protein S5-1	91	5,00E-79	69.62
		gen 28	2307	8	183637 - 189654	brak podobieństwa	brak podobieństwa			
		gen 29	3921	14	190844 - 200696	brak podobieństwa	brak podobieństwa			
S12p 7K11	18	gen 1	1677	3	1 - 2411	brak podobieństwa	brak podobieństwa			
		gen 2	2352	6	4321 - 12990	brak podobieństwa	brak podobieństwa			
		gen 3	279	2	14184 - 15658	brak podobieństwa	brak podobieństwa			
		gen 4	2262	8	15870 - 19922	brak podobieństwa	brak podobieństwa			
		gen 5	1821	6	28977 - 33552	T. aestivum gamma gliadin-D1	brak podobieństwa	89	0.0	95.88
		gen 6	2292	11	40700 - 45972	T. aestivum gamma gliadin-D1	brak podobieństwa	97	0.0	94.58
		gen7	1893	2	56874 - 58897	S. cereale clone HK-26 repeat region	brak podobieństwa	45	0.0	95.08
		gen8	2253	2	63362 - 65652	T. aestivum gamma gliadin-D1		99	0.0	82.14
							Anthranilate synthase component I-1, chloroplastic [H.vulgare]	93	0.0	60.25
		gen 9	4101	6	66013 - 71951	T. aestivum gamma gliadin-D1		68	0.0	90.91
							Ji1 putative pol protein [Z. mays]	68	0.0	53.44
		gen10	321	1	74420 - 74740	A. tauschii uncharacterized LOC109758429 (LOC109758429), ncRNA	brak podobieństwa	100	2,00E-129	93.21
		gen 11	687	5	75914 - 79450	S. cereale clone HK13-27 repeat region	brak podobieństwa	34	1,00E-60	97.28
		gen 12	171	1	81442 - 84053	brak podobieństwa	brak podobieństwa			
		gen 13	786	7	87125 - 93275	brak podobieństwa	brak podobieństwa			

		gen 14	1833	7	93352 - 98483	S. cereale putative cell wall invertase c mRNA, complete cds	brak podobieństwa	81	0.0	82.06
		gen 15	3456	6	99926 - 105006	A. tauschii chromosome 1Ds prolamin		28	0.0	84.70
		gen 16	225	2	106558 - 107128		putative cell wall invertase c [S.cereale]	75	3,00E-05	55.07
		gen 17	594	3	111461 - 113443	brak podobieństwa	brak podobieństwa			
		gen 18	309	1	121355 - 121663	T. aestivum gamma gliadin-B1	brak podobieństwa	100	3,00E-77	84.24
S19p 4N17	21	gen 1	219	2	1451 - 1778	A. tauschii chromosome 1Ds prolamin gene locu	brak podobieństwa	50	2,00E-37	93.75
		gen 2	729	1	7813 - 9155	brak podobieństwa	brak podobieństwa			
		gen 3	219	2	9495 - 9821	A. tauschii chromosome 1Ds prolamin gene locu	brak podobieństwa	100	3,00E-36	93.75
		gen 4	906	2	11448 - 12661	brak podobieństwa	brak podobieństwa			
		gen 5	3564	11	15897 - 24652	T. aestivum gamma gliadin-A1	brak podobieństwa	77	0.0	87.32
		gen 6	2562	7	27683 - 34918	brak podobieństwa	brak podobieństwa			
		gen 7	690	3	44913 - 46371	H. vulgare cultivar Dairokkaku ZCCT-Hb (VRN2)	brak podobieństwa	37	3,00E-99	91.92
		gen 8	3129	11	56457 - 63571	T. aestivum gamma gliadin-D1	brak podobieństwa	80	0.0	85.75
		gen 9	1914	5	64985 - 69626	brak podobieństwa	brak podobieństwa			
		gen 10	3084	8	72339 - 76579	S. cereale clone BAC956-D7		100	0.0	91.34
							putative retrotransposon protein [Phyllostachys edulis]	95	0.0	48.96
		gen 11	1047	2	78372 - 79570	S. cereale RAPD marker OPR5-727 genomic sequence	brak podobieństwa	48	8,00E-148	85.63
		gen 12	3411	11	81756 - 89054	T. aestivum gamma gliadin-D1	brak podobieństwa	58	0.0	92.99
		gen 13	639	1	89681 - 90319	T. aestivum isolate Dctg3 alpha-gliadin (Gli-D2) gene	brak podobieństwa	100	0.0	92.67
		gen 14	1746	2	93161 - 95143	T. aestivum gamma gliadin-D1	brak podobieństwa	95	0.0	97.05

							hypothetical protein PTTG_30483 [Puccinia triticina 1-1 BBBB Race 1]	75	0.0	93.89
		gen 15	900	4	95668 - 97346	T. aestivum omega gliadin-D1	brak podobieństwa	95	0.0	96.48
		gen 16	1920	5	98256 - 100857	T.aestivum omega gliadin-D1	brak podobieństwa	95	0.0	96.39
		gen 17	1989	5	102994 - 106212	T. monococcum nucleolin 2-like protein gene	brak podobieństwa	100	0.0	92.50
		gen 19	219	2	119906 - 120232	A. tauschii chromosome 1Ds prolamin gene locus	brak podobieństwa	50	2,00E-37	93.75
		gen 20	816	2	121932 - 122858	T. aestivum omega gliadin-D1		97	0.0	97.11
							hypothetical protein TRIUR3_14679 [T. urartu]	95	3,00E-133	80.69
		gen 21	2367	10	126307 - 130118	T. aestivum gamma gliadin-A1	brak podobieństwa	88	0.0	92.30
S18p 1K5	24	gen 1	603	5	353 - 2854	brak podobieństwa	brak podobieństwa			
		gen 2	369	4	5249 - 6897	S.cereale clone BAC956-D7	brak podobieństwa	79	1,00E-82	95.48
		gen 3	3246	6	7426 - 11730	S. cereale clone BAC956-D8	brak podobieństwa	93	0.0	95.08
		gen 4	285	1	12922 - 13206	S. cereale clone HK-52 repeat region	brak podobieństwa	100	8,00E-93	89.16
							MULTISPECIES: LEA type 2 family protein [unclassified Coralloccoccus]	57	6.0	41.07
		gen 5	396	1	13993 - 14388	S. cereale DNA for dispersed repeat sequence	brak podobieństwa	100	2,00E-85	81.84
		gen 6	2073	1	17023 - 19095	S. cereale clone HK-26 repeat region		54	0.0	94.15
							retrotransposon gag domain-containing protein, partial [A.urinae]	55	2,00E-168	67.18
		gen 7	2712	8	21395 -25811	brak podobieństwa	brak podobieństwa			
		gen 8	3186	5	26372 - 30193	T. monococcum actin (ACT-1) gene, putative resistance protein (RGA-2)		78	0.0	87.38
		gen 9			35713 - 38030	brak podobieństwa	brak podobieństwa			

		gen 10	447	1	48743 - 49189	T. aestivum omega gliadin-D1		100	3,00E-159	89.74
							Disease resistance protein RPM1 [H. vulgare]	98	6,00E-38	59.33
		gen 11	2466	6	66155 - 73072	brak podobieństwa	brak podobieństwa			
		gen 12	573	1	75540 - 77358	brak podobieństwa	brak podobieństwa			
		gen 13	1035	7	80429 - 86501	brak podobieństwa	brak podobieństwa			
		gen 14	3111	3	87873 - 91476	T.aestivum omega gliadin-D1	brak podobieństwa	100	0.0	90.04
						H. vulgare subsp. vulgare Lks2 gene	brak podobieństwa	100	0.0	86.13
						H. vulgare subsp. vulgare cultivar Nure frost resistance H2 gene	brak podobieństwa	100	0.0	86.03
		gen 15	969	8	102237 - 104524	T. aestivum floral homeotic protein (Q) mRNA		100	0.0	93.12
						H. vulgare APETALA2-like protein (AP2L1) mRNA		100	0.0	89.46
							APETALA2-like protein [A. tauschii]	96	9,00E-174	86.92
		gen 16	2052	5	117263 - 21445	brak podobieństwa	brak podobieństwa			
		gen 17	369	4	123844 - 125491	S. cereale clone BAC956-D7, complete sequence	brak podobieństwa	96	1,00E-82	95.48
		gen 18	2274	4	127654 - 130327	S. cereale clone BAC956-D7, complete sequence	brak podobieństwa	99	0.0	95.40
		gen 19	285	1	131523 - 131807	S. cereale clone HK-52 repeat region		100	2,00E-96	89.16
							MULTISPECIES: LEA type 2 family protein [unclassified Coralloccoccus]	57	6.1	41.07
		gen 20	387	1	132595 - 132981	S. cereale DNA for dispersed repeat sequence (R173-3)	brak podobieństwa	98	3,00E-97	82.82
		gen 21	2073	1	135629 - 137701	S. cereale clone HK-26 repeat region		54	0.0	94.15

						retrotransposon gag domain-containing protein [<i>Aerococcus urinae</i>]	55	2,00E-168	67.18	
		gen 22	2910	7	140179 - 143841	<i>T. monococcum</i> actin (ACT-1) gene, putative resistance protein (RGA-2)	brak podobieństwa	79	4,00E-99	71.91
		gen 23	2958	5	145038 - 149403	<i>S. cereale</i> clone BAC 17C17 pSc250 retrotransposons		99	0.0	87.03
						<i>T. monococcum</i> actin (ACT-1) gene, partial cds; putative chromosome condensation factor (CCF), putative resistance protein (RGA-2)	99	0.0	85.08	
		gen 24	786	3	155322 - 156383	<i>A. tauschii</i> subsp. <i>Strangulata</i> LOC120965180		43	1,00E-110	86.67
						DNA-binding protein HEXBP-like [<i>Ae. tauschii</i>]	77	8,00E-70	87.78	

a – zeskwencjonowane klony BAC; b – łączna liczba genów zidentyfikowanych w poszczególnych klonach; c – zidentyfikowane geny; d – długość mRNA każdego z genów; e – liczba egzonów przewidzianych w każdym z genów; f – pozycja genu w klonie BAC; g- wyniki porównania sekwencji przypuszczalnych genów do sekwencji zawartych w bazie nukleotydowej NCBI ("nucleotide nr"); h - wyniki porównania sekwencji przypuszczalnych genów do sekwencji zawartych w bazie białkowej NCBI ("protein RefSeq"); i - pokrycie [%]; j - E-wartość; k - poziom identyczności [%]

9. LITERATURA

Alachiotis, N., Stamatakis, A., Pavlidis, P., (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets, *Bioinformatics*, 28 (17), 2274–2275.

Alachiotis, N., & Pavlidis, P. (2018). RAI_{SD} detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications biology*, 1(1), 1-11.

Al-Beyroutiová, M., Sabo, M., Slezziak, P., Dušínský, R., Birčák, E., Hauptvogel, P., Švec, M. (2016). Evolutionary relationships in the genus *Secale* revealed by DArTseq DNA polymorphism. *Plant Systematics and Evolution*, 302(8), 1083–1091.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Akira, S., Takeda, K., Kaisho, T., (2001). Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol*; 2: 675680.

Ansorge, WJ. (2009). Next-generation DNA sequencing techniques. *N Biotechnol* 25(4):195–203.

Ariyadasa, R., & Stein, N. (2012). Advances in BAC-based physical mapping and map integration strategies in plants. *Journal of Biomedicine and Biotechnology*, 2012.

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., ... & Distelfeld, A. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, 357(6346), 93-97.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, 3(10), e3376.

Bakera, B., Makowska, B., Groszyk, J., Niziołek, M., Orczyk, W., Bolibok-Brągoszewska, Hromada-Judycka, A., Rakoczy-Trojanowska, M. (2015). Structural characteristics of ScBx genes controlling the biosynthesis of hydroxamic acids in rye (*Secale cereale* L.). *Journal of Applied Genetics*, 56(3), 287–298.

- Barkley, N., Wang, M. (2008). Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals. *Curr Genomics*, 9:212-226.
- Bartoš, J., Paux, E., Kofler, R., Havránková, M., Kopecký, D., Suchánková, P., ... & Doležel, J. (2008). A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. *BMC Plant Biology*, 8(1), 1-12.
- Bauer, E., Schmutzer, T., Barilar, I., Mascher, M., Gundlach, H., Martis, M. M., ... & Scholz, U. (2017). Towards a whole-genome sequence for rye (*Secale cereale* L.). *The Plant Journal*, 89(5), 853-869.
- Beales, J., Laurie, D.A., Devos, K.M. (2005). Allelic variation at the linked AP1 and PhyC loci in hexaploid wheat is associated but not perfectly correlated with vernalization response. *Theor Appl Genet*, 110(6):1099-1107.
- Becraft, P.W. (2002). Receptor kinase signaling in plant development. *Annu. Rev. Cell Dev. Biol.*, 18, 163–192.
- Benita, Y., Oosting, R. S., Lok, M. C., Wise, M. J., & Humphery-Smith, I. (2003). Regionalized GC content of template DNA as a predictor of PCR success. *Nucleic acids research*, 31(16), e99-e99.
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., ... & Roe, P. M. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *nature*, 456(7218), 53-59.
- Berlin, K., Koren, S., Chin, C. S., Drake, J. P., Landolin, J. M., & Phillippy, A. M. (2015). Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nature biotechnology*, 33(6), 623-630.
- Bhattacharai, G., Shi, A., Feng, C., Dhillon, B., Mou, B., & Correll, J. C. (2020). Genome wide association studies in multiple spinach breeding populations refine downy mildew race 13 resistance genes. *Frontiers in Plant Science*, 11, 563187.
- Blainey, P. C. (2013). The future is now: single-cell genomics of bacteria and archaea. *FEMS microbiology reviews*, 37(3), 407-427.

Bolibok-Bragoszewska, H., Heller-Uszyńska, K., Wenzl, P., Uszyński, G., Kilian, A., & Rakoczy-Trojanowska, M. (2009). DArT markers for the rye genome-genetic diversity and mapping. *Bmc Genomics*, 10(1), 1-11.

Bolibok-Bragoszewska, H., Targońska, M., Bolibok, L., Kilian, A., & Rakoczy-Trojanowska, M. (2014). Genome-wide characterization of genetic diversity and population structure in *Secale*. *BMC plant biology*, 14(1), 1-15.

Borzęcka, E., Hawliczek-Strulak, A., Bolibok, L., Gawroński, P., Tofil, K., Milczarski, P., Stojalowski, S., Myśków, B., Targońska-Karasek, M., Gradzielewska, A., Smolik, M., Kilian, A., Bolibok-Bragoszewska, H. (2018). Effective BAC clone anchoring with genotyping-by-sequencing and Diversity Arrays Technology in a large genome cereal rye. *Scientific Reports*, 8(1).

Cannon, SB., Zhu, H., Baumgarten, AM., Spangler, R., May, G., Cook, DR., Young, ND., (2002). Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *J Mol Evol*; 54: 548-562.

Cao, D., Wang, D., Li, S., Li, Y., Hao, M., & Liu, B. (2022). Genotyping-by-sequencing and genome-wide association study reveal genetic diversity and loci controlling agronomic traits in triticale. *Theoretical and Applied Genetics*, 135(5), 1705-1715.

Cavagnaro, P. F., Chung, S. M., Szklarczyk, M., Grzebelus, D., Senalik, D., Atkins, A. E., & Simon, P. W. (2009). Characterization of a deep-coverage carrot (*Daucus carota* L.) BAC library and initial analysis of BAC-end sequences. *Molecular Genetics and Genomics*, 281(3), 273-288.

Chandler, P. M., Marion-Poll, A., Ellis, M., & Gubler, F. (2002). Mutants at the Slender1 locus of barley cv Himalaya. Molecular and physiological characterization. *Plant physiology*, 129(1), 181-190.

Cheng, Z., Buell, C. R., Wing, R. A., & Jiang, J. (2002). Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. *Chromosome Research*, 10(5), 379-387.

Cheung, V. G., & Nelson, S. F. (1996). Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one

nanogram of genomic DNA. *Proceedings of the National Academy of Sciences*, 93(25), 14676-14679.

Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., ... & Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature methods*, 13(12), 1050-1054.

Cross, A., Li, J. B., Waugh, R., Golicz, A. A., & Pourkheirandish, M. (2022). Grain dispersal mechanism in cereals arose from a genome duplication followed by changes in spatial expression of genes involved in pollen development. *Theoretical and Applied Genetics*, 135(4), 1263-1277.

Cullum, R., Alder, O., Hoodless, P.A.. (2011). The next generation: using new sequencing technologies to analyse gene regulation. *Respirology*, 16:210-222.

Cviková, K., Cattonaro, F., Alaux, M., Stein, N., Mayer, K. F., Doležel, J., & Bartoš, J. (2015). High-throughput physical map anchoring via BAC-pool sequencing. *BMC plant biology*, 15(1), 1-13.

Czech, B., & Hannon, G. J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nature Reviews Genetics*, 12(1), 19-31.

Devos, K. M., Atkinson, M. D., Chinoy, C. N., Francis, H. A., Harcourt, R. L., Koebner, R. M. D., ... & Gale, M. D. (1993). Chromosomal rearrangements in the rye genome relative to that of wheat. *Theoretical and Applied Genetics*, 85(6), 673-680.

Dixon, M.S., Golstein, C., Thomas, C.M., van der Biezen, E.A., & Jones, J.D.G. (2000). Genetic complexity of pathogen perception by plants: The example of Rcr3, a tomato gene required specifically by Cf-2. *Proc. Natl. Acad. Sci. USA* 97, 8807–8814.

Doebley, J. (1989). Isozymic evidence and the evolution of crop plants. In *Isozymes in plant biology* (pp. 165-191). Springer, Dordrecht.

Doebley, J., & Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. *The Plant Cell*, 10(7), 1075-1082.

Doebley, J. (2004). The genetics of maize evolution. *Annual review of genetics*, 38(1), 37-59.

Doebley, J. F., Gaut, B. S., & Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell*, 127(7), 1309-1321.

Dong, F., Song, J., Naess, S. K., Helgeson, J. P., Gebhardt, C., & Jiang, J. (2000). Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. *Theoretical and Applied Genetics*, 101(7), 1001-1007.

Dyda, M., Tyrka, M., Gołębiowska, G., Rapacz, M., & Wędzony, M. (2022). Genetic mapping of adult-plant resistance genes to powdery mildew in triticale. *Journal of applied genetics*, 63(1), 73-86.

Edwards, M.A. & Henry, R.J. (2011). DNA sequencing methods contributing to new directions in cereal research. *J.Cereal Sci.* 54, 395-400.

Evans, L. T., Evans, L. T., & Evans, L. T. E. (1998). *Feeding the ten billion: plants and population growth*. Cambridge University Press.

FAOSTAT. Food and Agriculture Data. Dostępne online: <https://fao.org> .

Fischer, S., Melchinger, A. E., Korzun, V., Wilde, P., Schmiedchen, B., Möhring, J., Piepho, H. P., Dhillon, S. B., Würschum, T., & Reif, J. C. (2010). Molecular marker assisted broadening of the Central European heterotic groups in rye with Eastern European germplasm. *Theoretical and applied genetics*, 120(2), 291-299.

Feng, C., Bluhm, B. H., & Correll, J. C. (2015). Construction of a spinach bacterial artificial chromosome (BAC) library as a resource for gene identification and marker development. *Plant molecular biology reporter*, 33(6), 1996-2005.

Flavell, R.B., Bennett, M.D., Smith, J.B., Smith, D.B., (1974). Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* 12, 257–269.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... & Venter, J. C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223), 496-512.

Flintham, J. E., Börner, A., Worland, A. J., & Gale, M. D. (1997). Optimizing wheat grain yield: effects of Rht (gibberellin-insensitive) dwarfing genes. *The Journal of Agricultural Science*, 128(1), 11-25.

Fransz, P. F., Armstrong, S., de Jong, J. H., Parnell, L. D., van Drunen, C., Dean, C., ... & Jones, G. H. (2000). Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell*, 100(3), 367-376.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Knaap, E. V. D., Cong, B., ... & Tanksley, S. D. (2000). fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289(5476), 85-88.

Frijters, A. C. J., Zhang, Z., Van Damme, M., Wang, G. L., Ronald, P., & Michelmore, R. W. (1997). Construction of a bacterial artificial chromosome library containing large Eco RI and Hin dIII genomic fragments of lettuce. *Theoretical and Applied Genetics*, 94(3), 390-399.

Furtado, A., Bundock, P. C., Banks, P. M., Fox, G., Yin, X., & Henry, R. J. (2015). A novel highly differentially expressed gene in wheat endosperm associated with bread quality. *Scientific reports*, 5(1), 1-14.

Galiba, G., Vagujfalvi, A., Li, C. X., Soltesz, A., Dubcovsky, J. (2009). Regulatory genes involved in the determination of frost tolerance in temperate cereals. *Plant Sci* 2009, 176(1):12-19.

Gan, S. T., Wong, W. C., Wong, C. K., Soh, A. C., Kilian, A., Low, E. T. L., Massawe, F., & Mayes, S. (2018). High density SNP and DArT-based genetic linkage maps of two closely related oil palm populations. *Journal of applied genetics*, 59(1), 23-34.

Gawroński, P., Pawełkowicz, M., Tofil, K., Uszyński, G., Sharifova, S., Ahluwalia, S., Tyrka, M., Wędzony, M., Kilian, A. & Bolibok-Brągoszewska, H. (2016). DArT markers effectively target gene space in the rye genome. *Frontiers in plant science*, 7, 1600.

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, 18, 9-19.

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., Chan, C. K. K., Severn-Ellis, A., McCombie R. W., Parkin, I. A. P., Paterson, A. H., Pires, C. J., Sharpe, A. G., Tang, H., Teakle, G. R., Town, C. D., Batley, J., & Edwards, D. (2016). The

pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature communications*, 7(1), 1-8.

Gómez, M. I., Islam-Faridi, M. N., Woo, S. S., Czeschin Jr, D., Zwick, M. S., Stelly, D. M., ... & Wing, R. A. (1997). FISH of a maize *sh2*-selected sorghum BAC to chromosomes of *Sorghum bicolor*. *Genome*, 40(4), 475-478.

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., & McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome research*, 25(11), 1750-1756.

Gore, M. A., Chia, J. M., Elshire, R. J., Sun, Q., Ersoz, E. S., Hurwitz, B. L., ... & Buckler, E. S. (2009). A first-generation haplotype map of maize. *Science*, 326(5956), 1115-1117.

Gråsten, S. M., Juntunen, K. S., Poutanen, K. S., Gylling, H. K., Miettinen, T. A., & Mykkänen, H. M. (2000). Rye bread improves bowel function and decreases the concentrations of some compounds that are putative colon cancer risk markers in middle-aged women and men. *The Journal of nutrition*, 130(9), 2215-2221.

Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., Odden, A.R., Comai, L., Henikoff, S. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*, 164: 731-740.

Grzebelus, D., Iorizzo, M., Senalik, D., Ellison, S., Cavagnaro, P., Macko-Podgorni, A., ... & Baranski, R. (2014). Diversity, genetic mapping, and signatures of domestication in the carrot (*Daucus carota* L.) genome, as revealed by Diversity Arrays Technology (DArT) markers. *Molecular breeding*, 33(3), 625-637.

Hackauf, B., Korzun, V., Wortmann, H., Wilde, P., & Wehling, P. (2012). Development of conserved ortholog set markers linked to the restorer gene *Rfp1* in rye. *Molecular breeding*, 30(3), 1507-1518.

Haseneyer, G., Schmutzer, T., Seidel, M., Zhou, R., Mascher, M., Schön, C. C., ... & Bauer, E. (2011). From RNA-seq to large-scale genotyping-genomics resources for rye (*Secale cereale* L.). *BMC Plant Biology*, 11(1), 1-13.

Hawliczek, A., Bolibok, L., Tofil, K., Borzęcka, E., Jankowicz-Cieślak, J., Gawroński, P., ... & Bolibok-Braęoszewska, H. (2020). Deep sampling and pooled amplicon sequencing reveals hidden genic variation in heterogeneous rye accessions. *BMC genomics*, 21(1), 1-16.

Himmelbach, A., Ruban, A., Walde, I., Šimková, H., Doležel, J., Hastie, A., Stein, N., & Mascher, M. (2018). Discovery of multi-megabase polymorphic inversions by chromosome conformation capture sequencing in large-genome plant species. *The Plant Journal*, 96(6), 1309-1316.

Holma, R., Hongisto, S. M., Saxelin, M., & Korpela, R. (2010). Constipation is relieved more by rye bread than wheat bread or laxatives without increased adverse gastrointestinal effects. *The Journal of nutrition*, 140(3), 534-541.

Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., Peñagaricano, F., Lindquist, E., Pedraza A. M, Barry, K., de Leon, N., Shawn M. Kaeppler., S. M., & Buell, C. R. (2014). Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell*, 26(1), 121-135.

Holušová, K., Vrána, J., Šafář, J., Šimková, H., Balcárková, B., Frenkel, Z., ... & Bartoš, J. (2017). Physical map of the short arm of bread wheat chromosome 3D. *The plant genome*, 10(2), plantgenome2017-03.

Howden, B. P., McEvoy, C. R., Allen, D. L., Chua, K., Gao, W., Harrison, P. F., ... & Stinear, T. P. (2011). Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. *PLoS pathogens*, 7(11), e1002359.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res*, 19(6):1068–76.

Hulse-Kemp, A. M., Bostan, H., Chen, S., Ashrafi, H., Stoffel, K., Sanseverino, W., Li, L., Cheng, S., Schatz, M. C., Garvin, T., Toit, L. J., Tseng, E., Chin, J., Iorizzo, M., Van Deynze, A. (2021). An anchored chromosome-scale genome assembly of spinach improves annotation and reveals extensive gene rearrangements in euasterids. *The Plant Genome*, e20101.

Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., & Harris, S. R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome biology*, 16, 294.

Hwang, CF., Williamson VM. (2003). Leucine-rich repeat-mediated intramolecular interactions in nematode recognition and cell death signaling by the tomato resistance protein Mi. *Plant J*; 34: 585-593.

Iorizzo, M., Ellison, S., Senalik, D., (2016). A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat Genet* 48, 657–666.

Iovene, M., Cavagnaro, P. F., Senalik, D., Buell, C. R., Jiang, J., & Simon, P. W. (2011). Comparative FISH mapping of *Daucus* species (Apiaceae family). *Chromosome Research*, 19(4), 493-506.

Jackson, S. A., Cheng, Z., Li Wang, M., Goodman, H. M., & Jiang, J. (2000). Comparative fluorescence in situ hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. *Genetics*, 156(2), 833-838.

Jia, QJ., Zhang, JJ., Westcott, S., Zhang, XQ., Bellgard, M. (2009). GA-20 oxidase as a candidate for the semidwarf gene *sdw1/denso* in barley. *Funct Integr Genomics* 9:255–262.

Jiang, J., Gill, B. S., Wang, G. L., Ronald, P. C., & Ward, D. C. (1995). Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proceedings of the National Academy of Sciences*, 92(10), 4487-4491.

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., ... & Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, 546(7659), 524-527.

Jonsson, K., Andersson, R., Knudsen, K. E. B., Hallmans, G., Hanhineva, K., Katina, K., ... & Landberg, R. (2018). Rye and health-Where do we stand and where do we go?. *Trends in Food Science & Technology*, 79, 78-87.

Jordan, B., Charest, A., Dowd, J. F., Blumenstiel, J. P., Yeh, R. F., Osman, A., Housman, D. E., & Landers, J. E. (2002). Genome complexity reduction for SNP genotyping analysis. *Proceedings of the National Academy of Sciences*, 99(5), 2942-2947.

Jørgensen, J.H. (1994). Genetics of powdery mildew resistance in barley. *Crit. Rev. Plant Sci.* 13, 97–119.

Jung, H., Winefield, C., Bombarely, A., Prentis, P., & Waterhouse, P. (2019). Tools and strategies for long-read sequencing and de novo assembly of plant genomes. *Trends in plant science*, 24(8), 700-724.

Kantar, M. B., Nashoba, A. R., Anderson, J. E., Blackman, B. K., & Rieseberg, L. H. (2017). The genetics and genomics of plant domestication. *BioScience*, 67(11), 971-982.

Khan, M. A., Han, Y., Zhao, Y. F., Troglio, M., & Korban, S. S. (2012). A multi-population consensus genetic map reveals inconsistent marker order among maps likely attributed to structural variations in the apple genome. *PloS one*, 7(11), e47864.

Kircher, M., Heyn, P., & Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC genomics*, 12(1), 1-14.

Klein, P. E., Klein, R. R., Cartinhour, S. W., Ulanich, P. E., Dong, J., Obert, J. A., ... & Mullet, J. E. (2000). A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. *Genome Research*, 10(6), 789-807.

Kochian, L. V., Hoekenga, O. A., & Pineros, M. A. (2004). How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorous efficiency. *Annual review of plant biology*, 55(1), 459-493.

Konishi S, Lin SY, Fukuta Y, Izawa T, Sasaki Tet al (2005). Molecular cloning of a major QTL, qSH-1, controlling seed shattering habit in rice. In: *Plant and animal genome XIII abstract*, W306.

Koren, S., Walenz, BP., Berlin, K., Miller, JR., Phillippy, AM. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*.

Korlach, J., Bjornson, K. P., Chaudhuri, B. P., Cicero, R. L., Flusberg, B. A., Gray, J. J., ... & Turner, S. W. (2010). Real-time DNA sequencing from single polymerase molecules. In *Methods in enzymology* (Vol. 472, pp. 431-455). Academic Press.

Kosova, K., Prasil, I. T., Vitamvas, P. (2008). The relationship between vernalization-and photoperiodically-regulated genes and the development of frost tolerance in wheat and barley. *Biol Plant*, 52(4):601-615.

Krishnan, S. & Dayanandan, P. (2003). Structural and histochemical studies on grain-filling in the caryopsis of rice (*Oryza sativa* L.). *J. Biosci.* 28, 455–469.

Krzyszowska, M. (2019). Odpowiedzi roślin na czynniki biotyczne. W: J. Kopcewicz. S. Lewak (red.), *Fizjologia roślin* (s. 710-722). Warszawa: Wydawnictwo Naukowe PWN

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, 19(9), 1639-1645.

Kurata, S., Kanagawa, T., Magariyama, Y., Takatsu, K., Yamada, K., Yokomaku, T., & Kamagata, Y. (2004). Reevaluation and reduction of a PCR bias caused by reannealing of templates. *Applied and Environmental Microbiology*, 70(12), 7545-7549.

Laehnemann, D., Borkhardt, A., & McHardy, A. C. (2016). Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Briefings in bioinformatics*, 17(1), 154-179.

Lenser, T., & Theißen, G. (2013). Molecular mechanisms involved in convergent crop domestication. *Trends in plant science*, 18(12), 704-714.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *methods*, 25(4), 402-408.

Li, C., Zhou, A. & Sang, T. (2006). Rice domestication by reducing shattering. *Science* 311, 1936–1939.

Li, C., Lin, F., An, D., Wang, W., & Huang, R. (2017). Genome sequencing and assembly by long reads in plants. *Genes*, 9(1), 6.

Li, B., Tian, L., Zhang, J., Huang, L., Han, F., Yan, S. (2014). Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in *Glycine max*. *BMC Genomics*. 15(1):1086.

Li, H., Vikram, P., Singh, R. P., Kilian, A., Carling, J., Song, J., ... & Singh, S. (2015). A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC genomics*, 16(1), 1-15.

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.

Liu, J., Liu, X., Dai, L., & Wang, G. (2007). Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. *Journal of genetics and genomics*, 34(9), 765-776.

Liu, H., McNicol, J., Bayer, M., Morris, J. A., Cardle, L., Marshall, Schulte D., Stein N., Shi BJ., Taudien S., Hedley, P. E. (2011). Highly parallel gene-to-BAC addressing using microarrays. *BioTechniques*, 50(3), 165–174.

Liu, J., Chen, J., Zheng, X., Wu, F., Lin, Q., Heng, Y., ... & Wan, J. (2017). GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nature Plants*, 3(5), 1-7.

Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, R., ... & Tedersoo, L. (2019). Relative Performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) Third-Generation Sequencing Instruments in Identification of Agricultural and Forest Fungal Pathogens. *Applied and Environmental Microbiology*, 85(21), e01368-19.

Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature methods*, 12(8), 733-735.

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics, proteomics & bioinformatics*, 14(5), 265-279.

- Luo, M. C., Xu, K., Ma, Y., Deal, K. R., Nicolet, C. M., & Dvorak, J. (2009). A high-throughput strategy for screening of bacterial artificial chromosome libraries and anchoring of clones on a genetic map constructed with single nucleotide polymorphisms. *BMC genomics*, 10(1), 1-10.
- Mantri, N.L., Ford, R., Coram, T.E. & Pang, E.C. (2007). Transcriptional profiling of chickpea genes differentially regulated in response to high-salinity, cold and drought, *BMC Genomics*, 8, 303.
- Marek, L.F. & Shoemaker, R.C. (1997). BAC contig development by fingerprint analysis in soybean. *Genome*, 40, 420-427.
- Mares, D. J., & Mrva, K. (2014). Wheat grain preharvest sprouting and late maturity alpha-amylase. *Planta*, 240(6), 1167-1178.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., ... & Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376-380.
- Maron, L. G., Guimarães, C. T., Kirst, M., Albert, P. S., Birchler, J. A., Bradbury, P. J., ... & Kochian, L. V. (2013). Aluminum tolerance in maize is associated with higher MATE1 gene copy number. *Proceedings of the National Academy of Sciences*, 110(13), 5241-5246.
- Martis, M. M., Zhou, R., Haseneyer, G., Schmutzer, T., Vrána, J., Kubaláková, M., ... & Stein, N. (2013). Reticulate evolution of the rye genome. *The Plant Cell*, 25(10), 3685-3698.
- McIntosh, R.A.; Ducovsky, W.J.; Appels, R.; Xia, X.C. (2017). Catalogue of Gene Symbols for Wheat: Supplement. KOMUGI Integrated Wheat Science Database. Available online: <https://shigen.nig.ac.jp/wheat/komugi/about/about.jsp> .
- Michelmore, R. W., & Meyers, B. C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome research*, 8(11), 1113-1130.
- Milczarski, P., Bolibok-Braęoszewska, H., Myśków, B., Stojalowski, S., Heller-Uszyńska, K., Góralaska, M., ... & Rakoczy-Trojanowska, M. (2011). A high density consensus map of rye (*Secale cereale* L.) based on DArT markers. *PLoS one*, 6(12), e28495.

- Milczarski, P., Hanek, M., Tyrka, M. & Stojalowski, S. (2016). The application of GBS markers for extending the dense genetic map of rye (*Secale cereale* L.) and the localization of the Rfc1 gene restoring male fertility in plants with the C source of sterility-inducing cytoplasm. *J. Appl. Genet.* 57, 439–451
- Meksem, K., Shultz, J., Tebbji, F., Jamai, A., Henrich, J., Kranz, H., ... & Lightfoot, D. A. (2005). A bacterial artificial chromosome based physical map of the *Ustilago maydis* genome. *Genome*, 48(2), 207-216.
- Messing, J., Crea, R., & Seeburg, P. H. (1981). A system for shotgun DNA sequencing. *Nucleic acids research*, 9(2), 309-321.
- Messing, J., & Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene*, 19(3), 269-276.
- Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., & Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *The Plant Journal*, 20(3), 317-332.
- Monaghan, J., & Zipfel, C. (2012). Plant pattern recognition receptor complexes at the plasma membrane. *Current opinion in plant biology*, 15(4), 349-357.
- Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92(5), 255-264.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNASeq. *Nat Methods*, 5:621-628.
- Muñoz-Amatriaín, M., Lonardi, S., Luo, M., Madishetty, K., Svensson, J. T., Moscou, M. J., ... & Close, T. J. (2015). Sequencing of 15 622 gene-bearing BAC s clarifies the gene-dense regions of the barley genome. *The Plant Journal*, 84(1), 216-227.
- Muterko, A. F., Balashova, I. A., Fayt, V. I., & Sivolap, Y. M. (2015). Molecular-genetic mechanisms of regulation of growth habit in wheat. *Cytology and Genetics*, 49(1), 58-71.
- Myśków, B., Stojalowski, S., Łań, A., Bolibok-Bragoszewska, H., Rakoczy-Trojanowska, M., & Kilian, A. (2012). Detection of the quantitative trait loci for α -amylase activity on a

high-density genetic map of rye and comparison of their localization to loci controlling preharvest sprouting and earliness. *Molecular breeding*, 30(1), 367-376.

Myśków, B., Hanek, M., Banek-Tabor, A., Maciorowski, R., & Stojałowski, S. (2014). The application of high-density genetic maps of rye for the detection of QTLs controlling morphological traits. *Journal of applied genetics*, 55(1), 15-26.

Naranjo, T., & Fernández-Rueda, P. (1991). Homoeology of rye chromosome arms to wheat. *Theoretical and applied genetics*, 82(5), 577-586.

Niedziela, A., Jarska W., Bednarek, PT., (2015). Wybrane elementy nowoczesnych rozwiązań wpływające na skuteczność programów. *Biul. IHAR 275*: 3 — 15.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome research*, 15(11), 1566-1575.

Nicholls, S.M.; Quick, J.C.; Tang, S.; Loman, N.J. (2019). Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *GigaScience*, 8, 1–9.

Noczeń, J., Puchta, M., & Czembor, J. H. (2018). Using the next generation DNA sequencing (technology NGS) in gene banks and plant breeding. A review. *Agronomy Science*, 73(1), 5-17.

Nowicka, A., Grzebelus, E., & Grzebelus, D. (2012). Fluorescent in situ hybridization with arbitrarily amplified DNA fragments differentiates carrot (*Daucus carota* L.) chromosomes. *Genome*, 55(3), 205–213. doi:10.1139/g2012-003.

Orban, L., Elshire, RJ., Glaubitz, JC., Sun, Q., Poland, JA., Kawamoto, K. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6(5):e19379.

Osipowski, P., Pawełkiewicz, M., Wojcieszek, M., Skarzyńska, A., Przybecki, Z., & Pląder, W. (2020). A high-quality cucumber genome assembly enhances computational comparative genomics. *Molecular Genetics and Genomics*, 295(1), 177-193.

Ozata, D.M., Gainetdinov, I., Zoch, A. (2019). PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* 20, 89–108.

Pachota, K., Niedziela, A., Orłowska, R., & Bednarek, P. (2016). Nowoczesne metody genotypowania DArT i GBS w hodowli gatunków roślin użytkowych. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, (279).

Pavlidis, P., Jensen, J.D., Stephan, W., (2010). Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics* 185:907–922.

Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, 30(9), 2224-2234.

Pavlidis, P., & Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research-Thessaloniki*, 24(1), 1-17.

Pearce, S., Saville, R., Vaughan, S. P., Chandler, P. M., Wilhelm, E. P., Sparks, C. A., ... & Thomas, S. G. (2011). Molecular characterization of Rht-1 dwarfing genes in hexaploid wheat. *Plant physiology*, 157(4), 1820-1831.

Pedrosa, A., Sandal, N., Stougaard, J., Schweizer, D., & Bachmair, A. (2002). Chromosomal map of the model legume *Lotus japonicus*. *Genetics*, 161(4), 1661-1672.

Pellicer, J., Hidalgo, O., Dodsworth, S., & Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes*, 9(2), 88.

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., ... & Harberd, N. P. (1999). ‘Green revolution’ genes encode mutant gibberellin response modulators. *Nature*, 400(6741), 256-261.

Peterson, D. G., Tomkins, J. P., Frisch, D. A., Wing, R. A., & Paterson, A. H. (2000). Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. *Journal of Agricultural genomics*, 5, 1-100.

Petroli, C. D., Sansaloni, C. P., Carling, J., Steane, D. A., Vaillancourt, R. E., Myburg, A. A., ... & Grattapaglia, D. (2012). Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus* genome. *PLoS One*, 7:e44684.

- Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., ... & Lagudah, E. (2013). The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science*, 341(6147), 786-788.
- Piasecka, I., Knozowski, P., Ropińska, P., Tomporowski, A., & Ignaszak, P. (2017). Badanie i ocena możliwości wykorzystania na cele energetyczne rozdrobnionych ziaren zbóż wiechlinowatych. *Acta Scientiarum Polonorum. Technica Agraria*, 16(1-2).
- Pinard, R., de Winter, A., Sarkis, G. J., Gerstein, M. B., Tartaro, K. R., Plant, R. N., Egholm, M., Rothberg J. M., & Leamon, J. H. (2006). Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics*, 7(1), 1-21.
- Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J., & Dickau, R. (2009). Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences*, 106(13), 5019-5024.
- Poland, JA., Rife, TW. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome* 5: 92-102.
- Poland, JA., Brown, PJ., Sorrells, ME., Jannink, JL., (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping by-sequencing approach. *PLoS One*, 7(2):e32253.
- Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M., ... & Komatsuda, T. (2015). Evolution of the grain dispersal system in barley. *Cell*, 162(3), 527-539.
- Priyadarshan, P. M. (2019). Molecular Breeding. In *PLANT BREEDING: Classical to Modern* (pp. 509-539). Springer, Singapore.
- Purugganan, M. D., & Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature*, 457(7231), 843-848.
- Rabanus-Wallace, M. T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., ... Stein, N. (2019). Chromosome-scale genome assembly provides insights into rye biology,

evolution, and agronomic potential. *BioRxiv*, 2019.12.11.869693.
<https://doi.org/10.1101/2019.12.11.869693> .

Rakoczy-Trojanowska, M., Krajewski, P., Bocianowski, J., Schollenberger, M., Wakuliński, W., Milczarski, P., ... & Kilian, A. (2017). Identification of single nucleotide polymorphisms associated with brown rust resistance, α -amylase activity and pre-harvest sprouting in rye (*Secale cereale* L.). *Plant molecular biology reporter*, 35(3), 366-378.

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, 19(1), 1-11.

Rangan, P., Furtado, A., & Henry, R. J. (2017). The transcriptome of the developing grain: a resource for understanding seed development and the molecular control of the functional and nutritional properties of wheat. *BMC genomics*, 18(1), 1-9.

Röder, M. S., Huang, X. Q., & Börner, A. (2008). Fine mapping of the region on wheat chromosome 7D controlling grain weight. *Functional & integrative genomics*, 8(1), 79-86.

Ronaghi, M. (2001). Pyrosequencing sheds light on DNA sequencing. *Genome research*, 11(1), 3-11.

Roselli, S., Olry, A., Vautrin, S., Coriton, O., Ritchie, D., Galati, G., ... & Hehn, A. (2017). A bacterial artificial chromosome (BAC) genomic approach reveals partial clustering of the furanocoumarin pathway genes in parsnip. *The Plant Journal*, 89(6), 1119-1132.

Rothberg, J., & Myers, J. (2011). Semiconductor sequencing for life. *Journal of Biomolecular Techniques: JBT*, 22(Suppl), S41.

Roux, S. R., Hackauf, B., Linz, A., Ruge, B., Klocke, B., & Wehling, P. (2004). Leaf-rust resistance in rye (*Secale cereale* L.). 2. Genetic analysis and mapping of resistance genes Pr3, Pr4, and Pr5. *Theoretical and Applied Genetics*, 110(1), 192-201.

Roux, S. R., Hackauf, B., Ruge-Wehling, B., Linz, A., & Wehling, P. (2007). Exploitation and comprehensive characterization of leaf-rust resistance in rye. *Vortr Pflanzenzüchtung*, 71, 144-150.

Šafář, J., Simkova, H., Kubaláková, M., Suchankova, P., Cihalikova, J., Bartos, J., ... & Lucretti, S. (2007). Generating resources for genomics of wheat homoeologous chromosome group 3: 3AS-and 3DS-specific BAC libraries. *Journal of Genetics and Breeding*, 61(1-2), 151-160.

Sahoo, N. (2017). Sequence base-calling through Albacore software: a part of the Oxford Nanopore technology (Doctoral dissertation).

Saisho, D., & Purugganan, M. D. (2007). Molecular phylogeography of domesticated barley traces expansion of agriculture in the Old World. *Genetics*, 177(3), 1765-1776.

Salamov, A. A., & Solovyev, V. V. (2000). Ab initio Gene Finding in Drosophila Genomic DNA, 516–522.

Santos, E., Benito, C., Silva-Navas, J., Gallego, F. J., Figueiras, A. M., Pinto-Carnide, O., & Matos, M. (2018). Characterization, genetic diversity, phylogenetic relationships, and expression of the aluminum tolerance MATE1 gene in *Secale* species. *Biologia plantarum*, 62(1), 109-120.

Sambrook, J., (2001). *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y. :Cold Spring Harbor Laboratory Press.

Sang, T. (2009). Genes and mutations underlying domestication transitions in grasses. *Plant Physiology*, 149(1), 63-70.

Sanger, F., Nicklen, S., Coulson, AR. (1977). DNA sequencing with chainterminating inhibitors. *Proc Natl Acad Sci USA*,74:5463-5467.

Saville, R. J., Gosman, N., Burt, C. J., Makepeace, J., Steed, A., Corbitt, M., ... & Nicholson, P. (2012). The ‘Green Revolution’ dwarfing genes play a role in disease resistance in *Triticum aestivum* and *Hordeum vulgare*. *Journal of experimental botany*, 63(3), 1271-1283.

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human molecular genetics*, 19(R2), R227-R240.

Scheunert, A., Dorfner, M., Lingl, T., & Oberprieler, C. (2020). Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing. *PLoS One*, 15(3), e0226234.

Schmidt, M. H. W., Vogel, A., Denton, A. K., Istace, B., Wormit, A., van de Geest, H., ... & Usadel, B. (2017). De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *The Plant Cell*, 29(10), 2336-2348.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science*. 2009;326:1112–5.

Schnable, J. C. (2015). Genome evolution in maize: from genomes back to genes. *Annual review of plant biology*, 66, 329-343.

Schreiber, M., Himmelbach, A., Börner, A., & Mascher, M. (2019). Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. *Evolutionary applications*, 12(1), 66-77.

Sekhwai, M. K., Li, P., Lam, I., Wang, X., Cloutier, S., & You, F. M. (2015). Disease resistance gene analogs (RGAs) in plants. *International journal of molecular sciences*, 16(8), 19248-19290.

Sharma, M. K., Sharma, R., Cao, P., Harkenrider, M., Jenkins, J., Grimwood, J., ... & Ronald, P. C. (2016). Targeted switchgrass BAC library screening and sequence analysis identifies predicted biomass and stress response-related genes. *BioEnergy Research*, 9(1), 109-122.

Shcherban, A. B., Khlestkina, E. K., Efremova, T. T., & Salina, E. A. (2013). The effect of two differentially expressed wheat VRN-B1 alleles on the heading time is associated with structural variation in the first intron. *Genetica*, 141(4), 133-141.

She, H., Qian, W., Zhang, H., Liu, Z., Wang, X., Wu, J., ... & Xu, Z. (2018). Fine mapping and candidate gene screening of the downy mildew resistance gene RPF1 in Spinach. *Theoretical and Applied Genetics*, 131(12), 2529-2541.

- Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., & Yano, M. (2008). Deletion in a gene associated with grain size increased yields during rice domestication. *Nature genetics*, 40(8), 1023-1028.
- Shneider, M. N., & Pekker, M. (2015). Pre-breakdown processes in a dielectric fluid in inhomogeneous pulsed electric fields. *Journal of Applied Physics*, 117(22), 224902.
- Silva-Navas, J., Benito, C., Téllez-Robledo, B., El-Moneim, A., & Gallego, F. J. (2012). The ScaACT1 gene at the Q alt5 locus as a candidate for increased aluminum tolerance in rye (*Secale cereale* L.). *Molecular Breeding*, 30(2), 845-856.
- Simkova H, Rakoczy-Trojanowska M. (2010). Konstrukcja biblioteki BAC żyta ozimego. III Polski Kongres Genetyki, Lublin, 12-15 września 2010.
- Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., & Faris, J. D. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics*, 172(1), 547-555.
- Sinnett, D., Richer, C., & Baccichet, A. (1998). Isolation of stable bacterial artificial chromosome DNA using a modified alkaline lysis method. *Biotechniques*, 24(5), 752-754.
- Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23-35.
- Soderlund, C., Humphray, S., Dunham, A., & French, L. (2000). Contigs built with fingerprints, markers, and FPC V4. 7. *Genome Research*, 10(11), 1772-1787.
- Solovyev, V. (2001). Statistical approaches in eukaryotic gene prediction. In: D.J. Balding, M. Bishop and C. Cannings (Eds.), *Handbook of Statistical Genetics*, John Wiley & Sons, Ltd, New York, pp. 83–127.
- Sosnowski, O., Charcosset, A. & Joets, J. (2012). Biomecator V3: An upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics* 28, 2082–2083.
- Suzuki, G., Ura, A., Saito, N., Do, G. S., Seo, B. B., Yamamoto, M., & Mukai, Y. (2001). BAC FISH analysis in *Allium cepa*. *Genes & genetic systems*, 76(4), 251-255.

Stein, N., Perovic, D., Kumlehn, J., Pellio, B., Stracke, S., Streng, S., ... & Graner, A. (2005). The eukaryotic translation initiation factor 4E confers multiallelic recessive BYmovirus resistance in *Hordeum vulgare* (L.). *The Plant Journal*, 42(6), 912-922.

Steuernagel, B., Witek, K., Krattinger, S. G., Ramirez-Gonzalez, R. H., Schoonbeek, H. J., Yu, G., Baggs, e., I. Witek, A., Yadav, I., Krasileva, K. V., Jones J. D. G., Uauy, C., Keller, B., J. Ridout, C., & Wulff, B. B. (2020). The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiology*, 183(2), 468-482.

Swaminathan, K., Alabady, M. S., Varala, K., De Paoli, E., Ho, I., Rokhsar, D. S., ... & Hudson, M. E. (2010). Genomic and small RNA sequencing of *Miscanthus × giganteus* shows the utility of sorghum as a reference genome sequence for Andropogoneae grasses. *Genome biology*, 11(2), 1-18.

Święcicka, M., Siedlecka, E., Bakera, B., Rakoczy-Trojanowska, M. (2019). Analiza ekspresji żytnich ortologów genów Lr w siewkach inokulowanych *Puccinia recondita f.sp. secalis*. XIV Ogólnopolska Konferencja Naukowa 'Nauka dla hodowli i nasiennictwa roślin uprawnych', Zakopane 5-8.02.2019.

Święcicka, M., Dmochowska-Boguta, M., Orczyk, W., Grądzielewska, A., Stochmal, A., Kowalczyk, M., ... & Rakoczy-Trojanowska, M. (2020). Changes in benzoxazinoid contents and the expression of the associated genes in rye (*Secale cereale* L.) due to brown rust and the inoculation procedure. *PLoS one*, 15(5), e0233807.

Takumi, S., Koyama, K., Fujiwara, K., & Kobayashi, F. (2011). Identification of a large deletion in the first intron of the *Vrn-D1* locus, associated with loss of vernalization requirement in wild wheat progenitor *Aegilops tauschii* Coss. *Genes & genetic systems*, 86(3), 183-195.

Targońska-Karasek, M., Bolibok-Braęoszewska, H., & Rakoczy-Trojanowska, M. (2017). DArTseq genotyping reveals high genetic diversity of polish rye inbred lines. *Crop Science*, 57(4), 1906-1915.

Tjin-Wong-Joe, A. F. (2004). Early Grain Filling Environment and Pre-Maturity α -Amylase Formation in UK Winter Wheat Genotypes. Open University (United Kingdom).

Tucker, E. J., Baumann, U., Kouidri, A., Suchecki, R., Baes, M., Garcia, M., Okada, T., Dong, C., Wu, Y., Sandhu, A., Singh, M., Langridge, P., Wolters, P., Albertsen, C. M., Cigan, A. M., & Whitford, R. (2017). Molecular identification of the wheat male fertility gene Ms1 and its prospects for hybrid breeding. *Nature Communications*, 8(1), 1-10.

Tulpová, Z., Luo, M. C., Toegelová, H., Visendi, P., Hayashi, S., Vojta, P., ... & Šimková, H. (2019). Integrated physical map of bread wheat chromosome arm 7DS to facilitate gene cloning and comparative studies. *New biotechnology*, 48, 12-19.

Vendelbo, N. M., Mahmood, K., Sarup, P., Hovmøller, M. S., Justesen, A. F., Kristensen, P. S., ... & Jahoor, A. (2021). Discovery of a Novel Leaf Rust (*Puccinia recondita*) Resistance Gene in Rye (*Secale cereale* L.) Using Association Genomics. *Cells*, 11(1), 64.

Venter, J.C., Adams, M.D., Sutton, G.G., Kerlavage, A.R., Smith, H.O. and Hunkapiller, M. (1998). Shotgun sequencing of the human genome. *Science*, 280, 1540-1542.

Visendi, P., Berkman, P. J., Hayashi, S., Golicz, A. A., Bayer, P. E., Ruperao, P., ... & Edwards, D. (2016). An efficient approach to BAC based assembly of complex genomes. *Plant Methods*, 12(1), 1-9.

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., ... & He, Z. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nature genetics*, 40(11), 1370-1374.

Wei, F., Gobelmann-Werner, K., Morroll, S.M., Kurth, J., Mao, L., Wing, R., Leister, D., Schulze-Lefert, P., and Wise, R.P. (1999). The Mla (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* 153, 1929–1948.

Wei, F., Wing, R. A., & Wise, R. P. (2002). Genome dynamics and evolution of the Mla (powdery mildew) resistance locus in barley. *The Plant Cell*, 14(8), 1903-1917.

Wehling, P., Linz, A., Hackauf, B., Roux, S. R., Ruge, B., & Klocke, B. (2003). Leaf-rust resistance in rye (*Secale cereale* L.). 1. Genetic analysis and mapping of resistance genes Pr1 and Pr2. *Theoretical and Applied Genetics*, 107(3), 432-438.

- Wen, W., Deng, Q., Jia, H., Wei, L., Wei, J., Wan, H., ... & Ma, Z. (2013). Sequence variations of the partially dominant DELLA gene Rht-B1c in wheat and their functional impacts. *Journal of experimental botany*, 64(11), 3299-3312.
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome biology*, 20(1), 1-10.
- Wise, R.P. (2000). Disease resistance: What's brewing in barley genomics. *Plant Dis.* 84, 1160–1170
- Wright, S. I., & Gaut, B. S. (2005). Molecular population genetics and the search for adaptive evolution in plants. *Molecular biology and evolution*, 22(3), 506-519.
- Wu, H. B., Wang, L., Wang, Q. S., Han, Y. J., Li, H. S., Zhou, W. L., & Tian, Y. (2015). Adenocarcinoma with BAC features presented as the nonsolid nodule is prone to be false-negative on 18F-FDG PET/CT. *BioMed Research International*, 2015.
- Xu, Y., Jia, Q., Zhou, G., Zhang, X. Q., Angessa, T., Broughton, S., ... & Li, C. (2017). Characterization of the sdw1 semi-dwarf gene in barley. *BMC plant biology*, 17(1), 1-10.
- Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., ... & Zhang, Q. (2008). Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nature genetics*, 40(6), 761-767.
- Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., & Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene VRN1. *Proceedings of the National Academy of Sciences*, 100(10), 6263-6268.
- Yang, S., Li, J., Zhang, X., Zhang, Q., Huang, J., Chen, J. Q., ... & Tian, D. (2013). Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. *Proceedings of the National Academy of Sciences*, 110(46), 18572-18577.
- Yao, H., Guo, L., Fu, Y., Borsuk, L. A., Wen, T. J., Skibbe, D. S., Cui, X., Scheffler, B. E., Cao, J., Emrich, S. J., Ashlock, D. A., & Schnable, P. S. (2005). Evaluation of five ab initio gene prediction programs for the discovery of maize genes. *Plant molecular biology*, 57(3), 445-460.

- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, 13(1), 1-11.
- Yim, Y. S., Davis, G. L., Duru, N. A., Musket, T. A., Linton, E. W., Messing, J. W., ... & Coe Jr, E. H. (2002). Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. *Plant Physiology*, 130(4), 1686-1696.
- Yoon, J., Cho, L. H., Kim, S. L., Choi, H., Koh, H. J., & An, G. (2014). The BEL 1-type homeobox gene SH 5 induces seed shattering by enhancing abscission-zone development and inhibiting lignin biosynthesis. *The Plant Journal*, 79(5), 717-728.
- You, F. M., Luo, M. C., Xu, K., Deal, K. R., Anderson, O. D., & Dvorak, J. (2010). A new implementation of high-throughput five-dimensional clone pooling strategy for BAC library screening. *BMC genomics*, 11(1), 1-10.
- Yu, Y., Tomkins, J.P., Waugh, R., Frisch, D.A., Kudrna, D., Kleinhofs, A., Brueggeman, R.S., Muehlbauer, G.J., Wise, R.P. & Wing, R.A. (2000). A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. *Theor. Appl. Genet.* 101, 1093–1099
- Yu, J., Hu, S., Wang, J., Wong, G. K. S., Li, S., Liu, B., ... & Yang, H. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *science*, 296(5565), 79-92.
- Zeng, X., Mishina, K., Jia, J., Distelfeld, A., Maughan, P. J., Kikuchi, S., ... & Komatsuda, T. (2020). The brittle rachis trait in species belonging to the Triticeae and its controlling genes *Btr1* and *Btr2*. *Frontiers in plant science*, 11, 1000.
- Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18(5), 821-829.
- Zhang, H. B., & Wu, C. (2001). BAC as tools for genome sequencing. *Plant Physiology and Biochemistry*, 39(3-4), 195-209.

Zhang, P., Li, W., Fellers, J., Friebe, B., & Gill, B. S. (2004). BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements. *Chromosoma*, 112(6), 288-299.

Zhang, W., Ciclitira, P., & Messing, J. (2014). PacBio sequencing of gene families—A case study with wheat gluten genes. *Gene*, 533(2), 541-546.

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., ... & Huang, X. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature genetics*, 50(2), 278-284.

Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J. (2010). The next-generation sequencing technology and application. *Protein Cell*, 1:520-536.

Wyrażam zgodę na udostępnienie mojej pracy w czytelniach Biblioteki SGGW w tym w Archiwum Prac Dyplomowych SGGW

Ewa Borzęcka
.....
(czytelny podpis autora pracy)