

Instytut Chemii Bioorganicznej, Polska Akademia Nauk

Zakład Biologii RNA



Joanna Gracz

Alternatywny splicing jako mechanizm odpowiedzi na
stres herbicydowy u kukurydzy zwyczajnej

Rozprawa doktorska

Promotor: prof. dr hab. Tomasz Twardowski

Promotor pomocniczy: dr Agata Tyczewska

Poznań, 2014

Panu prof. Tomaszowi Twardowskiemu składam serdeczne podziękowania za powierzenie mi ciekawego tematu badań, opiekę naukową oraz życzliwość i wyrozumiałość w trakcie realizacji pracy.

Pani dr Agacie Tyczewskiej za wszystkie rady, niezwykle wsparcie, cenne dyskusje i wiarę we mnie serdecznie dziękuję.

Pani dr Kamilli Bąkowskiej-Żywickiej bardzo dziękuję za okazaną życzliwość, wskazówki i konstruktywną krytykę.

Panu prof. Markowi Figlerowiczowi i prof. Jackowi Błażewiczowi dziękuję za stworzenie możliwości współpracy naukowej.

Dziękuję za współpracę, dyskusje i pomoc

dr Agnieszce Żmieńko

dr Aleksandrze Świercz

Arkadiuszowi Hoffa

Za wszelką pomoc i niezapomnianą atmosferę dziękuję wszystkim pracownikom Zakładu Biologii RNA i Samodzielnego Zakładu Epigenetyki.

Szczególnie dziękuję mojej mamie, siostrze i wszystkim, którzy wspierali mnie w dążeniu do celu.

Spis treści

1. Wstęp	3
2. Wprowadzenie literaturowe	5
2.1 Dojrzewanie cząsteczek pre-mRNA.....	5
2.1.1 Splicing, czyli składanie mRNA	8
2.1.2 Alternatywny splicing	11
2.1.3 Rola AS w ekspresji genów.....	13
2.2 Kukurydza zwyczajna – pochodzenie, charakterystyka	16
2.2.1 Zastosowanie i hodowla kukurydzy zwyczajnej	19
2.3 Stres herbicydowy.....	22
2.3.1 Herbicyd Roundup®	24
3. Cel pracy.....	27
4. Materiały i metody.....	28
4.1. Materiały	28
4.1.1 Materiał roślinny	28
4.1.2 Stosowane odczynniki chemiczne.....	31
4.1.3 Oligonukleotydy	31
4.1.5 Zestawy komercyjne	32
4.1.6 Roztwory	33
4.1.7 Bazy danych.....	33
4.2 Metody	40
4.2.1 Hodowla wybranych linii kukurydzy zwyczajnej	40
4.2.2 Izolacja całkowitego RNA oraz trawienie DNA.....	41
4.2.3 Analiza ilościowa i jakościowa całkowitego RNA	42
4.2.3 Izolacja DNA i reakcja PCR w gradiencie temperatur.....	43
4.2.4 Reakcja odwrotnej transkrypcji sprzężona z reakcją PCR	45
4.2.4 Oczyszczanie i analiza produktów PCR w żelu agarozowym.....	46
4.2.5 Sekwencjonowanie transkryptomu	48
4.2.6 Analiza bioinformatyczna.....	52
5. Wyniki i dyskusja	57

5.1 Analiza bioinformatyczna sekwencji chromosomu 1 kukurydzy zwyczajnej pod kątem zdarzeń AS	57
5.1.2 Analiza ekspresji transkryptów dla wybranych genów z zastosowaniem reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR	69
5.1.3 Dyskusja	74
5.2 Analiza wyników sekwencjonowania transkryptomów	79
5.2.1 Sekwencjonowanie i analiza jakości uzyskanych odczytów	79
5.2.2 Analiza różnic w ekspresji genów między badanymi liniami pod wpływem stresu herbicydowego.....	85
5.2.3 Badanie różnic w profilu składania mRNA między analizowanymi bibliotekami.....	94
5.2.4 Dyskusja	108
6. Podsumowanie.....	119
7. Bibliografia	121
Streszczenie.....	127
Tytuł i streszczenie w języku angielskim	130
Wykaz ważniejszych skrótów	132
Spis rycin i tabel	133
Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS	135
Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS	151
Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych	160
Aneks 4 Dorobek naukowy	176

1. Wstęp

Kukurydza zwyczajna jest jedną z roślin stosunkowo niedawno wprowadzonych do hodowli na terenie Polski, a w naszej strefie klimatycznej, ze względu na swoje pochodzenie może być nawet uznana za roślinę egzotyczną. Ogromny postęp w jej hodowli jest możliwy przede wszystkim dzięki zastosowaniu nowych odmian (opartych głównie o wykorzystanie zjawiska heterozji) coraz lepiej dostosowanych do warunków środowiska. O dużym sukcesie i znaczeniu ekonomicznym kukurydzy przesądziły jej cechy użytkowe, możliwość wykorzystania całej biomasy rośliny jako paszy, żywności i surowca przemysłowego.

Poznanie sekwencji genomów roślin uprawnych miało dać jednoznaczną odpowiedź na pytanie jak zwiększyć plonowanie, jakość i ilość uzyskiwanego ziarna czy też w jaki sposób uzyskać odmiany odporniejsze na warunki stresu biotycznego czy abiotycznego. Dziś wiemy, że otrzymanie holistycznego obrazu zmian warunkujących postęp hodowlany, w tym także odmiany odporne na warunki stresowe, obejmuje nie tylko znajomość genomu rośliny, ale także całej ścieżki ekspresji genów oraz bogatego spektrum sposobów regulacji przepływu informacji genetycznej w organizmie, aż po badanie poziomu i różnic w ilości syntetyzowanych białek czy metabolitów wtórnych. Dotychczasowe badania genomiczne pozwoliły określić wiele genów oraz ich produktów biorących udział w regulacji ekspresji informacji genetycznej, przewodzeniu sygnałów czy wytworzeniu stanu tolerancji lub odporności w organizmie na czynnik stresowy. Dopiero jednak rozszerzenie pojęcia genomiki funkcjonalnej o dziedziny takie jak transkryptomika, proteomika czy metabolomika uzmysłowiło nam w pełni jak skomplikowana i wielopoziomowa sieć zależności warunkuje przystosowanie się rośliny, jej dalszy wzrost i rozwój pod wpływem nieoptymalnych warunków zewnętrznych.

Opracowanie nowych, wysokowydajnościowych metod sekwencjonowania drugiej generacji (NGS, ang. *next-generation sequencing*) pozwala obecnie na szybkie i coraz tańsze określenie sekwencji genomów różnych organizmów, ale także badanie puli cząsteczek RNA z precyzją i dokładnością przewyższającą stosowane dotychczas w tym celu mikromacierze. Badania transkryptomu z wykorzystaniem metod NGS określane jest mianem RNA-Seq (ang.

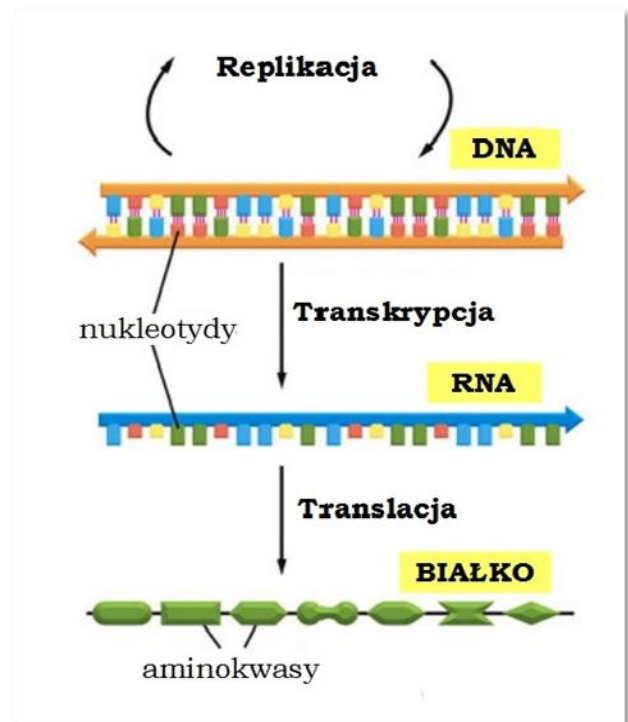
RNA sequencing) i umożliwia ono nie tylko mapowanie zsekwencjonowanych transkryptów do genomu, ale także precyzyjne określenie ich ilości. Dzięki wysokiej czułości i rozdzielczości przy pomocy tej metody możliwe jest znacznie lepsze scharakteryzowanie niejednorodnej puli transkrybowanych RNA, m. in. opisanie nowych, aktywnych transkrypcyjnie rejonów genomów, poznanie wielu nowych izoform transkryptów znanych genów czy dokładne określenie 5' i 3' miejsc splicingowych. Generowana ogromna ilość danych stanowi wyzwanie i jednocześnie wymusza ciągły postęp w analizie, przechowywaniu i zarządzaniu tymi informacjami, czyniąc tym samym transkryptomikę nowatorskim obszarem badań i zachęca do ekscytującej próby zgłębienia tej dziedziny wiedzy.

2. Wprowadzenie literaturowe

2.1 Dojrzewanie cząsteczek pre-mRNA

W 1957 roku na posiedzeniu Towarzystwa Biologii Eksperymentalnej Francis Crick zaproponował hipotezę przepływu informacji w komórkach, którą określił mianem centralnego dogmatu biologii molekularnej. Dogmat ten stanowi podstawę zrozumienia przekazywania informacji pomiędzy trzema grupami biopolimerów występującymi w komórce - między DNA, RNA i białkami. Zakłada, że w najbardziej powszechnym i ogólnym przypadku, informacja przechowywana przez DNA w procesie replikacji może być skopiuowana do kolejnej cząsteczki DNA lub w procesie transkrypcji do cząsteczki matrycowego RNA (mRNA). Ta ostatnia z kolei stanowi podstawę procesu translacji w wyniku którego powstają białka (Ryc. 1).

Już z przyjętych przez centralny dogmat założeń wynika, że synteza RNA (transkrypcja) jest zasadniczym etapem ekspresji informacji genetycznej. U organizmów eukariotycznych transkrypcja zachodzi w jądrze i dopiero dojrzały transkrypt jest uwalniany do cytoplazmy, gdzie zachodzi proces translacji, w przeciwieństwie do organizmów prokariotycznych, u których translacja może zacząć się na nie do końca zsyntetyzowanym mRNA. Przestrzenne i czasowe rozdzielenie transkrypcji i translacji pozwala organizmom eukariotycznym regulować ekspresję genów w znacznie bardziej skomplikowany sposób, przyczyniając się do zwiększenia bogactwa ich postaci i funkcji. Drugą istotną różnicę między tymi grupami organizmów stanowi zakres obróbki jakiej jest poddawany transkrypt RNA. Zarówno prokarioty, jak i eukarioty modyfikują cząsteczki tRNA i rRNA, jednakże tylko u tych ostatnich



Ryc. 1

Schemat założeń centralnego dogmatu biologii molekularnej zaproponowany przez F. Cricka w 1957 roku [1].

wszystkie pierwotne produkty transkrypcji prowadzonej przez polimerazę II RNA podlegają intensywnej obróbce potranskrypcyjnej. Wśród nich największą grupę stanowią prekursorowe mRNA (pre-mRNA, ang. *precursor mRNA*), a proces ich modyfikacji określany jest jako dojrzewanie mRNA (ang. *mRNA processing*). Obejmuje on cztery podstawowe procesy: tworzenie struktury czapeczki (kap, ang. *cap*) na 5' końcu transkryptu, cięcie i poliadenylację końca 3', redagowanie RNA (ang. *RNA editing*) oraz splicing, zwany także składaniem RNA (rozdział 2.1.1).

Kiedy pierwotny transkrypt syntetyzowany przez polimerazę II osiągnie długość ok. 25-30 nukleotydów, wówczas koniec 5' transkryptu jest modyfikowany. Pierwsze dwa etapy reakcji katalizowane są przez dimeryczny enzym kapujący (ang. *capping enzyme*), który jest zasocjowany z ufosforylowaną karboksy-terminalną domeną CTD (ang. *carboxy-terminal tail domain*) polimerazy II (transkrypty syntetyzowane przez polimerazę I i III nie posiadają na końcu 5' struktury kapu). W pierwszym etapie jedna reszta fosforanowa GTP jest uwalniana na drodze hydrolizy. Następnie difosforan na końcu 5' transkryptu atakuje atom fosforu α cząsteczki GTP tworząc wiązanie 5',5'-trifosforanowe. Kolejne dwa etapy reakcji katalizowane są przez dwie metylotransferazy, które umożliwiają przeniesienie grupy metylowej z S-adenozylometioniny na azot w pozycji 7 końcowej guaniny (N-7) - powstaje tzw. kap 0. Grupy 2' hydroksylowe rybozy dwóch przyległych nukleotydów mogą także zostać zmetylowane - powstają wówczas struktury typu kap 1 i kap 2. W ten sposób zmodyfikowany koniec transkryptu stymuluje translację oraz wpływa na stabilność cząsteczki mRNA zabezpieczając ją przed działaniem fosfataz i nukleaz [2].

Dojrzewanie końca 3' obejmuje rozcięcie pierwotnego transkryptu przez specyficzną endonukleazę oraz dodanie ogona poli(A), który nie jest kodowany przez DNA. Proces poliadenylacji wymaga obecności dwóch podstawowych czynników - działających w sposób *trans* białek, które przeprowadzają reakcję cięcia i poliadenylacji oraz czynników *cis*, czyli sygnałów sekwencyjnych obecnych w pre-mRNA. Dotychczas zidentyfikowano około 25-30 wysoce konserwatywnych białek biorących udział w rozpoznawaniu sygnału, cięciu i syntezie ogona poli(A) [3]. W większości transkryptów u ssaków sygnał poliadenylacji obejmuje bogatą w reszty adeniny sekwencję znajdującą się 10-30 nukleotydów powyżej miejsca cięcia (najczęściej jest to sekwencja AAUAAA) oraz bogatą w reszty GU lub U sekwencję ok. 50

nukleotydów poniżej miejsca cięcia [4]. Natomiast w organizmach roślinnych każde miejsce poliadenylacji jest definiowane przez interakcje między trzema typami elementów *cis*:

- miejscem cięcia i poliadenylacji (CS, ang. *cleavage/polyadenylation site*) - najczęściej jest to dinukleotyd CA albo UA zlokalizowany w rejonie bogatym w reszty U,
- elementem zlokalizowanym 13-30 nukleotydów powyżej miejsca poliadenylacji (NUE, ang. *near-upstream element*)- zazwyczaj jest to 6-10 nukleotydowa sekwencja bogata w reszty A,
- elementem znajdującym się nawet do 100 nukleotydów powyżej CS (FUE, ang. *far-upstream element*), bogatym w GU (często jest to sekwencja UUGUAA).

Roślinne geny charakteryzują się ponadto obecnością więcej niż jednego miejsca poliadenylacji w obrębie transkryptów - każde z tych miejsc posiada oddzielny elementy CS oraz NUE, ale często kilka z nich może być kontrolowane przez pojedynczy sygnał FUE [5].

Po rozpoznaniu odpowiednich sygnałów sekwencyjnych przez czynniki białkowe, następuje utworzenie wielobiałkowego kompleksu, do którego przyłączana jest także polimeraza poli(A) (PAP, ang. *poly(A) polymerase*). Tuż po cięciu pierwotnego transkryptu przez endonukleazę i uwolnieniu jego końca 3', PAP dodaje do niego około 200 reszt A, których donorem jest ATP. Wykazano, że ogon poli(A) zwiększa stabilność mRNA, wydajność jego translacji, wpływa także na eksport z jądra do cytoplazmy oraz na długość życia cząsteczki mRNA, która zależy od szybkości degradacji jej ogona poli(A) [6].

Proces dojrzewania RNA może obejmować także potranskrypcyjną zmianę informacji zawartej w niektórych mRNA w reakcji zwanej redagowaniem RNA (ang. *RNA editing*). U roślin dotyczy ona przede wszystkim transkryptów powstających w mitochondriach i chloroplastach. Opisano kilka typów redagowania: deaminacja adenozyiny do inozyiny w tRNA, cytydiny do urydiny w mRNA i tRNA (najczęściej występująca) oraz zamiana urydiny na cytydynę w kilku przypadkach opisanych u roślin nagonasiennych. Zamiana reszty C do U występuje najczęściej w sekwencji kodującej i powoduje zmianę na kodon niosący inny aminokwas niż to wynika z sekwencji DNA, co może mieć znaczący wpływ na funkcje pełnione przez kodowane białko. Często zmiany te powodują zwiększenie hydrofobowości polipeptydu i dotyczą białek błonowych. Uszkodzenie białek zaangażowanych w proces

redagowania RNA może nawet prowadzić do letalnego zaburzenia rozwoju rośliny - przykładem może być mutacja w genie kodującym 19 białko biogenezy chloroplastów (clb19, ang. *chloroplast biogenesis 19*) [7]. Redagowanie RNA jest więc procesem niezbędnym do prawidłowego wzrostu i rozwoju rośliny, który dodatkowo powoduje zwiększenie molekularnej różnorodności cząsteczek mRNA, a tym samym zwiększa różnorodność kodowanych przez nie białek.

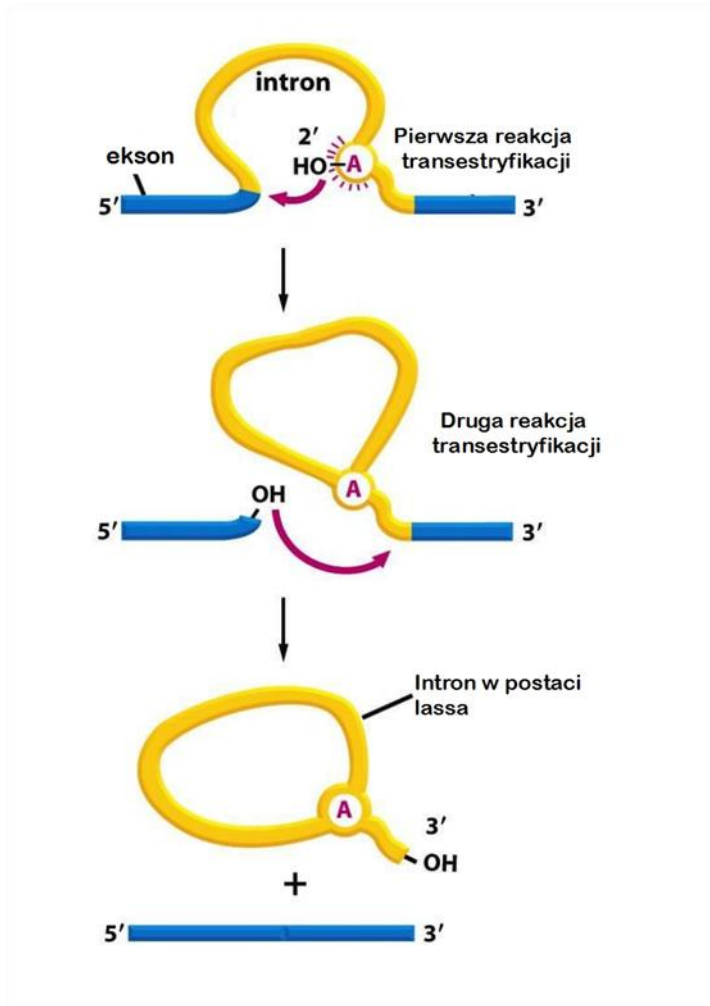
2.1.1 Splicing, czyli składanie mRNA

Kolejnym procesem umożliwiającym powstanie w pełni funkcjonalnej i dojrzałej cząsteczki mRNA jest splicing, czyli składanie mRNA. Podczas splicingu następuje wycinanie intronów z pierwotnego transkryptu oraz ligacja eksonów, czyli sekwencji mRNA kodujących białko. Sekwencje intronów i eksonów muszą zostać prawidłowo rozpoznane, wycięte i połączone, aby powstała dojrzała cząsteczka mRNA, która będzie mogła ulec translacji. Odpowiedni splicing umożliwia także ekspresję informacji dotyczących: transportu cząsteczek mRNA z jądra do cytoplazmy, związanych z ich lokalizacją w komórce oraz z degradacją mRNA zawierających sekwencje nonsensowe (NMD, ang. *nonsense mediated mRNA decay*).

Reakcja splicingu jest katalizowana przez strukturę będącą dużym kompleksem rybonukleoproteinowym, zwanym spliceosomem, w którego skład wchodzi 5 cząsteczek snRNP (ang. *small nuclear ribonucleoprotein*). RNA, z którego są zbudowane, jest bogate w uracyl, stąd cząsteczki te określane są jak U1, U2 itd. snRNP. Dużą część spliceosomu stanowią rozmaite białka (około 300), które należą do różnych klas, m.in. białka wiążące RNA (RBP, ang. *RNA - binding proteins*), zależne od RNA ATPazy (ang. *RNA -dependent ATPases*), helikazy RNA i kinazy [8]. Struktura spliceosomu jest tworzona *de novo* przy każdym intronie i umożliwia przyjęcie odpowiedniej konformacji przez pre - mRNA - następuje przestrzenne zbliżenie eksonów znajdujących się powyżej i poniżej intronu, który w ten sposób ulega wypętlaniu. Rozpoznanie intronów jest związane z obecnością w pierwotnym transkrypcie charakterystycznych i wysoce konserwatywnych sekwencji:

2. Wprowadzenie literaturowe

- miejsca splicingowego 5' z konserwatywną dwójką nukleotydów GU, do których wiąże się U1snRNP;
- miejsca splicingowego 3', w którego skład wchodzi konserwatywna sekwencja AG rozpoznawana przez U2AF35 (ang. *U2 auxiliary factor 35*);



Ryc. 2

Schematyczne przedstawienie reakcji splicingu z zaznaczeniem kluczowej reszty adenozy w sekwencji rozgażeniowej [9].

Najpierw grupa fosforanowa guanozyny w miejscu splicingowym 5' jest atakowana przez grupę 2' hydroksylową adenozy z sekwencji rozgażeniowej tworząc w rezultacie cząsteczkę w formie lassa oraz uwalniając ekson. W kolejnym etapie zachodzi rozcięcie mRNA za resztą guaniny z sekwencji AG w miejscu splicingowym 3' oraz połączenie eksonów. Intron jest uwalniany w postaci lassa, a następnie degradowany [11] (Ryc.2).

- traktu polipirymidynowego na końcu 3' intronu;
- miejsca rozgażeniowe, czyli sekwencji CURAY (gdzie R jest puryną, a Y pirymidyną), które jest zlokalizowane 17 do 40 nukleotydów powyżej miejsca splicingowego 3'; jest rozpoznawane przez U2 snRNP [10].

Składanie spliceosomu jest wysoce złożoną i dynamiczną reakcją, która wymaga hydrolizy kilku cząsteczek ATP oraz wielu rearanżacji strukturalnych. Po utworzeniu spliceosomu i wypętleniu intronu, przebiega dwustopniowa reakcja splicingu.

Ogólna organizacja genów i proces dojrzewania pre-mRNA u roślin jest zbliżona do tego występującego u kręgowców, jednak jest w znacznie mniejszym stopniu poznana i opisana. Mechanizm splicingu oraz skład spliceosomu został określony na modelu zwierzęcym i jak dotąd nie udało się uzyskać spliceosomu z komórek roślinnych. Wiemy jednak, że tak jak w innych organizmach na proces składania RNA u roślin składają się dwie reakcje transestryfikacji oraz etap formowania struktury lassa. Wiele analogów komponentów spliceosomu zwierzęcego udało się także zidentyfikować w genomach roślinnych. Pomimo wielu podobieństw w komórkach roślinnych nie zachodzi proces splicingu heterologicznych cząsteczek pre-mRNA, co sugeruje, że proces ten u roślin jest przeprowadzany i regulowany w nieco odmienny sposób niż u zwierząt [12]. Do tej pory zidentyfikowano przede wszystkim istotne różnice na etapie rozpoznawania intronów. Stwierdzono, że sekwencje roślinnych miejsc splicingowych zarówno 3' jak i 5' wykazują większą zmienność, a miejsce rozgałęzienia jest gorzej zdefiniowane w porównaniu z drożdżami i zwierzętami wielokomórkowymi (dotychczas eksperymentalnie udało się wyznaczyć jedynie dwa roślinne miejsca rozgałęzienia) [13]. Cechą charakterystyczną intronów roślinnych jest wysoka zawartość nukleotydów UA oraz U - zarówno u jedno - jak i dwuliściennych introny są o ok. 15% bogatsze w sekwencje UA niż eksony (za tę różnicę odpowiada przede wszystkim wysoka zawartość U). Różnica ta jest istotna przy rozpoznawaniu 5' i 3' miejsc splicingowych oraz przy usuwaniu intronów [14]. Wybór poszczególnych miejsc splicingowych jest ponadto warunkowany obecnością dodatkowych elementów *cis* w sekwencji transkryptu, które określane są jako elementy regulujące splicing (SREs, ang. *splicing regulatory elements*). Mogą być one zlokalizowane zarówno w obrębie intronów jak i eksonów oraz pełnić funkcje wzmacniaczy lub wyciszaczy splicingu, stąd podział na eksonowe sekwencje wzmacniające/wyciszające splicing (ESE i ESS, ang. *exonic splicing enhancer; exonic splicing silencer*) i intronowe sekwencje wzmacniające/wyciszające splicing (ISE i ISS, ang. *intronic splicing enhancer; intronic splicing silencer*). Rozpoznano także kilka białek, które specyficznie wiążą się z tymi elementami, np. białka bogate w serynę i argininę (SR, ang. *serine/arginine-rich proteins*). Białka te są ważnymi czynnikami splicingowymi i uczestniczą w aranżacji spliceosomu pośrednicząc w oddziaływaniach białko – białko oraz białko – mRNA. Łączą się one poprzez swoją domenę wiążącą RNA (RRM, ang. *RNA-recogniton motif*) z odpowiednim motywem sekwencyjnym w obrębie pierwotnego

transkryptu. W zależności od rozpoznawanej sekwencji, przyłączenie się białek SR promuje lub hamuje włączanie danego elementu (intronu lub eksonu) do dojrzewającej cząsteczki mRNA [15].

2.1.2 Alternatywny splicing

W 1978 roku, tuż po odkryciu, że sekwencje kodujące białko (eksony) są rozdzielone przez fragmenty usuwane z pierwotnego transkryptu (introny), zaproponowano mechanizm, w którym różne eksony mogą być łączone ze sobą prowadząc do powstania kilku cząsteczek mRNA z pojedynczego genu. Zjawisko selektywnego wyboru i łączenia różnych eksonów zostało nazwane alternatywnym splicingiem (AS). Jest on powszechnym mechanizmem zwiększającym różnorodność powstających transkryptów, który w znaczący sposób przyczynia się do zwiększenia potencjału kodującego genomów roślinnych i zwierzęcych oraz kompleksowości proteomu. Początkowo jednak proces AS uważano za odosobnione zjawisko występujące tylko w rzadkich przypadkach (pierwsze szacunki donosiły, że jedynie 5% genów człowieka podlega temu procesowi [16]), jednak w ostatnich latach wykorzystanie wysokowydajnościowych metod badania transkryptomów pozwoliło na nowo oszacować skalę tego zjawiska i obecnie uważa się, że około 95% genów człowieka [17] i ok. 61% genów *Arabidopsis thaliana* [18], podlega AS.

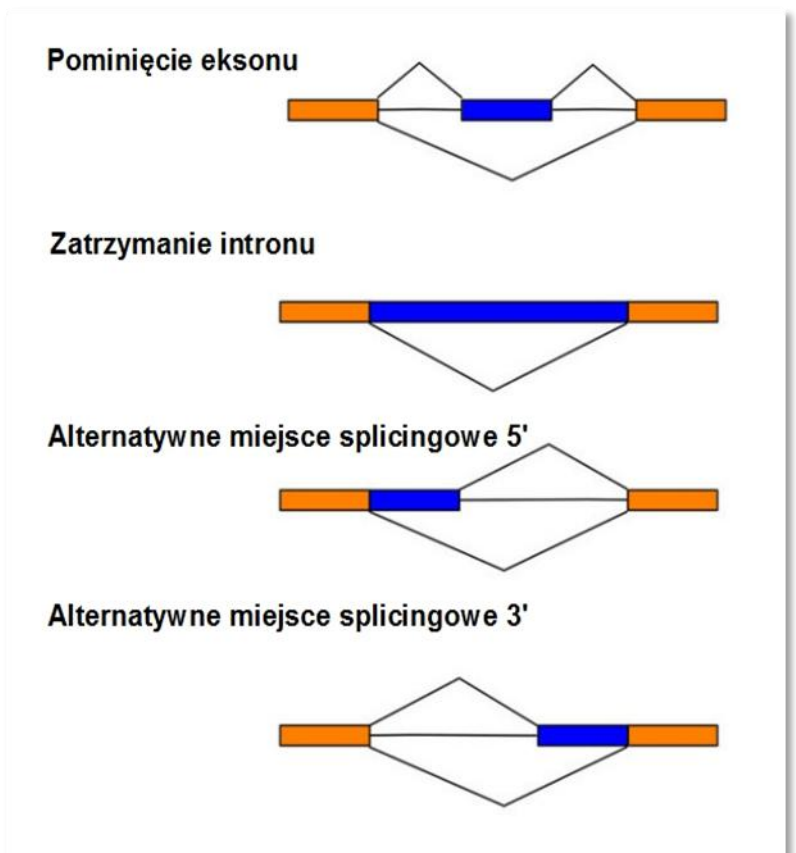
W przeciwieństwie do kontroli potranskrypcyjnej AS zmienia strukturę transkryptów, przez co może wpływać na prawie wszystkie aspekty działania białek, dlatego też alternatywny splicing jest jednym z ważniejszych mechanizmów odpowiedzialnych za regulację ekspresji genów w komórkach eukariotycznych. Wiele zdarzeń alterantywnego splicingu występuje kotranskrypcyjnie i jest uzależnione od DNA, a także struktury i modyfikacji chromatyny [19]. Powstałe w wyniku alternatywnego splicingu mRNA mogą różnić się stabilnością czy lokalizacją w komórce. Mogą ulegać translacji, w wyniku której syntetyzowane są białka o krótszym lub dłuższym łańcuchu aminokwasowym, o zmienionej aktywności (zwiększonej, zmniejszonej lub może zostać ona całkowicie utracona), o odmiennym sposobie regulacji czy stabilności. Alternatywny splicing jest także zaangażowany w regulację poziomu transkryptów w komórce poprzez sprzężenie z procesem

rozpadu cząsteczek mRNA - NMD, w których pojawia się przedwczesny sygnał terminacji translacji (PTC, ang. *premature termination codon*)[20].

Zdarzenia alternatywnego splicingu można pogrupować w kilka głównych typów (Ryc. 3). Jednym z nich jest pominięcie eksonu (ang. *exon skipping*), w którym następuje usunięcie całego eksonu, tzw. eksonu kasetowego (ang. *cassette exon*) wraz z przyległymi intronami z dojrzewającego transkryptu. Kolejne dwa typy AS związane są z rozpoznaniem dwóch lub większej liczby miejsc splicingowych przy końcu eksonu, co prowadzi do wyboru różnych miejsc splicingowych 5' (ang. *acceptorsite*) lub 3' (ang. *donor site*) i przesunięcia granicy ekson/intron. Ostatnim z głównych typów zdarzeń splicingowych jest zatrzymanie intronu (ang. *intron retention*), który skutkuje pozostawieniem części lub całego intronu w obrębie dojrzałego transkryptu.

Analiza ilości i typów zdarzeń splicingowych wykazała różnice w ich rozkładzie między gatunkami. Najlichniesze są zdarzenia splicingowe u wyższych kręgowców, których 40% to pominięcie eksonu [21]. Śledząc rozkład typów AS na drzewie

ewolucyjnym eukariontów można zauważyć wzrastający udział tego typu splicingu, stąd teoria, że właśnie on w największym stopniu przekłada się na zróżnicowanie fenotypowe (sugeruje się, że w organizmach roślinnych większe znaczenie miały zjawiska duplikacji genomów, dlatego obserwuje się u nich mniejszy udział tego typu zdarzeń splicingu) [22]. U *A. thaliana* i kilku innych gatunków roślin, dominującym typem AS okazało się zatrzymanie



Ryc. 3

Główne typy zdarzeń alternatywnego splicingu [18].

intronu. Początkowo zakładano, że transkrypty z zatrzymanym intronem są głównie celem NMD. Jednak ostatnie badania wykazały, że produkty tego typu AS nie są degradowane w tym procesie, co ponowiło pytanie o ich przeznaczenie i funkcję [23]. Wskazówką do odpowiedzi na nie może być zjawisko zaobserwowane podczas badania dojrzewania gametofitu mszaka *Marsilea vestia*, u którego transkrypty z pozostawionym intronem są gromadzone w sporach do czasu kiedy ich końcowy splicing i translacja nie są zaindukowane w trakcie spermatogenezy [24]. Inną przesłanką do określenia roli transkryptów z zatrzymanym intronem są badania nad neurogenezą u ssaków. Wykazano, że regulowane zatrzymanie intronu w transkryptach genów specyficznych dla neuronów hamuje ich eksport do cytoplazmy i powoduje degradację w obrębie jądra w sposób niezależny od maszynerii białek zaangażowanych w proces NMD [25]. Wydaje się możliwe, że proces zatrzymywania intronów w transkryptach znajdujących się w jądrze może być rozpowszechnionym mechanizmem umożliwiającym magazynowanie prekursorowych RNA oraz regulację końcowych etapów ich dojrzewania w sposób zależny od fazy rozwojowej lub czynników zewnętrznych.

2.1.3 Rola AS w ekspresji genów

Zwiększenie zróżnicowania proteomu przy ograniczonej pojemności kodującej genomów jest podstawową funkcją alternatywnego splicingu. Jednak rozmiar tego zjawiska nadal trudno oszacować. Niektóre geny ulegają alternatywnemu splicingowi konstytutywnie, inne natomiast są tkankowo specyficzne czy też zależne od fazy rozwojowej lub warunków środowiska. Powstawanie izoform białek zależy także od typu zdarzeń AS. Niektóre z nich nie prowadzą do przesunięcia ramki odczytu - analiza wzbogacona o ontologię genów wykazała, że zdarzenia AS tego typu szczególnie często dotyczą białek wiążących RNA - RBP, które biorą udział w regulacji splicingu, a także czynników transkrypcyjnych [26]. Z kolei skutkiem przesunięcia ramki odczytu na drodze alternatywnego splicingu jest często translacja białek krótszych, które w niektórych przypadkach mogą zachować swoje domeny funkcjonalne, w innych - sekwencje kodujące te rejony mogą zostać usunięte z dojrzałego transkryptu lub mogą pojawić się dodatkowe sekwencje niosące informacje o nowych domenach. Wykazano, że niektóre alternatywnie włączane eksony kodują domeny o nieokreślonej ściśle strukturze,

które jednak pojawiają się na powierzchni białka i wchodzi w oddziaływania białko - białko. W ten sposób AS może modyfikować całe sieci interakcji białek, a także szlaki przekazywania sygnału [27]. Sugeruje się również inną, potencjalną rolę regulacyjną jaką takie krótkie izoformy białka mogą pełnić w komórkach roślinnych. W kilku przypadkach powstające skrócone peptydy zachowują zdolność do dimeryzacji i tworzenia homo- lub heteromerów, jednak nie posiadają domen funkcjonalnych. Przypuszcza się, że w czasie dimeryzacji mogą one konkurować z peptydami pełnej długości, w ten sposób pełniąc rolę negatywnych regulatorów [28]. Ponadto takie skrócone peptydy mogą mieć zmienione miejsca modyfikacji potranslacyjnych, np. fosforylacji, co także może mieć wpływ na zależny od kinaz przepływ sygnału w komórkach [29]. Wykazano na przykład, że u *A. thaliana* białka RS41, CypRS64 i SR45a posiadają miejsca fosforylacji specyficzne dla pojedynczych wariantów splicingowych [30].

Alternatywny splicing bierze udział w regulacji ilości transkryptów w komórce poprzez sprzężenie z procesem rozkładu cząsteczek mRNA - NMD. NMD jest cytoplazmatycznym systemem degradacji RNA, który włączany jest podczas pierwszej rundy translacji, wtedy gdy terminacja translacji z danego transkryptu zostaje zakłócona. Transkrypty będące celem NMD posiadają długie sekwencje 3' UTR, z obecnymi w nich intronami, przedwczesny sygnał terminacji translacji (PTC) około 50 - 55 nukleotydów powyżej miejsca splicingowego oraz znajdującą się powyżej otwartą ramką odczytu (uORF, ang. *upstream open reading frame*). Wszystkie te cechy mogą być wprowadzone lub zmienione na drodze AS, przez co proces ten może wpływać bezpośrednio na stabilność transkryptu. Analizy transkryptomu *Arabidopsis thaliana* z zastosowaniem kolekcji ESTów (ang. *expressed sequence tags*) oraz reakcji RT-PCR pozwoliły oszacować, że od 11 do 18% transkryptów genów zawierających introny jest celem dla NMD [23, 31].

Proces alternatywnego splicingu jest często sprzężony z NMD i wpływa na ekspresję genów na zasadzie autoregulującej negatywnej pętli zwrotnej (zwiększenie poziomu danego białka w komórce prowadzi do zwiększenia ilości powstałych na drodze AS transkryptów zawierających PTC, które są celem dla NMD). Taką zależność AS - NMD zaobserwowano m. in. dla czynników splicingowych [32], genów kodujących GRP8 i GRP7, które są częścią zegara circadiálnego (ang. *circadian clock*) [33, 34] czy dla genu AFC2 kodującego wysoce

konserwatywną kinazę LAMMER [18]. Proces takiego sprzężenia zwrotnego między splicingiem a NMD dotyczy wielu genów, o czym świadczą podane wyżej przykłady.

Kolejnym przykładem wpływu AS na ekspresję genów jest jego związek z miRNA. W procesie alternatywnego splicingu może dochodzić do tworzenia izoform mRNA lub niekodujących RNA, które posiadają lub nie miejsce docelowego wiązania miRNA. AS może modyfikować także proces wycinania miRNA z pri-miRNA i w ten sposób wpływać na ilość dojrzałych miRNA w komórce, które z kolei odpowiednio zmieniają ilość docelowego mRNA. Modyfikacja ścieżki miRNA przez proces splicingu obejmuje także wpływ na wzór splicingowy pre-miRNA oraz mRNA genów, kodujących enzymy zaangażowane w biogenezę miRNA, np. genu kodującego kluczową endonukleazę szlaku miRNA - DCL1 (ang. *Dicer-like 1*). Na drodze AS, związanego z wycinaniem intronu 12, powstają 4 izoformy tego transkryptu, z których tylko jeden posiada miejsce wiązania miRNA162 [35]. U roślin miejsca wiązania miRNA (MBS, ang. *miRNA binding sites*) znajdują się zarówno w sekwencji kodującej jak i w rejonie UTR. Stosując analizę bioinformatyczną stwierdzono, że u *A. thaliana* proces AS wpływa na 12% MBS (44 z 354 mRNA z określonymi MBS) [36].

Jednym z ciekawszych przykładów zależności między AS a miRNA jest ekspresja miRNA400 u *A. thaliana* pod wpływem stresu ciepła. To miRNA znajduje się w obrębie intronu genu kodującego białko (At1g32853) i wraz z nim ulega kotranskrypcji. U roślin z nadekspresją tego miRNA pod wpływem stresu ciepła następuje spadek wzrostu i siły kiełkowania. Natomiast w typie dzikim pod wpływem wysokiej temperatury następuje wycinanie intronu kodującego to miRNA, czego rezultatem jest zwiększenie ilości pierwotnego transkryptu miR400, a spadek jego dojrzałej, funkcjonalnej formy. Zwiększa to szanse rośliny na przetrwanie warunków stresowych [37].

W ostatnich latach znacznie powiększyła się wiedza na temat ogólnej roli alternatywnego splicingu w procesie ekspresji genów oraz przykładów genów, które ulegają temu procesowi, jednakże tylko w niewielu przypadkach zaproponowano funkcję biologiczną powstałych transkryptów i ewentualnych białek, które mogą być syntetyzowane na ich podstawie. Niedawno opisano rolę alternatywnego splicingu genu N tytoniu w nabywaniu odporności rośliny na infekcję wirusem mozaiki tytoniu TMV (ang. *tabacco mosaic virus*).

Roślina syntetyzuje dwa transkrypty z genu N - Ns i NI, przy czym NI powstaje w wyniku włączenia alternatywnego eksonu znajdującego się w obrębie intronu 3, a jego pula zwiększa się w czasie infekcji. Wykazano ponadto, że obecność obu transkryptów genu N jest niezbędna dla zapewnienia roślinie odporności na tego wirusa [38]. Innym przykładem ważnego procesu zależnego od alternatywnego splicingu jest kontrola kwitnienia u *Arabidopsis thaliana* związana z ekspresją genu *FCA* (ang. *flowering locus C*). U roślin transgenicznych z nadekspresją tego genu zaobserwowano przyspieszenie kwitnienia, co jest związane z niewielkim wzrostem puli transkryptu *FCA* z usuniętymi wszystkimi intronami w stosunku do pozostałych 3 transkryptów, które posiadają fragment intronu 3. Sugeruje się, że negatywna autoregulacja dojrzewania transkryptów tego genu zabezpiecza roślinę przed przedwczesnym kwitnieniem [39].

Badania nad genomem kukurydzy ujawniły, że część genów biorących udział w procesach takich jak kiełkowanie czy synteza antocyjanin podlega alternatywnemu splicingowi. W czasie ekspresji genu *Knox7*, który koduje czynniki transkrypcyjne regulujące kiełkowanie, powstaje aż 7 alternatywnych transkryptów [40]. Natomiast pod wpływem czynników stresowych takich jak obecność kadmu, aż 50 – krotnie wzrasta poziom alternatywnego transkryptu genu *Bronze2*, z którego nie zostały usunięte introny. Gen ten jest odpowiedzialny za końcowy etap syntezy antocyjanin (koduje transferazę glutationu GST, która przenosi resztę glutationu na cyjanidynę umożliwiając w ten sposób jej transport do wakuoli) [41]. Również geny kodujące białka SR, wchodzące w skład spliceosomu i biorące udział w wyborze miejsc splicingowych, podlegają u kukurydzy alternatywnemu splicingowi – zidentyfikowano 4 transkrypty dla genu *zmRSp31A* oraz 3 transkrypty dla genu *zmRSp31B*, które powstają z użyciem niekanonicznych miejsc splicingowych [42].

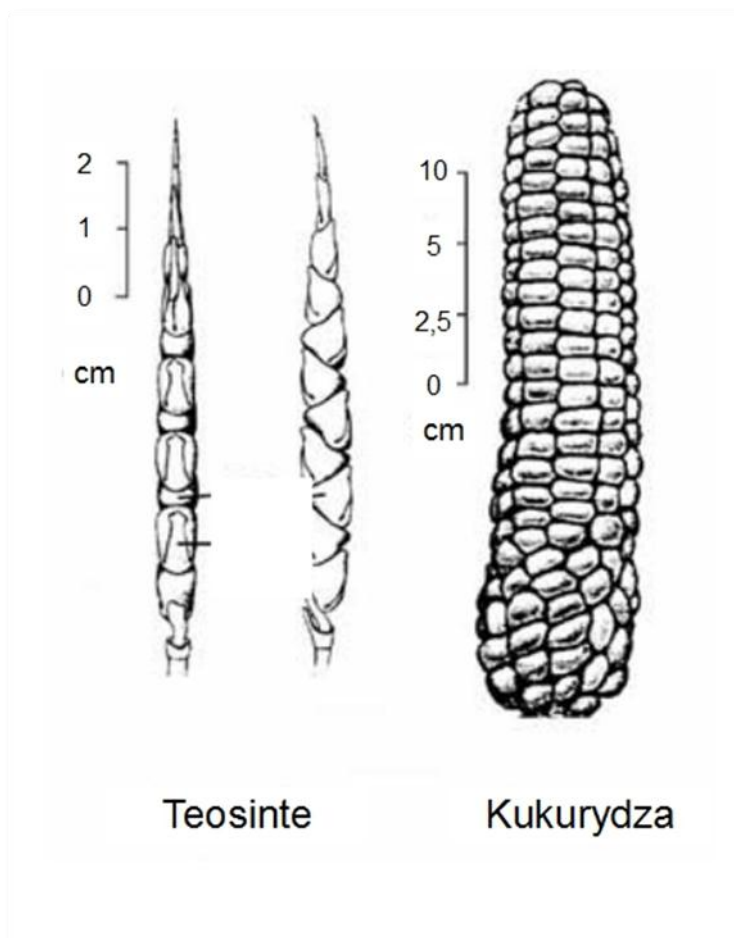
2.2 Kukurydza zwyczajna – pochodzenie, charakterystyka

Domestykacja jest często rozważana jako wyjątkowy przypadek ewolucji – jest to koewolucyjna interakcja między dwoma gatunkami, która prowadzi do morfologicznych i fizjologicznych zmian, a w konsekwencji do ustalenia nowego, udomowionego gatunku, którego wzrost i reprodukcja jest kontrolowana głównie przez korzyści, które czerpie z niego

inny gatunek [43]. Proces domestykacji zarówno roślin jak i zwierząt był jednym z najważniejszych w historii gatunku *Homo sapiens* i uważa się, że jest w dużej mierze odpowiedzialny za rewolucję neolityczną (13 000 – 10 000 lat temu), w czasie której plemiona zbieraczy/myśliwych zaczęły prowadzić osiadły tryb życia, dając tym samym początek kulturze argotechnicznej. Nadwyżki żywności, które powstawały dzięki uprawie roślin i hodowli zwierząt, pozwoliły na powstanie takich dziedzin ludzkiej aktywności jak

sztuka, rozwój rzemiosł, pisma, hierarchii społecznej, a w końcu doprowadziły do urbanizacji [44].

Proces domestykacji kukurydzy (*Zea mays* spp. *mays* L.) rozpoczął się około 9 000 lat temu w dolinie rzeki Balsas w południowo-środkowym Meksyku. Selekcja hodowlana, która nastąpiła później doprowadziła do wykształcenia ogromnej ilości odmian, która znacznie przekracza zróżnicowanie fenotypowe występujące u jej dzikiego przodka – teosinte (*Zea mays* ssp. *parviglumis*). Nacisk selekcyjny spowodował ostatecznie przekształcenie twardych, małych ziaren teosinte w duże, łatwe do pozyskania i mielenia bezokrywowe ziarna współczesnych odmian (Ryc. 4). Kłosek teosinte może pomieścić



Ryc. 4

Porównanie wielkości i ogólnego pokroju teosinte i jednej z odmian kukurydzy [45].

od 5 do 12 nasion, ma około 6 cm długości, natomiast kłos kukurydzy ma około 30 cm długości i może pomieścić nawet do 500 ziaren. Ziarna kukurydzy są trwale przymocowane do kolby, która się nie rozpada w trakcie dojrzewania nasion - gdy pozostawimy ją na

roślinie, po skończonym okresie wegetacji kolba wraz z nasionami opadnie na ziemię. Ponieważ setki nasion z pojedynczej kolby nie są w stanie się rozdzielić, w następnym sezonie wykiełkują bardzo blisko siebie, siewki kukurydzy nie będą miały dostatecznej ilości światła i gleby, aby prawidłowo wzrastać i się rozwijać, dlatego mówi się, że przetrwanie tego gatunku jest w pełni uzależnione od człowieka. Zmiany związane z domestykacją obejmowały również architekturę i pokrój rośliny - pojedyncza roślina teosinte ma wiele łodyg, zwanych odrostami, u kukurydzy występuje pojedyncza łodyga z kolbami. Udało się zidentyfikować i sklonować gen, który odpowiadał za tę zmianę (tb1, ang. *teosinte branched 1*), a rośliny, u których dochodziło do jego ekspresji posiadały, podobnie jak przodek kukurydzy, liczne łodygi i kłosa [46]. Rozprzestrzenienie uprawy kukurydzy na teren obu kontynentów amerykańskich (do południowo-zachodniej części obecnych Stanów Zjednoczonych dotarła ok. 3200 lat temu, a do granic Kanady - 1300 lat temu [47]) oraz wprowadzenie jej do uprawy w Europie i Azji, przyczyniło się do stopniowej adaptacji do nowych warunków środowiska i progresywnego powstawania coraz liczniejszych odmian lokalnych. Na przykład w prekolumbijskim Peru zidentyfikowano 35 odmian, w tym odmiany szkliste, których nasiona posiadają twardszą okrywę (ang. *flint corn*) czy odmiany, które były hodowane do konkretnych zastosowań jak produkcja piwa chicha, barwników tekstylnych czy mąki [48]. W setkach odmian, które powstały do czasów współczesnych możemy wyróżnić także odmiany karłowate, które nie osiągają więcej niż ok. 20 cm, inne są gigantyczne – mogą mieć ok. 3m wysokości. Za olbrzymią różnorodność odmian kukurydzy odpowiada przede wszystkim plastyczność jej genomu oraz łatwość z jaką dochodzi do zapylenia między różnymi odmianami.

Do tej pory zostały zsekwencjonowane dwa genomy kukurydzy (w 2009 roku genom linii wsobnej B73 oraz meksykańskiej odmiany *Palomero toluqueno*) [49, 50]. Genom tej rośliny ulegał kilkakrotnym duplikacjom i obejmuje 10 chromosomów złożonych z $2,3 \cdot 10^9$ nt (ok. 80% wielkości genomu człowieka). Szacuje się, że koduje ok. 32 tys. genów dających produkty białkowe, z czego od 2 - 4% genów było obiektem selekcji związanej z udomowieniem kukurydzy [51]. Jednak ogromna większość genomu kukurydzy, bo aż 85%, stanowią sekwencje różnego typu elementów powtórzonych [49]. Badania porównawcze uzyskanych do tej pory sekwencji genomowych potwierdziły olbrzymią zmienność, także na

poziomie genomów, nawet blisko ze sobą spokrewnionych linii kukurydzy, która znacznie przewyższa różnorodność spotykaną u innych organizmów eukariotycznych. Wykazano, że między dwiema dowolnie wybranymi liniami kukurydzy polimorfizm pojedynczych nukleotydów (SNP, ang. *single nucleotide polymorphism*) występuje średnio co 100 nt. Genom kukurydzy cechuje występowanie nie tylko polimorfizmów typu pojedynczych nukleotydów czy typu insercja/delecja (IDP, ang. *insertion/deletion polymorphism*), ale także różnice w strukturze chromosomów czy zmieniona lokalizacja genów i elementów powtórzonych [52].

2.2.1 Zastosowanie i hodowla kukurydzy zwyczajnej

O sukcesie trwającego tysiące lat procesu hodowli kukurydzy świadczy fakt, że zajmuje ona trzecie miejsce w światowej strukturze upraw i drugie miejsce pod względem wysokości plonu [53]. Obecnie nadal jest podstawą diety wielu mieszkańców Ameryki Południowej czy Afryki, ale znajduje także coraz liczniejsze zastosowania przemysłowe. Areał uprawy kukurydzy wzrastał w Polsce w ostatnich latach (Tab. 1) i obecnie szacuje się, że uprawy kukurydzy zajmują ok. 1 mln ha, z czego połowa uprawiana jest na ziarno. Jest ono podstawowym surowcem dla przemysłu paszowego, ale szerokie zastosowanie może mieć także cała biomasa kukurydzy, jak i jej poszczególne frakcje, np. rdzenie kolbowe, słoma, zarodki. Kukurydzę wykorzystuje się w przemyśle spożywczym do produkcji kaszy, mąki, olejów, ale także do bezpośredniego spożycia lub jako półprodukty. Zwiększa się zainteresowanie kukurydzą jako typową rośliną przemysłową, którą stosuje się w procesie fermentacji alkoholowej do pozyskania etanolu (zarówno spożywczego jak i paliwowego), otrzymuje się z niej wysokiej jakości skrobię, wykorzystuje się ją do celów energetycznych poprzez spalanie (słoma, rdzenie kolbowe) lub produkcję biogazu (biomasa). Słoma kukurydziana jest poza tym surowcem do produkcji papieru i materiałów izolacyjnych. Dominującym celem uprawy pozostaje jednak przemysł paszowy, gdzie kukurydza jest wykorzystywana jako materiał na kiszonki z całych roślin lub pasze z ziarna i kolb. Krajowa hodowla nie jest jeszcze w stanie zaspokoić potrzeb rynku paszowego, dlatego znaczny udział w produkcji pasz w Polsce ma kukurydza importowana.

	2000	2002	2005	2010	2011	2012
Powierzchnia zasiewów [tys. ha]						
• kukurydza na ziarno	152	319	339	334	333	544
• kukurydza na paszę	162	196	326	396	426	508
Suma	314	358	665	730	759	1052
Zbiory [tys. t]						
• kukurydza na ziarno	923,3	1962,0	1945,4	1994,4	2392,1	3995,9
Plon [dt/ha]						
• kukurydza na ziarno	60,6	61,6	57,7	59,7	71,8	73,5

Tabela 1

Zestawienie powierzchni zasiewów, wielkości zbiorów i plonu kukurydzy w latach 2000 - 2012, przygotowano według Rocznika Statystycznego Rolnictwa 2013 [54].

Coraz liczniejsze możliwości wykorzystania kukurydzy wymagają od hodowców ciągłej pracy nad udoskonalaniem odmian, także dedykowanych do konkretnych zastosowań. Przykładem mogą być dwie odmiany wyprowadzone w Polsce - Vitras i SMH 27108 Ułan, których bardzo wysoka produkcja biomasy czyni je idealnymi kandydatami do hodowli na kiszonki z całych roślin, które mogą być wykorzystywane w produkcji biogazu [55]. Hodowla odmian kukurydzy w Polsce ma dopiero 60-letnią tradycję i dużą konkurencję zarówno ze strony zagranicznych ośrodków naukowych jak i koncernów chemiczno-rolniczych. Początkowo w hodowli kukurydzy stosowano odmiany populacyjne, najsilniej rozwijane na terenie Stanów Zjednoczonych w obrębie tzw. Pasa Uprawy Kukurydzy (ang. *US Corn Belt*). Stamtąd pochodziła odmiana Reid Yellow Dent, która była uprawiana na 75% areału w USA do 1890 roku, kiedy zaczęto wprowadzać odmiany mieszańcowe. Współcześnie uprawiane odmiany są to także głównie odmiany mieszańcowe, które powstały w oparciu o wykorzystanie zjawiska heterozji, czyli tzw. wigoru mieszańców heterozygotycznych. Jest to efekt bujności pierwszego pokolenia mieszańca uzyskanego ze skrzyżowania celowo dobranych form rodzicielskich. Po sformułowaniu naukowych podstaw tego zjawiska przez Shull'a w roku 1908 nastąpił szybki rozwój tej technologii - początkowo eksperymentowano z mieszańcami pojedynczymi (utworzonymi ze skrzyżowania dwóch linii wsobnych - A x B), lecz ze względu na niską plenność linii matecznych zaczęto stosować odmiany populacyjne jako formy rodzicielskie. Przełomem było wprowadzenie w 1918 roku koncepcji mieszańców czteroliniowych (A x B/C x D), których formami rodzicielskimi były nie linie wsobne, a

mieszańce pojedyncze. Wyższy potencjał tych mieszańców został szybko zauważony i już na przełomie lat dwudziestych i trzydziestych zaczęto je powszechnie uprawiać. Dzięki temu w latach 1930-1960 nastąpił w USA wzrost roczny plonu średnio o 0,65 dt/ha [56].

W Europie badania nad wprowadzeniem mieszańców liniowych do programów hodowlanych zaczęto po drugiej wojnie światowej, głównie w oparciu o materiał genetyczny pochodzący z USA i Kanady. Ponieważ europejskie warunki klimatyczne są mniej sprzyjające uprawie kukurydzy ocenia się, że 2/3 uprawianych odmian to odmiany typu "semident" uzyskane ze skrzyżowania form rodzicielskich o różnym typie endospermu. W odmianach mieszańcowych tego typu łączy się wyższy potencjał plonowania form zębokształtnych (dent, które najczęściej są komponentem matecznym) z lepszą tolerancją na niekorzystne warunki termiczne form szklistych (flint, komponent ojcowski). Formy szkliste kukurydzy rosną szybciej w warunkach chłodnej wiosny i szybciej osiągają dojrzałość fizjologiczną (35% zawartości wody w ziarnie). Formy zębokształtne dla wykazania wyższego potencjału plonowania potrzebują dłuższego i cieplejszego sezonu wegetacyjnego [57].

W Polsce pierwsze prace nad wprowadzeniem kukurydzy do uprawy rozpoczęły się w 1953 roku w trzech ośrodkach: Pustkowie, Kobierzycach i w Smolicach. Dopiero jednak w 1967 r. zarejestrowano pierwszą polską odmianę będącą mieszańcem liniowym. Był to czteroliniowy mieszańiec KB 260, którego formułę tworzyły 4 linie wsobne pochodzące z hodowli zagranicznej. Kolejne odmiany, IHAR 280 i IHAR 262, rejestrowane odpowiednio w 1971 i 1972 r., były też mieszańcami czteroliniowymi, ale w ich formułach były wyłącznie linie wsobne wyhodowane w Smolicach. Szacuje się, że od lat 90-tych ubiegłego wieku 40% odmian kukurydzy na polskim rynku to odmiany rodzime [58].

Obecnie do pozyskania linii wsobnych najczęściej wykorzystuje się technikę podwajania haploidów (linie DH), a do wyboru odpowiednich linii wsobnych, charakteryzujących się wysoką zdolnością kombinacyjną, a więc możliwością uzyskania dużego efektu heterozji, stosuje się metody biologii molekularnej, np. technikę RAPD (ang. *Random Amplified Polymorphic DNA*) czy AFLP (ang. *Amplified Fragment Length Polymorphism*), które mają skuteczniej ocenić dystans genetyczny między formami rodzicielskimi. Inną drogą do uzyskania nowych odmian i wzrostu plonowania jest tworzenie

organizmów genetycznie modyfikowanych (GMO, ang. *genetically modified organism*). Jednym z bardziej znanych przykładów jest szeroko stosowana w uprawie kukurydza odporna na herbicyd Roundup (Roundup Ready®) czy na omacnicę prosowiankę - kukurydza Bt-176.

2.3 Stres herbicydowy

Uzyskanie postępu biologicznego jest jedną z możliwości zwiększenia plonowania. Równoległe z doskonaleniem metod hodowli duże znaczenie dla wzrostu plonów kukurydzy miało wprowadzenie nowych rozwiązań agrotechnicznych, np. zaproponowane w 1965 r. zwiększenie obsady roślin na jednostce powierzchni czy wprowadzenie w 1967 r. atrazyny do zwalczania chwastów [59]. Nadal mimo wprowadzania odmian coraz lepiej przystosowanych do różnych warunków środowiskowych, szczególnie w pierwszych fazach wzrostu (do osiągnięcia stadium 4-6 liścia), należy kukurydzy zapewnić dostateczną ilość światła, wody i składników mineralnych do prawidłowego wzrostu, którą znacznie zmniejsza obecność chwastów na polu. Do gatunków szczególnie konkurencyjnych w tym okresie należą chwasty o dużym współczynniku rozmnażania, których setki tysięcy nasion zalega w powierzchniowych warstwach gleby. Należą do nich: komosa biała, ognicha polna, tobołki, przytulia czepna, gwiazdnica pospolita, przetaczniki rumianowate oraz szarłat szorstki, ostrożeń polny, chwastnica jednostronna i perz właściwy.

Aby zmniejszyć konkurencję i zapewnić kukurydzy optymalne warunki do wzrostu, w uprawie stosuje się różnego rodzaju zabiegi agrotechniczne takie jak odchwaszczanie przedplonu, pełen zespół uprawek późniwnych, wczesna orka przedzimowa, wczesne wykonanie uprawek przedsiewnych z użyciem bron i włók oraz mechaniczne pielęgnowanie zasiewów. Uzupełnieniem tych zabiegów jest stosowanie herbicydów (łac. *herba* - trawa, *caedo* - zabijać), czyli środków chemicznych, które zakłócają wzrost i rozwój roślin. Mogą działać jako inhibitory wzrostu siewek, produkcji pigmentów, fotosyntezy, syntezy aminokwasów czy regulatorów wzrostu. Powinny one być traktowane jako uzupełnienie prawidłowo wykonanych zabiegów agrotechnicznych, jednak ponieważ kukurydza jest rośliną, która dobrze znosi uproszczenia uprawowe, oprysk herbicydami jest często

podstawowym sposobem odchwaszczania pól. Dostępny asortyment środków chemicznych jest bardzo szeroki. Aby utrzymać plantację kukurydzy czystą przez cały okres wegetacji powinien on być przede wszystkim właściwie dobrany do występujących na danym polu chwastów oraz ich wrażliwości na substancję czynną.

Wprowadzenie do środowiska uprawy egzogenego środka, jakim są herbicydy, jest silnym czynnikiem stresowym dla rozwijającej się rośliny. Stosowane herbicydy mogą być fitotoksyczne, co może przejawiać się nekrozą liści, przebarwieniami lub zahamowaniem wzrostu. Takie zmiany morfologii są często przemijające, jednak w skrajnych przypadkach mogą prowadzić do obniżenia plonu o 20%. Oprócz bezpośredniego negatywnego wpływu herbicydu na wznrastającą siewkę, oprysk herbicydem może prowadzić do uszkodzeń mechanicznych, które mogą powodować szereg skutków pośrednich np. zakażenia oportunistyczne spowodowane przez bakterie, wirusy, grzyby lub pierwotniaki chorobotwórcze, które zwykle nie wywołują chorób u zdrowych gospodarzy.

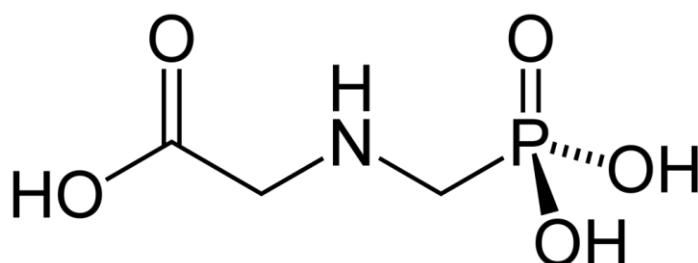
W ostatnich latach pojawiło się wiele prac związanych z poznaniem szerokiego spektrum odpowiedzi roślin na zmieniające się warunki środowiska i czynniki stresowe, które się w nim pojawiają. Wiadomo m.in., że u ryżu pod wpływem wysokiego zasolenia geny kodujące małą i dużą podjednostkę karboksydysmutazy rybulozo-1,5-bisfosforanu (RuBisCo), jednego z najważniejszych enzymów biorących udział w fotosyntezie, czy reduktazy enoyl-ACP (ang. *enoyl-acyl carrier protein reductase*) ulegają zwiększonej ekspresji. U pszenicy, także w warunkach wysokiego zasolenia, stwierdzono podniesiony poziom ekspresji kilku innych genów kodujących m. in. białko G, receptor etylenu, białko biorące udział w biosyntezie giberelin i kilka białek związanych z transdukcją sygnałów [60]. Udało się także potwierdzić rolę mechanizmów epigenetycznych w odpowiedzi na stresy abiotyczne [61]. Ponadto w eksperymentach z wykorzystaniem transgenicznych roślin z nadekspresją czynników transkrypcyjnych takich jak DREB1A (ang. *dehydration-responsive element binding-1A*) [62] czy CBF (ang. *C-box binding factor-1*) [63] stwierdzono zwiększoną tolerancję tych roślin na stresy związane z suszą, zasoleniem i zimnem. W wielu laboratoriach na świecie prowadzi się intensywne badania mające na celu identyfikację czynników zaangażowanych w odpowiedź na różne warunki stresowe, zarówno biotyczne jak i abiotyczne, jednakże nadal

nie zostały poznane molekularne mechanizmy związane z odpowiedzią roślin uprawnych na stres herbicydowy.

2.3.1 Herbicyd Roundup®

Jednym z powszechniej i częściej stosowanych tzw. herbicydów nieselektywnych lub totalnych (działających zarówno na chwasty jak i rośliny uprawne) jest produkowany przez firmę Monsanto herbicyd Roundup®, którego składnik aktywny - glifosat został wprowadzony do sprzedaży w 1974 roku. Obecnie jest dopuszczony do sprzedaży w 130 krajach, a jego roczna produkcja

szacowana jest na około 600 tys. ton. Zgodnie z informacjami podanymi na etykiecie herbicyd Roundup® skutecznie niszczy 100 jednorocznych oraz 60 wieloletnich gatunków chwastów, co sprawia, że jest jednym z herbicydów o najszerszym spektrum działania dostępnych na rynku [64]. Jest



Ryc. 5

Wzór strukturalny cząsteczki glifosatu

stosowany nie tylko w uprawach, ale często także do oczyszczania z materiału roślinnego ziemi odpadowej, torowisk, nasypów kolejowych czy obiektów przemysłowych. W uprawach najczęściej stosowany jest przedwschodowo, w celu przygotowania ziemi pod uprawę i zwalczania uciążliwych chwastów. W latach 90-tych zyskał rozgłos i szczególne znaczenie przy opracowaniu odmian roślin GMO odpornych na jego działanie (Roundup®Ready) - umożliwia to opryski upraw tym środkiem w celu selektywnego niszczenia chwastów. Ze względu na korzyści ekonomiczne wynikające z ich stosowania, w krajach, które dopuściły odmiany transgeniczne do hodowli, stanowią one obecnie większość uprawianej kukurydzy czy soi.

Glifosat, czyli N-fosfometyloglicyna (Ryc. 5), hamuje działanie enzymu EPSPS (ang. 5 - *enolpyruvylshikimate* - 3 - *phosphatesynthase*). Jest on aktywny w postaci soli (np. soli

sodowych lub izopropyloaminowych) oraz w obecności kationów. Efektywnie działa na zielone, wzrastające części roślin, przez które jest wchłaniany [65], a następnie ulega translokacji do pozostałych tkanek rośliny. Hamuje tam działanie enzymu EPSPS, który jest częścią szlaku kwasu szikimowego odpowiedzialnego za syntezę aminokwasów aromatycznych (nie występujący u ssaków). Brak tych aminokwasów powoduje zablokowanie procesu biosyntezy białek, przez co uniemożliwia podział komórek i wzrost roślin wrażliwych na glifosat. Po oprysku nadziemne części roślin wrażliwych żółkną, brązowieją, a w końcu więdną. Po 10-14 dniach od oprysku rośliny umierają i ulegają rozkładowi dzięki czemu wzbogacają glebę. Przenikanie i transport glifosatu wewnątrz roślin zapewnia także niszczenie merystemów korzeniowych, podziemnych kłączy i bulw, które mogłyby służyć do rozmnażania wegetatywnego chwastów.

Tolerancja na składnik aktywny herbicydu Roundup® jest uwarunkowana genetycznie i tylko w niewielkim stopniu jest modyfikowana czynnikami klimatyczno - glebowymi czy zabiegami agrotechnicznymi. Mimo ogromnych nakładów finansowych obecnie komercyjnie dostępnych jest tylko kilka technologii wykorzystujących transgeniczne odmiany odporne na glifosat. Jedną ze strategii jest zastosowanie transgeny z dodatkową kopią genu EPSPS, która nie jest inhibowana przez ten składnik aktywny. W większości przypadków jest to gen *cp4 epsps* z *Agrobacterium tumefaciens* lub zmutowany gen *zm-2mepsps* z kukurydzy zwyczajnej. Inna skomercjalizowana strategia wykorzystuje w transgenach geny, których produkty białkowe biorą udział w usuwaniu i rozkładzie glifosatu, np. gen oksydoreduktazy glifosatu (GOX, ang. *glyphosate oxidoreductase*) z *Ochrobactrum anthropi* (szczep LBAA) czy dwa jego homologii - gen *pat* z bakterii *Streptomyces viridoc* i gen *bar* ze *Streptomyces hygrosopicus* [66].

Wprowadzenie do praktyki rolniczej odmian transgenicznych niewrażliwych na glifosat, doprowadziło do wielu zmian, jednak najbardziej dramatyczne w skutkach okazało się uproszczenie taktyki kontroli zachwaszczenia. Można powiedzieć, że glifosat stał się ofiarą własnego sukcesu - od momentu wprowadzenia go do sprzedaży przez 20 lat uważano, że niemożliwe jest wykształcenie tolerancji czy odporności na ten składnik. Jednak uproszczenia uprawowe, na które pozwalają uprawy transgeniczne (brak orki, a przede wszystkim wielokrotne opryski dużą dawką herbicydu w czasie jednego sezonu

wegetacyjnego) zwiększyły presję selekcyjną i w 1996 roku w Australii pojawiły się pierwsze doniesienia o gatunku życicy (*Lolium rigidum* Gaudin) odpornej na działanie herbicydu Roundup® [67]. Według portalu internetowego weedscience.org, który zrzesza naukowców z 80 państw i gromadzi informacje o występowaniu gatunków chwastów odpornych na herbicydy, do tej pory zarejestrowano 213 przypadków odporności na glifosat u takich gatunków jak szkarłat Palmera (*Amaranthus plameri*), przymiotno kanadyjskie (*Conyza canadensis*), ambrozja trójdzielna (*Ambrosia trifida*), życica trwała (*Lolium perenne*) czy mietlnik żakula (*Kochia scoparia*). Podstawą odporności jest najczęściej zmiana miejsca docelowego wiązania glifosatu w obrębie enzymu EPSPS, amplifikacja jego genu, zmniejszona translokacja glifosatu w obrębie całej rośliny lub zmniejszony transport do plastydów, które są miejscem działania enzymu EPSPS, a także sekwestracja glifosatu w wakuolach [68].

Alternatywą dla wprowadzania do uprawy odmian transgenicznych, które budzą liczne kontrowersje i opór społeczny, jest wyprowadzenie metodami tradycyjnymi odmian o zwiększonej tolerancji na warunki stresowe, w tym także stresu herbicydowego. Zrozumienie mechanizmów molekularnych leżących u podstaw cechy odporności na herbicydy u roślin uprawnych może przyczynić się do szybszego postępu hodowlanego i wzbogaci nasze zrozumienie interakcji roślin ze środowiskiem.

3. Cel pracy

Celem pracy była identyfikacja różnic w typie i ilości zdarzeń alternatywnego splicingu oraz ekspresji poszczególnych izoform transkryptów powstających u kukurydzy zwyczajnej pod wpływem oprysku herbicydem Roundup®. Zaobserwowane wariacje zdarzeń splicingowych miały ponadto pomóc w wytypowaniu genów biorących udział w kształtowaniu cechy tolerancji na herbicyd. Aby zidentyfikować różnice w mechanizmie alternatywnego splicingu między wytypowanymi dwoma liniami kukurydzy (tolerancyjnej oraz wrażliwej na działanie herbicydu Roundup®) zastosowano metodę sekwencjonowania mRNA drugiej generacji (RNA- seq).

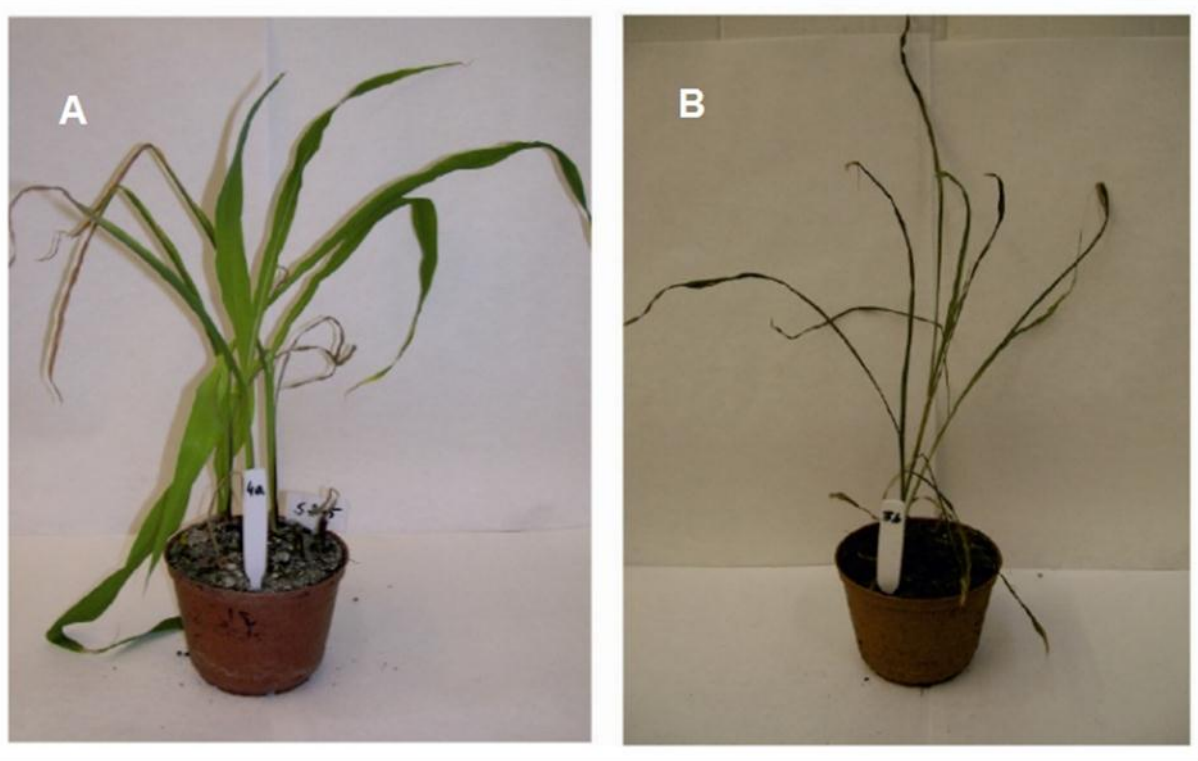
Z przedstawionych w poprzednim rozdziale informacji wynika, że u roślin warunki stresowe powodują modyfikację ekspresji genów, która umożliwia dostosowanie do pojawiających się w środowisku zmian. Stres herbicydowy, który można zaliczyć do stresów abiotycznych, dotyczy większości roślin uprawnych i ma negatywny wpływ na ich plonowanie. Mimo wszystko nadal niewiele wiadomo o mechanizmach biochemicznych leżących u podstaw odpowiedzi roślin na stosowane opryski herbicydami, dlatego postanowiono sprawdzić jak zmienia się jeden z etapów ekspresji genów - proces alternatywnego splicingu, pod wpływem powszechnie stosowanego herbicydu Roundup® u ważnej gospodarczo rośliny jaką jest kukurydza zwyczajna. Postawiona hipoteza badawcza zakłada, że na tak ważnym etapie ekspresji genów mogą istnieć różnice w wyborze miejsc splicingowych w genach istotnych dla przetrwania rośliny podczas występowania warunków stresowych.

4. Materiały i metody

4.1. Materiały

4.1.1 Materiał roślinny

Obiektem badań były dwie linie homozygotyczne kukurydzy, które różniły się reakcją fenotypową na stosowanie herbicydu Roundup®. Linie te zostały wytypowane do badań w oparciu o wyniki projektu badawczego PBZ-MNiSW-2/3/2006 pt. „Stres wywołany herbicydami stosowanymi w roślinach transgenicznych (glifosat i glifosynat) oraz rimsulfuronem - zmiany i detekcja w kukurydzy na poziomie morfologicznym DNA i RNA; poszukiwanie diagnostycznych markerów molekularnych”, który zakładał zarówno ocenę fenotypową jak i stopnia pokrewieństwa oraz zróżnicowania genetycznego 25 linii homozygotycznych kukurydzy zwyczajnej poddanej działaniu 3 herbicydów (Roundup®, Titus i Basta). Linie wytypowane do badań pochodziły z kolekcji spółki Hodowla Roślin Smolice działającej przy Instytucie Hodowli i Aklimatyzacji Roślin PAN. Wybrane linie kukurydzy (linia S79757 oraz linia S245) charakteryzowały się skrajnymi reakcjami na zastosowane dawki herbicydu Roundup®. Linia S245 jest linią najbardziej tolerancyjną spośród przebadanych na ten herbicyd (TL, ang. *tolerant line*). Charakteryzuje się dłuższymi, ciemniejszymi i bardziej sztywnymi liśćmi niż wybrana do porównań linia S79757, która wykazuje największą wrażliwość na oprysk (SL, ang. *sensitive line*). Ma drobniejsze, jaśniejsze i bardziej wiotkie liście, które łatwiej ulegają uszkodzeniom mechanicznym. Tydzień po oprysku herbicydem u roślin z obu linii obserwowano zmiany na liściach (usychanie końcówek, plamki nekrotyczne), które u linii tolerancyjnej nasilały się aż do uschnięcia całych liści już 14 dni po oprysku (Ryc.6). W tym samym punkcie czasowym rośliny z linii tolerancyjnej mimo zmian na liściach, zachowały niezmienny pierwotny pokrój oraz zdolność wzrostu.



Ryc. 6

Zdjęcia roślin linii tolerancyjnej S245 (A) i linii wrażliwej S79757 (B) 14 dni po zabiegu oprysku herbicydem.

Do badania reakcji fenotypowej obu linii kukurydzy na herbicyd Roundup®, która została przeprowadzona w Instytucie Ochrony Roślin, Państwowy Instytut Badawczy, Poznań, pod nadzorem prof. K. Adamczewskiego i prof. J. Adamczyka w ramach wspomnianego projektu badawczego PBZ-MNiSW-2/3/2006, zastosowano subletalne dla kukurydzy dawki herbicydu. Już przy oprysku najniższą dawką (Roundup®0,5 l/ha, co odpowiada 150 g glifosatu) zaobserwowano znaczne zróżnicowanie fenotypowe badanych linii w odpowiedzi na stres herbicydowy.

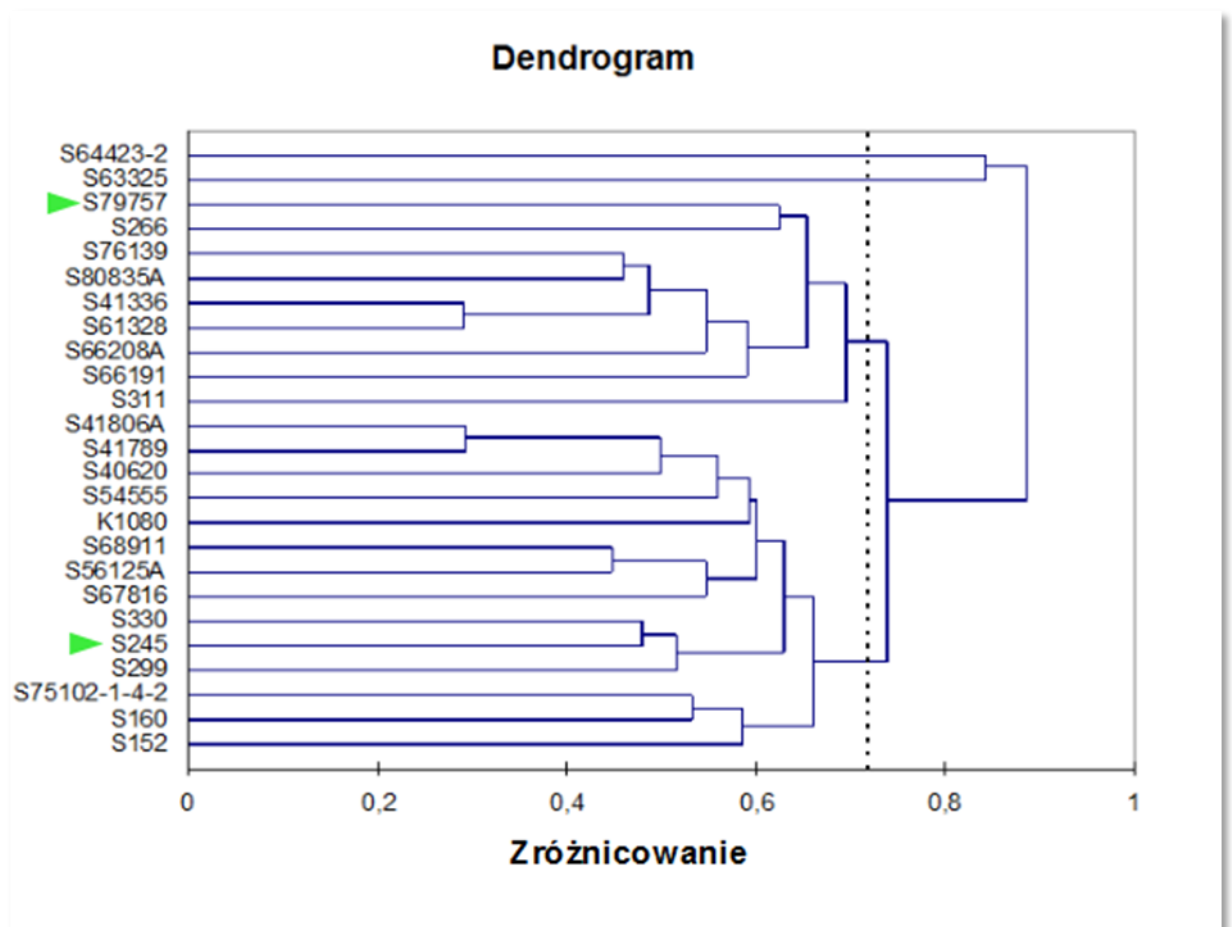
W Tabeli 2 zestawiono procentową wartość uszkodzeń powodowanych przez herbicyd w wybranych do badań liniach kukurydzy, a na Ryc. 7 dendrogram sporządzony na podstawie analizy polimorfizmów AFLP.

4. Materiały i metody

Nazwa linii hodowlanej	Dawka herbicydu Roundup® 360 SL [l/ha]		
	1,0	0,75	0,5
S245 [% uszkodzeń]	40	30	25
S79757 [% uszkodzeń]	85	75	75

Tabela 2

Uszkodzenia roślin kukurydzy po oprysku różnymi dawkami herbicydu Roundup® wyrażona w procentach. Dane pochodzą z badań fenotypowych wykonanych w Instytucie Hodowli i Aklimatyzacji Roślin.



Ryc. 7

Dendrogram pokrewieństwa genotypów kukurydzy analizowanych w projekcie PBZ-MNiSW-2/3/2006 sporządzony na podstawie wyników analizy AFLP. Zielonymi strzałkami zaznaczono linie wybrane do badań charakteryzujące się skrajnymi reakcjami roślin na herbicyd Roundup®.

4.1.2 Stosowane odczynniki chemiczne

Sigma - etanol, izopropanol, fenol, chloroform, błękit bromofenolowy, cyjanol ksylenu, β -merkaptoetanol

POCH- chlorek sodu (NaCl), kwas borowy (H_2BO_3), wodorotlenek sodu (NaOH), octan sodu (CH_3COONa), kwas solny (HCl)

BioShop - agarozą, kwas etylenodiamino-tetraoctowy (EDTA), Tris

Ambion - woda wolna od RNaz

Thermo Scientific - marker wielkości DNA (GeneRuler 50 bp i 100 bp DNA Ladder)

Invitrogen - odczynnik umożliwiający detekcję kwasów nukleinowych po rozdiale elektroforetycznym SYBR[®] Safe DNA gel stain

4.1.3 Oligonukleotydy

Użyte w badaniach jednoniciowe cząsteczki DNA zakupiono w firmie Genomed. Sekwencje nukleotydowe cząsteczek DNA przedstawiono w Tabeli 3.

Nazwa genu	Starter (5'-3')
Białko P związane z czynnikiem transkrypcyjnym Myb (ang. <i>Myb-related protein P</i>)	
Transkrypt 1	F: CGTGCTGCGAGAAGGTGGGG R: CGAGGGTGGCGTGGAGCTTG
Transkrypt 2	F: CCGGCCGGTGGTCCCTGAT R: CCAAGCCCTTCGAGGGCAGC
Transkrypt 3	F: AAGAATGCAGGCCTGCTCCG R: CCAAGCCCTTCGAGGGCAGC
S-metylotransferaza 2 homocysteiny (ang. <i>homocysteine S-methyltransferase 2</i>)	
Transkrypt 1	F: CACCCGGTTCTGGTCGCAGC R: TTGCCTGCAGCAGCCACCG
Transkrypt 2	F: CCTGCTCTTCGCCACTATTC R: TTGCCTGCAGCAGCCACCG
Dehydrogenaza glutationu (ang. <i>glutamate dehydrogenase</i>)	
Transkrypt 1	F: AAGCTGCTGGGCCTCGACTC R: CCCAAGCGCTGCTGGGATGA
Transkrypt 2	F: ATCCAGCCCCCTCGTCTGCGA R: GGTGCGGATGCTCGCAGTGA

Tabela 3

Zestawienie sekwencji jednoniciowych DNA stosowanych jako startery w reakcjach PCR (F - starter komplementarny do końca 5' powielanego fragmentu; R - starter komplementarny do końca 3' powielanego fragmentu)

4.1.5 Zestawy komercyjne

Ambion - TURBO™ DNase (2 U/μL)

Agilent - zestaw RNA 6000 Nano

4. Materiały i metody

Life Technologies - SuperScript III Reverse Transcriptase

Illumina - TruSeq™ RNA Sample Prep Kit, TruSeq SBS Kit v5

Marchery-Nagel - NucleoSpin Plant II

Thermo Scientific - PCR Master Mix

Qiagen - RNeasy Plant Mini Kit, RNase-Free DNase Set, MiniElute PCR Purification Kit, MiniElute Gel Extraction Kit, OneStep RT-PCR Kit

4.1.6 Roztwory

Roztwór do nakładania kwasów nukleinowych na żele agarozowe

0,05% błękit bromofenolowy w/v, 0,05% cyjanol ksylenu w/v, 0,05% pomarańcz G (orange G) w/v, 30% glicerol

Roztwór do elektroforezy kwasów nukleinowych w żelu poliakrylamidowym (TBE) 10x

108 g Tris/HCl ($C_k=890$ mM)

55 g kwas borowy ($C_k=890$ mM)

40 ml 0,5M EDTA pH 8.0 ($C_k=20$ mM)

dopełniano H₂O do 1000 ml

4.1.7 Bazy danych

W Tabeli 5 na końcu podrozdziału znajduje się podsumowanie wykorzystywanych w pracy baz danych oraz ich adresy internetowe, poniżej znajduje się krótka charakterystyka każdej z nich.

Baza danych NCBI (ang. *National Center for Biotechnology Information*)

Wzrastające znaczenie komputerowych metod przetwarzania informacji w naukach biomedycznych doprowadziło do utworzenia NCBI 4 listopada 1988 roku jako części

Narodowej Biblioteki Medycznej (NLM, ang. *National Library of Medicine*) działającej przy amerykańskim Narodowym Instytucie Zdrowia (NIH, ang. *National Institutes of Health*). NLM zostało wybrane do tego celu ze względu na swoje doświadczenie w budowaniu i utrzymywaniu baz danych, a jako część NIH mogło ustanowić projekty badawcze związane z tematyką obliczeniowej biologii molekularnej. Początkowo zadania NCBI koncentrowały się wokół rozwijania nowych technologii informatycznych mających wspomóc zrozumienie zjawisk biologii molekularnej leżących u podstaw chorób człowieka, jednak wkrótce poszerzono obszar zagadnień, którego dotyczy działalność NCBI. Obecnie zautomatyzowane systemy NCBI służą do przechowywania i analizy wiedzy z dziedziny biologii molekularnej, genetyki i biochemii wielu gatunków roślin i zwierząt. NCBI prowadzi także ciągłe prace nad ułatwieniem dostępu do stworzonych baz danych, ich utrzymaniem, zdobywaniem informacji w kraju, jak i zagranicą. Realizuje także badania związane z zaawansowanymi komputerowymi metodami przetwarzania informacji służących do analizy struktur i funkcji biologicznie ważnych cząsteczek.

W skład NCBI wchodzi obecnie 60 szczegółowych baz danych. W badaniach wykorzystano następujące:

- Genome - zawiera sekwencje i mapy genomów ponad 1000 organizmów, których projekty sekwencjonowania zostały już ukończone oraz tych, które są jeszcze w toku. Zebrane dane genomowe reprezentują trzy podstawowe domeny życia (archea, bakterie i organizmy eukariotyczne), a także wiele sekwencji pochodzących z wirusów, wiroidów, plazmidów i organelli komórkowych;
- Nucleotide - jest to zbiór sekwencji nukleotydów stanowiących transkrypty, geny i inne części genomów. Zebrane w niej informacje pochodzą z kilku źródeł m. in.:
 - GenBank - zawiera wszystkie publicznie dostępne sekwencje DNA [69];
 - RefSeq - kolekcja sprawdzonych i unikalnych (ang. *non-redundant*) sekwencji genomowego DNA, transkryptów RNA (*refseq_rna*) i białek opisanych przez NCBI, zapewnia stabilne odniesienie dla opisywania genomu, identyfikacji

genów, mutacji, analizy polimorfizmów, badań ekspresji i analiz porównawczych;

- TPA (ang. *Third Party Annotation*) - składa się z sekwencji, które powstały w oparciu o wcześniej istniejące dane zdeponowane w GenBank, których nowa anotacja jest udowodniona eksperymentalnie i została opublikowana w recenzowanym czasopiśmie naukowym;
- PDB (ang. *Protein Data Bank*) - zawiera informacje na temat struktury przestrzennej białek i kwasów nukleinowych, które zostały zebrane na podstawie rentgenografii strukturalnej i spektroskopii NMR, zawierają informacje o strukturze drugorzędowej białek i kwasów nukleinowych wraz z współrzędnymi atomów;
- Conserved Domain Database (CDD) - jest bazą danych domen białkowych konserwatywnych ewolucyjnie; opisuje także funkcje, które te domeny pełnią, a dzięki możliwości modelowania 3D zapewnia wgląd w relacje sekwencja/struktura/funkcja.

Baza danych UniProtKB (ang. *Universal Protein Knowledgebase*)

Bazą danych białek, z której korzystałam w czasie wykonywania badań, jest baza danych UniProt Knowledgebase. Jest to centralny ośrodek gromadzenia informacji na temat białek i pełnionych przez nie funkcji. Baza ta powstała jako konsorcjum trzech jednostek zajmujących się badaniami białek - Europejskiego Instytutu Bioinformatyki (EBI, ang. *European Bioinformatic Institute*), Szwajcarskiego Instytutu Bioinformatyki (SBI, ang. *Swiss Bioinformatic Institute*) i jednego z najstarszych ośrodków badawczych związanych z proteomiką PIR (ang. *Protein Information Resource*), którego podstawą założenia był opublikowany w 1965 roku przez Margaret Dayhoff "Atlas sekwencji i struktur białek". Oprócz podstawowych danych, takich jak sekwencja aminokwasów, nazwa białka, dane taksonomiczne organizmu, w którym zostało zidentyfikowane czy dane bibliograficzne, wejście (ang. *entry*) w bazie UniProt zawiera jak najwięcej informacji, które zostały określone dla danego białka. Obejmuje to ontologię, klasyfikacje czy wszelkie możliwe nawiązania i odsyłacze obecne w literaturze czy

innych, zdeponowanych rekordach. Ponadto wyraźnie zaznaczona jest jakość anotacji dla poszczególnych białek - czy zostało ono zidentyfikowane na drodze analizy informatycznej czy też dowodów eksperymentalnych. Baza ta jest podzielona na dwie części - UniProtKB/SwissProt, która zawiera rekordy z ręcznie sprawdzonymi anotacjami pochodzącymi z literatury jak i analizy obliczeniowej, a druga część - UniProtKB/TrEMBL gromadzi dane niezrecenzowane, które czekają na ręczne sprawdzenie. Około 98% sekwencji znajdujących się w bazie UniProt zostało przetłumaczonych z sekwencji kodujących (CDS, ang. *coding sequence*), które zostały zdeponowane w publicznych bazach danych kwasów nukleinowych - EMBL-Bank, GenBank, DDBJ (ang. *DNA Data Bank of Japan*). Wszystkie te sekwencje wraz z danymi umieszczonymi przez ich autorów są gromadzone w UniProtKB/TrEMBL. Następnie dane te są ręcznie sprawdzane przez zespół ekspertów, którzy uzupełniają rekord o dane literaturowe, wyniki eksperymentów czy odnośniki i w ten sposób rekord trafia do UniProtKB/SwissProt. Informacje tam zdeponowane są aktualizowane, gdy pojawią się nowe doniesienia o danym białku.

Baza danych PlantGDB (ang. *Plant Genome Database*)

Wzrastająca ilość danych pochodzących z wysokoprzepustowych projektów sekwencjonowania przewyższa naszą zdolność do analizy i anotacji uzyskanych sekwencji. Wychodząc z tego założenia twórcy PlantGDB pragną stworzyć platformę nie tylko do przechowywania danych, ale także ich analizy. Ich celem jest opracowanie dokładnych, dających się skalować, wysoce zautomatyzowanych i kompleksowych metod analizy genomów roślinnych. Rezultatami projektu są nowe algorytmy predykcyjne, portale internetowe do wizualizacji danych, a także cyberinfrastruktura służąca rozwojowi narzędzi do obliczeń rozproszonych, udostępniania protokołów czy gromadzenia zapisów dotyczących pochodzenia danych (ang. *data provenance*). W dłuższej perspektywie projekt jakim jest PlantGDB dąży do rozwinięcia biologii genomowej z nauki bardziej opisowej jaką jest teraz, do nauki predykcyjnej, której przewidywania będą oparte o algorytmy i obliczenia pochodzące z wielu pomiarów ilościowych [70]. Ostatnia wersja tej bazy danych została udostępniona 15.12.2011 roku i zawiera dane zestawione w Tabeli 4.

Typ danych	Opis	Liczba sekwencji	Liczba gatunków
Białka	Unikatowe sekwencje białek	768 327	59 165
EST	Sekwencyjne znaczniki ekspresji (ang. <i>Expressed Sequence Tag</i>)	23 579 550	870
GSS	Zbadane sekwencje genomowe (ang. <i>Genomic Survey Sequence</i>)	13 986 498	356
STS	Miejsca w genomie oznaczone sekwencją (ang. <i>Sequence-Tagged Site</i>)	159 869	273
TSA	Złożone transkryptomy z sekwencjonowania typu shotgun (ang. <i>Transcriptome Shotgun Assembly</i>)	1 672 820	46
HTG	Wysokowydajnościowe DNA genomowe (ang. <i>High Throughput Genomic DNA</i>)	25 458	87
HTC	Wysokowydajnościowe cDNA (ang. <i>High Throughput cDNA</i>)	66 543	120
PLN	Dział roślinnego mRNA (ang. <i>Plant Division mRNA</i>)	1 749 128	107 076
DNA_PLN	Dział roślinnego DNA (ang. <i>Plant Division DNA</i>)	1 109 744	98 395

Tabela 4

Zestawienie liczby i typów danych w bazie PlantGDB.

Baza danych Maize GDB (ang. *Maize Genetics/Genomics Database*)

Jest to główne repozytorium informacji dotyczących kukurydzy zwyczajnej, które powstało z połączenia dwóch istniejących wcześniej baz danych - MaizeDB i ZmDB. Integracja danych w nich zebranych umożliwiła uporządkowanie i ujednoczenie schematu ich zapisu, a także opracowanie praktycznego interfejsu odpowiedzialnego za dostęp do informacji, który ma

ułatwić ich przeszukiwanie. Przede wszystkim jednak MaizeGDB jest bazą danych tworzoną przez społeczność naukowców zajmujących się badaniami nad tym gatunkiem rośliny uprawnej i modelowej; jest repozytorium danych pochodzących z prawie 50 projektów dotyczących m. in. opisywania genów kukurydzy, mapowania genomu, stworzenia map cytogenetycznych, lokalizowania wszystkich elementów powtórzonych czy badania endospermu i fenotypu różnych odmian kukurydzy. Dlatego też głównymi założeniami twórców MaizeGDB jest uczynienie tej bazy danych jak najbardziej użyteczną i łatwo dostępną dla wszystkich członków społeczności naukowej poprzez:

- udostępnianie prostych w użyciu narzędzi bioinformatycznych, które są jednak dostosowane do ekstrakcji dużej ilości danych,
- szybką i skuteczną reakcją na otrzymywane uwagi i propozycje,
- ciągłe poszukiwanie nowych danych i informacji o kukurydzy, które wzbogacają zgromadzony zasób wiedzy.

Jedną z ważniejszych cech tej bazy danych jest spójny interfejs, który na każdej stronie zawiera identyczne narzędzia i skróty do najważniejszych stron, wspomagające wyszukiwanie konkretnych danych. Obszerny dział pomocy, w którym znajdują się także liczne samouczki dodatkowo ułatwia poruszanie się po bazie danych. Każdy rekord zawiera również liczne odsyłacze do stron z danymi źródłowymi. Informacje są zamieszczane w postaci oddzielonych tabulatorami danych (o ile to możliwe), które są łatwe do eksportowania i użycia w różnych programach bioinformatycznych. W przyszłości twórcy MaizeGDB chcą także zwiększyć nacisk na współpracę z naukowcami zajmującymi się innymi gatunkami roślin, aby wspólnie ulepszać zarówno ontologię roślin jak i poszczególnych genów, a także przepływ i wymianę danych między różnymi repozytoriami.

Baza danych Gramene

Baza danych Gramene została utworzona przez prof. Doreen Ware z laboratorium Cold Spring Harbor. Pierwotnie zbierała ona informacje na temat genomów traw, a przede wszystkim zbóż. Obecnie Gramene jest bardzo dobrym źródłem informacji dla porównawczej genomiki funkcjonalnej zarówno zbóż jak i roślin modelowych. Zgromadzono w niej

informacje na temat 27 w pełni i 10 częściowo zsekwencjonowanych genomów referencyjnych. Jedną z głównych zalet tej bazy danych jest zastosowanie metod analizy filogenetycznej do porównań genomów, a także użycie ontologii, które pozwalają na integrację danych strukturalnych i funkcjonalnych. Annotacja genomów w oparciu o skonstruowane drzewa filogenetyczne pozwala wnioskować o relacjach syntenicznych i ortologicznych badanych genów. Dla 10 gatunków, w tym *Arabidopsis*, kukurydzy i ryżu, dostępne dane na temat wariacji genetycznej oraz mapy genomowe pomagają przewidywać skutki zmienności na poziomie genów i transkryptów. Gramene posiada również sekcję, w której gromadzone są informacje na temat szlaków metabolicznych (w 2014 roku dla 10 gatunków) oraz reaktomów roślinnych (ang. *plant reactome*), w którym do tej pory zamieszczono informacje o 133 szlakach metabolicznych ryżu. Ta część bazy danych ma być rozbudowywana i powiększana o kolejne gatunki roślin. Gramene gromadzi także informacje na temat markerów genetycznych i mapowania loci cech ilościowych (QTL, ang. *quantitative trait loci*), które są szczególnie przydatne naukowcom związanym z wyprowadzaniem i hodowlą nowych odmian roślin uprawnych. Projekt związany z rozwojem bazy danych Gramene zapewnia platformę do integracji danych ewolucyjnych, biochemicznych, genetycznych, związanych z biologią molekularną, systemową, jak i hodowlą roślin [71].

Nazwa bazy danych	Adres strony internetowej
NCBI Genome	http://www.ncbi.nlm.nih.gov/genome
NCBI Nucleotide	http://www.ncbi.nlm.nih.gov/nucleotide
NCBI Conserved Domain	http://www.ncbi.nlm.nih.gov/cdd
UniProtKB	http://www.uniprot.org/
Plant GDB	http://www.plantgdb.org/
Maize GDB	http://www.maizegdb.org/
Gremene	http://www.gramene.org

Tabela 5

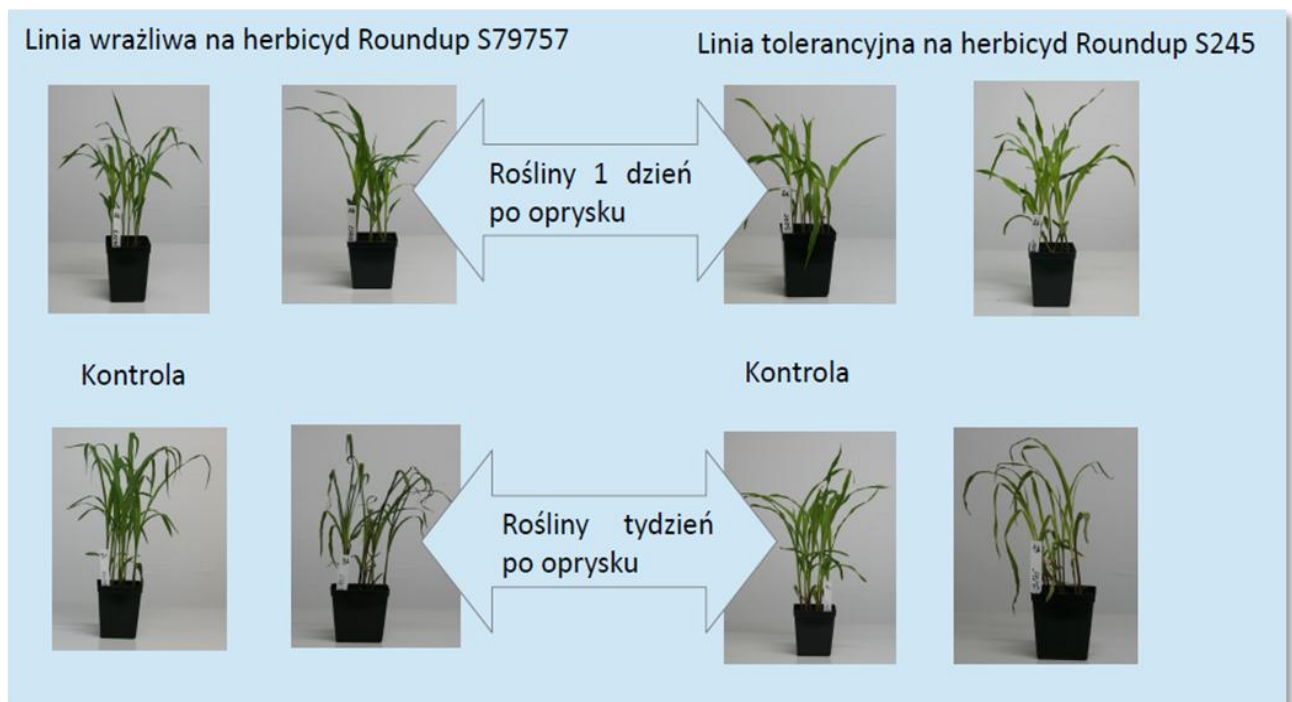
Wykaz wykorzystanych baz danych i ich adresy internetowe.

4.2 Metody

4.2.1 Hodowla wybranych linii kukurydzy zwyczajnej

Hodowlę kukurydzy przeprowadzono w Instytucie Ochrony Roślin w Poznaniu. Ziarna kukurydzy do badań pochodziły ze stacji Hodowli Roślin Smolice Sp. z o. o. Do hodowli użyto wazonów o pojemności 1,2 litra wypełnione glebą ogrodniczą. Wysiano po 8 ziarniaków do każdego z 10 wazonów (dla każdej linii kukurydzy). Hodowlę prowadzono w warunkach szklarniowych przy zastosowaniu naprzemiennie fazy 16h oświetlenia i 8h ciemności oraz temperatury 22°C i stałej wilgotności. Rośliny z 6 wazonów poddano opryskowi herbicydem, a rośliny z 4 wazonów, nietraktowane herbicydem, służyły w badaniach jako materiał kontrolny. Zabieg opryskiwania wykonano w fazie 4-5 liści kukurydzy opryskiwaczem szklarniowych wyposażonym w rozpylacze płaskostrumieniowa TeeJet TT 11002 i stosując ciśnienie 3 bary, ilość wody do zabiegu wynosiła 250 l/ha. Glifosat zastosowano w dawce 360 g (herbicyd Roundup® 360 SL - 1,0 l/ha + adiuwant AS 500 SL - 4,0 l/ha).

Dzień i tydzień po wykonaniu zabiegu oprysku zebrano materiał roślinny (liście) roślin zarówno poddanych zabiegowi oprysku herbicydem jak i kontrolnych z obu linii (na Ryc. 8 przedstawione są zdjęcia roślin przed zebraniem materiału). Następnie przeprowadzono ich homogenizację w ciekłym azocie, a otrzymany proszek zamrożono w -80°C jako materiał wyjściowy do izolacji kwasów nukleinowych.



Ryc. 8

Zdjęcia roślin z obu badanych linii, zarówno kontrolnych oraz traktowanych herbicydem, dzień i tydzień po zabiegu oprysku herbicydem.

4.2.2 Izolacja całkowitego RNA oraz trawienie DNA

Izolację całkowitego RNA z zamrożonych, rozdrobnionych liści kukurydzy zwyczajnej przeprowadzano przy użyciu zestawu komercyjnego QiagenRNeasy Plant Mini Kit. W trakcie izolacji wykonywano także trawienie DNA stosując RNase-Free DNase Set tego samego producenta. Obie procedury wykonywano zgodnie z zaleceniami producenta z modyfikacjami:

- 100 mg utartej uprzednio w mrożeniu tkanki liści kukurydzy zalewano 450 μ l buforu RLT, mieszano przy użyciu wytrząsarki laboratoryjnej i inkubowano przez 3 min w temperaturze 56°C;
- lizat komórek przenoszono na kolumnę z filtrem umieszczoną w 1,5 ml probówce i wirowano przez 2 min, 13 000g w temperaturze pokojowej;
- supernatant przenoszono do nowej 1,5 ml probówki i dodawano 0,5 objętości etanolu, mieszano przez pipetowanie;

4. Materiały i metody

- roztwór przenoszono na kolejną kolumnę z filtrem umieszczoną w 1,5 ml probówce, wirowano 15 s, 8 000g w temperaturze pokojowej, przesącz usuwano;
- następnie przeprowadzano trawienie DNA:
 - na kolumnę dodawano 350 μ l buforu RW1 i ponownie ją wirowano 15 s, 8 000g w temperaturze pokojowej, przesącz usuwano,
 - bezpośrednio na filtr kolumnienki podawano roztwór DNazy (10 μ l DNase I stock solution i 70 μ l buforu RDD) i inkubowano w temperaturze pokojowej 15 min,
 - dodawano 350 μ l buforu RW1 na kolumnę i wirowano 15 s, 13000 g w temperaturze pokojowej, przesącz usuwano,
- w kolejnym etapie dodawano 500 μ l buforu RPE i ponownie wirowano kolumnienki 15 s, 13 000g w temperaturze pokojowej, przesącz usuwano;
- aby osuszyć membranę z resztek etanolu obecnego w buforze RPE, po dodaniu 500 μ l tego buforu na kolumnienkę całość wirowano 2 min, 13 000g w temperaturze pokojowej;
- po wirowaniu kolumnienkę przekładano do nowej probówki o objętości 1,5 ml; aby wymyć RNA związane z filtrem, dodawano 30 μ l wody wolnej od RNaz bezpośrednio na membranę;
- wirowano 1 min, 13 000g, w temperaturze pokojowej;
- po analizie ilościowej i jakościowej uzyskany przesącz przechowywano w -80°C.

4.2.3 Analiza ilościowa i jakościowa całkowitego RNA

Stężenie otrzymanego po izolacji całkowitego RNA wyznaczano poprzez pomiar jego absorbancji przy długości fali $\lambda = 260$ nm (1 OD/ml= 40 μ g/ml) przy pomocy spektrofotometru Nano Drop 2000c. Wstępnej oceny jakości wyizolowanych preparatów dokonywano na podstawie wyglądu krzywej absorbancji w zależności od długości fali w spektrum λ 190 - 840 nm.

By dokładnie ocenić jakość otrzymanego RNA przeprowadzano elektroforezę kapilarną z wykorzystaniem aparatu Bioanalyzer 2100 firmy Agilent (program Plant RNA Nano) oraz zestawu Agilent RNA Nano 6000. Na przygotowaną płytkę do elektroforezy nakładano po 1

μl preparatu i przez 30 min prowadzono elektroforezę w obecności markera wielkości. Uzyskane obrazy rozdziału elektroforetycznego oraz wyznaczony współczynnik integralności RNA (RIN, ang. *RNA integrity number*) analizowano pod kątem jakości rozdziału (wyraźne prążki, brak smug) oraz wielkości współczynnika RIN.

4.2.3 Izolacja DNA i reakcja PCR w gradiencie temperatur

Aby przeprowadzić reakcję łańcuchowej polimerazy (PCR, ang. *polymerase chain reaction*) z zamrożonych liści kukurydzy izolowano DNA. Izolację przeprowadzano przy użyciu zestawu NucleoSpin Plant II zgodnie z zaleceniami producenta z modyfikacjami:

- odważano 100 mg wcześniej utartych w moździerzu i zamrożonych liści do 1,5 ml probówki;
- dodawano 300 μl buforu PL2 i mieszano z użyciem wytrząsarki laboratoryjnej
- następnie aby usunąć RNA dodawano 10 μl RNazy A, mieszano i inkubowano 10 min w temperaturze 65°C;
- w kolejnym etapie dodawano 75 μl buforu PL3, mieszano i inkubowano 5 min na lodzie;
- nanoszono lizat na kolumnę z filtrem umieszczoną w 2 ml probówce, którą wirowano przez 2 min., 11 000g w temperaturze pokojowej;
- klarowny przesącz przenoszono do nowej 1,5 ml probówki, dodawano 450 μl buforu PC i mieszano przez pipetowanie;
- całość nanoszono na kolumnę z filtrem umieszczoną w 2 ml probówce, wirowano przez 1 min, 11 000g w temperaturze pokojowej; powstały przesącz usuwano;
- na kolumnę dodawano 400 μl buforu PW1, wirowano 1 min, 11 000g w temperaturze pokojowej; powstały przesącz usuwano;
- dodawano 700 μl buforu PW2, wirowano 1 min, 11 000g w temperaturze pokojowej; powstały przesącz usuwano;
- aby usunąć resztki buforów i osuszyć membranę dodawano 200 μl buforu PW2 i wirowano kolumnienki 2 min, 11 000g w temperaturze pokojowej;

4. Materiały i metody

- kolumnę umieszczano następnie w nowej probówce 1,5 ml i dodawano 50 μ l buforu PE podgrzanego do 65°C, prowadzono 5 minutową inkubację w temperaturze 65°C;
- aby wymyć DNA związane z membraną wirowano kolumnienki przez 1 min, 11 000g w temperaturze pokojowej.

Ilość uzyskanego DNA mierzono spektrofotometrycznie przy użyciu spektrofotometru Nano Drop 2000ci wyznaczano z zależności absorbancji przy długości fali $\lambda = 260$ nm (1 OD/ml = 50 μ g/ml).

Po zaprojektowaniu i otrzymaniu zsyntetyzowanych starterów przeprowadzono reakcję PCR w gradiencie temperatur dla każdej pary starterów. Miała ona na celu dobranie odpowiedniej temperatury hybrydyzacji, która umożliwi optymalne łączenie starterów do matrycy zapobiegając jednocześnie ich renaturacji. Zbyt wysoka temperatura może powodować brak wiązania starterów do matrycowego DNA, zbyt niska może skutkować niespecyficznym wiązaniem starterów i pojawianiem się niespecyficzných produktów reakcji. Skład typowej pojedynczej reakcji PCR przedstawiono w Tabeli 6, a schemat reakcji PCR w gradiencie temperatur w Tabeli 7.

Składnik	Ilość
Matryca	1-100 ng
Starter komplementarny do końca 3' powielanego fragmentu (7,5 μ M)	0,75 μ l
Starter komplementarny do końca 5' powielanego fragmentu (7,5 μ M)	0,75 μ l
Bufor reakcyjny 10x	1 μ l
Polimeraza DNA Taq (5U/ μ l)	1 μ l
Mieszanina dNTP (10 mM każdego dNTP)	0,2 μ l
MgCl ₂ (25mM)	0,5 μ l
H ₂ O	Do 10 μ l

Tabela 6

Skład typowej reakcji PCR (w nawiasach podano stężenia końcowe składników reakcji).

Etap	Czas	Temperatura	Ilość powtórzeń
Denaturacja wstępna	5 min	95°C	1x
Denaturacja	0,5 min	94°C	35x
Hybrydyzacja	0,5 min	50-60°C	
Elongacja	0,5 min	72°C	
Elongacja końcowa	10 min	72°C	1x

Tabela 7

Schemat reakcji PCR w gradiencie temperatur

4.2.4 Reakcja odwrotnej transkrypcji sprzężona z reakcją PCR

W celu detekcji transkryptów powstałych na drodze alternatywnego splicingu dla wybranych genów przeprowadzono reakcję odwrotnej transkrypcji (RT, ang. *reverse transcriptase*) sprzężoną z reakcją PCR. Do tego celu wykorzystywano zestaw OneStep RT-PCR Kit, który zawiera specjalnie dobraną mieszaninę enzymów umożliwiającą przeprowadzenie zarówno reakcji odwrotnej transkrypcji, jak i reakcji łańcuchowej polimeryzacji w jednej probówce. Mieszanina enzymów zawiera odwrotną transkryptazę Omniscript i Superscript, które charakteryzują się wysokim powinowactwem do matrycy RNA i zapewniają tym samym wysoką czułość i wydajność transkrypcji nawet z niewielkich ilości RNA. Po etapie reakcji odwrotnej transkrypcji (przebiegającej w 50°C przez 30 min), mieszanina reakcyjna jest inkubowana przez 15 min w temperaturze 95°C, aby aktywować polimerazę DNA - HotStart Taq i jednocześnie inaktywować odwrotne transkryptazy. Ten etap eliminuje niespecyficzne produkty amplifikacji, zmniejsza tło, zapewniając tym samym wysoką czułość i powtarzalność reakcji. Ogólny skład reakcji przedstawiono w Tabeli 8, natomiast schemat przebiegu reakcji w Tabeli 9. Zarówno reakcję PCR w gradiencie temperatur jak i reakcję odwrotnej transkrypcji sprzężoną z reakcją PCR prowadzono przy użyciu termocyklera T100 firmy Bio-Rad.

Składnik	Ilość
Matryca RNA	1pg-2 µg
Mieszanina dNTP (400µM każdego z dNTP)	2 µl
Bufor reakcyjny (1x Qiagen One Step RT-PCR buffer)	10 µl
QiagenOneStep RT-PCR Enzyme Mix	2 µl
Starter komplementarny do końca 3' powielanego fragmentu (0,6 µM)	2 µl
Starter komplementarny do końca 5' powielanego fragmentu (0,6 µM)	2 µl
Woda wolna od RNaz	do 50 µl

Tabela 8

Składniki reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR (w nawiasach podano stężenia końcowe składników reakcji).

Etap	Czas	Temperatura	Ilość powtórzeń
Reakcja odwrotnej transkrypcji	30 min	50°C	1x
Aktywacja polimerazy i inaktywacja odwrotnej transkryptazy	15 min	95°C	
Denaturacja	0,5 min	94°C	30x
Hybrydyzacja	0,5 min	Zależna od starterów	
Wydłużanie	0,5 min	72°C	
Wydłużanie końcowe	10 min	72°C	1x

Tabela 9

Ogólny schemat przebiegu reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR

4.2.4 Oczyszczanie i analiza produktów PCR w żelu agarozowym

Po zakończonej reakcji PCR, produkty oczyszczano przy użyciu zestawu MiniElute PCR Purification Kit zgodnie z zaleceniami producenta z modyfikacjami:

4. Materiały i metody

- mieszaninę reakcyjną przenoszono do 1,5 ml probówki i dodawano 5 objętości buforu PB, następnie całość nanoszono na kolumnę z membraną umieszczoną w 2 ml probówce;
- wirowano 1 min, 13 000g w temperaturze pokojowej, przesącz usuwano;
- 750 μ l buforu PE dodawano na kolumnę i wirowano 1 min, 13 000g w temperaturze pokojowej;
- aby pozbyć się resztek etanolu, po usunięciu przesączu, kolumnę wirowano 2 min, 13 000g w temperaturze pokojowej, a następnie przenoszono do nowej probówki 1,5 ml;
- aby wymyć DNA 10 μ l wody nanoszono bezpośrednio na membranę kolumny; po 1 minutowej inkubacji, wirowano kolumnę 1 min, 13 000g w temperaturze pokojowej.

Oczyszczone produkty amplifikacji reakcji PCR (w gradiencie temperatur oraz reakcji sprzężonej z odwrotną transkrypcją) rozdzielano w 1,5% - 2% żelu agarozowym o wymiarach 15cm x 15cm i grubości 0,5 cm (Tabela 10). Do żelu dodawano odczynnik wiążącego się do kwasów nukleinowych - SYBR[®] Safe DNA gel stain, a rozdział elektroforetyczny prowadzono w obecności markera wielkości, dzięki któremu można było oszacować wielkość powstałych produktów reakcji PCR. Przed nałożeniem na żel do preparatów dodawano 1/6 objętości buforu obciążającego, a rozdział prowadzono w buforze 1xTBE przy natężeniu prądu 40-120mA, w temperaturze pokojowej. Żele obrazowano naświetlając je promieniowaniem UV przy zastosowaniu systemu dokumentacji żeli UV GelDoc-It 310 ImagingSystem firmy UVP.

Składnik	1,5% żel agarozowy	2% żel agarozowy
agarozą	0,75g	1 g
1 x bufor TBE	50 ml	50 ml
SYBR® Safe DNA gel stain	0,1 µl/ml	0,1 µl/ml

Tabela 10

Skład żeli agarozowych

4.2.5 Sekwencjonowanie transkryptomu

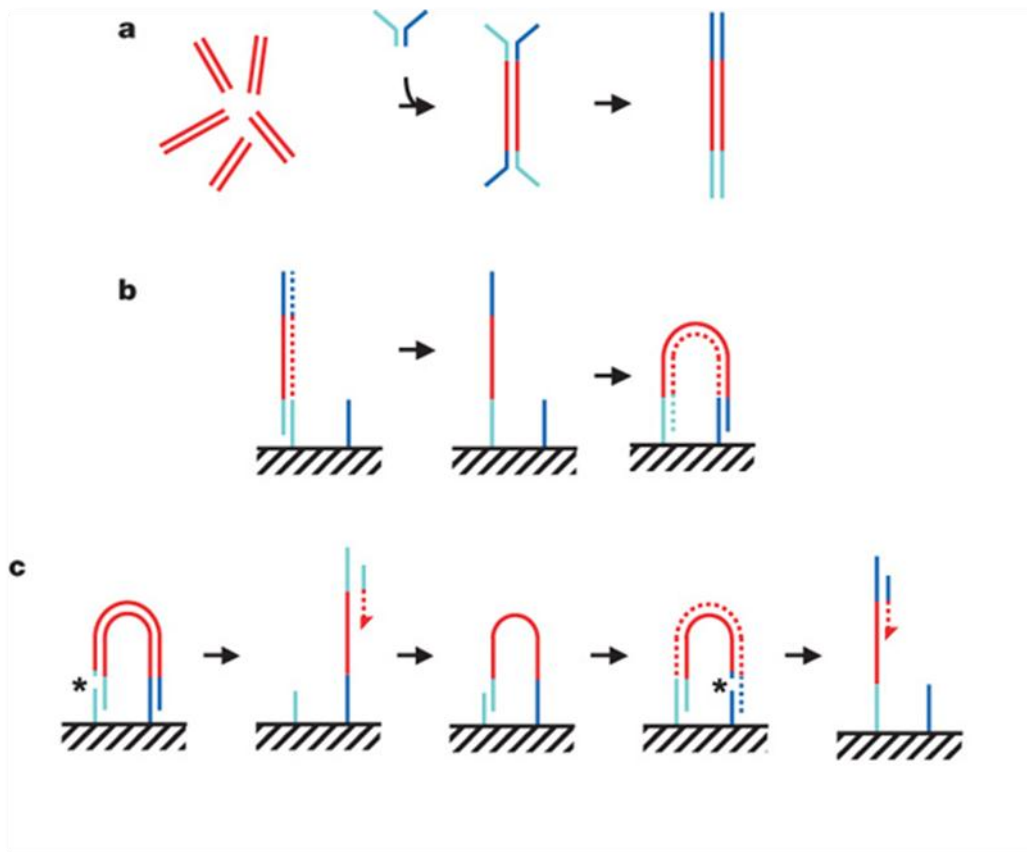
Po wyizolowaniu całkowitego RNA próbki zostały przygotowane do sekwencjonowania przy użyciu zestawu TruSeq™ RNA Sample Prep Kit, a do przeprowadzenia reakcji sekwencjonowania zastosowano zestaw TruSeq SBS Kit v5 (oba firmy Illumina). Sekwencjonowanie typu paired-end odcinków cDNA o długości 50 nt przeprowadzono we współpracy z dr A. Żmieńko, na aparacie Illumina Genome Analyzer II, który znajduje się w Europejskim Centrum Bioinformatyki i Genomiki.

Przygotowanie próbek RNA do sekwencjonowania wymagało jego konwersji do biblioteki cDNA i hybrydyzacji z płytką sekwencyjną. W pierwszym etapie izolowano frakcję mRNA z całkowitego RNA z użyciem kulek magnetycznych opłaszczonych oligo dT. Przyłączają się do nich tylko cząsteczki mRNA posiadające ogon poli(A). Następnie mRNA poddawano fragmentacji przy użyciu kationów dwuwartościowych i podwyższonej temperatury. Stosując startery o przypadkowej sekwencji i odwrotną transkryptazę SuperScript II na matrycy pofragmentowanego mRNA przeprowadzano syntezę pierwszej nici cDNA. Następnie przy pomocy RNazy H usuwano matrycę RNA, a polimerazę DNA I używano do syntezy nici komplementarnej - w ten sposób otrzymywano dwuniciowe sekwencje cDNA. W kolejnym etapie usuwano lepkie końce 3' (aktywność egzonukleazowa) i dobudowywano brakujące nukleotydy na końcu 5', dzięki zastosowaniu polimerazy T4 i fragmentu Klenowa. Zmodyfikowane w ten sposób fragmenty posiadały tępe końce zarówno 3' jak i 5'. Wykorzystując aktywność polimerazową fragmentu Klenowa do końca 3' dodawano

pojedynczą resztę adenozyliny. Do tak przygotowanego DNA można było doligować sekwencje adaptorowe, które posiadały na końcu 3' wystającą resztę tymidyny. Specjalnie dobrane sekwencje adapterowe musiały zostać dołączone, aby umożliwić amplifikację fragmentów w reakcji PCR, a przede wszystkim hybryzację dwuniciowej biblioteki cDNA (ds cDNA, ang. *double-stranded complementary DNA*) z płytką sekwencyjną. Sekwencje adaptorowe umożliwiają sekwencjonowanie więcej niż jednej biblioteki w pojedynczej ścieżce na płycie sekwencyjnej (ang. *multiplexing*). Zawierają specyficzne sekwencje, które służą jako swoiste kody kreskowe, dzięki którym można rozpoznać, która sekwencja pochodzi z jakiej biblioteki. W kolejnym etapie produkty ligacji rozdzielano w żelu agarozowym, z którego wycinano prążki o odpowiedniej wielkości. Pozwala to na usunięcie niezligowanych adaptorów czy ich dimerów i na selekcję fragmentów odpowiedniej wielkości, która umożliwia wydajne tworzenie klastrów w czasie sekwencjonowania. Po elucji dwuniciowego cDNA z wyciętego prążka przeprowadzano wzbogacanie biblioteki (ang. *enrichment*) o te fragmenty, które miały odpowiednio przyłączone do obu końców sekwencje adaptorowe. Zastosowano do tego reakcję PCR i startery, które wiązały się do końców sekwencji adaptorowych. Ilość cykli została tak dobrana, aby zminimalizować ryzyko preferencyjnego powielania niektórych sekwencji, co mogłoby zmniejszyć reprezentatywność przygotowanej biblioteki. W ostatnim etapie sprawdzano jakość otrzymanych bibliotek cDNA poprzez elektroforezę w żelu agarozowym. Sprawdzano czystość, szacowano wielkość i stężenie otrzymanego cDNA - końcowy produkt powinien być wyraźnym prążkiem na wysokości ok. 250 - 300 nt.

Zastosowana technologia Illumina jest to jedna z metod sekwencjonowania nowej generacji NGS zapewniająca wysokowydajne odczytywanie sekwencji DNA. Technika Illumina (dawniej Solexa) została skomercjalizowana w 2006 roku i opiera się na sekwencjonowaniu przez syntezę z zastosowaniem nukleotydów znakowanych fluorescencyjnie jako terminatorów reakcji polimeryzacji DNA oraz specjalnej polimerazy DNA, która jest zdolna do wbudowywania takich nukleotydów w rosnący łańcuch DNA. Technika sekwencjonowania Illumina składa się z kilku etapów. Po otrzymaniu biblioteki cDNA przeprowadzana jest denaturacja dwuniciowych fragmentów DNA, pozwalająca na ich immobilizację na płycie sekwencyjnej, która jest podłożem stałym gęsto pokrytym sekwencjami adaptorowymi i

sekwencjami do nich komplementarnymi. Jednoniciowe fragmenty przyłączają się jednym końcem do podłoża i tworzą strukturę przypominającą mostek – drugi koniec hybryduje do sekwencji komplementarnej do adaptora. Następnie dodawane są reagenty potrzebne do reakcji amplifikacji DNA – po kilku cyklach reakcji PCR (starterami są sekwencje adaptorowe) powstają skupiska ok. 1000 kopii jednoniciowych fragmentów DNA, tzw. klastrów (ang. *clusters*). W kolejnym etapie dodawane są startery sekwencyjne, polimeraza DNA oraz nukleotydy znakowane różnymi barwnikami fluorescencyjnymi, które mają zablokowany koniec 3'. Po inkorporacji znakowanego nukleotydu do łańcucha DNA, zostaje on wykryty i zidentyfikowany przy pomocy kamery z matrycą CCD (ang. *charge coupled device*). Następnie koniec 3' nukleotydu zostaje odblokowany, a barwnik fluorescencyjny odłączony i synteza łańcucha może przebiegać dalej. W ten sposób w czasie trwania pojedynczej reakcji możliwe jest odczytanie ok. 40 mln sekwencji DNA o długości 35 nt. Technika Illumina jest jedną z nowszych metod sekwencjonowania i umożliwia szybkie poznanie całych genomów lub transkryptomów, co przy zastosowaniu tradycyjnej metody Sangera trwało wiele lat. Schematy przebiegu reakcji sekwencjonowania w oparciu o technologię Illumina przedstawiono na Ryc. 9.



Ryc. 9

Schemat sekwencjonowania z użyciem technologii Illumina

a) wyizolowane DNA (lub cDNA) zostaje poddane fragmentacji, a następnie do krótkich odcinków dołączane są sekwencje adaptorowe o charakterystycznej rozwidłonej budowie; tak powstałe fragmenty są amplifikowane przy użyciu starterów komplementarnych do sekwencji adaptorowych i dają w rezultacie pulę dwuniciowych odcinków DNA o tępych końcach, które mają doligowane zarówno na końcu 3' jak i 5' sekwencje adaptorowe; b) tak przygotowane dwuniciowe odcinki poddaje się denaturacji i umieszcza na płycie sekwencyjnej pokrytej sekwencjami adaptorowymi i komplementarnymi do nich, co umożliwia hybryzację fragmentów do powierzchni płytki; nowa nić (zaznaczona kropkowaną linią) jest syntetyzowana w reakcji wydłużania startera, którym jest sekwencja adaptorowa; oryginalna nić jest usuwana, a sekwencja adaptorowa na 3' końcu nowo zsyntetyzowanej nici hybryduje do kolejnej sekwencji komplementarnej na płycie tworząc tzw. strukturę mostka i umożliwiając tym samym syntezę następnej nici (wykropkowana linia); wiele następujących po sobie takich cykli hybryzacji i wydłużania prowadzi do powstania skupisk klonów pojedynczej sekwencji o średnicy ok. 1 μm, czyli tzw. klastrów; c) aby otrzymać sekwencję odczytu, DNA we wszystkich klastrach jest linearyzowane poprzez cięcie w obrębie sekwencji adaptorora (oznaczone gwiazdką) - umożliwia to wygenerowanie jednoniciowej matrycy do sekwencjonowania przez syntezę (produkt sekwencjonowania zaznaczono kropkowaną linią); aby otrzymać odczyty z obu końców odcinka (typu paired-end) produkty sekwencjonowania denaturuje się i wymywa z płytki; matryca używana jest do syntezy nici komplementarnej (zaznaczona wykropkowaną linią), a następnie jest przecinana (miejsce cięcia zaznaczono gwiazdką), odmywana i zachodzi sekwencjonowanie na matrycy nowo zsyntetyzowanej nici [72].

4.2.6 Analiza bioinformatyczna

W pierwszym etapie badań oszacowano skalę zjawiska AS na podstawie sekwencji największego chromosomu kukurydzy zwyczajnej - chromosomu pierwszego. Do tego celu zastosowano program AUGUSTUS (dostępny przez serwer [www http://augustus.gobics.de](http://augustus.gobics.de)). Jest to jeden z pierwszych programów do przewidywania struktury genów *ab initio* z podanej sekwencji, który uwzględnia możliwość zajścia alternatywnego splicingu. Do przewidywania struktury genów wykorzystuje on ukryte modele Markova (HMM, ang. *hidden Markov model*) i był testowany również na sekwencjach pochodzących z kukurydzy zwyczajnej. Program ten wyszukuje sekwencje, które spełniają następujące wymogi:

- sekwencja eksonu nie zawiera kodonu stop,
- miejsca splicingowe zawierają sekwencje konsensusowe,
- wszystkie zaproponowane geny zaczynają się od kodonu ATG, a kończą kodonem stop,
- każdy gen kończy się przed miejscem rozpoczęcia nowego (geny nie zachodzą na siebie),
- długość intronów i eksonów jest równa lub przewyższa wartości minimalne określone dla poszczególnych gatunków.

Z zaproponowanych struktur genów AUGUSTUS wybiera te, które najlepiej pasują do modelu i spełniają jak najwięcej zadanych kryteriów.

Korzystając z serwera udostępnionego dla programu AUGUSTUS można przeanalizować sekwencję w formacie FASTA o długości nieprzekraczającej 3 mln nt. Sekwencja chromosomu 1 kukurydzy (GK000031.1) posiada ponad 300 mln nt, dlatego też podzielono ją na 100 odcinków o długości 3 mln nt i jeden o długości 239 tys. nt. Aby nie ominąć genów, które mogły się znajdować na granicy fragmentów analizę uzupełniono o 100 sekwencji, które miały po 40 tys. nt i zachodziły na graniczące ze sobą fragmenty (po 20 tys. nt z każdej strony). Dane wyjściowe generowane przez program AUGUSTUS oprócz sekwencji nukleotydowej przewidzianych genów zawierają także jej tłumaczenie na sekwencję aminokwasową i są udostępniane w formacie GFF (ang. *general feature format*). Każdy ekson

zapisany jest w osobnej linii wraz z danymi takimi jak jego koordynaty, nazwa sekwencji i genu oraz to czy został przewidziany na nici kodującej czy komplementarnej.

W kolejnym etapie porównywano otrzymane wyniki z kolekcją sekwencyjnych znaczników ekspresji (EST, ang. *Expressed Sequence Tags*) stosując program Geneious Basic [73] i algorytm megablast, a sekwencje aminokwasowe zestawiono z danymi zdeponowanymi w bazie UniProtKB (zastosowano algorytmu blastp). Program Geneious umożliwia między innymi jednoczesne przeszukiwanie baz danych dla dużego zestawu sekwencji i wyświetla wyniki w przyjazny dla użytkownika sposób ułatwiając tym samym ich interpretację. Zawężoną pulę wyników porównano także z sekwencjami zdeponowanymi w bazie danych Gramene oraz z wynikami programu F-genesh, które znajdują się także w tej bazie danych (szczegóły analizy w rozdziale 5.1) oraz z cząsteczkami referencyjnych mRNA z bazy danych Nucleotide (refseq_rna).

Po otrzymaniu danych z sekwencjonowania transkryptomu pierwszym krokiem analizy bioinformatycznej było przeprowadzenie kontroli jakości otrzymanych wyników z zastosowaniem programu FastQC. Analizę bioinformatyczną prowadzono we współpracy z dr A. Świercz i mgr A. Hoffa z Politechniki Poznańskiej. Program FastQC pozwala na ocenę jakości surowych danych pochodzących z wysokoprzepustowych metod sekwencjonowania drugiej generacji. Zapewnia zestaw modułowych analiz i wizualizacji ich wyników, które pozwalają szybko ocenić jakość uzyskanych sekwencji, co jest kluczowe przed podjęciem dalszych etapów analizy. Podstawowe funkcje tego programu umożliwiają:

- import danych z plików BAM, SAM lub plików FastQ (dowolna wersja),
- sporządzenie wykresów i tabel zbiorczych,
- eksport wyników do raportu zapisanego w formacie HTML,
- analizę danych i generowanie raportów offline, bez konieczności uruchamiania interaktywnej aplikacji.

Aby określić różnice w ekspresji poszczególnych genów i ich izoform pomiędzy badanymi liniami potrzebne były odpowiednie narzędzia bioinformatyczne, które opierają się o wydajne i uzasadnione statystycznie algorytmy. Zdecydowano się na zastosowanie ścieżki analizy zwanej Tuxedo (z ang. *smoking*) [74], której schemat przedstawia Ryc. 9.

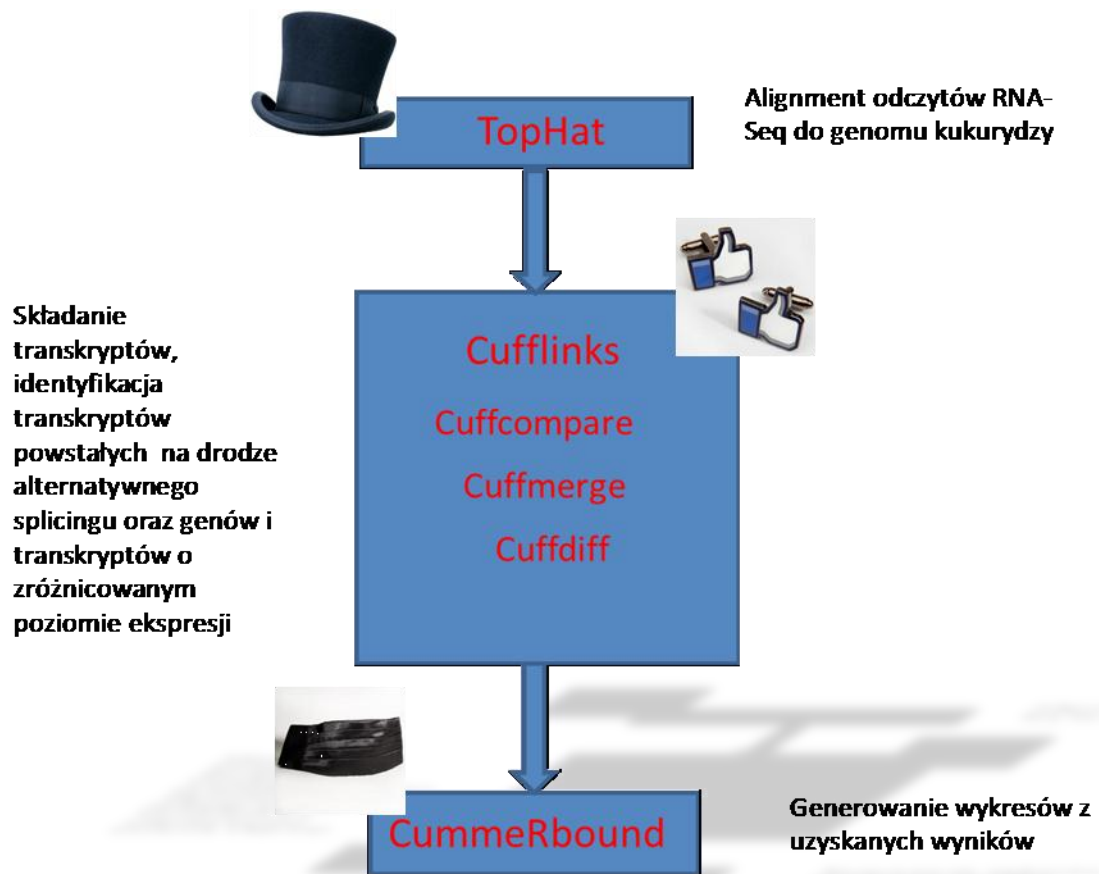
W pierwszym etapie odczyty z każdego warunku zmapowano do genomu referencyjnego przy użyciu programu TopHat (z ang. *cylinder*). Jest to krok niezbędny i podstawowy wielu analiz takich jak np. CHIP-seq (ang. *chromatin immunoprecipitation-sequencing*) czy RNA-seq. Dopasowanie sekwencji jest często pojawiającym się problemem w bioinformatyce, dlatego (szczególnie w ostatnich latach) pojawiło się wiele różnych programów próbujących zmierzyć się z tym zagadnieniem. Jednym z najbardziej popularnych jest program Bowtie (z ang. *muszka*), który wykorzystuje niezwykle wydajną strukturę danych do przechowywania genomu referencyjnego, dzięki czemu zapewnia bardzo szybkie jego przeszukiwanie (umożliwia dopasowanie dziesiątek milionów odczytów na godzinę pracy procesora). Jednakże Bowtie nie zezwala na obecność dużych przerw pomiędzy dopasowywanym odczytem a genomem referencyjnym, stąd też może nie radzić sobie z odczytami, które obejmują introny. TopHat używa Bowtie jako „silnika” dopasowującego, lecz rozbija odczyty, których Bowtie nie mógł dopasować na mniejsze fragmenty zwane segmentami. Takie krótsze fragmenty traktowane pojedynczo często zostają zmapowane do genomu referencyjnego. Jeżeli kilka segmentów dopasuje się do genomu w oddaleniu od siebie (pomiędzy 100 a kilkaset tysięcy par zasad) TopHat interpretuje taki odczyt jako obejmujący miejsce splicingowe i oszacowuje gdzie to miejsce splicingowe występuje. Ta zdolność jest kluczowa dla zlokalizowania nowych transkryptów. Jak pokazały liczne eksperymenty głębokiego sekwencjonowania transkryptomów opisany do tej pory katalog zdarzeń AS pozostaje nadal niekompletny i przy każdym dodatkowym sekwencjonowaniu (nawet dla organizmów modelowych) identyfikowane są ciągle nowe izoformy transkryptów.

W kolejnym kroku użyto programu Cufflinks (z ang. *spinki do mankietów*), który składa indywidualne odczyty, które zostały dopasowane do genomu referencyjnego w pojedyncze transkrypty. Ponieważ próbka może, albo prawie na pewno będzie zawierać odczyty z różnych wariantów splicingowych dla poszczególnych genów, Cufflinks musi być w stanie wywnioskować i zasugerować wzór splicingu. Nie wiadomo ile różnych form transkryptów pojawi się dla danego genu, dlatego Cufflinks zgłasza wszystkie złożenia transkryptów, które najlepiej i najpełniej tłumaczą otrzymane dane sekwencyjne. Używając rygorystycznego modelu statystycznego Cufflinks ocenia także poziom ekspresji dla wszystkich zaproponowanych transkryptów.

Przed rozpoczęciem analizy różnicowej pomiędzy próbkami RNA-seq konieczne jest zebranie i złożenie sekwencji w kompleksowy zestaw transkryptów. Do tego celu użyliśmy programu Cuffmerge, który łączy razem złożone transkrypty. Jeżeli jest dostępny może także włączyć do analizy transkrypty referencyjne. Taka pula transkryptów może być następnie przeszukiwana w poszukiwaniu genów i transkryptów, których ekspresja jest zróżnicowana pomiędzy próbkami. Program Cuffdiff oblicza poziom ekspresji i testuje czy otrzymana różnica ma znaczenie statystyczne. Tabularyczne pliki wyjściowe z tego programu mogą być przeglądane w rozmaitych aplikacjach, m. in. Microsoft excel. Pliki te zawierają informacje statystyczne dla poszczególnych genów i izoform transkryptów takie jak wartość P czy zlogarytmowaną wielokrotność zmiany poziomu ekspresji.

W ostatnim etapie, który pozwala na lepszą wizualizację otrzymanych wyników, zastosowano program Cummerbund (z ang. *szeroki pas, szarfa do stroju wieczorowego*), który z plików wyjściowych programu Cuffdiff generuje wykresy korzystając ze środowiska R. R jest to zarówno język, jak środowisko programistyczne dedykowane do obliczeń statystycznych i generowania grafiki. Działa na wielu platformach UNIX, Windows czy MacOS. Kod źródłowy tej platformy jest dostępny jako Wolne Oprogramowanie (ang. *Free Software*) na warunkach Fundacji Wolnego Oprogramowania Powszechnej Licencji Publicznej GNU (ang. *Free Software Foundation's GNU General Public License*).

Aby lepiej zobrazować funkcje genów wybranych w czasie analizy skorzystano z programu MapMan. Jest to program, który umożliwia wizualizację zmian w profilu ekspresji dotycząca zarówno białek, genów jak i transkryptów. Składa się z kilku modułów, które pozwalają przypisać dane ontologiczne do informacji podanych w pliku wejściowym, zobrazowanie zaobserwowanych zmian na szlakach metabolicznych i ocenę statystyczna istotności badanych różnic. Początkowo program ten został stworzony do badania danych pochodzących z *A. thaliana*, jednak obecnie może być używany dla wielu różnych gatunków, m. in. kukurydzy [75].



Ryc. 10

Schemat przebiegu analizy bioinformatycznej wyników sekwencjonowania transkryptomów badanych linii kukurydzy zwyczajnej.

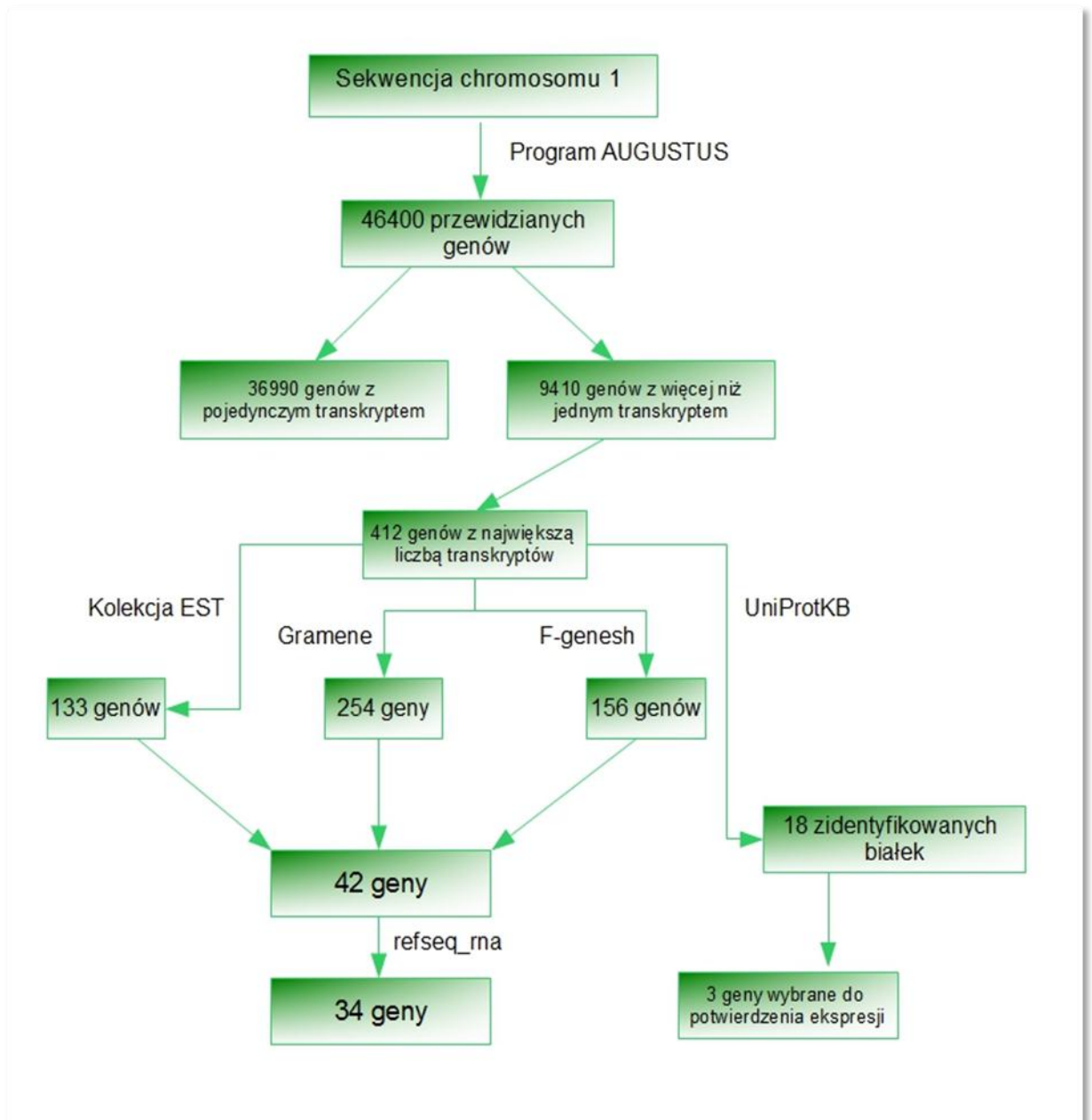
5. Wyniki i dyskusja

5.1 Analiza bioinformatyczna sekwencji chromosomu 1 kukurydzy zwyczajnej pod kątem zdarzeń AS

Aby oszacować skalę AS u kukurydzy zwyczajnej oraz ocenić zdolności do przewidywania tego zjawiska tylko na podstawie sekwencji nukleotydowej, przeprowadzono analizę sekwencji chromosomu pierwszego kukurydzy z wykorzystaniem programu AUGUSTUS (rozdział 4.2.6). Sekwencja chromosomu pochodziła z projektu sekwencjonowania linii wsobnej kukurydzy zwyczajnej B73 i została zdeponowana w bazie NCBI pod numerem dostępu GK000031.1 (obecnie jest dostępna także zaktualizowana wersja tego rekordu GK000031.2). Z czterech dostępnych w programie AUGUSTUS opcji przewidywania genów wybrano wariant, który raportował kilka transkryptów dla pojedynczego genu. Schemat przeprowadzonej analizy przedstawiono na Ryc. 11.

Opierając się jedynie o informacje zawarte w sekwencji genomowej, po przeanalizowaniu wszystkich fragmentów o długości 3 mln nt oraz 40 tys. nt, program AUGUSTUS znalazł 46 tys. genów. Jest to wynik znacząco zawyżony i sugeruje wiele fałszywie dodatnich przewidywań - ilość genów w całym genomie kukurydzy szacowana jest jedynie na 32 tys. Dla około 1/5 z przewidzianych genów (9410) program zaraportował zdarzenia alternatywnego składania transkryptów. Z tej puli do dalszych analiz wybrano 412 genów, dla których AUGUSTUS przewidział największą liczbę transkryptów, a więc najwięcej zdarzeń AS. Założono, że istnieje najwyższe prawdopodobieństwo, że w tej grupie transkryptów znajdują się takie, których ekspresja została potwierdzona. Dlatego te sekwencje porównywano z kolekcją EST kukurydzy zdeponowaną w bazie danych NCBI oraz z wynikami przewidywań z dwóch innych źródeł - bazy danych Gramene i programu F-genes, którego wyniki również zdeponowano w tej bazie danych. Program F-genes (ang. *Find GENES using HMM*), podobnie jak AUGUSTUS do przewidywania struktur genów wykorzystuje ukryte modele Markowa, jednak nie uwzględnia możliwości zajścia procesu alternatywnego składania transkryptów. Wcześniej był także używany do przewidywania genów kukurydzy i w porównaniu z czterema innymi programami (GeneMark.hmm, GENESCAN, GlimmerR, Grail) generował najbardziej dokładne i trafne wyniki [76].

Do porównań sekwencji zastosowano program Geneious Basic i wybrano algorytm megablast, który uwzględnia tylko dopasowania sekwencji o wysokim podobieństwie. Jako dopasowanie znaczące uznawano takie, dla którego wartość oczekiwana (E-value, ang. *Expect value*) była mniejsza lub równa $1e-100$, a procent identyczności (ang. *percentage of identities*) wynosił 85% lub więcej.



Ryc. 11

Schemat analizy bioinformatycznej sekwencji chromosomu 1 kukurydzy zwyczajnej

Wartość oczekiwana jest parametrem, który określa ile rekordów może zostać dopasowanych przez przypadek do zadanej sekwencji i maleje wykładniczo wraz ze wzrostem wartości wyniku dopasowania (ang. *match score*). Zasadniczo wartość E opisuje szum tła, np. wartość E równa 1 oznacza, że w bazie danych o określonej wielkości dla zadanej sekwencji możemy oczekiwać, że jedno dopasowanie pojawi się po prostu przez przypadek. Im niższa wartość E (bliższa zeru) tym dopasowanie ma większy poziom istotności. Należy jednak pamiętać, że wartość oczekiwana jest szacowana także na podstawie długości zadanej sekwencji - im jest ona krótsza tym wyższa wartość E, ponieważ jest wyższe prawdopodobieństwo wystąpienia przez przypadek identycznej, krótkiej sekwencji. Aby dodatkowo zawęzić pulę wyników do tych o największym znaczeniu biologicznym przyjęto wysoki procent identyczności, który opisuje w ilu procentach dwie sekwencje nukleotydów (lub aminokwasów), mają takie same nukleotydy w tych samych pozycjach dopasowania.

Po dopasowaniu około 1200 sekwencji transkryptów z wybranej puli 421 genów do zbioru sekwencyjnych znaczników ekspresji kukurydzy zwyczajnej, udało się potwierdzić prawdopodobną ekspresję dla 272 genów, w tym:

- dla 113 genów znaleziono dopasowania dla wszystkich przewidzianych przez program AUGUSTUS transkryptów,
- dla 129 genów dopasowano EST dla dwóch z kilku przewidzianych przez AUGUSTUS transkryptów,
- dla 30 genów tylko dla jednego transkryptu z kilku zaproponowanych przez program AUGUSTUS dopasowano sekwencyjny znacznik ekspresji (dla tych genów najprawdopodobniej nie zachodzi proces AS).

Porównanie z wynikami zdeponowanymi w bazie danych Gramene umożliwiło dopasowanie 254 genów przewidzianych przez program AUGUSTUS, jednak tylko dla 23 dopasowane geny, zdeponowane w tej bazie danych, posiadały adnotacje o więcej niż jednym transkrypcie, który był wynikiem zajścia zdarzeń alternatywnego składania. Dopasowanie do wyników działania programu F-genesz pozwoliło potwierdzić przewidzianą sekwencję 156 genów. Na wszystkich etapach, czyli zarówno w kolekcji EST, wynikach

programu F-genesh i w bazie danych Gramene, udało się potwierdzić alternatywne składanie transkryptów dla 42 genów. Dla tak zawężonej puli genów przeprowadzono także dopasowanie do sekwencji zawartych w bazie danych referencyjnych sekwencji mRNA kukurydzy zwyczajnej. Udało się zidentyfikować cząsteczki mRNA dla 34 genów.

W znajdującej się poniżej tabeli (Tabela 11) zestawiono numery dostępów GenBank do rekordów zawierających zidentyfikowane sekwencyjne znaczniki ekspresji oraz sekwencje referencyjnych cząsteczek mRNA, a także loci, w którym są one zmapowane dla puli 42 genów, które uznano za najbardziej prawdopodobne cele procesu AS. W Aneksie 1 załączone zostały wyniki programu AUGUSTUS w formacie GFF dla tej puli genów.

5. Wyniki i dyskusja

Nr genu	Nr dostępu GenBank dla kolekcji EST	Nr dotępu GenBank dla referencyjnych mRNA	Locus
1.1	ES700631	NM_001154387	LOC100281469(umc1292)
1.2	DR789872	NM_001151597	LOC100278265 (umc2523)
1.3	EE173725	NM_001154387	LOC100281469 (umc1292)
2.1	FL133048	NM_001149410	LOC100275291
2.2	DY687486		
2.3	EE295362		
3.1	FK992155	NM_001138663	LOC100193552
3.2	EE013166		
3.3	CF675154		
4.1	AI586830	NM_001111911	Brak danych
4.2	FL431278	NM_001148450	LOC100274071
4.3	DV516499		
5.1	DN219456		
5.2	CO459182	NM_001147392	LOC100272940
5.3	DV492877	NM_001150745	LOC100277079
6.1	CD44134		
6.2	DY397626		
6.3	CD440514	NM_001154468	LOC100281549
7.1	CF032355		
7.2	DT641341		
7.3	BM498806	NM_001146992	LOC100272523
8.1	FL458621	NM_001152336	LOC100279315 (umc2185)
8.2	FL182719		
8.3	DY686105		
9.1	CD433356	NM_001175269	LOC100382534
9.2	FL452101		
9.3	CD433356		
10.1	FL452100	NM_001196985	LOC100502507
10.2	AI065585		
10.3	FL470663		
11.1	BE051540	Nie dopasowano sekwencji	
11.2	AI737422	Nie dopasowano sekwencji	
11.3	DV492877	Nie dopasowano sekwencji	
12.1	DV503653	NM_001196821	LOC100502343
12.2	DN230491	NM_001175269	LOC100382534
12.3	DV503652		
13.1	CO451512		
13.2	DV171091	NM_001175269	LOC100382534
13.3	CO451512		
14.1	FL041027	Nie dopasowano sekwencji	
14.2	FL460501	Nie dopasowano sekwencji	
14.3	FL103940	Nie dopasowano sekwencji	
15.1	CD437703	NM_001175269	LOC100382534

5. Wyniki i dyskusja

15.2	CF384417		
15.3	AI770397		
16.1	FL449334	NM_001151416	LOC100277992
16.2	CF92014		
16.3	CD526614		
17.1	EE036851	NM_001157149	LOC100284254
17.2	EE159977		
17.3	CO525111		
18.1	EB705358	NM_001137956	LOC100192758
18.2	CO456711		
18.3	EB707230		
19.1	CF021246	NM_001154263	LOC100281345
19.2	DR803317		
19.3	EB822172		
20.1	EE153174	NM_001147885	LOC100273448
20.2	EC878750		
20.3	EE187811		
21.1	FL071738	Nie dopasowano sekwencji	
21.2	FL071761		
21.3	DN219554		
22.1	FL470298	NM_001196821	LOC100502343
22.2	BG265396		
22.3	DV492877		
23.1	EB819056	NM_001112440	LOC732830
23.2	FL475343		
23.3	EE178254		
24.1	DT939585	NM_001156438	LOC100283538
24.2	DV518063		
24.3	DR964496	NM_001196029	LOC100501266
25.1	CA405225	Nie dopasowano sekwencji	
25.2	CK701087		
25.3	AI861641		
26.1	FL183072	NM_001148535	LOC100274157
26.2	FL183076		
26.3	DY688374		
27.1	DR962736	NM_001177286	LOC100384850
27.2	FL155278		
27.3	FL119451		
28.1	FL071743	NM_001147983	LOC100273565
28.2	EE045605		
28.3	FL071759		
29.1	DV522358	NM_001143521	LOC100217156
29.2	DT939972		
29.3	EE042896		
30.1	FL043545	Nie dopasowano sekwencji	
30.2	CK701087		
30.3	FL470298		

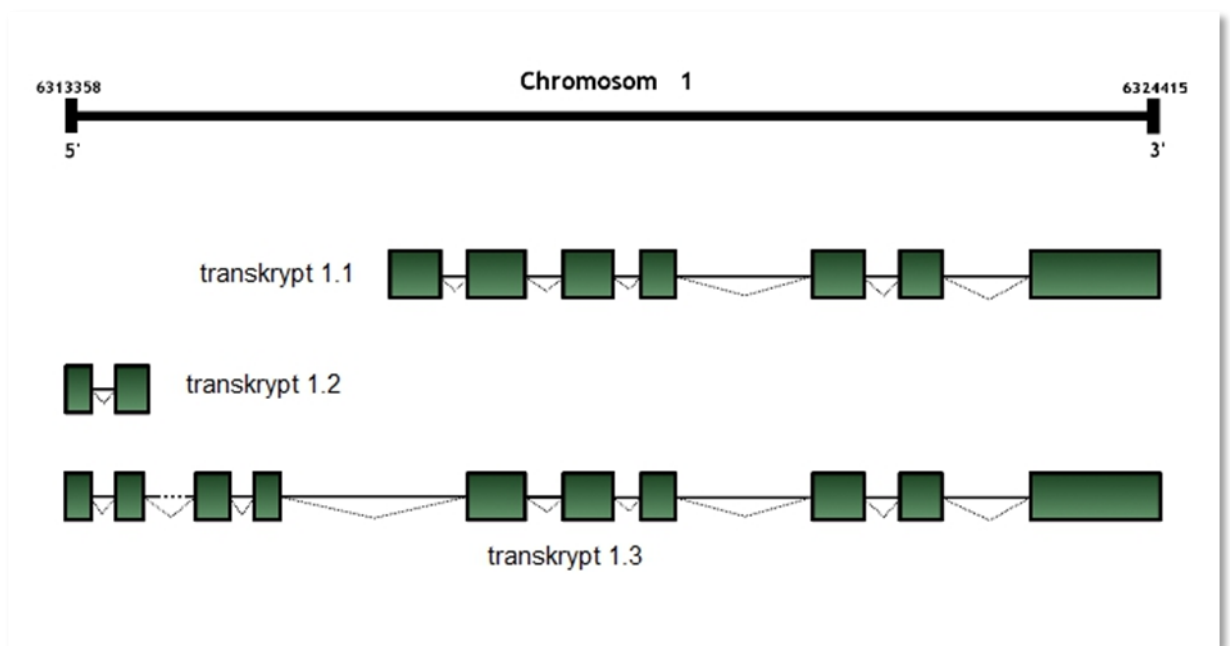
31.1	EE038035	NM_001137175	LOC100191748
31.2	EC886908		
31.3	EB816122		
32.1	FL451896	Nie dopasowano sekwencji	
32.2	FL163211		
32.3	FL192768		
33.1	EE158277	NM_001196539	LOC100501900
33.2	EE188625		
33.3	EE158276		
34.1	FL346542	NM_001143225	LOC100216831
34.2	FL347070	NM_001156490	LOC100283589
34.3	DN228312		
35.1	AW017875	Nie dopasowano sekwencji	
35.2	CF631655		
35.3	EE680983		
36.1	EB405807	NM_001176131	LOC100383483
36.2	BM032586		
36.3	BM417288		
37.1	CK368211	NM_001157319	LOC100284424
37.2	BG841552		
37.3	DV020738	NM_001150808	LOC100277159
38.1	DV026704	NM_001139290	LOC100194249
38.2	DR813331		
38.3	DR803682		
39.1	CF675160	Nie dopasowano sekwencji	
39.2	CA405225		
39.3	BG265653		
40.1	DY234241	NM_001149596	LOC100275540
40.2	FL418346		
40.3	DV025759		
41.1	FL062547	NM_001158454	LOC100285563
41.2	FL062554		
41.3	FL069420	NM_001149278	LOC100275095
42.1	DN230491	NM_001196821	LOC100502343
42.2	GH201244		
42.3	BG265649		

Tabela 11

Zestawienie wyników dopasowania sekwencji 42 genów przewidzianych przez program AUGUSTUS, których ekspresja została potwierdzona przez porównanie z kolekcją EST i referencyjnymi cząsteczkami mRNA

Aby lepiej zobrazować uzyskane wyniki na Ryc. 12 zaprezentowano schemat struktury genu pierwszego (1) z przewidzianych przez program AUGUSTUS. Dla tego genu zostały

przewidziane 3 transkrypty, z których najdłuższy (1.3) składa się z 10 eksonów i bardzo długiej sekwencji intronu (6782 nt) i znajduje się między eksonem 8 i 9 (eksony liczone od prawej do lewej). Najkrótszy jest transkrypt 1.2, który składa się tylko z dwóch eksonów i powstał w wyniku użycia alternatywnego miejsca splicingowego 3' w obrębie eksonu 9. Natomiast transkrypt 1.1 posiada dodatkowy ekson, który znajduje się w obrębie intronu 6 transkryptu 1.3.



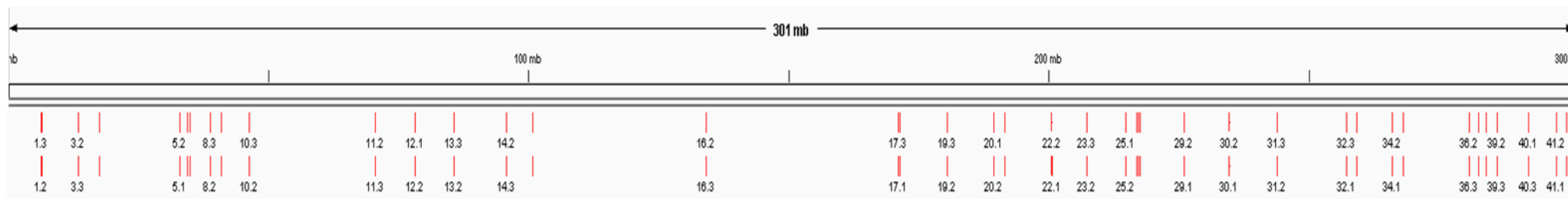
Ryc. 12

Schemat organizacji genu nr 1 przewidywanego przez program AUGUSTUS, dla którego udało się dopasować sekwencje EST, rekord z bazy danych Gramene i sekwencję referencyjnego mRNA.

Wykorzystując program IGV [77] dla wszystkich 42 genów wygenerowano ryciny ilustrujące wzajemne położenie przewidzianych transkryptów, sekwencji EST i mRNA w odniesieniu do chromosomu 1 (Aneks 2), natomiast na Ryc. 13 przedstawiono ogólną dystrybucję i lokalizację transkryptów z tej zawężonej puli. Jak wynika z przedstawionych w Aneksie 2 rycin, dla 16 genów (o numerach: 1, 2, 3, 4,5, 8, 18, 20, 23, 24, 27, 29, 31, 34, 38 i 41) uzyskano wysoki stopień korelacji przewidzianych przez AUGUSTUS transkryptów z dopasowanymi sekwencjami EST i mRNA. Dla pozostałych genów wzajemna relacja koordynatów poszczególnych sekwencji jest raczej niska. Może mieć na to wpływ niskie

wysycenie kolekcji mRNA i EST kukurydzy, szczególnie w porównaniu z organizmami modelowymi takimi jak *A. thaliana*.

5. Wyniki i dyskusja



Ryc. 13

Wzajemna dystrybucja i lokalizacja zbioru 42 genów na chromosomie 1 kukurydzy zwyczajnej przewidzianych przez program AUGUSTUS, które zostały uznane za najbardziej prawdopodobne cele AS.

Równocześnie z analizą sekwencji nukleotydowej przeprowadzono analizę na poziomie białek, które mogą powstawać z przewidzianych genów. Sekwencję aminokwasową przetłumaczoną z sekwencji nukleotydowej zaproponowanych przez AUGUSTUS genów, stosowano jako zapytanie do bazy danych UniProtKB przy użyciu algorytmu blastp. Podobnie jak w przypadku sekwencji nukleotydowej, wybrano tylko te geny, które posiadały największą ilość przewidzianych transkryptów (412 genów), czyli najprawdopodobniej były celem alternatywnego składania (Ryc. 11). Za wynik znaczący uznawano takie dopasowanie, którego wartość oczekiwana E wynosiła $1e-100$ (podobnie jak w przypadku sekwencji nukleotydowej) i w którym sekwencje były przynajmniej w 40% identyczne.

W ten sposób udało się zidentyfikować 18 białek zdeponowanych w bazie danych UniProtKB. Większość z tych białek została opisana dla *A. thaliana*, a tylko 3 białka pochodziły z kukurydzy zwyczajnej. Szczegółowe wyniki zebrano w Tabeli 12 zamieszczonej poniżej.

Numer genu	Wartość E	% identyczności	Zidentyfikowane białko	Nr dostępu do bazy UniProtKB
1	2,92e-107	44,70	DNA (cytozyno-5) metylotransferaza 1 (ang. <i>DNA (cytosine-5)-methyltransferase DRM1</i>) z <i>Arabidopsis thaliana</i>	Q9LXE5
2	7,11e-114	45,70	DNA (cytozyno-5)metylotransferaza 2 (ang. <i>DNA (cytosine-5)-methyltransferase DRM2</i>) z <i>Arabidopsis thaliana</i>	Q9M548
3	8,92e-169	65,80	Białko podobne do białka FIZZY 2 (ang. <i>protein FIZZY-RELATED 2</i>) z <i>Arabidopsis thaliana</i>	Q8L3Z8
4	5,23e-175	62,10	Prawdopodobna kinaza serynowo/treoninowa (ang. <i>Probable serine/threonine-protein kinase</i>) z <i>Arabidopsis thaliana</i>	Q3EDL4
5	0	68,50	Dehydrogenaza fitoenu (ang. <i>Phytoene dehydrogenase</i>) z <i>Solanum lycopersicum</i>	P28554
6	1,66e-131	63,40	Niezwiązany z ATP regulator podjednostki 4 26S proteasome (ang. <i>26S proteasome non-ATPase regulatory subunit 4</i>) z <i>Arabidopsis thaliana</i>	P55034
7	2,01e-156	79,90	Syntetaza asparaginy (ang. <i>Asparagine synthetase</i>) z <i>Sandersonia aurantiaca</i>	O24338
8	1,27e-134	96,90	Białko P związane z czynnikiem transkrypcyjnym Myb (ang. <i>Myb-related protein P</i>) z <i>Zea mays</i>	P27898
9	1,73e-151	81,70	S-metylotransferaza 2 homocysteiny (ang. <i>Homocysteine S-methyltransferase 2</i>) z <i>Zea mays</i>	Q9FUM9
10	0	78,10	Dkcarboksylaza 2 glutaminianu (ang. <i>Glutamate decarboxylase 2</i>) z <i>Arabidopsis thaliana</i>	Q42472
11	5,48e-106	69,50	Prawdopodobne białko o masie 125 kDa związane z kinezyną (ang. <i>Probable 125 kDa kinesin-related protein</i>) z <i>Arabidopsis thaliana</i>	P82266
13	1,33e-108	76,20	Reduktaza 3 12-oksofitodienowa (ang. <i>12-oxophytodienoate reductase 3</i>) z <i>Arabidopsis thaliana</i>	Q9FUP0
13	0	83,30	Alfa-galaktozydaza (ang. <i>Alpha-galactosidase</i>) z <i>Oryza sativa subsp. japonica</i>	Q9FXT4
14	0	55,00	Nasienna lipooksygenaza 3 (ang. <i>Seed lipoxygenase-3</i>) z <i>Glycine max</i>	P09186
15	0	77,80	Lipooksygenaza 2 (ang. <i>Lipoxygenase 2</i>) z <i>Oryza sativa subsp. japonica</i>	P29250
16	2,39e-148	46,70	Białko 6 o masie 65-kDa związane z mikrotubulami (ang. <i>65-kDa microtubule-</i>	Q9SIS3

			<i>associated protein 6</i>) z <i>Arabidopsis thaliana</i>	
17	5,44e-168	76,90	Reduktaza ferredoksyna-NADP (ang. <i>Ferredoxin-NADP reductase</i>) z <i>Oryza sativa</i> subsp. <i>japonica</i>	P41345
18	7,01e-141	86,80	Dehydrogenaza glutationu (ang. <i>Glutamate dehydrogenase</i>) z <i>Zea mays</i>	Q43260

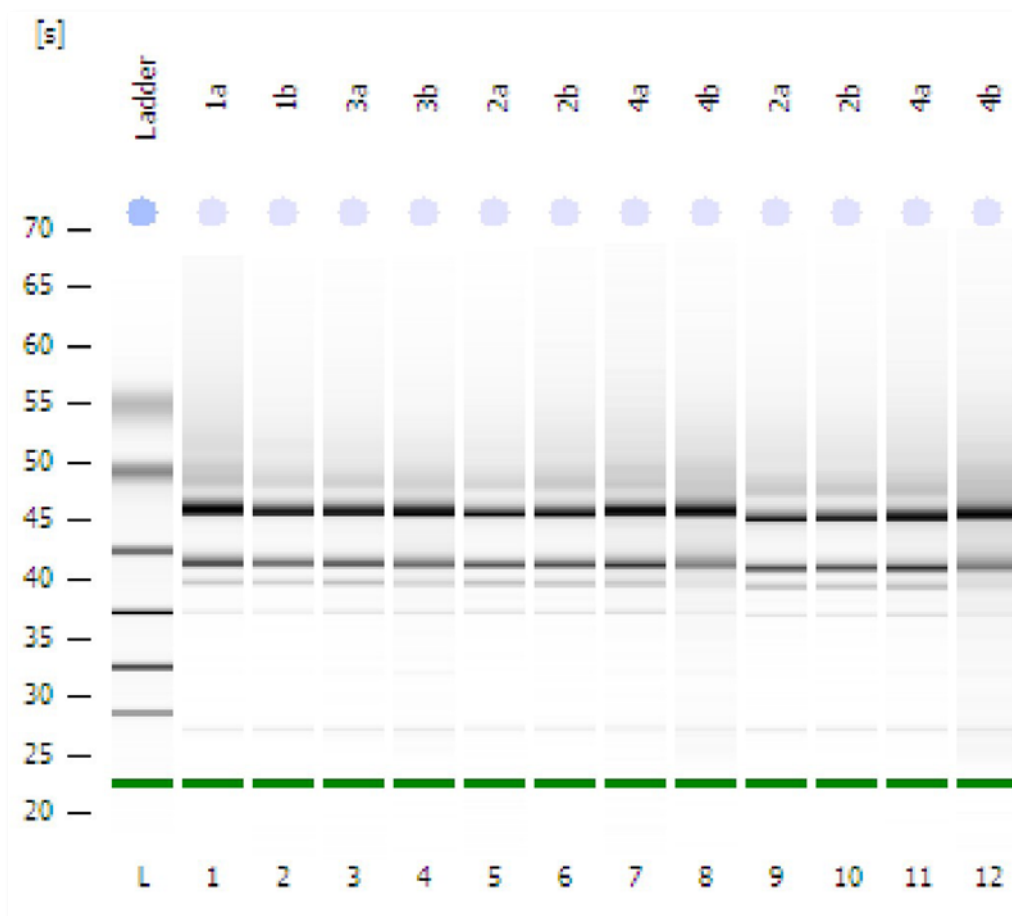
Tabela 12

Wyniki dopasowania sekwencji aminokwasowych przetłumaczonych z transkryptów przewidzianych przez program AUGUSTUS do danych zdeponowanych w bazie danych UniProtKB.

5.1.2 Analiza ekspresji transkryptów dla wybranych genów z zastosowaniem reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR

Z puli genów, dla których udało się zidentyfikować potencjalne produkty białkowe (18 genów przewidzianych przez program AUGUSTUS) wybrano trzy, aby potwierdzić ich ekspresję w dwóch liniach kukurydzy zwyczajnej (S245 i S79757). Wytypowano te geny, których produkty białkowe zostały zidentyfikowane u kukurydzy zwyczajnej: 1. gen kodujący białko P związane z czynnikiem transkrypcyjnym Myb, 2. S-metylotransferazę 2 homocysteiny, 3. dehydrogenazę glutationu (porównaj Tabela 12).

Z zebranych w dwóch punktach czasowych liści obu odmian kukurydzy zwyczajnej przeprowadzono izolację kwasów nukleinowych. Po pomiarze spektrofotometrycznym, wyizolowane RNA poddawano analizie jakości z wykorzystaniem elektroforezy kapilarnej. Oceniano jakość rozdziału - brak smug, wyraźne prążki oraz wartość współczynnika RIN. Do dalszych analiz wykorzystywano tylko te preparaty, których współczynnik integralności RNA wynosił 8 lub więcej w 10 stopniowej skali (10 - materiał niezdegradowany). Współczynnik ten jest wyznaczany na podstawie porównania obrazu rozdziału próbek oraz markera wielkości RNA, którego wielkość i stężenie jest ustalone. Poniżej zamieszczono przykładowy obraz rozdziału elektroforetycznego badanych próbek uzyskanych po rozdziale z wykorzystaniem aparatu Bioanalyzer 2100 firmy Agilent (Ryc. 14).



Ryc. 14

Obraz rozdziłu elektroforetycznego całkowitego RNA wyizolowanego z liści kukurydzy odmiany S245 (TL) i S79757 (SL) uzyskany przy pomocy aparatu Bioanalyzer 2100;

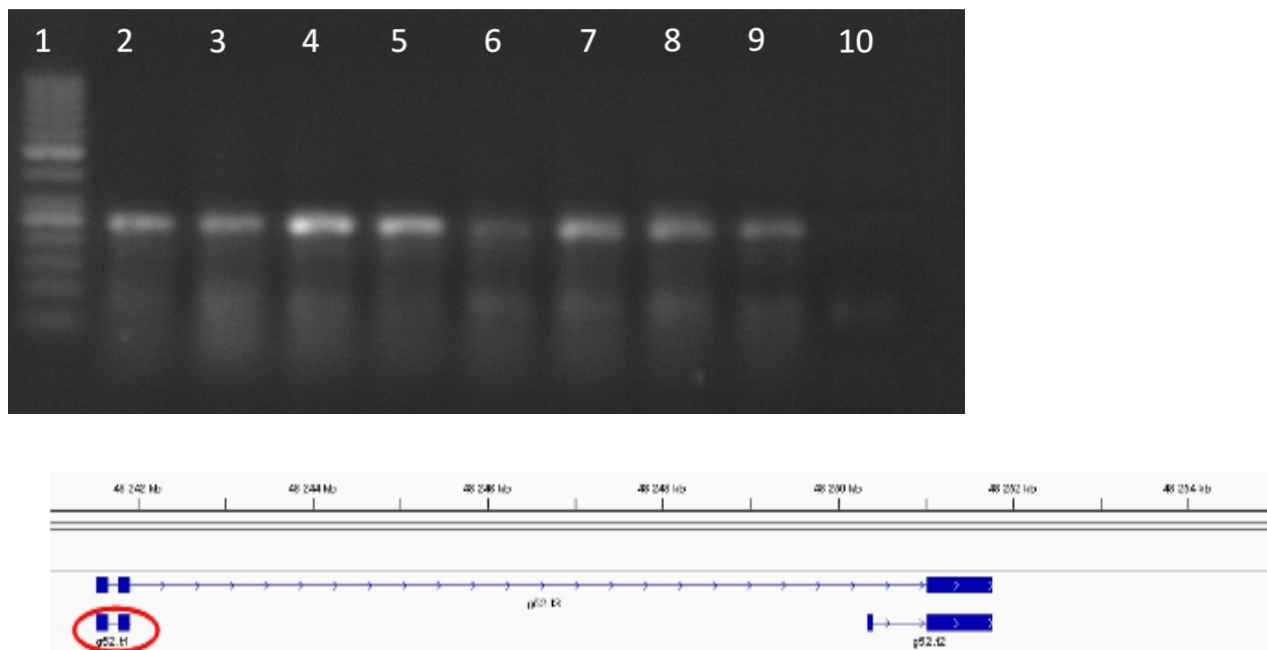
1a – linia wrażliwa kontrola po 24h (SL1C); 1b – linia wrażliwa traktowana po 24h (SL1); 3a – linia wrażliwa kontrola po 7 dniach (SL7C); 3b – linia wrażliwa traktowana po 7 dniach (SL7); 2a – linia tolerancyjna kontrola po 24 h (TL1C); 2b – linia tolerancyjna traktowana po 24 h (TL1); 4a – linia tolerancyjna kontrola po 7 dniach (TL7C); 4b – linia tolerancyjna traktowana po 7 dniach (TL7)

Próbki pochodzące z roślin poddawanych działaniu herbicydu, szczególnie te zebrane tydzień po oprysku, miały często niższą wartość RIN - RNA było mocniej zdegradowane. Izolację przeprowadzano z całych liści, także tych, na których obserwowano zmiany w postaci plamek nekrotycznych czy usychania. Było to przyczyną obniżenia ogólnej jakości uzyskiwanych preparatów.

Z uzyskanego materiału biologicznego przeprowadzano również izolację DNA. Wyizolowane DNA oceniano na podstawie pomiaru wartości absorbancji, a preparaty o odpowiednim stężeniu wykorzystywano jako matrycę do reakcji PCR w gradiencie

temperatur, która pozwoliła na optymalne dobranie temperatury hybrydyzacji starterów do reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR.

Po uzyskaniu preparatów RNA o odpowiedniej jakości przeprowadzono reakcję odwrotnej transkrypcji sprzężoną z reakcją PCR, aby sprawdzić czy przewidziane przez program AUGUSTUS transkrypty ulegają ekspresji w badanych warunkach. Produkty odwrotnej transkrypcji i amplifikacji oczyszczano i rozdzielano na żelu agarozowym. Na Ryc. 15 przedstawiono obraz rozdziału elektroforetycznego produktów reakcji RT-PCR dla genu kodującego białko P związane z czynnikiem transkrypcyjnym Myb, dla którego udało się potwierdzić ekspresję tylko jednego z przewidzianych przez program AUGUSTUS 3 transkryptów (brak alternatywnego składania mRNA). Na Ryc. 16 znajduje się obraz rozdziału elektroforetycznego produktów amplifikacji 2 transkryptów genu kodującego S-metylotransferazę 2 homocysteiny, które ulegają ekspresji w badanych odmianach. Również ekspresję dwóch przewidzianych transkryptów potwierdzono dla genu kodującego dehydrogenazę glutationu (obraz elektroforetyczny rozdziału produktów odwrotnej transkrypcji i amplifikacji przedstawiono na Ryc. 17). Poniżej obrazów elektroforetycznych zamieszczono również schemat struktury przewidzianych transkryptów wygenerowanych przez program IGV.

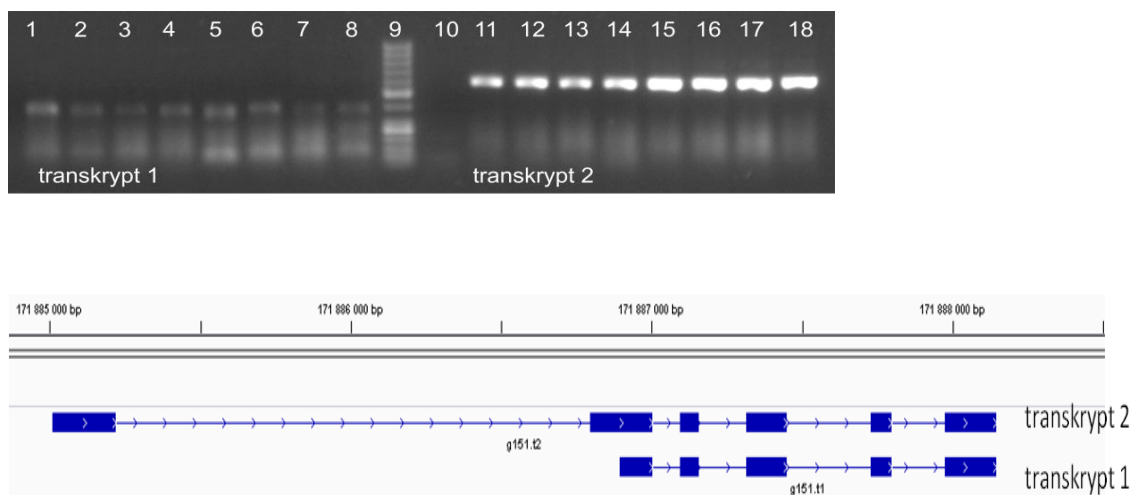


Ryc. 15

Obraz rozdziału elektroforetycznego produktów reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR dla dla transkryptu 1 genu kodującego białko P związane z czynnikiem transkrypcyjnym Myb. Dla tego genu udało się potwierdzić ekspresję tylko jednego (zaznaczonego na czerwono) z trzech transkryptów przewidzianych przez program AUGUSTUS. Strukturę wszystkich transkryptów przedstawiono na schemacie wygenerowanym przez program IGV.

1 – marker wielkości ; 2 – linia wrażliwa kontrola po 24h (SL1C); 3 – linia wrażliwa traktowana po 24h (SL1); 4 – linia wrażliwa kontrola po 7 dniach (SL7C); 5 – linia wrażliwa traktowana po 7 dniach (SL7); 6 – linia tolerancyjna kontrola po 24 h (TL1C); 7 – linia tolerancyjna traktowana po 24 h (TL1); 8 – linia tolerancyjna kontrola po 7 dniach (TL7C); 9 – linia tolerancyjna traktowana po 7 dniach (TL7); 10 – kontrola negatywna

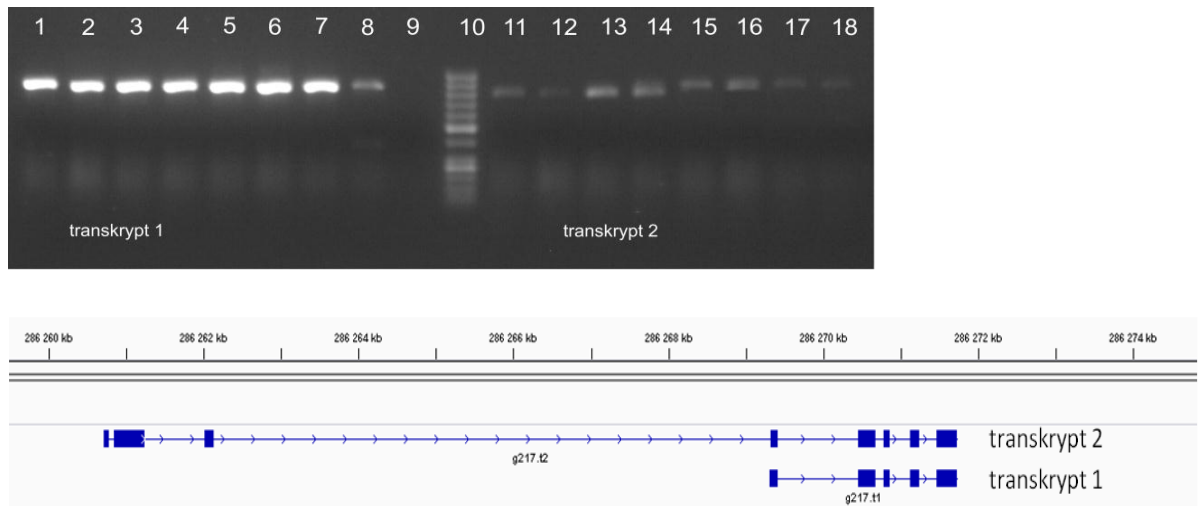
5. Wyniki i dyskusja



Ryc. 16

Obraz rozdziału elektroforetycznego produktów reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR dla dla transkryptu 1 i 2 genu kodującego S-metylotransferazę 2 homocysteiny. Dla tego genu udało się potwierdzić zdarzenie AS i ekspresję obu transkryptów przewidzianych przez program AUGUSTUS. Strukturę transkryptów przedstawiono na schemacie wygenerowanym przez program IGV.

1, 11 – linia wrażliwa kontrola po 24h (SL7C); 2, 12 – linia wrażliwa traktowana po 24h (SL1); 3, 13 – linia wrażliwa kontrola po 7 dniach (SL7C); 4, 14– linia wrażliwa traktowana po 7 dniach (SL7); 5, 15 – linia tolerancyjna kontrola po 24 h (TL1C); 6, 16 – linia tolerancyjna traktowana po 24 h (TL1); 7,17 – linia tolerancyjna kontrola po 7 dniach (TL7C); 8, 18 – linia tolerancyjna traktowana po 7 dniach (TL7); 9 – marker wielkości ; 10 – kontrola negatywna



Ryc. 17

Obraz rozdziału elektroforetycznego produktów reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR dla dla transkryptu 1 i 2 genu kodującego dehydrogenazę glutationu. Dla tego genu udało się potwierdzić zdarzenie AS i ekspresję obu transkryptów przewidzianych przez program AUGUSTUS. Strukturę transkryptów przedstawiono na schemacie wygenerowanym przez program IGV.

1, 11 – linia wrażliwa kontrola po 24h (SL7C); 2, 12 – linia wrażliwa traktowana po 24h (SL1); 3, 13 – linia wrażliwa kontrola po 7 dniach (SL7C); 4, 14– linia wrażliwa traktowana po 7 dniach (SL7); 5, 15 – linia tolerancyjna kontrola po 24 h (TL1C); 6, 16 – linia tolerancyjna traktowana po 24 h (TL1); 7,17 – linia tolerancyjna kontrola po 7 dniach (TL7C); 8, 18 – linia tolerancyjna traktowana po 7 dniach (TL7); 9 – marker wielkości ; 10 – kontrola negatywna

5.1.3 Dyskusja

Wraz ze wzrastającą ilością danych pochodzących z dużych projektów sekwencjonowania genomów, transkryptomów, degramów oraz puli małych, niekodujących cząsteczek RNA, rośnie potrzeba udostępnienia społeczności naukowej narzędzi bioinformatycznych, które umożliwią zarówno odpowiednie zarządzanie tymi danymi, ale przede wszystkim ich właściwą interpretację, zgodną z dostępną wiedzą biologiczną. Celem opisanych badań było oszacowanie liczby zdarzeń alternatywnego składania transkryptów pochodzących z największego chromosomu kukurydzy zwyczajnej (pierwszego) na podstawie przewidywań programu AUGUSTUS. Jest to jeden z pierwszych, ogólnodostępnych programów, który umożliwia przewidywanie zdarzeń AS z samej sekwencji nukleotydowej. Posiada także kilka opcji konfiguracyjnych, dzięki którym użytkownik może wybrać z jaką czułością program ma działać i ile transkryptów

przewidywać. Jednak, pomimo wyboru prawie najbardziej rygorystycznych ustawień, liczba genów zidentyfikowanych przy pomocy tego programu na jednym chromosomie znacznie przewyższała szacunkową liczbę genów dla całego genomu kukurydzy, co świadczy o wysokim poziomie fałszywie dodatnich wyników. Może być to związane z charakterystyką genomu kukurydzy, którego większość stanowią sekwencje powtórzone (rozdział 2.1). W ich obrębie często znajdują się pseudogeny lub fragmenty genów, które mogą przyczyniać się do zwiększenia liczby fałszywie pozytywnych rezultatów. Ponadto działanie modeli statystycznych takich jak HMM wymaga wstępnego ustalenia parametrów wejściowych w oparciu o zestaw sekwencji treningowych (ang. *training set*), które są przygotowywane i charakterystyczne dla każdego gatunku. Jednak niewłaściwy ich dobór - zbyt mała lub niereprezentatywna pula sekwencji, może być przyczyną ustalenia błędnych parametrów początkowych, które mogą zaburzać przewidywania programu i dawać niewłaściwe wyniki.

Kilkuetapowa weryfikacja wyników programu AUGUSTUS pozwoliła ocenić znaczenie i liczbę zdarzeń AS, a także zawęzić pulę badanych genów i skupić się na kilku konkretnych, których produkty białkowe zostały wcześniej opisane w bazie danych UniProtKB. Poza oceną skuteczności przewidywań programu AUGUSTUS, stworzona została prosta ścieżka analizy umożliwiająca zawężenie puli badanych genów. Ogólny odsetek genów przewidzianych przez program AUGUSTUS, które ulegały alternatywnemu składaniu (20%) okazał się zgodny z wcześniej postulowanymi [13]. Dla 58% (242) z 412 przewidzianych genów z największą liczbą transkryptów, udało się potwierdzić zajście zdarzeń alternatywnego składania poprzez dopasowanie różnych sekwencyjnych znaczników ekspresji. Ponadto sekwencje 42 genów zostały potwierdzone przez przewidywania z dwóch innych źródeł (program F-genesh i bazę danych Gramene), a dla 34 genów dopasowano referencyjne sekwencje mRNA pochodzące z kukurydzy. Te 42 geny uznano za nowe, wcześniej nieopisane zdarzenia splicingowe zachodzące u kukurydzy zwyczajnej.

Ponadto analiza na poziomie sekwencji aminokwasowej pozwoliła zidentyfikować 18 potencjalnych produktów białkowych powstających z zaproponowanych sekwencji. Dla czterech z genów kodujących zidentyfikowane białka, opisano w literaturze zajście zdarzeń alternatywnego splicingu. Trzy z tych genów zostały zidentyfikowane u *A. thaliana* i potwierdzono dla nich powstawanie dwóch izoform transkryptów.

Jednym z nich jest gen GAD kodujący dekarboksylazę glutationu (ang. *glutamate decarboxylase*), która odpowiedzialna jest za konwersję glutationu do kwasu γ -aminomasłowego (GABA). Cząsteczka ta jest akumulowana w komórkach roślinnych pod wpływem warunków stresowych, zarówno biotycznych jak i abiotycznych. W związku z tym postuluje się rolę GABA jako cząsteczki sygnałowej, biorącej udział w interakcjach rośliny z mikroorganizmami, zwierzętami czy metabolizmie azotu [78]. Dla genu GAD zidentyfikowano dwa transkrypty u *A. thaliana* - GAD1, który koduje białko o masie 57,1 kDa oraz GAD2, kodujący białko o masie 56,1 kDa. Oba białka mają sekwencje aminokwasowe identyczne w 85% i posiadają aktywność dekarboksylazy glutationu związaną z obecnością jonów wapnia. Analiza typu northern i western blot wykazała, że transkrypt GAD1 ulega ekspresji tylko w korzeniach, natomiast GAD2 w korzeniach, liściach, łodygach i kwiatach. Sugeruje się więc, że regulacja ekspresji poszczególnych izoform GAD jest organospecyficzna, a aktywność katalityczna jest zależna od jonów wapnia związanych z kalmoduliną [79].

Dwa inne geny, czyli białko związane z mikrotubulami o masie 65kDa (ang. *65 kDa microtubule associated protein*) i prawdopodobna białkowa kinaza serynowo/treoninowa (ang. *probable serine/threonine-protein kinase*) posiadają anotowane dwie izoformy transkryptów, które zostały zidentyfikowane odpowiednio na drodze sekwencjonowania pełnej długości EST [80] lub sekwencjonowania EST i analizy mikromacierzowej [81].

Czwartym genem, dla którego eksperymentalnie potwierdzono zjawisko alternatywnego składania mRNA, jest gen kodujący białko P związane z czynnikiem transkrypcyjnym Myb (ang. *Myb-related protein P*) z kukurydzy zwyczajnej. Białko to jest zaangażowane w szlak biosyntezy antocyjanin odpowiedzialnych m. in. za zabarwienie nasion kukurydzy. W obrębie owocni potwierdzono ekspresję dwóch izoform transkryptów - dłuższej (1802 nt) i krótszej (945 nt), które różnią się położeniem i długością eksonu trzeciego. Najprawdopodobniej to długi transkrypt koduje funkcjonalne białko, którego obecność jest konieczna do ekspresji dwóch innych enzymów biorących udział w procesie biosyntezy antocyjanin. N-terminalny koniec białka P wykazuje 40% homologię do czynników transkrypcyjnych z rodziny Myb, dlatego też sugeruje się, że działa ono na zasadzie aktywatora transkrypcji. Krótszy transkrypt, także posiada domenę funkcjonalną i może działać jako inhibitor kompetycyjny dla pełnej długości białka [82].

Dla genu kodującego białko P związane z czynnikiem transkrypcyjnym Myb program AUGUSTUS przewidział 3 transkrypty, których schemat organizacji odbiegał od wcześniej opisanego. W dwóch badanych liniach kukurydzy udało się potwierdzić ekspresję tylko jednego, najkrótszego z tych transkryptów. Brak przewidzianych produktów alternatywnego splicingu może być związany z różnicami genetycznymi między badanymi odmianami kukurydzy. Sekwencja genomowa, która była podstawą przewidywań programu AUGUSTUS pochodzi z linii wsobnej B73, natomiast w badaniach [82] stosowano odmiany australijskie o czerwonej i bezbarwnej owocni, w przeciwieństwie do analizowanych dwóch odmian rodzimych o owocni żółtej, w których biosynteza barwników może być regulowana w odmienny sposób.

Dla dwóch pozostałych genów przetestowanych w reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR udało się potwierdzić ekspresję obu przewidzianych transkryptów, a więc zajście zdarzeń alternatywnego składania mRNA. Wcześniej nie opisano możliwości zajścia AS dla genu kodującego S-metylotransferazę 2 homocysteiny (HMT2, ang. *homocysteine S-methyltransferase 2*). Wykazano, że u *A. thaliana* i kukurydzy zwyczajnej obecne są aż trzy geny kodujące HMT, które pochodzą z różnych klas i pojawiły się niezależnie u roślin jedno- i dwuliściennych. Białko kodowane przez ten gen jest to jeden z enzymów biorących udział w syntezie metioniny z homocysteiny i wraz z S-metylotransferazą metioniny reguluje poziom dostępności wolnej metioniny w cyklu SMM (ang. *S-methylometionine cycle*) u roślin okrytozalążkowych. Na podstawie badań aktywności enzymatycznej skonstruowano model, według którego prawie połowa produkowanej przez roślinę S-adenozylometioniny zużywana jest w cyklu przemian SMM, co zapobiega jej nadmiernemu gromadzeniu w komórce. Rośliny nie mają charakterystycznych dla innych organizmów eukariotycznych ujemnych pętli zwrotnych, regulujących poziom tego aminokwasu, jednak dzięki temu mechanizmowi mogą kontrolować ilość wolnej metioniny [83]. Program AUGUSTUS przewidział dwa transkrypty dla genu HMT2 i ekspresję obu udało nam się potwierdzić w badanych liniach kukurydzy. Ich ekspresja nie zależała od warunków stresowych - zachodziła zarówno u roślin po zabiegu oprysku herbicydem jak i roślin kontrolnych.

Drugim genem, którego ekspresję udało się potwierdzić w badanych liniach kukurydzy jest gen kodujący dehydrogenazę glutationu (GDH, ang. *glutamate dehydrogenase*). Jest to mitochondrialny enzym związany z metabolizmem glutationu i jonów amonowych. W kiełkujących nasionach bierze również udział w katabolizmie białek. Ponadto wiadomo, że poziom aktywności tego enzymu zależy od warunków stresowych związanych z deficytem wody czy substancji odżywczych, a także pojawieniem się w środowisku metali ciężkich czy herbicydów [84]. W genomie kukurydzy zidentyfikowano dwa geny dla tego enzymu, jednak do tej pory nie annotowano zdarzeń alternatywnego składania, jedynie różnice w aktywności izoform tego białka w obrębie komórek osłonki wiązki przewodzącej [85]. W badanych próbkach oba transkrypty ulegały ekspresji niezależnie od zastosowanego oprysku herbicydem Roundup® czy też wybranej linii kukurydzy.

Dla obu wyżej opisanych genów potwierdzone zdarzenia alternatywnego składania transkryptów mogą stanowić dodatkowy poziom regulacji ich ekspresji, który może także zależeć od zastosowanych warunków czy hodowanej odmiany. Aby to potwierdzić należałoby przeanalizować większe spektrum linii hodowlanych, a także włączyć do analizy materiał pochodzący z innych organów np. łodyg czy korzeni.

Opisana ścieżka analizy pozwoliła skutecznie zredukować pulę przewidzianych genów i potwierdzić ekspresję trzech z nich (białka P związane z czynnikiem transkrypcyjnym Myb, S-metylotransferazy 2 homocysteiny i dehydrogenazy glutationu), dla których został opisany produkt białkowy o znanej funkcji. Ewaluacja wyników programu AUGUSTUS umożliwiła określenie nowych zdarzeń alternatywnego składania dla 42 genów. Zidentyfikowano również AS w 18 genach kodujących znane białka, z których dla 4 udowodniono już eksperymentalnie ekspresję kilku izoform transkryptów. Przeprowadzona analiza wskazuje na duże znaczenie procesu AS w ekspresji genów u kukurydzy, wykazano, że dotyczy co najmniej 20% genów, jednak zastosowanie dostępnych programów i kolekcji sekwencji nie pozwala w pełni ocenić skali tego zjawiska, ani przewidzieć wpływu warunków stresowych na proces wyboru różnych miejsc splicingowych i wydajności ekspresji poszczególnych izoform transkryptów. Dlatego aby odpowiedzieć na te pytania i poznać zmiany w profilu AS u

badanych linii kukurydzy pod wpływem stresu herbicydowego, wykorzystano metodę sekwencjonowania mRNA drugiej generacji.

5.2 Analiza wyników sekwencjonowania transkryptomów

5.2.1 Sekwencjonowanie i analiza jakości uzyskanych odczytów

Po izolacji całkowitego RNA ze zhomogenizowanych liści kukurydzy zwyczajnej, przygotowano dla obu odmian i punktów czasowych biblioteki cDNA wg. opisanej w poprzednim rozdziale procedury (rozdział 4.2.5). Aby otrzymać odczyty o jak najlepszej jakości, do uzyskania bibliotek użyto tylko preparaty RNA o współczynniku RIN, który wynosił 9 lub więcej (porównaj 5.1.2). Zastosowanie materiału wyjściowego o niskiej jakości, który uległ degradacji, może skutkować uzyskaniem małej ilości odczytów, z których większość będzie pochodziła z końca 5' cząsteczek mRNA. W celu uzyskania reprezentatywnej puli sekwencji transkryptomu przeprowadzono dwukrotnie procedurę sekwencjonowania z obu końców (typu paired-end) fragmentów cDNA o długości 50 nt. Otrzymano od 61 mln do przeszło 110 mln odczytów dla poszczególnych linii z pojedynczego sekwencjonowania (szczegółowe dane zebrano w Tabeli 13).

	SL1C	SL1	SL7C	SL7	TL1C	TL1	TL7C	TL7C
Liczba odczytów z pierwszego sekwencjonowania	95,038,446	105,244,574	98,028,634	92,126,504	93,982,100	61,924,456	97,160,912	70,460,594
% odczytów z Q>=30	93	91,13	91,92	92,46	92,64	94,5	89,84	94,21
Liczba odczytów z pierwszego sekwencjonowania	110,744,012	106,890,424	106,162,428	98,363,356	99,578,068	65,020,556	103,788,630	105,913,506
% odczytów z Q>=30	80,68	88,51	80,94	69,47	55,40	68,59	61,77	90,28
Liczba odczytów po filtrowaniu i uwspólnieniu	68,133,883	72,371,196	47,368,031	25,435,263	42,720,339	3,818,536	55,960,902	18,158,043
% utraconych odczytów	33,78	6,42	32,64	53,98	10,04	89,13	27,19	61,79

Tabela 13

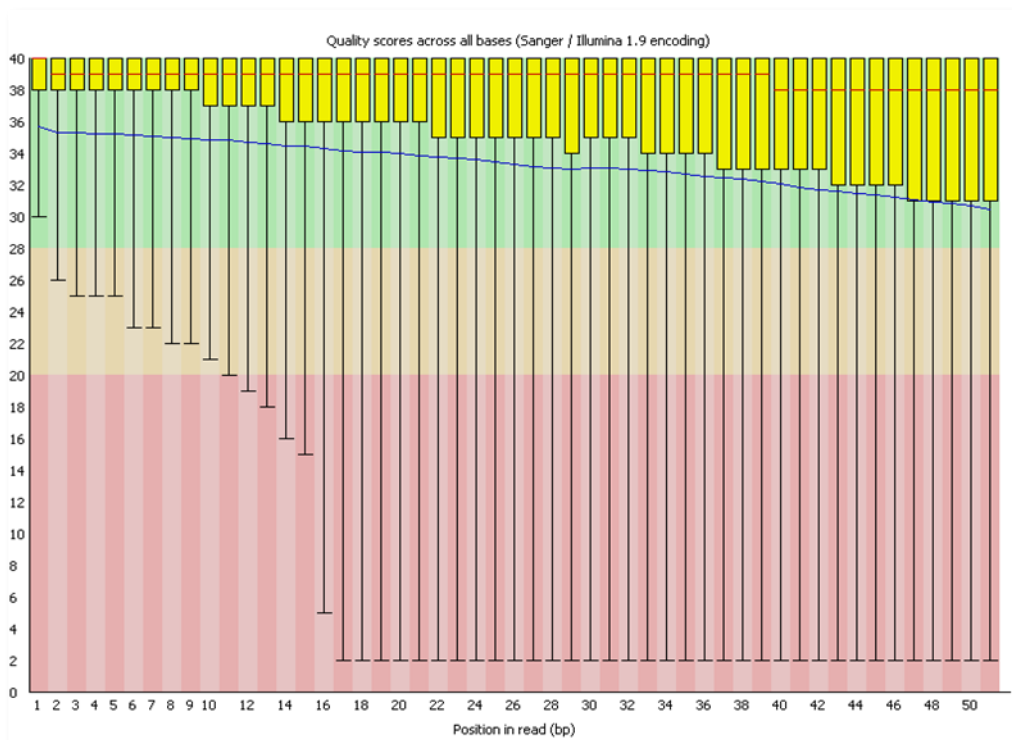
Zestawienie liczby uzyskanych surowych odczytów z obu sekwencjonowań oraz liczby odczytów po wykonaniu filtrowania i uwspólniania plików. Wyszczególniono także procent odczytów utraconych w wyniku analizy jakości.

W trakcie sekwencjonowania otrzymano pliki wyjściowe w formacie .bcl, które są plikami binarnymi zawierającymi informacje o nukleotydzie włączonym do sekwencjonowanego fragmentu wraz z przypisaną mu wartością jakości odczytu dla tego pojedynczego nukleotydu. Pliki te są tworzone na bieżąco wraz z postępem cyklu sekwencjonowania, jednak przed przystąpieniem do dalszych etapów analizy, poddaje się je konwersji do innych formatów, np. fastq. Fastq są to pliki tekstowe, które przechowują dane biologiczne (sekwencje nukleotydową), ale także zachowują informacje o jakości odczytu poszczególnych nukleotydów w sekwencji. Struktura tego pliku zawiera identyfikator sekwencji wraz z koordynatami klastra, z którego pochodzi, sekwencję nukleotydów, linię na dodatkowe informacje oraz kodowany za pomocą jednego znaku ASCII uzyskany dla danego nukleotydu wynik jakości. Jakość danego nukleotydu (ang. *quality score*) definiowana przez równanie $Q = -10 \log_{10}(e)$, gdzie e jest oszacowanym prawdopodobieństwem, że dany nukleotyd został błędnie odczytany. Im wyższa więc wartość Q , tym mniejsze prawdopodobieństwo, że nukleotyd został odczytany omyłkowo. Wartość $Q=30$ oznacza, że prawdopodobieństwo odczytania 1 błędnego nukleotydu na 1000 zsekwencjonowanych. Na takiej samej zasadzie jakość odczytów nukleotydów jest określana w sekwencjonowaniu metodą Sanger.

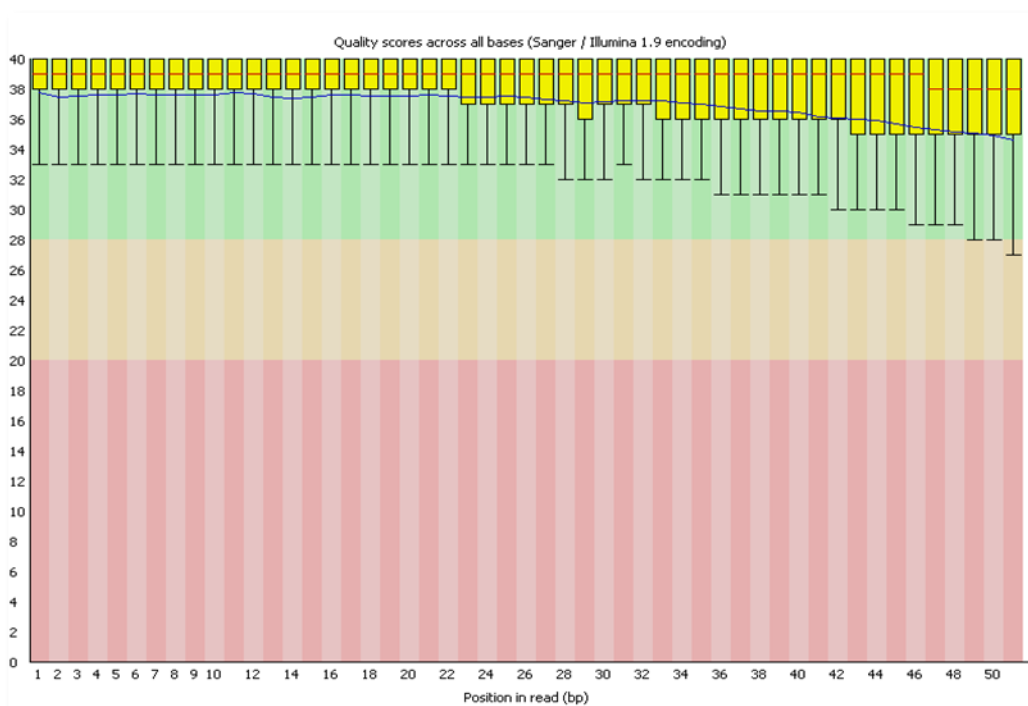
Po konwersji plików bcl, uzyskane odczyty poddano ocenie jakościowej z wykorzystaniem programu FastQC. Umożliwiło to zidentyfikowanie kilku sekwencji nadreprezentowanych występujących we wszystkich bibliotekach. Po przeszukaniu bazy danych NCBI nucleotide przy pomocy algorytmu blastn, udało się dopasować sekwencje nadreprezentowane do sekwencji pochodzenia mitochondrialnego, chloroplastowego i rybosomalnego. Zdeponowane w bazie danych rekordy dotyczące tych sekwencji pochodzących z kukurydzy zwyczajnej zastosowano do filtrowania odczytów. Wykorzystano do tego opcję programu Bowtie umożliwiającą oddzielenie odczytów, które zostaną dopasowane do zadanej sekwencji. W kolejnym etapie zastosowano jedną z opcji programu FastQC (Quality Filter), która pozwoliła usunąć odczyty o niskiej jakości (poniżej $Q=30$). Na Ryc. 18 przedstawiono wykresy wygenerowane przez program FastQC dla odczytów pochodzących z sekwencjonowania mRNA linii wrażliwej na Roundup® S79757 zebranej po 7

dniach od oprysku (SL7) tym herbicydem. W części A wykresy przedstawiają jakość otrzymanych odczytów (na osi Y) skorelowaną z poszczególnymi pozycjami nukleotydów w sekwencjonowanym 50 nt fragmencie (na osi X). Po etapie filtracji widać zmniejszony rozrzut wartości jakości odczytów dla wszystkich pozycji. Na wykresach w części B przedstawiona jest zależność ilości i jakości otrzymanych odczytów. Przed filtracją zaznaczono wyraźny pik pochodzący z odczytów o bardzo niskiej jakości, których brak po etapie odfiltrowania.

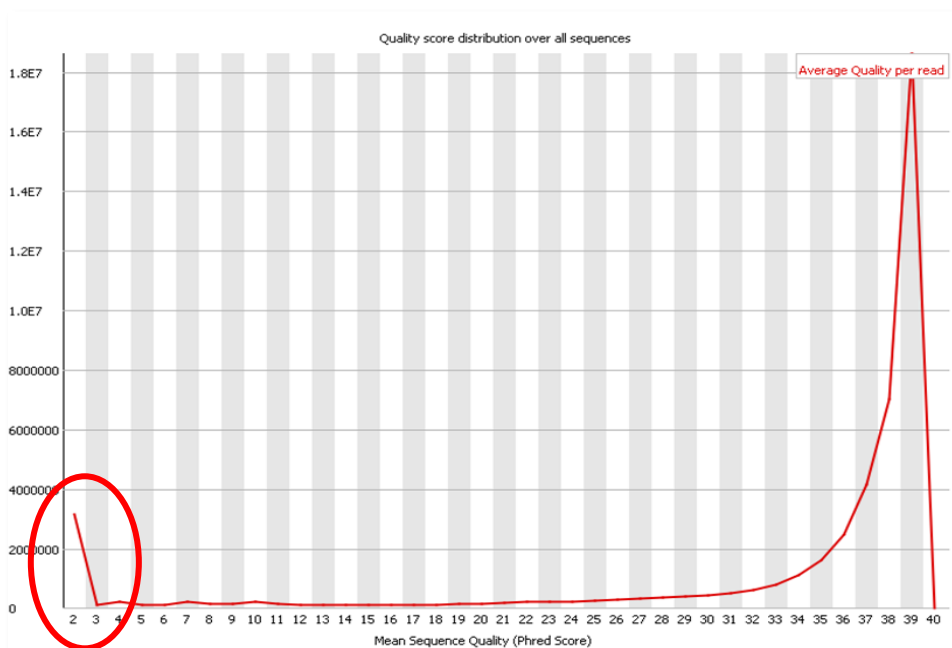
5. Wyniki i dyskusja



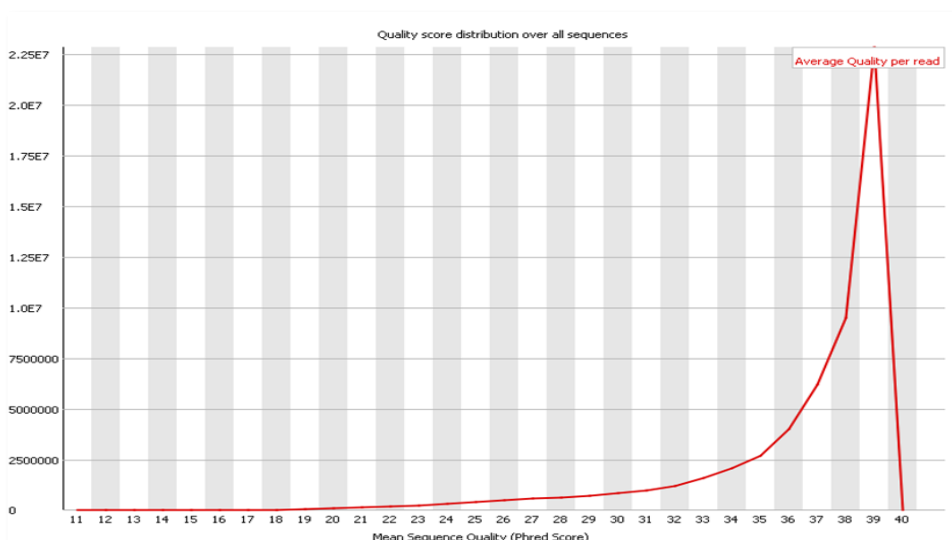
A1



A2



B1



B2

Ryc. 18

Wykresy wygenerowane przez program FastQC dla odczytów pochodzących z sekwencjonowania mRNA linii wrażliwej na działanie herbicydu Roundup® zebranej tydzień po oprysku herbicydem. Wykresy w części A przedstawiają zależność jakości (oś Y) od pozycji nukleotydu w otrzymanym po sekwencjonowaniu odczycie o długości 50 nt (oś X), na wykresie A1 dane przed odfiltrowaniem odczytów o jakości poniżej Q=30, na wykresie A2 dane po odiltrowaniu. W części B przedstawiono wykresy zależności liczby otrzymanych odczytów (oś Y) od jakości dopasowanego nukleotydu (oś X), na wykresie B1 na czerwono zaznaczono pik odczytów o bardzo niskiej jakości, który na wykresie B2 (dla danych po etapie filtrowania) został usunięty.

W ten sposób zawężono pulę odczytów tylko do tych o najwyższej jakości, czyli najmniejszym prawdopodobieństwie błędu, a odczyty z obu powtórzeń sekwencjonowań uwspólniono dla każdej z bibliotek. Tak przygotowane dane poddano kolejnym krokom analizy.

5.2.2 Analiza różnic w ekspresji genów między badanymi liniami pod wpływem stresu herbicydowego

Wykorzystanie programu Cuffcompare, który jest składową programy Cufflinks (zobacz Ryc. 10) umożliwiło porównanie uzyskanych sekwencji transkryptomów badanych linii kukurydzy z transkryptomem referencyjnym kukurydzy z linii B73. Dzięki temu zidentyfikowano od 166 tys. do 193 tys. transkryptów, z których potencjalnym celem dla AS było od 17 tys. do przeszło 24 tys. loci, do których został zmapowany więcej niż jeden transkrypt. Ponadto rozpoznano od 3,5 tys. do 27 tys. nowych eksonów, od 1,7 tys. do 11 tys. nowych intronów, które mapują do przeszło 11 tys. nowych loci. Najwięcej nowych, nieannotowanych w transkryptomie referencyjnym eksonów i intronów, znaleziono w linii wrażliwej zebranej dzień po zabiegu oprysku. Dane sekwencyjne z obu linii udało się zmapować prawie w 100% do transkryptomu referencyjnego. Jedynie od 3 do 6 eksonów i 2 do 4 intronów (w zależności od linii i warunku, z którego pochodziły dane), które były annotowane w transkryptomie referencyjnym nie zostały zidentyfikowane w transkryptomach badanych linii. Szczegółowe dane przedstawiono w Tabeli 14.

	SL1C	SL1	SL7C	SL7	TL1C	TL1	TL7C	TL7
Całkowita liczba mRNA	190647	192803	185576	172575	183597	146776	193320	166424
Loci z więcej niż 1 transkryptem	23 793	24 249	23 178	21 630	23 205	17 111	24 334	20 738
Nowe eksony	27585	27795	25731	17641	20818	3573	25846	12178
Nowe introny	10757	11180	9533	6957	8844	1795	10138	5502
Nowe loci	11726	11201	10767	7433	7582	1493	10161	4505

Tabela 14

Zestawienie porównania sekwencji transkryptomów uzyskanych dla badanych linii z transkryptomem linii referencyjnej kukurydzy B73. W transkryptomie referencyjnym kukurydzy annotowano do tej pory 136402 cząsteczki mRNA, zlokalizowane w 109846 loci.

SL1C – linia wrażliwa kontrola po 24h; SL1– linia wrażliwa traktowana po 24h; SL7C – linia wrażliwa kontrola po 7 dniach; SL7– linia wrażliwa traktowana po 7 dniach; TL1C – linia tolerancyjna kontrola po 24 h; TL1 – linia tolerancyjna traktowana po 24 h; TL7C – linia tolerancyjna kontrola po 7 dniach; TL7– linia tolerancyjna traktowana po 7 dniach

Z danych zebranych w Tabeli 13 wynika, że dla obu badanych linii kukurydzy nastąpił spadek aktywności transkrypcyjnej po tygodniu od zabiegu oprysku herbicydem – w linii wrażliwej S2757 13 tys. transkryptów nie ulegało ekspresji w roślinach poddanych zabiegowi oprysku w porównaniu do roślin kontrolnych. Dla linii tolerancyjnej S245 obniżenie ilości powstających transkryptów było bardziej znamienne – różnica liczby powstających transkryptów między roślinami kontrolnymi a poddanymi opryskowi wynosiła ponad 36 tys. transkryptów w przypadku roślin zebranych dzień po oprysku herbicydem i ponad 26 tys. transkryptów w przypadku roślin badanych tydzień po zabiegu oprysku.

Zastosowanie programu Cuffdiff wymaga użycia plików GTF (ang. *gene transfer format*), które są generowane przez program TopHat oraz plików SAM (pliki tekstowe przechowujące dane sekwencyjne w postaci kolumn ze znakami ASCII oddzielonych tabulatorami), w których zestawione są dopasowania dla danych sekwencyjnych pochodzących z poszczególnych linii, które zostały stworzone przy pomocy Cuffmerge. Dane wyjściowe Cuffdiff zapisuje w serii plików z zestawieniami wyników badania zmian w poziomie ekspresji (podanych jako FPKM, ang. *fragments per kilobase of transcript per*

million mapped reads) zarówno poszczególnych transkryptów jak i genów. Cuffdiff śledzi również zmiany względnej obfitości (ang. *abundance*) transkryptów dzielących wspólne miejsca startu transkrypcji i względną obfitość transkryptów pierwotnych każdego genu, które pozwala ocenić zmiany w składaniu mRNA oraz umożliwia rozpoznanie użycia różnych sekwencji promotorowych dla danego genu. Wyniki programu Cuffdiff zebrane są w szeregu następujących plików:

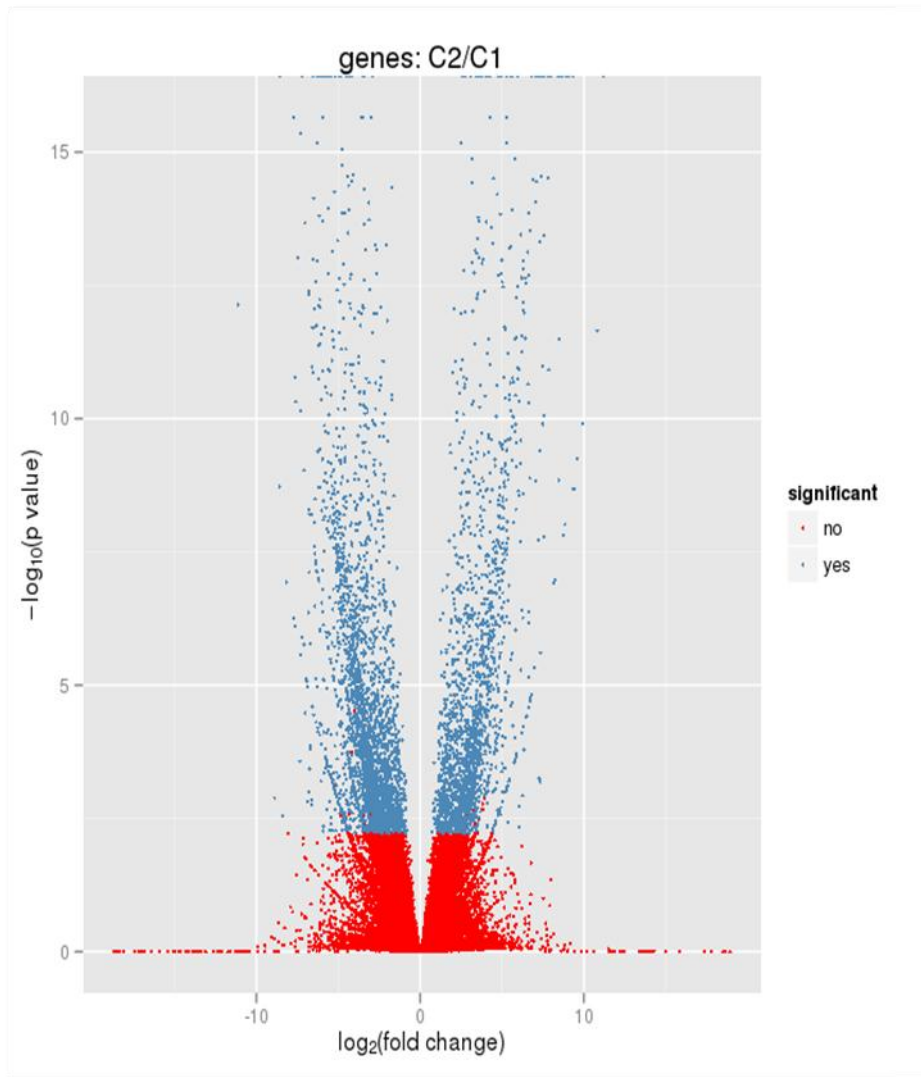
- pliki umożliwiające śledzenie wartości FPKM (ang. *FPKM tracking files*) pomiędzy porównywanymi próbkami dla izoform transkryptów - *isoforms.fpkm_tracking*, genów - *genes.fpkm_tracking*, sekwencji kodujących (CDS, ang. *coding sequences*) - *cds.fpkm_tracking* i pierwotnych transkryptów - *tss_groups.fpkm_tracking*; wartość FPKM dla genów i transkryptów pierwotnych jest obliczana jako suma FPKM dla wszystkich transkryptów zmapowanych w obrębie poszczególnych genów lub transkryptów pierwotnych;
- pliki umożliwiające śledzenie liczby odczytów między badanymi próbkami (ang. *count tracking files*) generowane są analogicznie, jak w przypadku plików z wartościami FPKM, dla izoform transkryptów - *isoforms.count_tracking*, genów - *genes.count_tracking*, sekwencji kodujących - *cds.count_tracking* i pierwotnych transkryptów - *tss_groups.count_tracking*; liczba odczytów dla genów i transkryptów pierwotnych jest sumą liczby odczytów dla wszystkich transkryptów zmapowanych w obrębie danego genu lub transkryptu pierwotnego
- pliki z danymi na temat ekspresji i liczby odczytów (ang. *read group tracking files*) dla każdego transkryptu (*isoforms.read_group_tracking*), genu (*genes.read_group_tracking*), sekwencji kodującej (*cds.read_group_tracking*) i pierwotnego transkryptu (*tss_groups.read_group_tracking*); pliki te generowane są osobno dla każdego z powtórzeń eksperymentu sekwencjonowania
- zestaw plików zawierających wyniki analizy różnicowej ekspresji izoform transkryptów, genów, sekwencji kodujących i transkryptów pierwotnych pomiędzy porównywanymi próbkami; dla każdej pary próbek generowane są cztery pliki (*isoform_exp.diff*, *gene_exp.diff*, *tss_group_exp.diff*, *cds_exp.diff*), które zawierają w poszczególnych kolumnach następujące dane:

- testowane id – unikalny identyfikator przypisany do każdego transkryptu, genu, sekwencji kodującej czy pierwotnego transkryptu w zależności od tego który z czterech plików jest rozpatrywany
 - nazwa lub identyfikator genu
 - locus genu
 - identyfikatory porównywanych próbek
 - status przeprowadzanego testu statystycznego (OK – został przeprowadzony, NOTEST – zbyt mała liczba dopasowań, która nie wystarczyła do jego przeprowadzenia, LOWDATA – zbyt mało danych sekwencyjnych, HIDATA – zbyt duża liczba odczytów zmapowana do danego locus, FAIL - źle skonstruowana macierz kowariancji lub inny wyjątek, który nie pozwala przeprowadzić testu)
 - FPKM sekwencji dla porównywanych próbek
 - zlogarytmowana różnica ekspresji między badanymi próbkami (logarytm binarny)
 - wartość poziomu istotności (ang. *p value*)
 - wartość poziomu istotności (wartość q , ang. *q value*) danego testu statystycznego z uwzględnieniem poziomu fałszywie pozytywnych wyników (FDR, ang. *False Discovery Rate*)
 - oznaczenie czy porównywane wartości są istotne statystycznie (oznaczone w kolumnie jako YES lub NO)
- plik zawierający wyniki testów oceniających różnice we wzorze AS między badanymi próbkami (splicing.diff); tylko transkrypty pierwotne, z których powstała więcej niż jedna izoforma są brane pod uwagę; pliki wyjściowe zawierają w kolejnych kolumnach:
 - testowane id – unikatowy identyfikator pierwotnego transkryptu
 - nazwa lub identyfikator genu
 - locus genu
 - identyfikatory porównywanych próbek
 - status przeprowadzanego testu (podobnie jak w przypadku plików z wynikami analizy różnicowej ekspresji)

- wartość pierwiastka kwadratowego z rozbieżności Jensen-Shannon obliczonej z względnej obfitości występowania danej izoformy transkryptu
 - wartość testowana, w tym wypadku równa wartości pierwiastka kwadratowego z rozbieżności Jensen-Shannon
 - wartość poziomu istotności
 - wartość q
 - oznaczenie czy porównywane wartości są istotne statystycznie (oznaczone w kolumnie jako YES lub NO)
- plik zawierający wyniki analizy różnic w ekspresji sekwencji kodujących (cds.diff); plik zawiera tylko te geny, dla których ekspresji ulega więcej niż jeden CDS; w kolejnych kolumnach zawiera informacje takie jak w przypadku plików splicing.diff, a rozbieżność Jensen-Shannon jest obliczana dla względnej obfitości rozpatrywanych sekwencji kodujących;
 - plik z danymi na temat wykorzystania różnych sekwencji promotorowych (promoters.diff), w którym są zestawione i porównane tylko te geny, z których powstał więcej niż jeden transkrypt pierwotny z wykorzystaniem różnych promotorów; kolumny w pliku opisują dane analogicznie jak w przypadku plików splicing.diff i cds. diff – rozbieżność Jensen-Shannon jest obliczana na podstawie obfitości występowania w badanych próbkach porównywanych transkryptów pierwotnych
 - dwa pliki zawierające informacje o kluczowych właściwościach programu, m. in. wartości czynników normalizujących użytych do określania wartości FPKM dla poszczególnych powtórzeń i bibliotek (read_groups.info) oraz informacje na temat przebiegu analizy (run.info).

Analiza danych wyjściowych z programu Cuffdiff umożliwiła zidentyfikowanie największych różnic w ekspresji genów pomiędzy transkryptomami badanych linii tydzień po oprysku herbicydem Roundup®. Linia wrażliwa różniła się od linii tolerancyjnej w tym punkcie czasowym ekspresją 599 genów oraz wzorem składania 424 transkryptów, a także 37 genami o różnym sposobie regulacji (wybór innej sekwencji promotorowej). Na Ryc. 19

przedstawiono wykres wygenerowany przez program CummeRbound z plików wyjściowych programu Cuffdiff dla genów o zróżnicowanej i znaczącej statystycznie ekspresji genów.

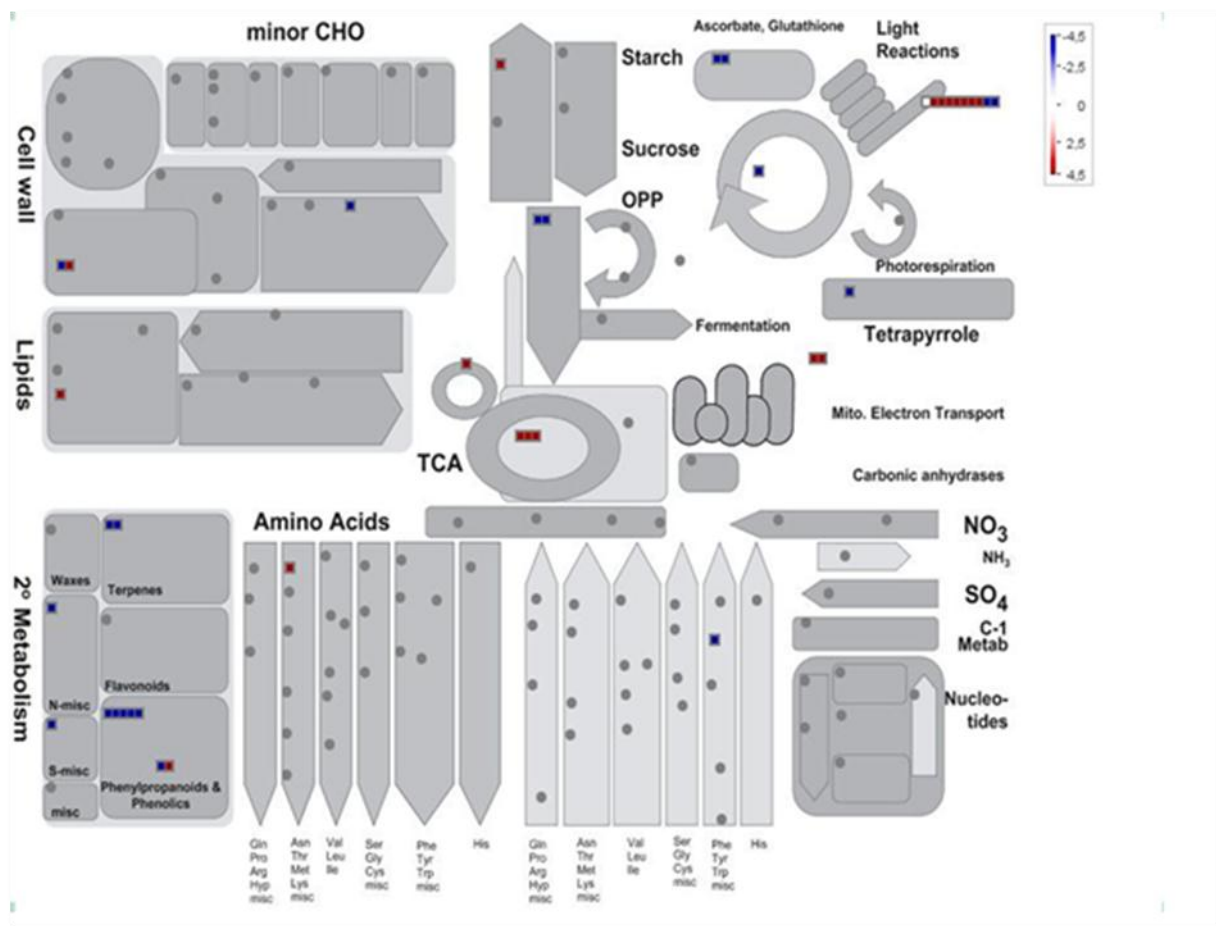


Ryc. 19

Porównanie ekspresji genów między linią tolerancyjną (na wykresie oznaczoną jako C2), a linią wrażliwą (C1), które były poddane opryskowi herbicydem i zebrane tydzień po zabiegu oprysku. Zobrazowano na wykresie istotne statystycznie różnice w poziomie ekspresji genów w obu próbkach (wielokrotność zmiany poziomu ekspresji) – niebieskie kropki reprezentują geny o znaczącej różnicy w ekspresji, czerwone geny o statystycznie nieznaczącej różnicy w ekspresji.

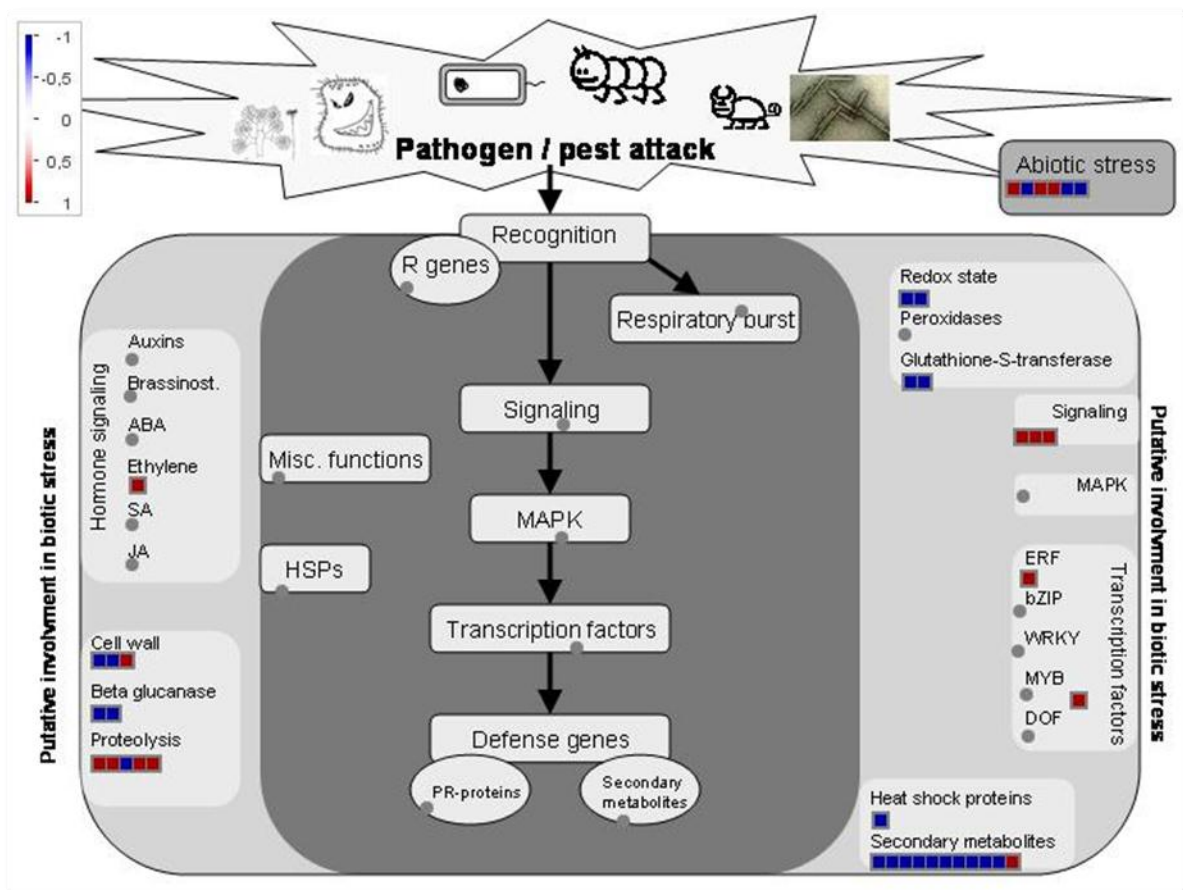
Porównanie zmian w ekspresji genów (plik gene_exp.diff), między roślinami kontrolnymi a poddanymi działaniu herbicydu pozwoliło zidentyfikować w obu liniach 914 genów, których ekspresja uległa zmianie po zastosowaniu oprysku herbicydem. W linii wrażliwej zmienionej ekspresji ulegało łącznie 71 genów, z czego tylko 6 genów miało zmieniony poziom ekspresji w obu punktach czasowych (wszystkie z nich ulegały podwyższonej ekspresji). W linii tolerancyjnej po zastosowaniu herbicydu poziom ekspresji 843 genów ulegał zmianie. Zarówno dzień jak i tydzień po oprysku zmieniony poziom ekspresji miało 61 genów - 19 z tych genów w obu punktach czasowych miało obniżony poziom ekspresji, a 42 podwyższony. Zidentyfikowane pule genów o zmienionej ekspresji pod wpływem oprysku herbicydem są unikalne dla obu badanych linii - oba zbiory są różne i nie znaleziono wspólnego genu, którego ekspresja zmieniałaby się po podaniu herbicydu Roundup® zarówno w linii wrażliwej jak i tolerancyjnej. W Aneksie 3 znajduje się tabela zawierająca szczegółowe informacje na temat zidentyfikowanych genów o zmienionej ekspresji w obu punktach czasowych (identyfikator genu, jego lokalizację na chromosomie oraz zlogarytmowaną wartość zmiany poziomu ekspresji i kodowaną domenę białkową).

Spośród genów o zmienionym poziomie transkrypcji wytypowano 140 transkryptów o największych różnicach w ekspresji po zastosowaniu oprysku herbicydem. Wykorzystując program MapMan dopasowano 96 spośród wybranych transkryptów do grup funkcjonalnych. Ryc. 20 przedstawia wyniki działania programu – przy szlakach metabolicznych zaznaczono geny o obniżonej (kolor niebieski) lub podwyższonej ekspresji (kolor czerwony). 23 z dopasowanych genów związane są z odpowiedzią rośliny na warunki stresowe (na Ryc. 21 przedstawiono szczegółowe wyniki). Liczną grupę genów o zróżnicowanej ekspresji po zastosowaniu oprysku herbicydem stanowią także geny kodujące białka biorące udział w fotosyntezie (14), metabolizmie wtórnym (11) czy metabolizmie białek (7). Duża ilość transkryptów została zmapowana do genów o nieznanym funkcjach (23) lub biorących udział w różnych szlakach metabolicznych (18).



Ryc. 20

Obraz głównych szlaków metabolicznych generowany przez program MapMan dla grupy 140 transkryptów różniących się ekspresją pomiędzy badanymi liniami po oprysku herbicydem. Czerwonymi kwadratami zaznaczono transkrypty o podwyższonej ekspresji pod wpływem zastosowanego oprysku herbicydem Roundup®, niebieskimi – transkrypty genów o obniżonej ekspresji, które biorą udział w poszczególnych reakcjach metabolicznych.



Ryc. 21

Graficzne przedstawienie szlaków metabolicznych biorących udział w reakcjach związanych z odpowiedzią roślin na warunki stresowe. Obraz został wygenerowany przy pomocy programu MapMan, czerwonymi kwadratami w poszczególnych grupach funkcjonalnych zaznaczono transkrypty o podwyższonej ekspresji, a niebieskimi geny o obniżonym poziomie ekspresji pod wpływem oprysku herbicydem Roundup®.

Jednym z genów o zmienionej ekspresji w linii wrażliwej na działanie herbicydu Roundup® był gen, którego sekwencja została zmapowana do dwóch modeli genów zawartych w transkryptomie referencyjnym. Jednym z nich jest gen o numerze identyfikacyjnym GRMZM5G833207, który koduje domenę o właściwościach ATPazowych i należy do rodziny białek - transporterów ABC (ang. *ATP-binding cassette transportes*). Według ostatnio opublikowanego spisu genów kukurydzy zwyczajnej kodujących białka z tej rodziny [86], ten gen należy do podgrupy ABCB. Białka wchodzące w skład tego podtypu posiadają tylko jedną domenę transmembranową i jedną domenę wiążącą ATP. Są

najczęściej zakotwiczone w błonach organelli komórkowych lub wewnątrz mitochondriów i są określane jako transportery związane z przetwarzaniem antygenów (TAP, ang. *transporters associated with antigen processing*). Drugi model genu o numerze identyfikacyjnym GRMZM5G808836 jest homologiem genu *A. thaliana* kodującego 5 glikoproteinę P (ang. *P-glycoprotein 5*), która także należy do podgrupy transporterów ABCB, a dokładniej do podtypu tych białek, które określane są jako białka związane z odpornością wielolekową (MDR, ang. *multi drug resistance protein*). Dzień po zabiegu oprysku w linii tolerancyjnej poziom ekspresji tego genu wzrósł 4-krotnie (w porównaniu do roślin kontrolnych), natomiast tydzień po oprysku aż 7-krotnie.

Również w linii tolerancyjnej na działanie herbicydu Roundup®, w puli 61 genów o podwyższonej ekspresji w obu badanych punktach czasowych, został zidentyfikowany gen kodujący białko z rodziny ABC transporterów - był to gen o numerze identyfikacyjnym GRMZM2G477872. W porównaniu do roślin niepoddanych działaniu herbicydu dzień po oprysku poziom jego ekspresji wzrósł 3 - krotnie, natomiast tydzień po zabiegu oprysku 4-krotnie. Gen ten koduje białko należące do białek związanych z plejotropową odpornością na leki (PDR, ang. *pleiotropic drug resistance protein*), które zbudowane jest z dwóch domen o właściwościach ATPazowych i dwóch domen transmembranowych.

5.2.3 Badanie różnic w profilu składania mRNA między analizowanymi bibliotekami

Aby ustalić jaki typ zdarzeń AS przeważał w badanych próbkach i czy ulegał zmianie pod wpływem zastosowanych czynników stresowych zastosowano skrypt korzystający ze środowiska java. Technologia java została zaprojektowana tak, aby umożliwić programowanie przenośnych, wysoce wydajnych aplikacji dla możliwie jak najszerszego spektrum platform przetwarzania cyfrowego, co umożliwia wykorzystanie stworzonego skryptu w różnych systemach operacyjnych. Korzysta on z informacji o koordynatach eksonów i intronów, które są zawarte w pliku wyjściowym z programu TopHat (plik GTF). Skrypt rozpatruje każdy gen wraz z jego transkryptami niezależnie i nie wymaga stosowania transkryptów referencyjnych. Posiadając informacje o transkryptach oraz adresach poszczególnych eksonów wyszukiwane są zakresy, w których występuje nałożenie się

eksonów. Dla każdego typu zdarzenia AS zdefiniowano reguły umożliwiające klasyfikację danego przypadku nakładania się koordynatów eksonów bądź intronów. Po przeanalizowaniu wszystkich zakresów tworzone jest podsumowanie informujące o ilości zdarzeń danego typu dla aktualnie analizowanego genu. Dodatkowo użytkownik otrzymuje zbiorczą informację o zdarzeniach alternatywnego składania we wszystkich przeanalizowanych genach. Podsumowanie wyników działania skryptu dla poszczególnych bibliotek znajduje się w Tabeli 15.

	SL1C	SL1	SL7C	SL7	TL1C	TL1	TL7C	TL7
Pominięcie eksonu	16%	16%	16%	17%	17%	18%	5%	18%
Dodatkowy ekson	22,3%	22%	23%	23%	21%	26%	7%	23%
Zatrzymanie intronu	16,2%	16%	16%	15%	16%	14%	18%	15%
Alternatywne miejsce splicingowe 3'	23%	23%	23%	23%	23%	22%	35%	22%
Alternatywne miejsce splicingowe 5'	22,5%	23%	22%	22%	23%	20%	35%	22%
% genów ulegających AS	31,15	32,2	32,16	33	33,32	34,76	32,92	33,77

Tabela 15

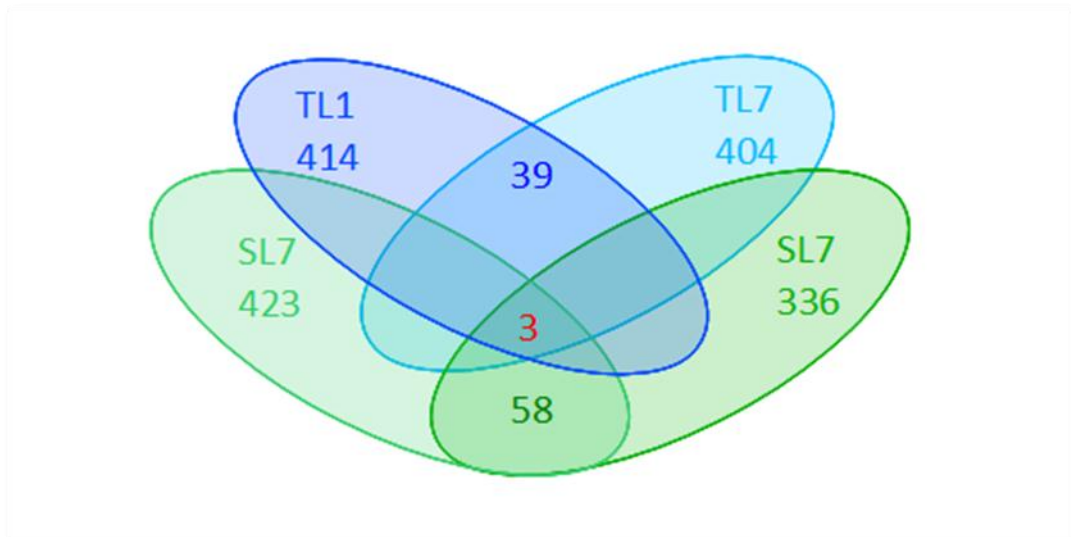
Typy zdarzeń alternatywnego składania określone w badanych liniach oraz procent genów ulegających AS określony przez opisany skrypt.

SL1C – linia wrażliwa kontrola po 24h; SL1– linia wrażliwa traktowana po 24h; SL7C – linia wrażliwa kontrola po 7 dniach; SL7– linia wrażliwa traktowana po 7 dniach; TL1C – linia tolerancyjna kontrola po 24 h; TL1 – linia tolerancyjna traktowana po 24 h; TL7C – linia tolerancyjna kontrola po 7 dniach; TL7– linia tolerancyjna traktowana po 7 dniach

Jednym z plików wyjściowych generowanych przez program Cuffdiff jest plik spliced.diff, w którym znajdują się dane dotyczące genów, które różniły się między analizowanymi próbkami wzorem składania oraz izoformą transkryptu, która ulegała najwyższej ekspresji. Analiza porównawcza tych plików między roślinami kontrolnymi, a poddanymi zabiegowi oprysku, pozwoliła więc wytypować te geny, które odpowiedziały na zastosowane warunki stresowe zmianą w składaniu transkryptów.

Po porównaniu danych dla roślin z linii S79757, czyli wrażliwych na herbicyd Roundup® zebranych dzień po zabiegu oprysku, zidentyfikowano 423 geny, które różniły się wzorem składania mRNA. Tydzień po zabiegu oprysku genów tych było mniej - 336. Dla roślin

z linii S245 tolerancyjnej na działanie herbicydu Roundup®, dzień po zabiegu oprysku rozpoznano 414 takich genów, natomiast tydzień po zabiegu ich liczba spadła nieznacznie - takich genów było 404. W obu punktach czasowych w linii wrażliwej różny wzór składania mRNA wykazywało 58 genów, natomiast w linii tolerancyjnej 39. Okazało się, że tylko 3 geny wykazywały różny wzór składania i różniły się dominującą izoformą w obu badanych liniach kukurydzy i punktach czasowych. Na Ryc. 22 przedstawiono schemat graficzny identyfikacji poszczególnych grup genów, a w Tabeli 16 zebrano szczegółowe informacje na temat ekspresji izoform 3 genów wspólnych dla badanych linii kukurydzy.



Ryc. 22

Diagram Venna obrazujący zależności między zbiorami genów różniących się wzorem splicingu po zastosowaniu herbicydu Roundup® (porównanie transkryptomów linii traktowanych herbicydem z transkryptomami roślin kontrolnych).

SL1 - zbiór 423 genów różniących się wzorem splicingu dzień po zabiegu oprysku zidentyfikowany w linii wrażliwej S79757; SL7 - zbiór 336 genów różniących się wzorem splicingu tydzień po zabiegu oprysku zidentyfikowany w linii wrażliwej S79757; TL1 - zbiór 414 genów różniących się wzorem splicingu dzień po zabiegu oprysku zidentyfikowany w linii tolerancyjnej S245; TL7 - zbiór 404 genów różniących się wzorem splicingu tydzień po zabiegu oprysku zidentyfikowany w linii tolerancyjnej S245

5. Wyniki i dyskusja

	SL1C	SL1	SL7C	SL7
Identyfikator genu	GRMZM5G832248 (4 anotowane transkrypty w bazie danych MaizeGDB)			
Liczba transkryptów ulegająca ekspresji	11	11	9	8
Transkrypt o najwyższej ekspresji	GRMZM5G832248_T02	GRMZM5G832248_T01	GRMZM5G832248_T02	SL7.3
Poziom ekspresji genu [FPKM]	24.4168	26.1938	19.9755	22.4602
Identyfikator genu	GRMZM2G366935 (10 anotowanych transkryptów w bazie danych MaizeGDB)			
Liczba transkryptów ulegająca ekspresji	15	16	17	17
Transkrypt o najwyższej ekspresji	GRMZM2G366935_T05	GRMZM2G366935_T09	GRMZM2G366935_T02	GRMZM2G366935_T05
Poziom ekspresji genu [FPKM]	25.8691	19.6482	26.5478	29.5137
Identyfikator genu	GRMZM2G124630 (1 anotowany transkrypt w bazie danych MaizeGDB) GRMZM2G124636 (2 anotowane transkrypt w bazie danych MaizeGDB)			
Liczba transkryptów ulegająca ekspresji	8	14	9	8
Transkrypt o najwyższej ekspresji	GRMZM2G124630_T01	GRMZM2G124636_T01	SL7C.7	SL7.3
Poziom ekspresji genu [FPKM]	25.5068	14.9765	31.9416	26.4446

A

	TL1C	TL1	TL7C	TL7
Identyfikator genu	GRMZM5G832248 (4 anotowane transkrypty w bazie danych MaizeGDB)			
Liczba transkryptów ulegająca ekspresji	11	6	8	11
Transkrypt o najwyższej ekspresji	GRMZM5G832248_T02	TL1.1	TL7C.3	GRMZM5G832248_T02
Poziom ekspresji genu [FPKM]	20.764	18.4148	28.8942	21.3362
Identyfikator genu	GRMZM2G366935 (10 anotowanych transkryptów w bazie danych MaizeGDB)			

B

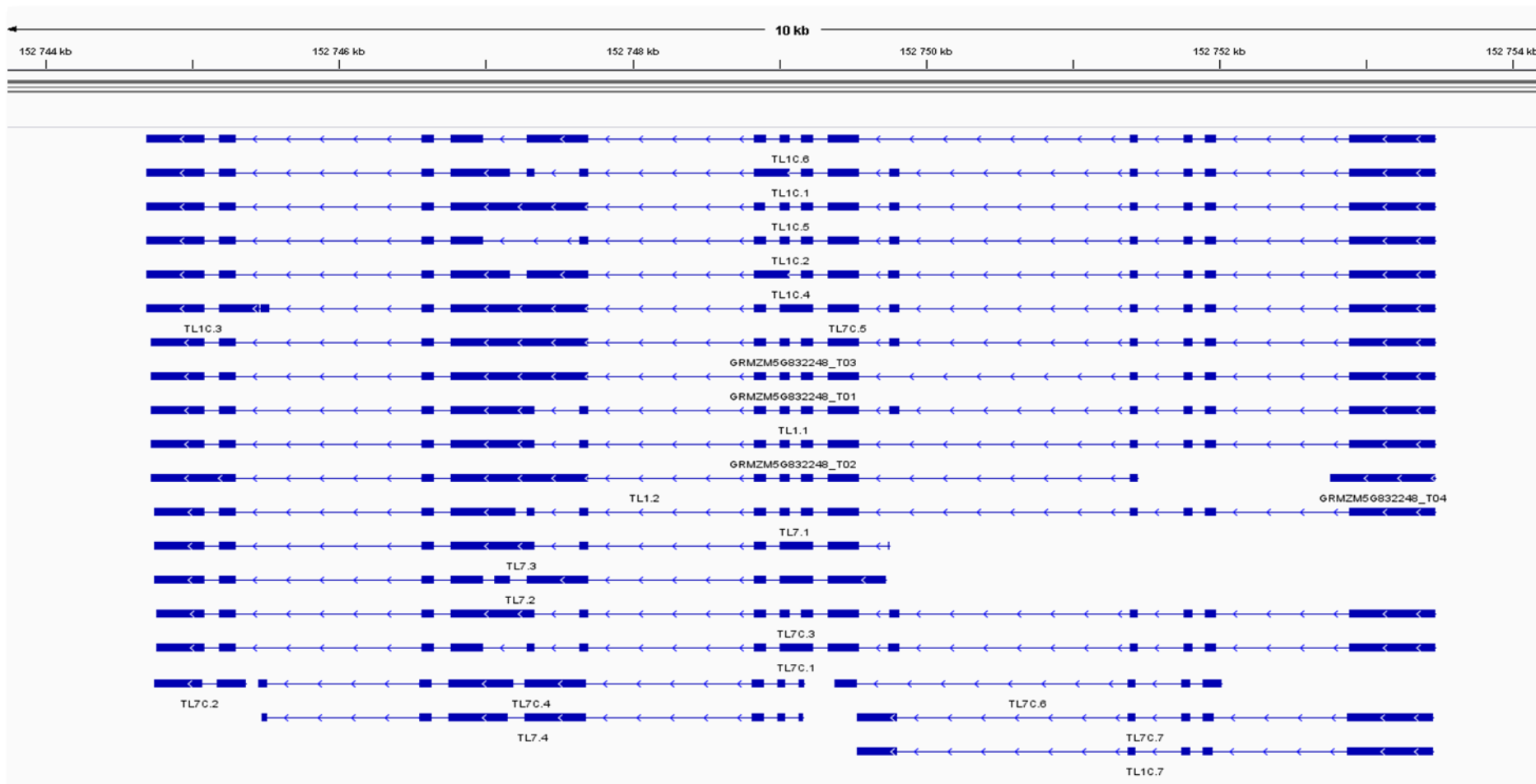
Liczba transkryptów ulegająca ekspresji	13	10	13	16
Transkrypt o najwyższej ekspresji	GRMZM2G366935_T05	GRMZM2G366935_T09	GRMZM2G366935_T05	GRMZM2G366935_T05
Poziom ekspresji genu [FPKM]	21.6864	21.3684	32.4226	21.8758
Identyfikator genu	GRMZM2G124630 (1 anotowany transkrypt w bazie danych MaizeGDB) GRMZM2G124636 (2 anotowane transkrypt w bazie danych MaizeGDB)			
Liczba transkryptów ulegająca ekspresji	12	10	7	11
Transkrypt o najwyższej ekspresji	TL1C.8	TL1.2	TL7C.6	TL7.6
Poziom ekspresji genu [FPKM]	21.1466	14.0882	48.3775	15.4694

Tabela 16

Zestawienie informacji o ekspresji o różnym wzorze splicingu i zróżnicowanej ekspresji poszczególnych izoform transkryptów trzech genów. W części A zebrano informacje dotyczące linii wrażliwej, a w części B linii tolerancyjnej.

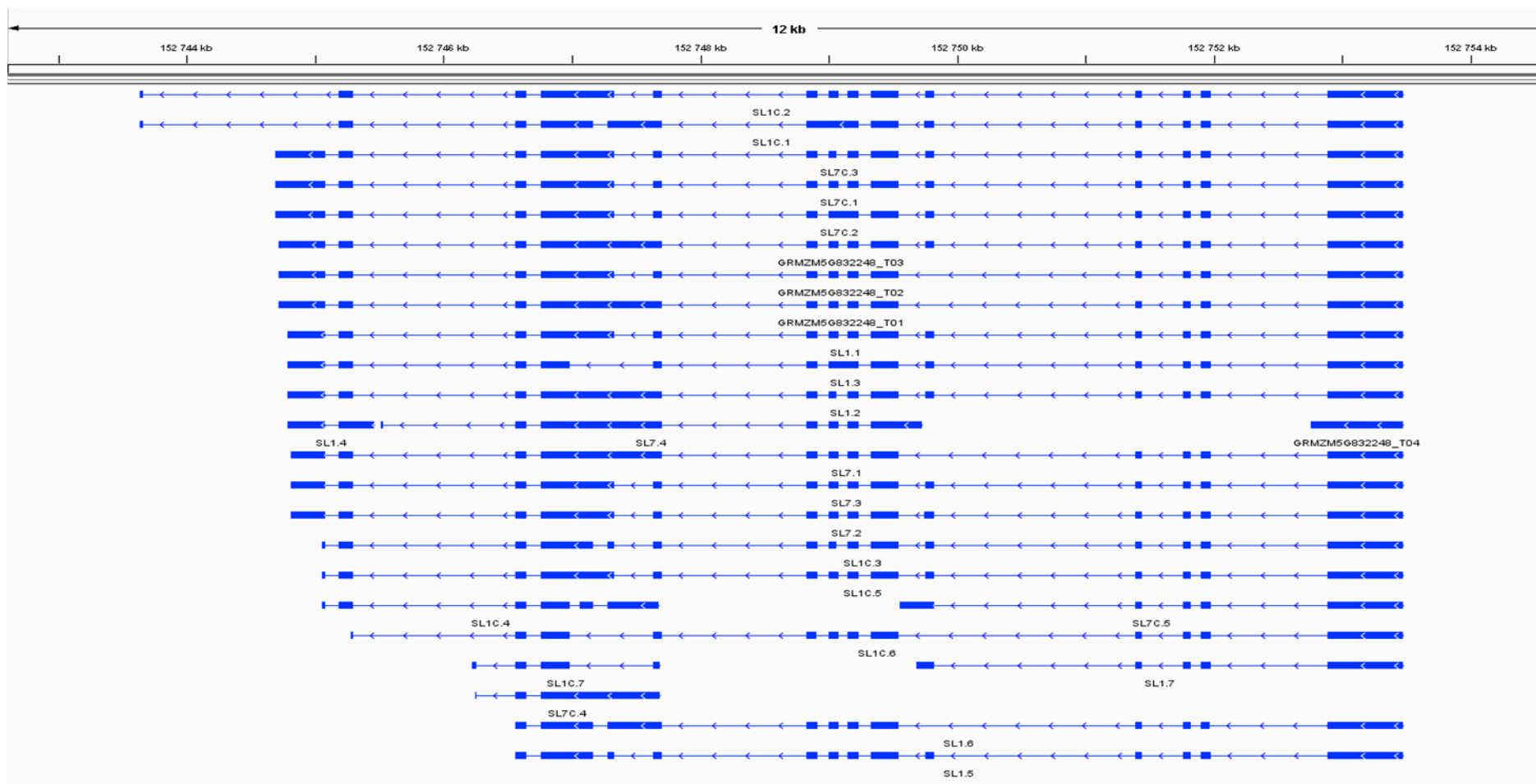
SL1C – linia wrażliwa kontrola po 24h; SL1 – linia wrażliwa traktowana po 24h; SL7C – linia wrażliwa kontrola po 7 dniach; SL7 – linia wrażliwa traktowana po 7 dniach; TL1C – linia tolerancyjna kontrola po 24 h; TL1 – linia tolerancyjna traktowana po 24 h; TL7C – linia tolerancyjna kontrola po 7 dniach; TL7 – linia tolerancyjna traktowana po 7 dniach

Jednym z 3 genów o zróżnicowanym wzorze składania pod wpływem działania herbicydu Roundup® w obu liniach kukurydzy i punktach czasowych był gen zdeponowany w bazie danych MaizeGDB pod nr dostępu GRMZM5G832248. Do tej pory zidentyfikowano 4 transkrypty dla tego genu. Po zastosowaniu analizy bioinformatycznej z danych otrzymanych po sekwencjonowaniu transkryptomu linii wrażliwej udało się dopasować i złożyć 23 dodatkowe transkrypty, natomiast w linii tolerancyjnej powstało 19 nowych izoform transkryptów dla tego genu. Wykorzystano program IGV aby przedstawić schemat organizacji złożonych transkryptów dla linii tolerancyjnej na Ryc. 23 i dla linii wrażliwej na Ryc. 24. Wcześniej opisany w bazie danych transkrypt GRMZM5G832248_T04 nie ulegał ekspresji w żadnej z badanych linii. Transkrypt zdeponowany pod nr dostępu GRMZM5G832248_T02 ulegał najwyższej ekspresji w roślinach kontrolnych (z wyjątkiem linii tolerancyjnej zebranej 7 dni po zabiegu oprysku). W tym transkrypcie w procesie składania mRNA użycie alternatywnych miejsc splicigowych 3' i 5' w obrębie eksonu 4 skutkowało powstaniem dodatkowego eksonu - ten transkrypt składa się z 13 eksonów w przeciwieństwie do 12-eksonowej izoformy GRMZM5G832248_T01. Sekwencja nukleotydów transkryptu GRMZM5G832248_T02 koduje dwie domeny zawierające motyw rozpoznający RNA (RRM, ang. *RNA recognition motif*), które określane są jako domeny wiążące RNA (RBD, ang. *RNA binding domain*) lub domena rybonukleoproteinowa (RNP, ang. *ribonucleoprotein domain*). Na C-końcu znajduje się ponadto 15 aminokwasów, które nie wykazują podobieństwa do żadnej z wcześniej opisanych domen białkowych. Z kolei w obrębie 100 pierwszych aminokwasów kodowanych przez transkrypt GRMZM5G832248_T01 znajduje się dodatkowa domena - hydrofilowa domena związana z 38 czynnikiem splicingowym (ang. *hydrophilic domain associated with splicing factor 38*). Ortologiem tego genu u *A. thaliana* jest czynnik splicingowy (ang. *splicing factor, CC1-like*) zdeponowany pod numerem dostępu AT5G09880.1. Powyższe dane wskazują, że białko kodowane przez gen GRMZM5G832248 może brać udział w procesie splicingu jako czynnik splicingowy.



Ryc. 23

Schemat organizacji izoform transkryptów genu GRMZM5G832248 ulegających ekspresji w linii tolerancyjnej na działanie herbicydu Roundup.

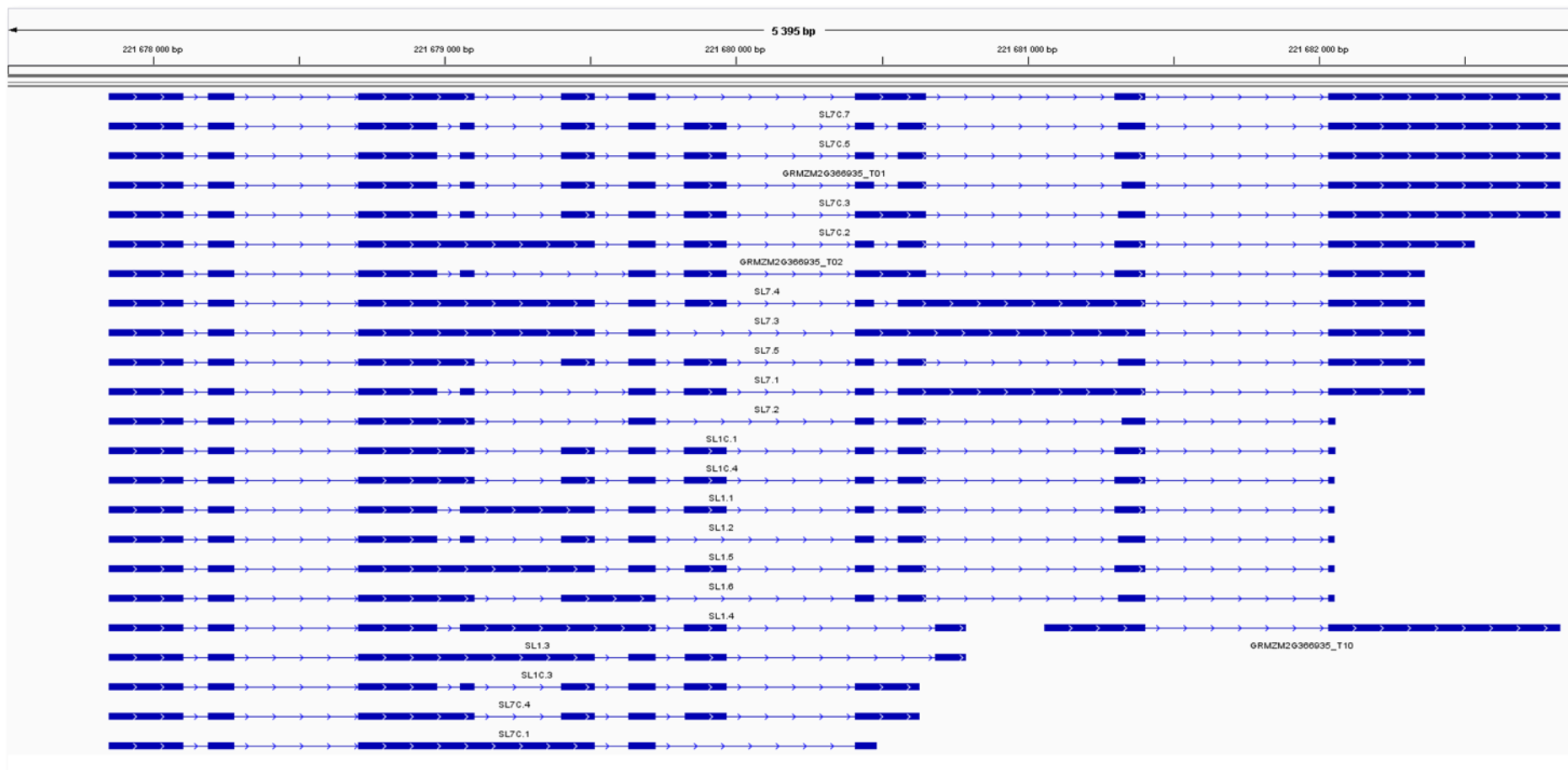


Ryc. 24

Schemat organizacji izoform transkryptów genu GRMZM5G832248 ulegających ekspresji w linii wrażliwej na działanie herbicydu Roundup.

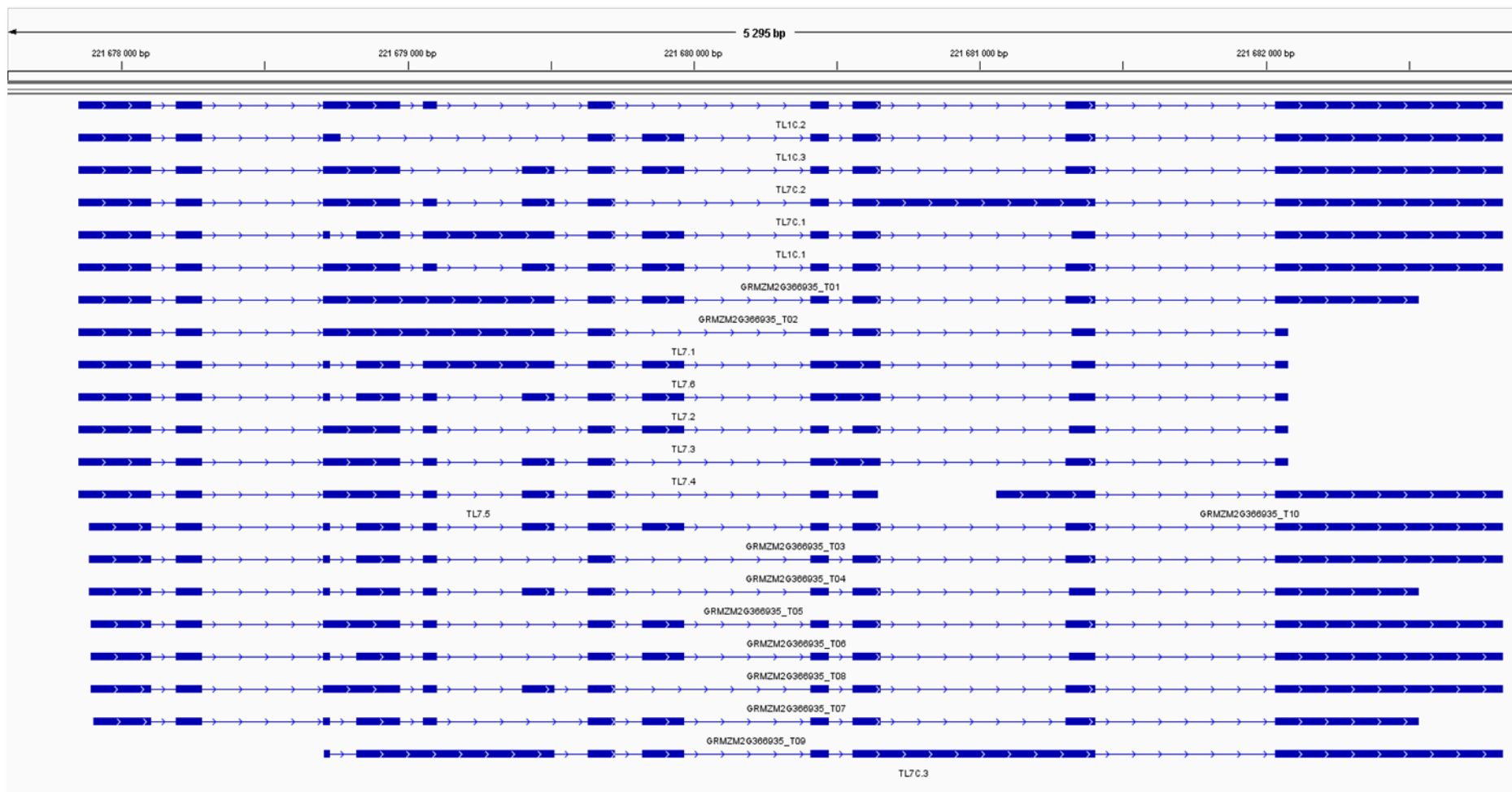
Drugim genem o zróżnicowanym wzorze AS powstającym w wyniku działania herbicydu Roundup®, zidentyfikowanym w czasie opisanej analizy, był gen zdeponowany pod numerem GRMZM2G366935 w bazie danych MaizeGDB. Do tej pory dla tego genu opisano 10 różnych izoform transkryptów, jednak w żadnej z badanych próbek wszystkie 10 nie ulegało transkrypcji. Co więcej w linii wrażliwej S245 dodatkowo pojawiło się 25 nowych izoform transkryptów, a w linii tolerancyjnej prawie o połowę mniej - 12. Schemat organizacji nowych transkryptów w relacji do wcześniej zdeponowanych w bazie danych przedstawiono na Ryc. 25 (linia wrażliwa) i na Ryc. 26 (linia tolerancyjna). Mimo dużej liczby nowych transkryptów (w obu liniach powstało aż 37 różnych izoform), najwyższej ekspresji ulegały transkrypty wcześniej opisane - w pięciu próbkach był to transkrypt GRMZM2G366935_T05, w dwóch transkrypt GRMZM2G366935_T09 i w jednej próbce izoforma GRMZM2G366935_T02 (szczegółowe dane zestawiono w Tabeli 14). Analiza na poziomie sekwencji aminokwasowej przetłumaczonej z sekwencji nukleotydowej transkryptów wykazała, że białko powstające z transkryptu GRMZM2G366935_T09 posiadałoby trzy dodatkowe aminokwasy na końcu C w porównaniu do informacji kodowanej przez izoformę GRMZM2G366935_T02. Z kolei białko kodowane przez transkrypt GRMZM2G366935_T05 jest o 14 aminokwasów krótsze, mimo to transkrypt ten zachowuje informację kodującą domenę kinazy proteinowej, która odpowiada za przeniesienie grupy fosforanowej z ATP na resztę seryny, treoniny lub tyrozyny. Część z nowych transkryptów ulegających ekspresji w badanych próbkach także koduje taką domenę. Z kolei izoformy GRMZM2G366935_T02 i GRMZM2G366935_T09 kodują domenę wykazującą podobieństwo do kinazy proteinowej regulowanej przez cykliny (CDKL, ang. *cyclin-dependent protein kinase like*), które fosforylują swój substrat w wyniku oddziaływaniu z odpowiednią cykliną. Cykliny są grupą białek biorących udział w regulacji cyklu procesów związanych z podziałem komórki, natomiast kinazy CDKL biorą udział także w regulacji transkrypcji i dojrzewania mRNA. Białka wchodzące w skład rodziny kinaz często pełnią rolę odwracalnych przełączników na szlakach przekazywania sygnałów ze środowiska do odpowiedniego białka wewnątrz komórki, odpowiedzialnego za reakcję na dany bodziec. W ten sposób kinazy biorą udział w tak ważnych procesach komórkowych jak proliferacja, różnicowanie, rearanżacje cytoszkieletu czy odpowiedź immunologiczna. Homolog tego genu z *Arabidopsis thaliana* to AME3, a z ryżu to gen AFC - oba kodują białka należące do rodziny kinaz.

5. Wyniki i dyskusja



Ryc. 25

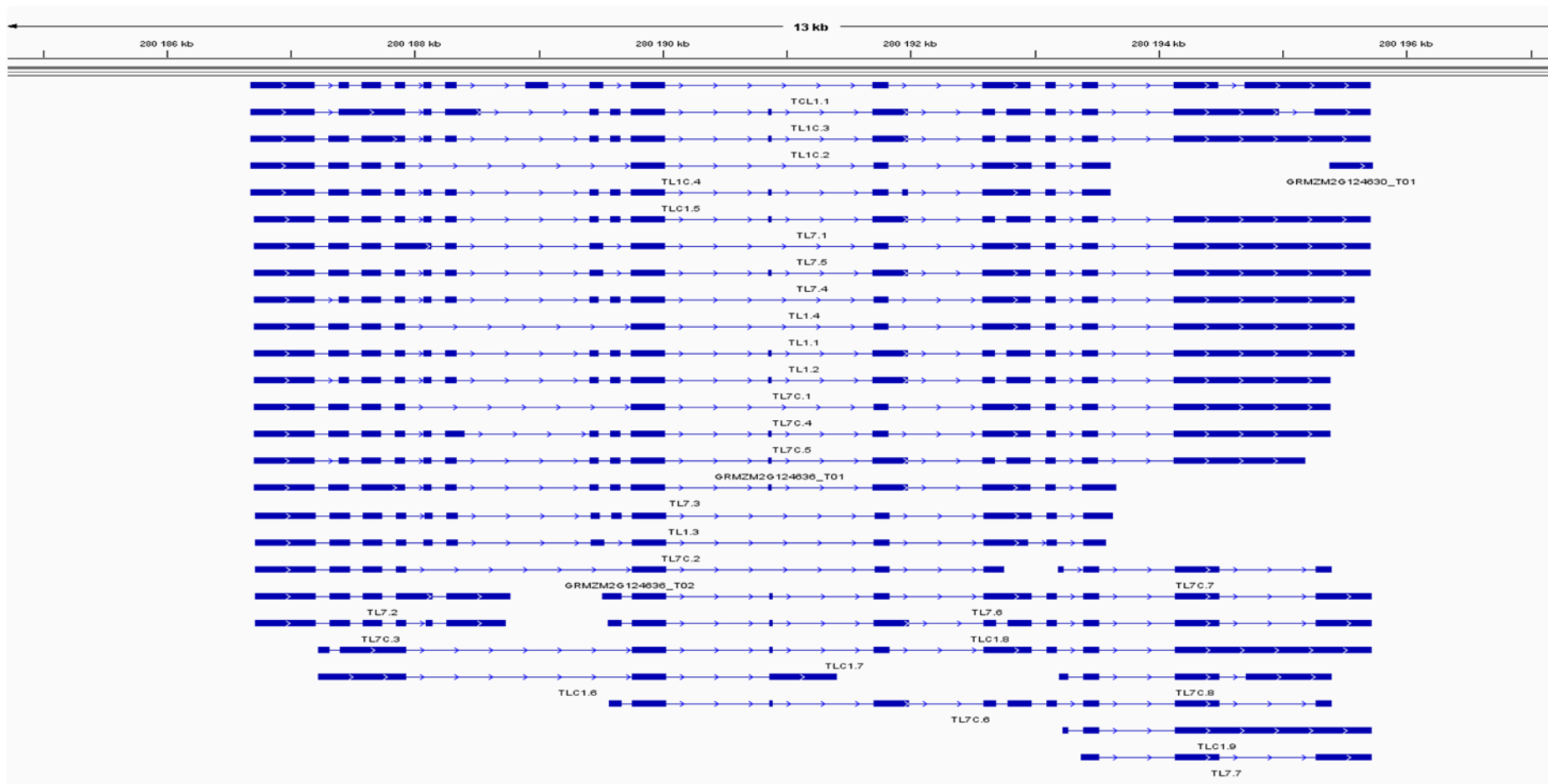
Schemat organizacji izoform transkryptów genu GRMZM2G366935 ulegających ekspresji w linii wrażliwej na działanie herbicydu Roundup.



Ryc. 26

Schemat organizacji izoform transkryptów genu GRMZM2G366935 ulegających ekspresji w linii tolerancyjnej na działanie herbicydu Roundup.

Trzecim genem różniącym się wzorem składania mRNA pod wpływem działania herbicydu Roundup® jest gen, który został zmapowany w obrębie locus zajmowanego przez dwa anotowane geny o numerach dostępu GRMZM2G124630 i GRMZM2G124636. Rekord dla pierwszego z nich (GRMZM2G124630) zawiera opisany 1 transkrypt, natomiast w rekordzie dla drugiego genu (GRMZM2G124636) pojawia się zdarzenie alternatywnego składania i anotowano dla niego 2 transkrypty, jednak dla żadnego z tych genów nie znaleziono homologów u *A. thaliana* czy ryżu. Nie udało się także dla tych genów zidentyfikować kodowanych przez nie domen funkcjonalnych. W badanych liniach ekspresji ulegało dodatkowo 55 transkryptów tego genu (28 w linii tolerancyjnej i 27 w linii wrażliwej; schemat powstałych transkryptów przedstawiono na Ryc. 27 i 28).



Ryc. 27

Schemat organizacji izoform transkryptów mapujących do genu GRMZM2G124630 i genu GRMZM2G124636, które ulegały ekspresji w linii tolerancyjnej na działanie herbicydu Roundup.



Ryc. 28

Schemat organizacji izoform transkryptów mapujących do genu GRMZM2G124630 i genu GRMZM2G124636, które ulegały ekspresji w linii wrażliwej na działanie herbicydu Roundup.

Przeprowadzono także porównanie plików wyjściowych splicing.diff i isoform_exp.diff. Wybrano 22 geny, które różniły się zarówno ekspresją jak i wzorem AS. Wśród nich dzięki zastosowaniu programu MapMan zidentyfikowaliśmy 6 genów biorących udział w metabolizmie białek, po 4 uczestniczące w metabolizmie RNA i lipidów oraz 5 genów związanych z odpowiedzią rośliny na warunki stresu abiotycznego. Dane na temat grupy genów związanych z odpowiedzią roślin na warunki stresu abiotycznego podsumowano w Tabeli 17.

Identyfikator genu	Lokalizacja na chromosomie	Ortolog z <i>A. thaliana</i> lub <i>O. sativa</i>	Prawdopodobna funkcja
GRMZM2G113875	1:19679506-19691660	Nie znaleziono	Element transposonowy
GRMZM2G377539	3:217788488-21780611	<i>A. thaliana</i> : AT5G64940.1 <i>O. sativa</i> : LOC_Os02g36570.1	Białko z rodziny ABC transporterów
GRMZM2G144196	5:646891-663760	<i>A. thaliana</i> : AT2G36990.1 <i>O. sativa</i> : LOC_Os08g14450.1	Czynnik sigma polimerazy RNA
GRMZM2G473511	8:8917987-9633821	<i>A. thaliana</i> : AT1G66920.2 <i>O. sativa</i> : LOC_Os01g02700.1	Kinaza tyrozynowa
GRMZM2G005633	10:127355791-1273718	<i>A. thaliana</i> : AT3G54420.1 <i>O. sativa</i> : LOC_Os04g41680.1	Prekursor endochitynazy B

Tabela 17

Grupa genów biorących udział w odpowiedzi na warunki stresu abiotycznego zidentyfikowana przez program MapMan wśród genów różniących się wzorem splicingu i poziomami ekspresji pomiędzy badanymi liniami.

5.2.4 Dyskusja

Podobnie jak kilkanaście lat temu sekwencjonowanie metodą Sanger'a umożliwiło identyfikację wielu nowych genów, tak teraz wykorzystanie nowych technologii sekwencjonowania pozwala poznać pełen repertuar genów i poszczególnych izoform transkryptów powstających nawet w niewielkich ilościach w różnych tkankach, na różnych etapach rozwoju czy pod wpływem określonych czynników zewnętrznych. Co więcej ponieważ liczba odczytów generowana dla poszczególnych transkryptów jest proporcjonalna

do ilości danego transkryptu w komórce, gęstość odczytów może zostać użyta do określenia poziomu ekspresji transkryptu czy genu z dokładnością równą lub przewyższającą stosowane do tej pory mikromacierze. Generowana w trakcie sekwencjonowania ogromna ilość danych wymaga precyzyjnie dobranych narzędzi analizy, które pozwolą odpowiedzieć na postawione pytania i nadać sens biologiczny otrzymanym wynikom.

Aby ocenić rolę i znaczenie zdarzeń alternatywnego splicingu mRNA pod wpływem warunków stresu herbicydowego u kukurydzy zwyczajnej, przeprowadzono sekwencjonowanie drugiej generacji i analizę różnicową uzyskanych transkryptomów dwóch linii kukurydzy różniących się wrażliwością na stosowany oprysk herbicydem Roundup®. Porównanie uzyskanych sekwencji z transkryptem referencyjnym kukurydzy linii B73 umożliwiło identyfikację dużej liczby nowych genów, które zostały zmapowane nawet do przeszło 11 tys. nowych loci (linia wrażliwa S79757 zebrana dzień po zabiegu oprysku). Naturalne populacje kukurydzy charakteryzują się bardzo wysokim poziomem heterozygotyczności, który przejawia się między innymi niskim stopniem nierównowagi sprzężeń (ang. *linkage disequilibrium*), które mierzy w jakim stopniu allele w dwóch loci są ze sobą sprzężone i determinuje przydatność mapowania asocjacyjnego cech jakościowych w obrębie kolekcji linii hodowlanych, odmian czy rodów. U kukurydzy zwyczajnej nierównowaga sprzężeń wynosi jedynie od kilkuset do kilkutysięcy nukleotydów [86]. Populacja kukurydzy zwyczajnej charakteryzuje się także dużym polimorfizmem sekwencji nawet blisko spokrewnionych odmian. Polimorfizmy pojedynczych nukleotydów występują z częstością od 10 do 20 razy większą niż w populacji ludzkiej (u człowieka polimorfizm typu SNP średnio występuje co tysiąc nukleotydów, u kukurydzy w zależności od porównywanych linii średnio co 44 do 100 nukleotydów). Ponadto duplikacje całego genomu jak i dużych jego regionów przyczyniają się do powstawania wielu bardzo podobnych sekwencyjnie genów, które są często paralogami [87]. Zlokalizowanie wielu nowych transkryptów w badanych liniach w porównaniu do linii referencyjnej B73, potwierdza wcześniejsze doniesienia o znaczących różnicach genetycznych między liniami i odmianami kukurydzy.

Analiza liczby powstających transkryptów u roślin poddanych opryskowi i roślin kontrolnych pozwoliła stwierdzić, że po zastosowaniu herbicydu Roundup® nastąpił spadek aktywności transkrypcyjnej – tydzień po przeprowadzeniu zabiegu oprysku zmniejszyła się

liczba powstających mRNA w obu badanych liniach. W linii tolerancyjnej S245 nastąpiła znacznie większa redukcja liczby powstających transkryptów – dzień po oprysku herbicydem Roundup® powstało aż o 36 tys. transkryptów mniej. Zmniejszenie aktywności transkrypcyjnej pod wpływem warunków stresowych może być związane z reorganizacją metabolizmu komórki - zatrzymaniem reakcji i szlaków metabolicznych, które nie są istotne dla przetrwania rośliny i aktywacją genów kodujących białka biorące udział w odpowiedzi na warunki stresowe. Spadek ogólnej liczby powstających transkryptów przyczynia się także do zmniejszenia zapotrzebowania energetycznego rośliny. Może być w ten sposób sprzężony z lepszym dostosowaniem linii tolerancyjnej, u której liczba powstających transkryptów obniżyła się bardziej znacząco po oprysku herbicydem (zarówno dzień jak i tydzień po zabiegu).

Wyznaczenie poziomu ekspresji genów w badanych transkryptomach kukurydzy umożliwiło zidentyfikowanie 599 genów o różniących się poziomie ekspresji między linią wrażliwą i tolerancyjną na herbicyd Roundup®. Większe różnice w poziomie transkrypcji genów zaobserwowano jednak w obrębie badanych linii po zastosowaniu herbicydu, porównując sekwencje transkryptomów roślin poddanych opryskowi i roślin kontrolnych. W przypadku roślin z linii tolerancyjnej odnotowano aż 843 geny o różnym poziomie ekspresji, w tym 61 genów różniących się ekspresją w obu rozpatrywanych punktach czasowych. W linii wrażliwej genów o zróżnicowanym poziomie ekspresji było znacznie mniej - jedynie 71, z czego zaledwie 6 miało podwyższoną ekspresję dzień i tydzień po zabiegu oprysku. Mniejsza liczba genów o zmienionym poziomie ekspresji może, podobnie jak opisana wcześniej mniej znacząco zmieniona aktywność transkrypcyjna, warunkować większą wrażliwość na działanie stresu herbicydowego roślin z linii S79757 w porównaniu do roślin z linii tolerancyjnej S245. W trakcie badania różnic w ekspresji genów nie udało się zidentyfikować żadnego genu, którego podwyższona lub obniżona ekspresja mogłaby być odpowiedzialna za uniwersalną odpowiedź na oprysk herbicydem - pule genów o zmienionej ekspresji dla obu linii są rozbieżne. Udało się jednak wskazać geny należące do rodziny ABC transporterów, które miały zmienioną ekspresję w obu badanych liniach po zastosowaniu oprysku herbicydem Roundup® - w linii wrażliwej był to gen GRMZM5G833207/ GRMZM5G808836, a

w linii tolerancyjnej gen GRMZM2G477872. Także wśród genów o zróżnicowanym wzorze splicingu znalazł się przedstawiciel tej rodziny białek - gen GRMZM2G377539 (Tabela 17).

ABC transportery są uważane za jedną z najliczniejszych rodzin białek, które występują we wszystkich żywych organizmach - zostały zidentyfikowane zarówno u bakterii, bezkręgowców, zwierząt jak i roślin. Wszystkie białka z tej rodziny posiadają co najmniej jedną wysoce konserwatywną domenę ATPazową, która zapewnia dostęp energii pochodzącej z hydrolizy ATP. Początkowo znane były ze swojej roli w importowaniu metabolitów pośrednich oraz eksportowaniu leków i produktów przemiany materii. Białka z tej rodziny pełniące funkcje transportowe są zakotwiczone w membranach komórkowych dzięki jednej lub dwóm domenom transmembranowym. Część transporterów ABC jest rozpuszczalna i znajduje się w cytoplazmie lub plastydach, gdzie pełni funkcje związane z regulacją ekspresji genów, funkcjonowaniem rybosomów, kanałów jonowych czy wiązaniem antygenów [88]. Zidentyfikowane geny, ulegające podwyższonej ekspresji w badanych liniach kukurydzy po oprysku herbicydem Roundup®, kodują ABC transportery posiadające domenę transmembranową. Wcześniejsze badania przeprowadzone u kukurydzy zwyczajnej i ryżu wykazały, że pod wpływem szerokiej gamy substancji o właściwościach fitotoksycznych, takich jak atarazyna, metolachlor, primisulfuron czy kwas 2,4-dichlorofenoksyoctowy (2,4-D)[89], a także w różnych warunkach stresowych (np. obecność metali ciężkich i niedotlenienie [90], zastosowanie kwasów organicznych i nieorganicznych [91]), poziom ekspresji genów kodujących transportery ABC ulegał podwyższeniu. Postuluje się, że białka z tej rodziny biorą udział w detoksyfikacji herbicydów poprzez wiązanie z produktami ich przemiany i rozpadu. Produkty metabolizmu herbicydów są często skoniugowane z cząsteczką glutationu, glukozy lub malonianu. W takiej postaci są rozpuszczalne w wodzie i mogą być transportowane przez znajdujące się w błonie tonoplastu białka ABC do wakuoli, gdzie zachodzą dalsze etapy ich rozkładu. Wykazano także, że transportery ABC znajdujące się w błonie komórkowej mogą brać udział w usuwaniu produktów rozkładu herbicydów związanych z cząsteczką glutationu [92].

Wykorzystanie programu MapMan i przyporządkowanie do grup funkcjonalnych transkryptów genów o największej różnicy w poziomie ekspresji pozwoliło zidentyfikować grupę 23 genów, które u roślin biorą udział w odpowiedzi na warunki stresowe. Zgodnie z danymi przedstawionymi na Ryc. 21 zastosowany oprysk herbicydem Roundup® zmienił ekspresję transkryptów związanych z syntezą ściany komórkowej, proteolizą, produkcją etylenu, przekazywaniem sygnału czy z odpowiedzią na czynniki abiotyczne. Ekspresja większości genów związanych z produkcją metabolitów wtórnych została zmniejszona, co może być związane ze zmianą metabolizmu komórki pod wpływem czynników stresowych. Wcześniejsze badania wykazały, że w zależności od pojawiającego się w środowisku czynnika zakłócającego wzrost i rozwój rośliny produkcja metabolitów wtórnych jest modyfikowana. Pod wpływem różnych czynników stresowych specyficzne metabolity wtórne są gromadzone, a z kolei produkcja innych zostaje zahamowana [93]. Na przykład stres zasolenia powoduje szereg zmian w stężeniu poliamin. Stwierdzono akumulację spermidyny i sperminy, a spadek ilości produkowanej putrescyny w korzeniach słonecznika, oliwki czy pszenicy pod wpływem wzrastającego zasolenia gleby [94]. Poliaminy są to dodatnio naładowane związki organiczne o małej masie molowej, które występują powszechnie we wszystkich organizmach żywych i są zaangażowane w procesy związane z dojrzewaniem i starzeniem, regulacją tempa wzrostu czy programowaną śmiercią komórki. Są także związane z odpowiedzią na warunki stresu abiotycznego, jednak ich dokładna funkcja jest nadal trudna do uchwycenia [95]. Zaobserwowany spadek ekspresji większości genów, których produkty białkowe zaangażowane są w syntezę metabolitów wtórnych, może być związany z charakterystyką odpowiedzi badanych linii kukurydzy na wprowadzenie do uprawy glifosatu. Wzrost poziomu transkrypcji nastąpił jedynie dla genu zaangażowanego w syntezę fenoli i fenylopropanoidów (Ryc. 20 i 21). Jest to największa grupa metabolitów wtórnych produkowanych przez rośliny głównie w odpowiedzi na czynniki takie jak promieniowanie UV, zranienia, infekcje, ekspozycja na ozon czy inne zanieczyszczenia. Uważa się, że podstawą działania ochronnego fenoli i fenylopropanoidów w warunkach stresowych są ich właściwości przeciwutleniające, związane z wychwytywaniem cząsteczek wolnych rodników przez te substancje [96]. Badania przeprowadzone na ryżu [97] oraz na zielenicach z gatunku *Chlorella kessleri* [98] wykazały, że toksyczne działanie glifosatu zawartego w herbicydzie Roundup® związane jest także z wywołaniem stresu oksydacyjnego.

Jego zastosowanie powoduje wzrost poziomu reaktywnych form tlenu (ROS, ang. *reactive oxygen species*), które powodują peroksydację i niszczenie lipidów, a także kwasów nukleinowych. Podwyższona ekspresja genów związanych z syntezą fenoli i fenylopropanoidów może być więc związana ze zwiększonym zapotrzebowaniem roślin poddanych opryskowi na związki o właściwościach antyoksydacyjnych.

W celu określenia liczby genów ulegających procesowi alternatywnego składania mRNA oraz rozkładu poszczególnych typów zdarzeń splicingowych w wybranych liniach kukurydzy pod wpływem zastosowanych warunków stresowych został stworzony skrypt w języku java (rozdział 5.2.3). Po przebadaniu uzyskanych po sekwencjonowaniu transkryptomów określono, że AS podlegało od przeszło 31% do prawie 35% genów, które ulegały ekspresji w poszczególnych liniach i punktach czasowych. Po przeprowadzeniu zabiegu oprysku herbicydem Roundup® w obu badanych liniach następował nieznaczny wzrost liczby genów ulegających alternatywnemu splicingowi w porównaniu do roślin kontrolnych niepoddanych opryskowi (około 1% genów więcej ulegało AS). Z przedstawionych w Tabeli 15 danych wynika także, że średni poziom AS był wyższy w linii tolerancyjnej S245 niż w linii wrażliwej S79757. Z kolei średni poziom zdarzeń alternatywnego składania mRNA w badanych liniach wynosi 32% i jest niewiele niższy niż określony poprzednio w badaniu dynamiki transkryptomu w obrębie różnych stref rozwojowych liści kukurydzy, gdzie wynosił około 37% [99]. Może być to związane zarówno z charakterystyką linii kukurydzy wybranych do badań (w [99] stosowano jako materiał do badań liście kukurydzy linii wsobnej B73), jak i różnymi warunkami hodowli, które także mogą mieć wpływ na proces splicingu.

Natomiast rozkład typów zdarzeń AS określony dla linii różniących się wrażliwością na glifosat, odbiega od wcześniej opisanych w literaturze (rozdział 2.1.2) [23]. Postuluje się, że najczęstszym typem AS u roślin jest zatrzymanie intronu. Jednak w badanych próbkach ten typ splicingu odpowiadał jedynie za 14 do 18% zdarzeń AS, a najczęściej dochodziło do wyboru alternatywnych miejsc splicingowych 3' i 5' (20 do 35% ogółu zdarzeń AS). Prawie równie często następowało włączenie dodatkowego eksonu (7 do 26%), a najrzadziej dochodziło do pominięcia eksonu (5 do 18%). Dla roślin kontrolnych z linii tolerancyjnej zebranej tydzień po zabiegu oprysku uzyskany rozkład typów splicingu zupełnie odbiega od pozostałych linii, co

utrudnia analizę i dostrzeżenie istotnych różnic w strukturze zdarzeń alternatywnego składania, jednak można zauważyć 5% wzrost ilości zdarzeń splicingowych związanych z włączaniem dodatkowego eksonu do transkryptu. Zaobserwowana rozbieżność dotycząca ogólnego rozkładu typów splicingu w badanych liniach może wynikać z charakterystyki metod sekwencjonowania drugiej generacji, które pozwalają wykryć transkrypty występujące nawet w niewielkich ilościach w komórce i nie wymagają przy tym wiedzy na temat istniejących w danej sekwencji połączeń intron/ekson (tak jak to było w przypadku macierzy dachówkowych wcześniej stosowanych do badania zdarzeń AS), dzięki czemu możliwe jest określenie dużej liczby nieopisanych wcześniej przypadków splicingu. Opisana różnica może być także związana z zastosowaniem skryptu do określania typów splicingu - aby potwierdzić otrzymane wyniki można by dodatkowo zaimplementować inne programy umożliwiające analizę tego typu, np. programu ASTALAVISTA [100]. Jednak badanie transkryptomu *A. thaliana* opisane w [18] sugeruje, że zdarzenia AS u roślin związane z zatrzymaniem intronu były dotychczas zawyżone i w rzeczywistości mogą dotyczyć mniejszej liczby transkryptów, dlatego tym bardziej uzasadniana wydaje się konieczność potwierdzenia uzyskanych wyników przy użyciu innego programu bioinformatycznego.

Kolejnym krokiem analizy wyników sekwencjonowania było porównanie zbiorów genów różniących się wzorem splicingu między roślinami poddanymi opryskowi a roślinami kontrolnymi. Pozwoliło ono zidentyfikować pule genów, których wzór alternatywnego składania był zmieniony po zastosowaniu oprysku herbicydem Roundup® w obu badanych punktach czasowych. W linii tolerancyjnej takich genów było 39, natomiast w linii wrażliwej 58. Ponadto okazało się, że 3 geny posiadają zmieniony wzór AS w obu badanych liniach zarówno dzień jak i tydzień po zabiegu oprysku herbicydem (rozdział 5.2.3, Tabela 15, Ryc. 17-22).

Jednym z tych genów jest gen zdeponowany w bazie danych MaizeGDB pod numerem dostępu GRMZM5G832248. Badanie transkryptomów wybranych linii kukurydzy pozwoliło poszerzyć wiedzę na temat transkryptów powstających z tego genu - poza 4 transkryptami anotowanymi dotychczas udało się zidentyfikować 23 dodatkowe izoformy w linii wrażliwej i 19 w linii tolerancyjnej. Gen ten koduje informacje o białku, które posiada domeny wiążące RNA oraz hydrofilową domenę związaną z 38 czynnikiem splicingowym.

Ponieważ ortologiem tego genu z *A. thaliana* jest czynnik splicingowy, wydaje się prawdopodobne, że produkt białkowy genu GRMZM5G832248 pełni podobną funkcję u kukurydzy zwyczajnej. Transkrypty kodujące czynniki splicingowe często podlegają zmienionemu składaniu pod wpływem warunków stresowych, jak np. produkty genu *At-SR30*, który koduje jedno z białek SR u *A. thaliana*. W warunkach stresowych związanych z nadmiernym oświetleniem i podwyższoną temperaturą wzrasta udział pełnej długości transkryptu, zawierającego całą ramkę odczytu, natomiast ilość transkryptu tego genu z przedwczesnym sygnał terminacji translacji (PTC) maleje [101]. Podobna sytuacja dotyczy transkryptów genu kodującego białko SR45a - w czasie stresu ciepła i suszy wzrasta ilość transkryptu pełnej długości, w porównaniu do innych izoform [102]. Ponadto ekspresja różnych izoform transkryptów genów kodujących czynniki splicingowe pod wpływem warunków stresowych zmienia lokalizację oraz determinuje ich fosforylację [103, 104]. Biorąc pod uwagę rolę czynników splicingowych w procesie rozpoznawania miejsc splicingowych oraz rekrutowania pozostałych białek biorących udział w procesie składania transkryptów, nawet niewielkie różnice w ilości, lokalizacji i potranslacyjnej regulacji aktywności (przede wszystkim fosforylacji) czynników splicingowych powstające pod wpływem warunków stresowych, mogą mieć znaczący wpływ na ekspresję genów docelowo regulowanych przez te czynniki. Zrozumienie regulacji aktywności czynników splicingowych u roślin w odpowiedzi na warunki stresowe jest jednym z ważniejszych elementów sieci wzajemnych zależności opisujących zarówno regulację samego procesu splicingu jak i regulację ekspresji genów, które temu procesowi ulegają [105].

Pod wpływem warunków stresu herbicydowego u kukurydzy powstaje wiele nowych izoform transkryptu genu GRMZM5G832248 kodującego najprawdopodobniej jeden z czynników splicingowych, jednak aby poznać dokładną funkcję powstałych transkryptów konieczne by były dodatkowe badania i eksperymenty związane np. z wyłączeniem lub wyciszeniem ekspresji tego genu. Podobne badania należało by przeprowadzić dla genu, którego transkrypty różniły się wzorem splicingu w badanych liniach po zastosowaniu oprysku herbicydem Roundup® i zostały zmapowane do dwóch genów opisanych w bazie danych MaizeGDB - genu GRMZM2G124630 i GRMZM2G124636. W trakcie przeprowadzonej analizy nie udało się znaleźć sekwencji homologicznych dla tego genu, a

przetłumaczona sekwencja aminokwasowa nie koduje znanych domen białkowych. Jednak duża liczba powstających w warunkach stresu herbicydowego nowych transkryptów (niektórych nawet dłuższych niż annotowane w bazie danych) zachęca do podjęcia próby określenia ich funkcji i roli w procesach komórkowych.

Trzecim z genów zróżnicowanym pod względem składania pod wpływem oprysku herbicydem jest gen GRMZM2G366935. Także dla tego genu zidentyfikowano nowe, wcześniej nieannotowane izoformy transkryptów, jednak we wszystkich badanych warunkach najwyższej ekspresji ulegały transkrypty już wcześniej opisane. Zachowują one informację o kodowanej przez ten gen domenie kinazy proteinowej, która może brać udział w fosforylacji białek docelowych i przez to regulacji ich aktywności w odpowiedzi na zastosowany oprysk herbicydem Roundup®. Również analiza z zastosowaniem programu MapMan pozwoliła zidentyfikować wśród transkryptów o zróżnicowanym sposobie składania oraz ekspresji gen GRMZM2G473511 kodujący kinazę tyrozynową.

Do tej pory jedną z lepiej poznanych i opisanych sieci przekazywania sygnału w warunkach stresu jest działanie kinaz aktywowanych mitogenem (MAPK, ang. *mitogen activated protein kinase*). Są one obecne we wszystkich komórkach eukariotycznych i regulują procesy związane z proliferacją, różnicowaniem, wzrostem i odpowiedzią na warunki stresowe. Zorganizowane są w moduły składające się z kinaz kinaz MAPK (MAPKKK, MAP3K, MEKK, ang. *MAPK kinase kinase*), kinaz MAPK (MAPKK, MAP2K, MEK, ang. *MAPK kinase*) i z białek MAPK. Postuluje się, że utrzymywane są przez szereg białek tworzących rusztowanie, które kotwiczy cały moduł w odpowiednim kompartmentcie komórkowym [106]. Udowodniono, że ścieżka sygnałowa związana z kinazami MAPK jest aktywowana przez szereg czynników abiotycznych takich jak stres zimna, suszy czy zranienia - aktywują one białka MAPK4 i MAPK6 u *A. thaliana* [107], a z kolei wysokie stężenia miedzi i kadmu wpływają na ekspresje genów kodujących białka MAPK u ryżu (*OsMPKA3* i *OsMPK6*) [108], a także na działanie czterech różnych kinaz (SIMK, MMK2, MMK3, SAMK) u lucerny siewnej [109]. Wiadomo ponadto, że kinazy MAPK są zaangażowane w odpowiedź na wyższe stężenie reaktywnych form tlenu w komórkach *Arabidopsis* [110]. Innym przykładem odpowiedzi na warunki stresowe poprzez aktywację kinaz proteinowych jest białko OST1 (ang. *open stomata 1*). Stwierdzono, że homolog tego białka z kapusty (BoOST1) jest

niezbędnym elementem dla przeżycia rośliny w warunkach stresu suszy i podwyższonego zasolenia oraz że oddziałuje z białkami związanymi z siecią przekazywania sygnałów zależną od kwasu abscysynowego [111].

Przedstawione dane sugerują, iż obecność substancji fitotoksycznej, jaką jest glifosat, może wpływać bezpośrednio lub pośrednio (np. poprzez zwiększone stężenie ROS) na przekazywanie sygnałów w komórce poprzez zmianę profilu ekspresji poszczególnych izoform kinaz proteinowych. Izoformy transkryptów ulegające najwyższej ekspresji w badanych warunkach posiadają zachowaną domenę funkcjonalną, jednak translacja białka z innej izoformy transkryptu może ingerować w lokalizację czy specyficzność oddziaływań powstałego enzymu z cząsteczkami docelowymi. Biorąc pod uwagę wielopoziomowość zależności w ścieżkach regulowanych przez kinazy proteinowe można przypuszczać, że nawet niewielka zmiana profilu ekspresji może przyczynić się do lepszej lub gorszej zdolności odpowiedzi badanych linii kukurydzy na stres herbicydowy.

Przeprowadzone sekwencjonowanie drugiej generacji oraz analiza bioinformatyczna uzyskanych wyników pozwoliła stwierdzić znaczny udział procesu AS w ekspresji genów kukurydzy zwyczajnej. Proces ten dotyczył przeszło 1/3 genów ulegających ekspresji w badanych liniach. Stwierdzono także, że pod wpływem zastosowanego oprysku herbicydem poziom AS wzrastał i zmieniał się udział poszczególnych typów zdarzeń AS między badanymi liniami. Wykazano ponadto, że proces alternatywnego składania dotyczył wielu genów, których udział w odpowiedzi na warunki stresowe był już potwierdzony wcześniejszymi badaniami. Czułość technologii sekwencjonowania Illumina umożliwiła wykrycie zarówno wielu nowych izoform transkryptów, jak i genów wcześniej nie opisanych w genomie referencyjnym kukurydzy zwyczajnej. Analiza zmian poziomu ekspresji genów między porównywanymi liniami pozwoliła wytypować 67 genów, których ekspresja ulegała zmianie po zastosowaniu herbicydu Roundup®. Zrealizowane badania potwierdzają obserwacje, że szeroko stosowane opryski herbicydami powodują liczne zmiany w metabolizmie roślin, nawet tolerancyjnych na składnik aktywny wybranego herbicydu. Wskazane w opisanej analizie geny i grupy białek mogą być obiektem dalszych badań funkcjonalnych jak i eksperymentów związanych z doskonaleniem nowych odmian hodowlanych kukurydzy, a być może także innych zbóż, które będą się charakteryzować lepszym dostosowaniem do

warunków stresowych. Wydaje się to szczególnie istotne w dobie zmian klimatycznych i coraz większego antropogenicznego zanieczyszczenia środowiska. Tradycyjne badania genetyczne nad poprawą plonowania i jakości upraw osiągnęły etap, na którym bardzo trudno o kolejne ulepszenia odmian. Jednak jak wynika z przedstawionego schematu analizy, badania genomów i transkryptomów roślin uprawnych z uwzględnieniem badania ekspresji genów i procesu splicingu z zastosowaniem nowoczesnych metod sekwencjonowania, mogą dostarczyć nowych spostrzeżeń i wiedzy na temat regulacji ekspresji genów i jej związku ze zwiększonym plonowaniem lub odpornością na czynniki stresowe.

6. Podsumowanie

Celem przeprowadzonych prac było określenie znaczenia procesu alternatywnego składowania transkryptów w odpowiedzi kukurydzy zwyczajnej na stres abiotyczny jakim jest oprysk herbicydem. Dzięki podjętym badaniom możliwe było poszerzenie zrozumienia podstaw molekularnych procesu odpowiedzi roślin na wprowadzenie do uprawy herbicydu Roundup®.

Najważniejsze wyniki i wnioski z przeprowadzonych analiz przedstawiono poniżej:

- zidentyfikowano nowe zdarzenia AS dla 42 genów pochodzących z chromosomu 1 kukurydzy zwyczajnej
- potwierdzono ekspresję pięciu przewidzianych transkryptów dla trzech wybranych genów o znanej funkcji
- obecne programy służące do przewidywania struktur genów są niewystarczającym narzędziem do weryfikacji zdarzeń AS
- sekwencjonowanie drugiej generacji cząsteczek mRNA pozwoliło zidentyfikować od 166 do 192 tys. transkryptów, w tym od 3,5 do 27 tys. nowych eksonów w badanych liniach kukurydzy
- pod wpływem działania herbicydu następuje spadek aktywności transkrypcyjnej w obu badanych liniach
- wytypowano 46 genów, które po oprysku wykazywały zróżnicowany wzór AS i poziom ekspresji (wśród nich 5 genów związanych z odpowiedzią na warunki stresu abiotycznego)
- analiza funkcjonalna wykazała, że geny te uczestniczą w odpowiedzi na warunki stresowe, fotosyntezie, metabolizmie wtórnym i metabolizmie białek
- alternatywny splicing dotyczy od 31 do 35% genów ulegających ekspresji u kukurydzy zwyczajnej pod wpływem warunków stresu herbicydowego
- najczęstszym typem AS był wybór różnych miejsc splicingowych zarówno 3' jak i 5'
- w linii tolerancyjnej po oprysku herbicydem udział zdarzeń AS związanych z włączeniem dodatkowego eksonu wzrasta

- wytypowano trzy geny, których profil AS zmienia się w obu badanych liniach i które są potencjalnie związane z różną wrażliwością badanych linii na herbicyd Roundup®.

7. Bibliografia

1. Bruce A., Bray D., Hopkin K. et al., *Essential Cell Biology*, Garland Science, 2010
2. Berg J.M., Tymoczko J.L., Stryer L., *Biochemia*, Wydawnictwo Naukowe PWN, Warszawa 2005, 798-99
3. Proudfoot N., *New perspectives on connecting messenger RNA 3' end formation to transcription*, *Curr Opin Cell Biol*, 2004, 16: 272-78
4. Li Q., Hunt A.G., *The polyadenylation of RNA in Plants*, *Plant Physiol.*, 1997, 115: 321-25
5. Loke J.C., Stahlberg E.A., Strenski D.G. et al., *Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures*, *Plant Physiol.*, 2005, 138: 1457-68
6. Proudfoot N.J., *Ending the message: poly(A) signals then and now*, *Genes Dev.*, 2011, 25: 1170-82
7. Chateigner-Boutin A., Small I., *Plant RNA editing*, *RNA Biology*, 2010, 7: 213-19
8. Lorkovic Z.J., Wieczorek Kirk D.A., Lambermon M.H.L., Filipowicz W., *Pre-mRNA splicing in higher plants*, *Trends in Plant Science*, 2000, 5: 160-67
9. Bruce A., Johnson A., Lewis J. Raff M. et al., *Molecular Biology of the Cell*, Gerland Science, 2008
10. Reddy A.S.N., *Alternative Splicing of Pre - Messenger RNAs in Plants in the Genomic Era*, *Annual Review of Plant Biology*, 2007, 58: 267-94
11. Turner, P.C., McLennan, A.G., Bates, A.D., White, M.R.H., *Krótkie wykłady. Biologia molekularna*, Wydawnictwo Naukowe PWN, Warszawa, 2004
12. Zhiguo E., Wang L., Zhou J., *Splicing and alternative splicing in rice and humans*, *BMB Rep.*, 2013, 46: 439-47
13. Severing E.I., van Dijk A.D.J., Stiekema W., van Ham R.C.H.J., *Comparative analysis indicates that alternative splicing in plants has a limited role in functional expansion of the proteome*, *BMC Genomics*, 2009, 10: 154
14. Ko C.H., Brendel V., Taylor R.D., Walbot V., *U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize*, *Plant Mol Biol.*, 1998, 36: 573-583
15. Reddy A.S.N., *Plant serine /arginine – rich proteins and their role in pre – mRNA splicing*, *Trends in Plant Sci.*, 2004, 9: 541-47
16. Gamazon E.R., Stranger B., *Genomics of alternative splicing: evolution, development and pathophysiology*, *Human Genet.*, 2014, 6: 679-87
17. Pan Q., Shai Q., Lee L.J., Frey B.J., Blencowe B.J., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing*, *Nat. Genet.*, 2008, 40: 1413-15
18. Marquez Y., Brown J.W.S., Simpson C., Barta A., Kalyna M., *Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis*, *Genome Res.*, 2012, 22: 1184-95
19. Braunschweig U., Gueroussov S., Plocik A.M., Graveley B.R., Blencowe B.J., *Dynamic integration of splicing within gene regulatory pathways*, *Cell*, 2013, 152: 1252-69
20. Kervestin S., Jacobson A., *NMD: A multifaceted response to premature translation termination*, *Nat. Rev. Mol. Cell Biol.*, 2012, 13: 700-712

21. Keren H., Lev-Maor G., Ast G., *Alternative splicing and evolution: diversification, exon definition and function*, Nat. Rev. Gen., 2010, 5: 345-55
22. Kim E., Goren A., Ast G., *Alternative splicing: current perspectives*, Bioessays, 2008, 30: 38-47
23. Kalyna M., Simpson C.G., Syed N.H., Lewandowska D., Maraquez Y. et al., *Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis*, Nucleic Acids Res., 2012, 6: 2454-69
24. Boothby T.C., Ziper R., van der Weele C.M., Wolniak S.M., *Removal of retained introns regulates translation in the rapidly developing gametophyte of Mersilea vestia*, Dev Cell., 2013, 5: 517-29
25. Yap K., Lin Z.Q., Khandelia P., Friedman B., Makeyev E., *Coordinated regulation of neuronal mRNA steady-state levels through developmentally controller intron retention*, Genes Dev., 2012, 11: 1209-23
26. Severing E.I., van Dijk A.D., Morabito G., Busscher-Lange J., Immink R.G., van Ham R.C., *Predicting the impact of alternative splicing on plant MADS domain protein function*, PLoS One, 2012, 7: e30524
27. Ellis J.D., Barrios-Rodiles M., Colak R., Irimia M., Kim T., Calarco J.A. et al., *Tissue-specific alternative splicing remodels protein-protein interaction networks*, Mol Cell., 2012, 6: 884-92
28. Seo P.J., Park M.J., Park C.M., *Alternative splicing of transcription factors in plant responses to low temperature stress: mechanisms and functions*, Planta, 2013, 6: 1415-24
29. Merkin J., Russell C., Chen P., Burge C.B., *Evolutionary dynamics of gene isoform regulation in Mammalian tissues*, Science, 2012, 338: 1593-9
30. de la Fuente van Bentem S., Anrather D., Roitinger E., Djamei A., Hufnagl T., Barta A. et al., *Phosphoproteomics reveals extensive in vivo phosphorylation of Arabidopsis proteins involved in RNA metabolism*, Nucleic Acids Res., 2006, 34: 3267-78
31. Drechsel G., Kahles A., Keserwani A.K., Stauffer E., Behr J., Derewe P. et al., *Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome*, Plant Cell, 2013, 25: 3726-42
32. Palusa S.G., Reddy A.S., *Extensive coupling of alternative splicing of pre-mRNA of serine/arginine (SR) genes with nonsense-mediated decay*, New Phytol., 2010, 185: 83-9
33. Steiger D., Zecca L., Wieczorek Kirk D.A., Apel K., Eckstein L., *The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA*, Plant J., 2003, 33: 361-71
34. Schöning J.C., Streitner C., Meyer I.M., Gao Y., Staiger D., *Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in Arabidopsis*, Nucleic Acids Res., 2008, 22: 6977-87
35. Yang G.D., Yan K., Wu B.J., Wang Y.H., Gao Y.X., Zheng C.C., *Genomewide analysis of intronic microRNAs in rice and Arabidopsis*, J. Genet., 2012, 91: 313-24
36. Yang X., Zhang H., Li L., *Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis*, Plant J., 2012, 70: 421-31
37. Yan K., Liu P., Wu C.A., Yang G.D., Xu R., Guo Q.H., Huang J.G., Zheng C.C., *Stress-induced alternative splicing provides a mechanism for the regulation of microRNA processing in Arabidopsis thaliana*, Mol Cell., 2012, 48: 521-31
38. Latijnhouwers M. J., Pairoba C.F., Brendel V., Walbot V., Carle-Urioste J.C., *Test of the combinatorial model of intron recognition in a native maize gene*, Plant Mol Biol, 1999, 41: 637-44

39. Quesada V., Macknight R., Dean C., Simpson G.G., *Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time*, The EMBO J, 2003, 22: 3142 - 52
40. Morere – Le Paven M., Anzala F., Rector A., Limami A.M., *Differential transcription initiation and alternative RNA splicing of Knox7 a class 2 homeobox gene of maize*, Gene, 2007, 401: 71 -9
41. Marrs, K.A., Wlabot, V., *Expression and RNA splicing of the maize glutathione S- transferase Bronze2 gene is regulated by Cadmium and other stresses*, Plant Physiol., 1997, 113: 93 – 102
42. Gupta S., Wang B., Stryker G.A., Zanetti M.E., Lal S.K., *Two novel arginine/serine (SR) proteins in maize are differentially spliced and utilize non – canonical splice sites*, Biochimica et Biophysica Acta, 2005, 1728: 105 – 114
43. Purugganan M.D., Fuller D.Q., *The nature of selection during plant domestication*, Nature, 2009, 457: 843-8
44. Diamond J., *Evolution, consequences and future of plant and animal domestication*, Nature, 2002, 418: 700-7
45. Doebley J., Stec A., Wendel J., Edwards M., *Genetic and morphological analysis of a maize-teosinte F₂ population: implications for the origin of maize*, Proc. Natl. Acad. Sci. USA, 1990, 87: 9888-92
46. Studer A., Zhao Q., Ross-Ibarra J., Doebley J., *Identification of a functional transposon insertion in the maize domestication gene tb1*, Nature Genetics, 2011, 43: 1160-3
47. Mir C., Zerjal T., Combes V., Dumas F., Madur D., Bedoya C. et al., *Out of America: tracking the genetic footprints of the global diffusion of maize*, Thero Appl Genet, 2013, 126: 2671-82
48. Freeling M., Walbot V., *The maize handbook*, Springer Lab Manuals, 1994, 66-77
49. Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F., Pasternak S. et al., *The B73 maize genome: complexity, diversity and dynamics*, Science, 2009, 326: 1112-5
50. Vielle-Calzada J.P., Martinez de la Vega O., Hernandez-Guzman G., Ibarra-Laclette E., Alvarez-Mejia C., Vega-Arrequin J.C. et al., *The Palomero genome suggests metal effects on domestication*, Science, 2009, 326: 1078
51. Zhao Q., Thuillet A., Uhlmann N.K., Weber A., Rafalski A., Allen S.M. et al., *The role of regulatory genes during maize domestication: evidence form nucleotide polymorphism and gene expression*, Genetics, 2008, 178: 2133-43
52. Messing J., Dooner H.K., *Organization and variability of the maize genome*, Curr Opin Plant Biol, 2006, 9: 157-63
53. *FAO Statistical Yearbook 2013. World food and agriculture*, Food and agriculture Organization of the United Nations, Rzym, 2013
54. *Rocznik statystyczny rolnictwa 2013*, Główny Urząd Statystyczny, Warszawa, 2013
55. Adamczyk J., Rogacki J., Cygert H., *Postęp w hodowli kukurydzy w Polsce*, Acta Sci. Pol., Agricultura, 2010, 9: 85-91
56. Wicki L., *Wykorzystanie postępu odmianowego w produkcji zbóż w polskim rolnictwie*, Rocznik Nauk Rolniczych, 2008, 94: 136-46
57. Lewandowska S., Kuriata R., Kałużebiec W., *Morfologiczne i genetyczne zróżnicowanie linii wsobnych kukurydzy*, Biuletyn IHAR, 2012, 264: 177-82
58. Agamczyk J., Cygert H., Czajczyński J., *50 lat hodowli kukurydzy mieszańcowej w Polsce - dorobek i perspektywy*, Biuletyn IHAR, 2003, 230: 423-31
59. Dubas A., *Udział naukowców i hodowców we wprowadzaniu nowoczesnych metod produkcji kukurydzy w Polsce*, Biuletyn IHAR, 1994, 191: 47-51

7. Bibliografia

60. Sobhanian H., Aghaei K. , Komatsu S., *Changes in the plant proteome resulting from salt stress: Toward the creation of salt-tolerant crops?*, Journal of Proteomics, 2011, 74: 1323-37
61. Chinnusamy V., Zhu J.K., *Epigenetic regulation of stress responses in plants*, Curr Opin in Plant Biol., 2009, 12: 133-9
62. Kasuga M., Miura S., Shinozaki M. et al, *A Combination of the Arabidopsis DREB1A Gene and Stress-Inducible rd29A Promoter Improved Drought- and Low-Temperature Stress Tolerance in Tobacco by Gene Transfer*, Plant Cell Physiol., 2004, 45: 346-50
63. Jaglo-Ottosen K. R., Gilmour S. J., Zarka D. G. et al, *Arabidopsis CBF1 Overexpression Induces COR Genes and Enhances Freezing Tolerance*, Science, 1998, 280: 104-6
64. Franz J., Mao M.K., Sikorski J.A., *Glyphosate: a unique global herbicide*, American Chemical Society, 1997
65. Duke S.O., Powles S.B., *Glyphosate: a once-in-a-century herbicide*, Pest Manage Sci., 2008, 64: 319-25
66. Green J.M., Owen M.D., *Herbicide-resistant crops: utilities and limitations for herbicide-resistant weed management*, J Agric Food Chem, 2011, 59: 5819-29
67. Powles S.B., *Evolved resistance to glyphosate in rigid ryegrass (Lolium rigidum) in Australia*, Weed Sci., 1998, 46: 604-7
68. Duke S.O., Powles S.B., *Glyphosate-resistant crops and weeds: now and in the future*, AgBioForum, 2009, 12: 346-57
69. Benson D.A., Cavanaugh M., Clark K., Karsch-Mizarachi I., Lipman D.J., Ostell J. et al., *GenBank*, Nucleic Acids Res., 2013, 41: D36-42
70. Dong Q., Schlueter S.D., Brendel V., *PlantGDB, plant genome database and analysis tools*, Nucleic Acids Res., 2004, 32: D354-9
71. Monaco M.K., Stein J., Naithani S., Wei S., Dharmawardhana P., Kumari S. et al., *Gramene 2013: comparative plant genomics resources*, Nucleic Acids Res., 2014, 42: D1193-9
72. Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G. at al., *Accurate whole human genome sequencing using reversible terminator chemistry*, Nature, 2008, 456: 53-9
73. Kearse M., Moir R., Drummond A. et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data*, Bioinformatics, 2012, 28: 1647-9
74. Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D. et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks*, Nature Prot., 2012, 3: 562-78
75. Usadel B., Poree F., Nagel A., Lohse M., Czedik-Eysenberg A., Sitt M., *A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize*, Plant Cell Environ., 2009, 32: 1211-29
76. Yao H., Guo L., Borsuk L.A., Wen T.J., Skibbe D.S., Cui X. et al., *Evaluation of five ab initio gene prediction programs for the discovery of maize genes*, Plant Mol Biol., 2005, 57: 445-60
77. Thorvaldsdottir H., Robinson J.T., Mesirov J.P., *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*, Brief Bioinform., 2013, 14: 178-92
78. Fait A., Fromm H., Walter D., Galili G., Fernie A.R., *Highway or byway: the metabolic role of the GABA shunt in plants*, Trends Plant Sci., 2008, 13: 14-9

79. Zik M., Arazi T., Snedden W.A., Fromm H., *Two isoforms of glutamate decarboxylase in Arabidopsis are regulated by calcium/calmodulin and differ in organ distribution*, Plant Mol Biol., 1998, 37: 967-75
80. Iida K., Fukami-Kobayashi K., Toyoda A., Sakaki Y., Kobayashi M., Seki M. et al., *Analysis of multiple occurrences of alternative splicing events in Arabidopsis thaliana using novel sequenced full-length cDNAs*, DNA Res., 2009, 16: 155-64
81. Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J. et al., *Empirical analysis of transcriptional activity in the Arabidopsis genome*, Science, 2003, 302: 842-6
82. Grotewold E., Athma P., Peterson T., *Alternatively spliced products of the maize P gene encode proteins with homology to the DNA-binding domain of myb-like transcription factors*, Proc Natl Acad Sci USA., 1991, 88: 4587-91
83. Ranocha P., McNeil S.D., Ziemak M.J., Li C., Tarczynski M.C., Hanson A.D., *The S-methylmethionine cycle in angiosperms: ubiquity, antiquity and activity*, Plant J., 2001, 25: 575-84
84. Forde B.G., Lea P.J., *Glutamate in plants: metabolism, regulation, and signaling*, J Exp Bot., 58: 2339-58
85. Becker T.W., Carrayol E., Hirel B., *Glutamine synthetase and glutamate dehydrogenase isoforms in maize leaves: localization, relative proportion and their role in ammonium assimilation or nitrogen transport*, Planta., 2000, 211: 800-6
86. Yan J., Shah T., Warburton M.L., Buckler E.S., McMullen M.D., Crouch J., *Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers*, PLoS one, 2009, 12: e8451
87. Pan Q., Ali F., Yang X., Li J., Yan J., *Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize*, PLoS One, 2012, 12: e52777
88. Pang K., Li Y., Liu M., Meng Z., Yu Y., *Inventory and general analysis of the ATP-binding cassette (ABC) gene superfamily in maize (Zea mays L.)*, Gene, 2013, 526: 411-28
89. Pang S., Duan L., Liu Z., Song X., Li X., Wang C., *Co-induction of a glutathione-S-transferase, a glutathione transporter and an ABC transporter in maize by xenobiotics*, PLoS One, 2012, 7: e40712
90. Moons A., *Ospdr9, which encodes a PDR-type ABC transporter, is induced by heavy metals, hypoxic stress and redox perturbations in rice roots*, FEBS Letters, 2003, 553: 370-6
91. Moons A., *Transcriptional profiling of the PDR gene family in rice roots in response to plant growth regulators, redox perturbations and weak organic acid stresses*, Planta., 2008, 229: 53-7
92. Cagnac O., Bourbouloux A., Chakrabaty D., Zhang M.Y., Delrot S., *AtOPT6 transports glutathione derivatives and is induced by rimsulfuron*, Plant Physiol., 2004, 135: 1378-87
93. Ramakrishna A., Ravishankar G. A., *Influence of abiotic stress signals on secondary metabolites in plants*, Plant Signal Behav., 2011, 6: 1720-31
94. Mutlu F., Bozcuk S., *Salinity-induced changes of free and bound polyamine levels in sunflower (Helianthus annuus L.) roots differing in salt tolerance*, Pak. J. Bot., 2007, 39: 1097-102
95. Minocha R., Majumdar R., Minocha S.C., *Polyamines and abiotic stress in plants: a complex relationship*, 2014, 5: 175
96. Korkina L.G., *Phenylpropanoids as naturally occurring antioxidants: from plant defense to human health*, Cell Mol Biol., 2007, 53: 15-25
97. Ahsan N., Lee D.G., Lee K.W., Alam I., Bahk J.D., Lee B.H., *Glyphosate-induced oxidative stress in rice leaves revealed by proteomic approach*, Plant Physiol Biochem., 2008, 46: 1062-70

98. Romero D.M., Rios de Molina M.C., Juarez A.B., *Oxidative stress induced by a commercial glyphosate formulation in a tolerant strain of Chlorella kessleri*, *Ecotoxicol Environ Saf.*, 2011, 74: 741-7
99. Li P., Ponnala L., Gandotra N., Wang L., Si y., Tausta S.L. et al., *The developmental dynamics of the maize leaf transcriptome*, *Nat Genet.*, 2010, 42: 1060-7
100. Foissac S., Sammeth M., *ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets*, *Nucleic Acids Res.*, 2007, 35: W297-9
101. Filichkin S.A., Priest H.D., Givan S.A., Shen R., Bryant D.W., Fox S.E. et al., *Genome-wide mapping of alternative splicing in Arabidopsis thaliana*, *Genome Res.*, 2010, 20: 45-58
102. Gullledge A.A., Roberts A.D., Vora H., Patel K., Loraine A.E., *Mining Arabidopsis thaliana RNA-seq data with Integrated Genome Browser reveals stress-induced alternative splicing of the putative splicing regulator SR45a*, *Am J Bot.*, 2012, 99: 219-31
103. Koroleva O.A., Calder G., Pendle A.F., Kim S.H., Lewandowska D., Simpson C.G. et al., *Dynamic behavior of Arabidopsis sIF4A-III, putative core protein of exon junction complex: fast relocation to nucleolus and splicing speckles under hypoxia*, *Plant Cell*, 2009, 21: 1592-606
104. Rausin G., Tillemans V., Stankovic N., Hanikenne M., Motte P., *Dynamic nucleocytoplasmic shuttling of an Arabidopsis SR splicing factor: role of the RNA-binding domains*, *Plant Physiol.*, 2010, 153: 273-84
105. Staiger D., Brown J.W., *Alternative splicing at the intersection of biological timing, development, and stress responses*, *Plant Cell*, 2013, 25: 3640-56
106. Samajova O., Plihal O., Al-Yosif M., Hirt H., Samaj J., *Improvement of stress tolerance in plants by genetic manipulation of mitogen-activated protein kinases*, *Biotechnol Adv.*, 2013, 31: 118-28
107. Teige M., Scheikl E., Eulgem T., Doczi R., Ichimura K., Shinozaki K. et al., *The MKK2 pathway mediates cold and salt stress signaling in Arabidopsis*, *Mol Cell*, 2004, 15: 141-52
108. Yeh C.M., Chien P.S., Huang H.J., *Distinct signaling pathways for induction of MAP kinase activities by cadmium and copper in rice roots*, *J Exp Bot.*, 2007, 58: 659-71
109. Jonak C., Nakagami H., Hirt H., *Heavy metal stress. Activation of distinct mitogen-activated protein kinase pathways by copper and cadmium*, *Plant Physiol.*, 2004, 136: 3276-83
110. Ortiz-Masia D., Perez-Amador M.A., Carbonell J., Marcote M.J., *Diverse stress signals activate the C1 subgroup MAP kinases of Arabidopsis*, *FEBS Letters*, 2007, 581: 1834-40
111. Wang M., Yuan F., Hao H., Zhang Y., Zhao H., Guo A. et al., *BolOST1, an ortholog of Open Stomata 1 with alternative splicing products in Brassica oleracea, positively modulates drought responses in plants*, *Biochem Biophys Res Commun.*, 2013, 442: 214-20

Streszczenie

W Zakładzie Biologii RNA od kilku lat badana jest reakcja roślin na warunki stresu herbicydowego (herbicyd systemowy Roundup®) u kukurydzy zwyczajnej. Składnikiem aktywnym herbicydu Roundup® jest glifosat, który inhibuje syntazę 5-enolopirogronianoszikimo-3-fosforanu, EPSPS (ang. *5-enolpyruvylshikimate-3-phosphate synthase*), biorącą udział w syntezie aminokwasów aromatycznych (enzym ten nie występuje u ssaków). Mimo wyprowadzenia wielu odmian roślin coraz lepiej przystosowanych do warunków klimatycznych Polski, kukurydza pozostaje rośliną ciepłolubną, której należy zapewnić dostateczną ilość światła, wody i składników mineralnych. Na plantacjach aby zmniejszyć konkurencję, w początkowych fazach wzrostu roślin stosuje się herbicydy, w tym różnego typu inhibitory wzrostu siewek, produkcji pigmentów, fotosyntezy, syntezy aminokwasów czy regulatorów wzrostu. Wprowadzenie ich do uprawy jest silnym czynnikiem stresowym dla rozwijających się roślin. Stosowane herbicydy mogą powodować powstawanie nekrozy liści, przebarwienia lub zahamować wzrost roślin, co w skrajnych przypadkach może prowadzić do obniżenia plonu nawet o 20%. Jest wysoce prawdopodobne, że część genów biorących udział w odpowiedzi na niekorzystne warunki środowiskowe, może ulegać procesowi alternatywnego splicingu (AS). Dotychczas nie zbadano w jakim stopniu stosowanie herbicydów wpływa na proces alternatywnego splicingu.

Celem prowadzonych badań było określenie różnic w profilach ekspresji wariantów transkryptów powstałych na drodze alternatywnego splicingu w dwóch liniach kukurydzy zwyczajnej, wykazujących zróżnicowaną wrażliwość na herbicyd Roundup – w linii S245, która charakteryzuje się tolerancją na oprysk herbicydem oraz linii S79757, która jest na niego wrażliwa.

W pierwszym etapie przeprowadzono analizę bioinformatyczną chromosomu 1 (pierwszego) kukurydzy zwyczajnej pod kątem genów ulegających alternatywnemu splicingowi, której celem było oszacowanie skali zjawiska AS w genomie kukurydzy. Do analizy wykorzystano sekwencję chromosomu 1 oraz program AUGUSTUS. Proces alternatywnego splicingu udało się potwierdzić dla 113 genów z chromosomu 1 i zidentyfikować nowe zdarzenia AS dla 42 genów. Dla wybranych trzech genów, kodujących białka o znanej funkcji, przeprowadzono reakcje odwrotnej transkrypcji sprzężoną z reakcją

PCR, dzięki czemu udało się potwierdzić zajście przewidzianych zdarzeń alternatywnego składania transkryptów dla dwóch genów (S-metylotransferazy 2 homocysteiny i dehydrogenazy glutationu).

W kolejnym przeprowadzono sekwencjonowanie mRNA drugiej generacji. Dla poszczególnych linii uzyskano od 3,5 do 76 mln 50 nt odczytów (1,75 do 3,8 Gbp). Porównanie uzyskanych sekwencji transkryptomów obu linii z transkryptomem referencyjnym kukurydzy (linia wsobna B73) pozwoliło zidentyfikować od 166 do 193 tys. transkryptów, w tym od 3,5 do 27 tys. nowych eksonów, od 1,7 do 11 tys. nowych intronów, które mapują do przeszło 11 tys. nowych loci. Porównanie zsekwencjonowanych transkryptomów z roślin poddanych działaniu stresu herbicydowego pozwoliło nam zidentyfikować 599 genów o znaczącej statystycznie różnicy w poziomie ekspresji, 424 transkryptów reprezentujących różne wzory splicingowe oraz 37 genów o zróżnicowanej regulacji (wybór innej sekwencji promotorowej) pomiędzy porównywanymi liniami zebranymi tydzień po oprysku herbicydem.

Ponadto zidentyfikowano 71 genów ulegających zmienionej ekspresji po oprysku herbicydem w linii wrażliwej oraz 843 w linii tolerancyjnej, z czego odpowiednio 6 genów w linii wrażliwej i 61 genów w linii tolerancyjnej miało zmienioną ekspresję w obu badanych punktach czasowych (dzień i tydzień po przeprowadzeniu oprysku). Wykonano także analizę funkcjonalną 140 transkryptów o największych różnicach w poziomie ekspresji po oprysku herbicydem Roundup® z zastosowaniem programu MapMan. Udało nam się zmapować 96 transkryptów, z których 23 związane są z odpowiedzią rośliny na warunki stresowe. Wśród zmapowanych genów, liczną grupę stanowią także geny kodujące białka, biorące udział w fotosyntezie (14), metabolizmie wtórnym (11) czy metabolizmie białek (7). Duża ilość transkryptów została zmapowana do genów o nieznanym funkcjach (23) lub biorących udział w różnych szlakach metabolicznych (18).

Aby określić liczbę genów ulegających AS zastosowano skrypt napisany w języku java. Po przebadaniu uzyskanych po sekwencjonowaniu transkryptomów określono, że AS dotyczył od 31 do 35% genów, które ulegały ekspresji w poszczególnych liniach i punktach czasowych. Rozkład typów splicingu w badanych liniach odbiegał od wcześniej opisanego – najliczniejsze zdarzenia AS dotyczyły wyboru alternatywnych miejsc splicingowych 3' i 5', a nie jak dotychczas postulowano zatrzymania intronu. Po porównaniu genów różniących się

wzorem AS między transkryptomami roślin poddanych opryskowi a roślinami kontrolnymi, wytypowano 58 genów w linii wrażliwej i 39 genów w linii tolerancyjnej, które różniły się sposobem składania transkryptów. Udało się zidentyfikować także 3 geny, które różniły się sposobem składania mRNA w obu liniach i punktach czasowych po oprysku herbicydem Roundup. Nie udało się zidentyfikować potencjalnego produktu jednego z tych genów, natomiast pozostałe dwa geny kodują najprawdopodobniej jeden z czynników splicingowych i kinazę białkową. W ten sposób zmiana we wzorze AS może wpływać na zmiany w przekazywaniu sygnałów i ekspresji genów docelowych regulowanych przez ten czynnik splicingowy.

Zrealizowane badania potwierdzają wpływ herbicydów na zmianę metabolizmu roślin poddawanych opryskowi, a wytypowane w przedstawionej analizie geny i grupy białek mogą być obiektem dalszych badań funkcjonalnych oraz hodowlanych.

Tytuł i streszczenie w języku angielskim

Title: Alternative splicing as a mechanism of herbicide stress response in maize

Summary: Plants, as sessile organisms, must adapt their growth and metabolic style to a changing environment. Splicing is one of the mechanisms which play an important role in plant adaptation and is an additional element of fitness benefit adjusted to the limited capacity of genome size. Studies of splicing and its role in diverse aspect of cell biology, pathology and stress response, has remained undescribed for many plant species, including maize. Through the mechanism of alternative splicing, exons from primary transcripts (pre-mRNA) with multiple introns may undergo ligation in many different ways generating multiple proteins from single gene. This process can affect mRNA stability and translation efficiency as well as activity, cellular localization, regulation and stability of coding protein.

For better characterization of alternative splicing in maize, one of the most important crop species, we used AUGUSTUS to predict multiple transcripts and alternative splicing events from maize chromosome 1 sequence. From over 300 mln bp of chromosome 1 AUGUSTUS software predicted ab initio 46 400 genes and 20% out of estimated genes have at least two RNA transcripts. For 412 genes with three transcripts we have performed additional analysis including EST's alignment, protein identification and comparative evaluation with outcome of two other programs. As a result we confirm alternative splicing events for 42 genes from maize chromosome 1 and we found 18 probable protein products. In next step of analysis we managed to confirm the presence of 4 out of 7 predicted transcripts in two maize varieties.

In next step for more detailed and global characterization of alternative splicing role in plant herbicide stress response, we sequenced transcriptomes of two maize breed lines – sensitive and tolerant to herbicide Roundup®. We used Illumina next-generation sequencer Genome Analyzer IIx and we conducted pair-end sequencing. As a result we obtained 35 to 76 mln 50 nt reads per sample. Using bioinformatics tools such as BowTie, TopHat, Cufflinks, Cuffdiff and CummRbund we identify differences both in expression levels and AS events between sensitive and tolerant maize line. After comparison to reference transcriptome we

have found 3,5 to 27 thousand of new exons, from 1,7 to 11 thousand of new introns, which were mapped to 11 thousand new loci. Differential analysis allowed us to select genes with different expression levels after herbicide spraying. In sensitive line we identified 71 of such genes, in contrast to 843 genes in tolerant line. We also managed to identify pool of genes with different splicing pattern and select 3 genes with changes in splicing pattern in both lines and time points for better characterization. Using java script we defined AS proportion in both lines – from 31 to 35% expressed genes undergo AS events. Most of them were related to alternative 3' and 5' splicing sites selection and the rarest AS events involved intron retention (in contrast to previously described distribution of types of AS events in plants). After functional analysis performed with MapMan we picked out group of genes with different splicing and expression pattern, which are involved in abiotic stress response, protein, lipid and RNA metabolism.

Wykaz ważniejszych skrótów

AS	alternatywny splicing
ATP	adenozyno-5'-trójfosforan
białka SR	białka bogate w serynę i argininę
białka RBP	białka wiążące RNA
cDNA	komplementarny DNA powstały na matrycy mRNA
CDS	sekwencje kodujące
ds cDNA	dwuniciowy komplementarny DNA
DNA	kwas deoksyrybonukleinowy
EST	sekwencyjne znaczniki ekspresji
FPKM	liczba fragmentów przypadających na tysiąc nukleotydów sekwencji transkryptu na milion zmapowanych odczytów
GABA	kwas γ -aminomasłowy
GDH	dehydrogenaza glutationu
GTP	guanozyno-5'-trójfosforan
HMT2	S-metylotransferaza 2 homocysteiny
mRNA	matrycowy RNA
NGS	sekwencjonowanie drugiej generacji
NMD	rozpad mRNA zawierających sekwencje nonsensowe
PCR	reakcja łańcuchowa polimerazy
PTC	przedwczesny sygnał terminacji translacji
RIN	współczynnik integralności RNA
RNA	kwas rybonukleinowy
RNA-Seq	sekwencjonowanie cząsteczek RNA z zastosowaniem technologii sekwencjonowania drugiej generacji
ROS	reaktywne formy tlenu
RRM	domena wiążąca RNA
RT-PCR	reakcja odwrotnej transkrypcji sprzężona z reakcją PCR
SL	linia wrażliwa S79757 na działanie herbicydu Roundup®
SNP	polimorfizm pojedynczego nukleotydu
TL	linia tolerancyjna S245 na działanie herbicydu Roundup®
SL	linia wrażliwa S79757 na działanie herbicydu Roundup®
Q	jakość nukleotydu

Spis rycin i tabel

Ryc. 1 Schemat założeń centralnego dogmatu biologii molekularnej.....	6
Ryc. 2 Schematyczne przedstawienie reakcji splicingu.....	10
Ryc. 3 Główne typy zdarzeń alternatywnego splicingu.....	13
Ryc. 4 Porównanie wielkości i ogólnego pokroju teosinte i jednej z odmian kukurydzy.....	18
Tabela 1 Zestawienie powierzchni zasiewów, wielkości zbiorów i plonu kukurydzy w latach 2000 - 2012.....	21
Ryc. 5 Wzór strukturalny cząsteczki glifosatu.....	25
Ryc. 6 Zdjęcia roślin linii tolerancyjnej S245 (A) i linii wrażliwej S79757 (B) 14 dni po zabiegu oprysku herbicydem.....	30
Tabela 2 Uszkodzenia roślin kukurydzy po oprysku różnymi dawkami herbicydu Roundup.....	31
Ryc. 7 Dendrogram pokrewieństwa genotypów kukurydzy.....	31
Tabela 3 Zestawienie sekwencji jednoniciowych DNA stosowanych jako startery w reakcjach PCR.....	33
Tabela 4 Zestawienie liczby i typów danych w bazie PlantGDB.....	38
Tabela 5 Wykaz wykorzystanych baz danych i ich adresy internetowe.....	40
Ryc. 8 Zdjęcia roślin badanych linii.....	42
Tabela 6 Skład typowej reakcji PCR.....	45
Tabela 7 Schemat reakcji PCR w gradiencie temperatur.....	46
Tabela 8 Składniki reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR.....	47
Tabela 9 Ogólny schemat przebiegu reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR.....	47
Tabela 10 Skład żeli agarozowych.....	47
Ryc. 9 Schemat sekwencjonowania z użyciem technologii Illumina.....	52
Ryc. 10 Schemat przebiegu analizy bioinformatycznej.....	57
Ryc. 11 Schemat analizy bioinformatycznej sekwencji chromosomu 1 kukurydzy zwyczajnej.....	59
Tabela 11 Zestawienie wyników dopasowania sekwencji 42 genów przewidzianych przez program AUGUSTUS.....	64
Ryc. 12 Schemat organizacji genu nr 1 przewidzianego przez program AUGUSTUS.....	65
Ryc. 13 Wzajemna dystrybucja i lokalizacja zbioru 42 genów przewidzianych przez program AUGUSTUS.....	67
Tabela 12 Wyniki dopasowania sekwencji aminokwasowych przetłumaczonych z transkryptów przewidzianych przez program AUGUSTUS.....	70
Ryc. 14 Obraz rozdziału elektroforetycznego całkowitego RNA.....	71
Ryc. 15 Obraz rozdziału elektroforetycznego produktów reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR dla transkryptu 1 genu kodującego białko P związane z czynnikiem transkrypcyjnym Myb.....	73
Ryc. 16 Obraz rozdziału elektroforetycznego produktów reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR dla transkryptu 1 i 2 genu kodującego S-metylotransferazę 2 homocysteiny.....	74
Ryc. 17 Obraz rozdziału elektroforetycznego produktów reakcji odwrotnej transkrypcji sprzężonej z reakcją PCR dla transkryptu 1 i 2 genu kodującego dehydrogenazę glutationu.....	75
Tabela 13 Zestawienie liczby uzyskanych surowych odczytów z sekwencjonowania mRNA drugiej generacji.....	81
Ryc. 18 Wykresy wygenerowane przez program FastQC.....	85

Tabela 14 Zestawienie porównania sekwencji transkryptomów uzyskanych dla badanych linii z transkryptomem linii referencyjnej kukurydzy B73.....	86
Ryc. 19 Porównanie ekspresji genów między linią tolerancyjną a linią wrażliwą.....	91
Ryc. 20 Obraz głównych szlaków metabolicznych generowany przez program MapMan.....	93
Ryc. 21 Graficzne przedstawienie szlaków metabolicznych biorących udział w reakcjach związanych z odpowiedzią roślin na warunki stresowe.....	94
Tabela 15 Typy zdarzeń alternatywnego składania określone w badanych liniach oraz procent genów ulegających AS.....	96
Ryc. 22 Diagram Venna obrazujący zależności między zbiorami genów różniących się wzorem splicingu po zastosowaniu herbicydu Roundup®.....	97
Tabela 16 Zestawienie informacji o ekspresji trzech genów o różnym wzorze splicingu i zróżnicowanej ekspresji poszczególnych izoform transkryptów.....	99
Ryc. 23 Schemat organizacji izoform transkryptów genu GRMZM5G832248 ulegających ekspresji w linii tolerancyjnej na działanie herbicydu Roundup.....	101
Ryc. 24 Schemat organizacji izoform transkryptów genu GRMZM5G832248 ulegających ekspresji w linii wrażliwej na działanie herbicydu Roundup.....	102
Ryc. 25 Schemat organizacji izoform transkryptów genu GRMZM2G366935 ulegających ekspresji w linii wrażliwej na działanie herbicydu Roundup.....	104
Ryc. 26 Schemat organizacji izoform transkryptów genu GRMZM2G366935 ulegających ekspresji w linii tolerancyjnej na działanie herbicydu Roundup.....	105
Ryc. 27 Schemat organizacji izoform transkryptów mapujących do genu GRMZM2G124630 i genu GRMZM2G124636, które ulegały ekspresji w linii tolerancyjnej na działanie herbicydu Roundup.....	107
Ryc. 28 Schemat organizacji izoform transkryptów mapujących do genu GRMZM2G124630 i genu GRMZM2G124636, które ulegały ekspresji w linii wrażliwej na działanie herbicydu Roundup.....	108
Tabela 17 Grupa genów biorących udział w odpowiedzi na warunki stresu abiotycznego zidentyfikowana przez program MapMan.....	109

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	6321054	6321170	0.48	-	0	gene 1.1
chr1	AUGUSTUS	CDS	6321275	6321484	0.88	-	0	gene 1.1
chr1	AUGUSTUS	CDS	6321578	6321707	1	-	1	gene 1.1
chr1	AUGUSTUS	CDS	6321806	6321976	0.97	-	1	gene 1.1
chr1	AUGUSTUS	CDS	6322778	6322900	1	-	1	gene 1.1
chr1	AUGUSTUS	CDS	6323014	6323118	0.96	-	1	gene 1.1
chr1	AUGUSTUS	CDS	6323727	6324415	1	-	0	gene 1.1
chr1	AUGUSTUS	CDS	6313358	6313468	0.91	-	0	gene 1.2
chr1	AUGUSTUS	CDS	6313568	6313681	0.36	-	0	gene 1.2
chr1	AUGUSTUS	CDS	6313358	6313468	0.91	-	0	gene 1.3
chr1	AUGUSTUS	CDS	6313568	6313664	0.2	-	1	gene 1.3
chr1	AUGUSTUS	CDS	6320446	6320470	0.2	-	2	gene 1.3
chr1	AUGUSTUS	CDS	6320611	6320674	0.19	-	0	gene 1.3
chr1	AUGUSTUS	CDS	6321275	6321484	0.88	-	0	gene 1.3
chr1	AUGUSTUS	CDS	6321578	6321707	1	-	1	gene 1.3
chr1	AUGUSTUS	CDS	6321806	6321976	0.97	-	1	gene 1.3
chr1	AUGUSTUS	CDS	6322778	6322900	1	-	1	gene 1.3
chr1	AUGUSTUS	CDS	6323014	6323118	0.96	-	1	gene 1.3
chr1	AUGUSTUS	CDS	6323727	6324415	1	-	0	gene 1.3
chr1	AUGUSTUS	CDS	6494605	6494898	0.95	-	0	gene 2.1
chr1	AUGUSTUS	CDS	6496095	6496629	0.69	-	1	gene 2.2
chr1	AUGUSTUS	CDS	6497166	6497382	1	-	2	gene 2.2
chr1	AUGUSTUS	CDS	6498699	6498735	0.66	-	0	gene 2.2
chr1	AUGUSTUS	CDS	6498886	6498962	0.61	-	2	gene 2.2
chr1	AUGUSTUS	CDS	6499126	6499255	0.5	-	0	gene 2.2
chr1	AUGUSTUS	CDS	6494605	6494898	0.95	-	0	gene 2.3
chr1	AUGUSTUS	CDS	6496101	6496629	0.31	-	1	gene 2.3
chr1	AUGUSTUS	CDS	6497166	6497382	1	-	2	gene 2.3
chr1	AUGUSTUS	CDS	6498699	6498735	0.66	-	0	gene 2.3
chr1	AUGUSTUS	CDS	6498886	6498962	0.61	-	2	gene 2.3
chr1	AUGUSTUS	CDS	6499126	6499255	0.5	-	0	gene 2.3
chr1	AUGUSTUS	CDS	13457950	13458432	0.49	+	0	gene 3.1

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	13458509	13458641	1	+	0	gene 3.1
chr1	AUGUSTUS	CDS	13458725	13458912	0.96	+	2	gene 3.1
chr1	AUGUSTUS	CDS	13458993	13459454	0.97	+	0	gene 3.1
chr1	AUGUSTUS	CDS	13457014	13457746	0.24	+	0	gene 3.2
chr1	AUGUSTUS	CDS	13457819	13458432	0.24	+	2	gene 3.2
chr1	AUGUSTUS	CDS	13458509	13458641	1	+	0	gene 3.2
chr1	AUGUSTUS	CDS	13458725	13458912	0.96	+	2	gene 3.2
chr1	AUGUSTUS	CDS	13458993	13459454	0.97	+	0	gene 3.2
chr1	AUGUSTUS	CDS	13457014	13457781	0.44	+	0	gene 3.3
chr1	AUGUSTUS	CDS	17591888	17592148	0.76	+	0	gene 4.1
chr1	AUGUSTUS	CDS	17596441	17596573	0.84	+	0	gene 4.1
chr1	AUGUSTUS	CDS	17596692	17596780	0.92	+	2	gene 4.1
chr1	AUGUSTUS	CDS	17596879	17596936	0.89	+	0	gene 4.1
chr1	AUGUSTUS	CDS	17597749	17597897	0.85	+	2	gene 4.1
chr1	AUGUSTUS	CDS	17598749	17598962	0.95	+	0	gene 4.1
chr1	AUGUSTUS	CDS	17599035	17599137	0.95	+	2	gene 4.1
chr1	AUGUSTUS	CDS	17599262	17599309	1	+	1	gene 4.1
chr1	AUGUSTUS	CDS	17599450	17599495	1	+	1	gene 4.1
chr1	AUGUSTUS	CDS	17600391	17600578	1	+	0	gene 4.1
chr1	AUGUSTUS	CDS	17600688	17600850	0.73	+	1	gene 4.1
chr1	AUGUSTUS	CDS	17603474	17603671	0.77	+	0	gene 4.1
chr1	AUGUSTUS	CDS	17603756	17603852	0.64	+	0	gene 4.1
chr1	AUGUSTUS	CDS	17604247	17604356	0.88	+	2	gene 4.1
chr1	AUGUSTUS	CDS	17591888	17592148	0.76	+	0	gene 4.2
chr1	AUGUSTUS	CDS	17596441	17596573	0.84	+	0	gene 4.2
chr1	AUGUSTUS	CDS	17596692	17596780	0.92	+	2	gene 4.2
chr1	AUGUSTUS	CDS	17596879	17596936	0.89	+	0	gene 4.2
chr1	AUGUSTUS	CDS	17597749	17597897	0.85	+	2	gene 4.2
chr1	AUGUSTUS	CDS	17598749	17598962	0.95	+	0	gene 4.2
chr1	AUGUSTUS	CDS	17599035	17599137	0.95	+	2	gene 4.2
chr1	AUGUSTUS	CDS	17599262	17599309	1	+	1	gene 4.2
chr1	AUGUSTUS	CDS	17599450	17599495	1	+	1	gene 4.2
chr1	AUGUSTUS	CDS	17600391	17600578	1	+	0	gene 4.2
chr1	AUGUSTUS	CDS	17600688	17600850	0.73	+	1	gene 4.2
chr1	AUGUSTUS	CDS	17600945	17600956	0.3	+	0	gene 4.2
chr1	AUGUSTUS	CDS	17603263	17603325	0.5	+	0	gene 4.3
chr1	AUGUSTUS	CDS	17603474	17603671	0.77	+	0	gene 4.3
chr1	AUGUSTUS	CDS	17603756	17603852	0.64	+	0	gene 4.3
chr1	AUGUSTUS	CDS	17604247	17604356	0.88	+	2	gene 4.3

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	33066932	33066982	0.43	-	0	gene 5.1
chr1	AUGUSTUS	CDS	33067887	33067955	0.69	-	0	gene 5.1
chr1	AUGUSTUS	CDS	33068048	33068145	0.39	-	2	gene 5.1
chr1	AUGUSTUS	CDS	33068950	33069073	0.62	-	0	gene 5.1
chr1	AUGUSTUS	CDS	33070194	33070497	0.99	-	1	gene 5.1
chr1	AUGUSTUS	CDS	33070578	33070787	1	-	1	gene 5.1
chr1	AUGUSTUS	CDS	33070869	33070983	1	-	2	gene 5.1
chr1	AUGUSTUS	CDS	33071150	33071255	0.94	-	0	gene 5.1
chr1	AUGUSTUS	CDS	33071366	33071479	0.99	-	0	gene 5.1
chr1	AUGUSTUS	CDS	33071565	33071576	0.53	-	0	gene 5.1
chr1	AUGUSTUS	CDS	33066932	33066982	0.43	-	0	gene 5.2
chr1	AUGUSTUS	CDS	33067887	33067955	0.69	-	0	gene 5.2
chr1	AUGUSTUS	CDS	33068048	33068145	0.39	-	2	gene 5.2
chr1	AUGUSTUS	CDS	33068950	33069073	0.62	-	0	gene 5.2
chr1	AUGUSTUS	CDS	33070194	33070497	0.99	-	1	gene 5.2
chr1	AUGUSTUS	CDS	33070578	33070787	1	-	1	gene 5.2
chr1	AUGUSTUS	CDS	33070869	33070983	1	-	2	gene 5.2
chr1	AUGUSTUS	CDS	33071150	33071255	0.94	-	0	gene 5.2
chr1	AUGUSTUS	CDS	33071366	33071479	0.99	-	0	gene 5.2
chr1	AUGUSTUS	CDS	33071565	33071624	0.41	-	0	gene 5.2
chr1	AUGUSTUS	CDS	33075868	33075890	0.35	-	2	gene 5.2
chr1	AUGUSTUS	CDS	33076123	33076335	0.58	-	2	gene 5.2
chr1	AUGUSTUS	CDS	33077550	33077907	0.39	-	0	gene 5.2
chr1	AUGUSTUS	CDS	33079355	33079898	1	-	1	gene 5.2
chr1	AUGUSTUS	CDS	33080079	33080220	0.99	-	2	gene 5.2
chr1	AUGUSTUS	CDS	33080443	33080710	0.99	-	0	gene 5.2
chr1	AUGUSTUS	CDS	33075800	33075855	0.23	-	2	gene 5.3
chr1	AUGUSTUS	CDS	33076123	33076335	0.58	-	2	gene 5.3
chr1	AUGUSTUS	CDS	33077550	33077907	0.39	-	0	gene 5.3
chr1	AUGUSTUS	CDS	33079355	33079898	1	-	1	gene 5.3
chr1	AUGUSTUS	CDS	33080079	33080220	0.99	-	2	gene 5.3
chr1	AUGUSTUS	CDS	33080443	33080710	0.99	-	0	gene 5.3
chr1	AUGUSTUS	CDS	34457112	34457309	1	-	0	gene 6.1
chr1	AUGUSTUS	CDS	34457402	34457530	1	-	0	gene 6.1
chr1	AUGUSTUS	CDS	34457680	34458213	0.99	-	0	gene 6.1
chr1	AUGUSTUS	CDS	34470689	34470988	0.93	-	0	gene 6.1
chr1	AUGUSTUS	CDS	34457112	34457309	1	-	0	gene 6.2
chr1	AUGUSTUS	CDS	34457402	34457530	1	-	0	gene 6.2
chr1	AUGUSTUS	CDS	34457680	34458213	0.99	-	0	gene 6.2

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	34459415	34459453	0.27	-	0	gene 6.2
chr1	AUGUSTUS	CDS	34467652	34468049	0.54	-	2	gene 6.3
chr1	AUGUSTUS	CDS	34469041	34469095	0.4	-	0	gene 6.3
chr1	AUGUSTUS	CDS	34470689	34470988	0.93	-	0	gene 6.3
chr1	AUGUSTUS	CDS	34899138	34899425	0.38	+	0	gene 7.1
chr1	AUGUSTUS	CDS	34899553	34899644	1	+	0	gene 7.1
chr1	AUGUSTUS	CDS	34901343	34901424	0.47	+	1	gene 7.1
chr1	AUGUSTUS	CDS	34901667	34901807	0.93	+	0	gene 7.1
chr1	AUGUSTUS	CDS	34901986	34902081	1	+	0	gene 7.1
chr1	AUGUSTUS	CDS	34902180	34902321	1	+	0	gene 7.1
chr1	AUGUSTUS	CDS	34902640	34902767	0.49	+	2	gene 7.1
chr1	AUGUSTUS	CDS	34902858	34903034	0.45	+	0	gene 7.1
chr1	AUGUSTUS	CDS	34899138	34899249	0.58	+	0	gene 7.2
chr1	AUGUSTUS	CDS	34899331	34899425	0.53	+	2	gene 7.2
chr1	AUGUSTUS	CDS	34899553	34899644	1	+	0	gene 7.2
chr1	AUGUSTUS	CDS	34901343	34901424	0.47	+	1	gene 7.2
chr1	AUGUSTUS	CDS	34901667	34901807	0.93	+	0	gene 7.2
chr1	AUGUSTUS	CDS	34901986	34902081	1	+	0	gene 7.2
chr1	AUGUSTUS	CDS	34902180	34902321	1	+	0	gene 7.2
chr1	AUGUSTUS	CDS	34905632	34905765	0.32	+	2	gene 7.2
chr1	AUGUSTUS	CDS	34908122	34908288	1	+	0	gene 7.2
chr1	AUGUSTUS	CDS	34909087	34909263	0.45	+	1	gene 7.2
chr1	AUGUSTUS	CDS	34909826	34909899	0.74	+	1	gene 7.2
chr1	AUGUSTUS	CDS	34910878	34911296	0.6	+	2	gene 7.2
chr1	AUGUSTUS	CDS	34905232	34905765	0.27	+	0	gene 7.3
chr1	AUGUSTUS	CDS	34908122	34908288	1	+	0	gene 7.3
chr1	AUGUSTUS	CDS	34909087	34909263	0.45	+	1	gene 7.3
chr1	AUGUSTUS	CDS	34909826	34909899	0.74	+	1	gene 7.3
chr1	AUGUSTUS	CDS	34910878	34911296	0.6	+	2	gene 7.3
chr1	AUGUSTUS	CDS	38940906	38941559	0.54	+	0	gene 8.1
chr1	AUGUSTUS	CDS	38938902	38939216	0.98	+	0	gene 8.2
chr1	AUGUSTUS	CDS	38939333	38939767	0.45	+	0	gene 8.2
chr1	AUGUSTUS	CDS	38939952	38940119	0.23	+	0	gene 8.2
chr1	AUGUSTUS	CDS	38938902	38939216	0.98	+	0	gene 8.3
chr1	AUGUSTUS	CDS	38939333	38939725	0.4	+	0	gene 8.3
chr1	AUGUSTUS	CDS	38940927	38941559	0.43	+	0	gene 8.3

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	40999736	41000473	0.67	-	0	gene 9.1
chr1	AUGUSTUS	CDS	41000927	41001268	0.56	-	0	gene 9.2
chr1	AUGUSTUS	CDS	40999736	41000473	0.67	-	0	gene 9.3
chr1	AUGUSTUS	CDS	41000972	41001268	0.23	-	0	gene 9.3
chr1	AUGUSTUS	CDS	46298698	46299711	0.85	+	0	gene 10.1
chr1	AUGUSTUS	CDS	46294503	46294826	0.94	+	0	gene 10.2
chr1	AUGUSTUS	CDS	46295270	46295963	0.88	+	0	gene 10.2
chr1	AUGUSTUS	CDS	46298095	46298180	0.48	+	2	gene 10.2
chr1	AUGUSTUS	CDS	46294503	46294826	0.94	+	0	gene 10.3
chr1	AUGUSTUS	CDS	46295270	46295963	0.88	+	0	gene 10.3
chr1	AUGUSTUS	CDS	46297830	46297900	0.25	+	2	gene 10.3
chr1	AUGUSTUS	CDS	46298698	46299711	0.85	+	0	gene 10.3
chr1	AUGUSTUS	CDS	70561414	70561730	0.47	-	2	gene 11.1
chr1	AUGUSTUS	CDS	70561914	70563138	0.38	-	0	gene 11.1
chr1	AUGUSTUS	CDS	70560117	70560812	0.64	-	0	gene 11.2
chr1	AUGUSTUS	CDS	70561489	70561730	0.42	-	2	gene 11.2
chr1	AUGUSTUS	CDS	70561914	70563138	0.38	-	0	gene 11.2
chr1	AUGUSTUS	CDS	70560117	70560812	0.64	-	0	gene 11.3
chr1	AUGUSTUS	CDS	70561067	70561261	0.45	-	0	gene 11.3
chr1	AUGUSTUS	CDS	78144243	78145521	1	-	1	gene 12.1
chr1	AUGUSTUS	CDS	78145600	78145936	0.96	-	2	gene 12.1
chr1	AUGUSTUS	CDS	78146943	78146999	0.81	-	2	gene 12.1
chr1	AUGUSTUS	CDS	78152036	78152308	0.46	-	2	gene 12.1
chr1	AUGUSTUS	CDS	78153069	78153426	0.53	-	0	gene 12.1
chr1	AUGUSTUS	CDS	78155384	78156823	0.28	-	0	gene 12.1
chr1	AUGUSTUS	CDS	78144243	78145521	1	-	1	gene 12.2
chr1	AUGUSTUS	CDS	78145600	78145936	0.96	-	2	gene 12.2
chr1	AUGUSTUS	CDS	78146943	78146999	0.81	-	2	gene 12.2
chr1	AUGUSTUS	CDS	78152036	78152424	0.33	-	1	gene 12.2
chr1	AUGUSTUS	CDS	78152530	78152687	0.14	-	0	gene 12.2
chr1	AUGUSTUS	CDS	78155279	78156823	0.28	-	0	gene 12.3
chr1	AUGUSTUS	CDS	85692313	85693332	0.54	-	0	gene 13.1

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	85700313	85700335	0.48	-	2	gene 13.1
chr1	AUGUSTUS	CDS	85700567	85700769	0.87	-	1	gene 13.1
chr1	AUGUSTUS	CDS	85700938	85701096	0.49	-	1	gene 13.1
chr1	AUGUSTUS	CDS	85701930	85702306	0.81	-	0	gene 13.1
chr1	AUGUSTUS	CDS	85704076	85704567	0.77	-	0	gene 13.1
chr1	AUGUSTUS	CDS	85687358	85687599	0.97	-	2	gene 13.2
chr1	AUGUSTUS	CDS	85690068	85690352	0.38	-	2	gene 13.2
chr1	AUGUSTUS	CDS	85690683	85691916	0.67	-	0	gene 13.2
chr1	AUGUSTUS	CDS	85687358	85687599	0.97	-	2	gene 13.3
chr1	AUGUSTUS	CDS	85690068	85690352	0.38	-	2	gene 13.3
chr1	AUGUSTUS	CDS	85690683	85691916	0.67	-	0	gene 13.3
chr1	AUGUSTUS	CDS	85692967	85693332	0.33	-	0	gene 13.3
chr1	AUGUSTUS	CDS	85700313	85700335	0.48	-	2	gene 13.3
chr1	AUGUSTUS	CDS	85700567	85700769	0.87	-	1	gene 13.3
chr1	AUGUSTUS	CDS	85701004	85701096	0.39	-	1	gene 13.3
chr1	AUGUSTUS	CDS	85701930	85702306	0.81	-	0	gene 13.3
chr1	AUGUSTUS	CDS	85704076	85704567	0.77	-	0	gene 13.3
chr1	AUGUSTUS	CDS	95796359	95796508	0.67	+	0	gene 14.1
chr1	AUGUSTUS	CDS	95796970	95797470	1	+	0	gene 14.1
chr1	AUGUSTUS	CDS	95797627	95798820	0.91	+	0	gene 14.1
chr1	AUGUSTUS	CDS	95785160	95785447	0.82	+	0	gene 14.2
chr1	AUGUSTUS	CDS	95785658	95786629	0.83	+	0	gene 14.2
chr1	AUGUSTUS	CDS	95788584	95788941	0.85	+	0	gene 14.2
chr1	AUGUSTUS	CDS	95796237	95796508	0.24	+	2	gene 14.2
chr1	AUGUSTUS	CDS	95796970	95797470	1	+	0	gene 14.2
chr1	AUGUSTUS	CDS	95797627	95798820	0.91	+	0	gene 14.2
chr1	AUGUSTUS	CDS	95785160	95785447	0.82	+	0	gene 14.3
chr1	AUGUSTUS	CDS	95785658	95786629	0.83	+	0	gene 14.3
chr1	AUGUSTUS	CDS	95788584	95788941	0.85	+	0	gene 14.3
chr1	AUGUSTUS	CDS	95790518	95790558	0.56	+	2	gene 14.3
chr1	AUGUSTUS	CDS	100753342	100754016	0.91	-	0	gene 15.1
chr1	AUGUSTUS	CDS	100757560	100758447	0.53	-	0	gene 15.2
chr1	AUGUSTUS	CDS	100758644	100759607	0.87	-	1	gene 15.2
chr1	AUGUSTUS	CDS	100759829	100760763	0.96	-	0	gene 15.2
chr1	AUGUSTUS	CDS	100753342	100754016	0.91	-	0	gene 15.3
chr1	AUGUSTUS	CDS	100755067	100755432	0.67	-	0	gene 15.3

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	100757701	100758447	0.34	-	0	gene 15.3
chr1	AUGUSTUS	CDS	100758644	100759607	0.87	-	1	gene 15.3
chr1	AUGUSTUS	CDS	100759829	100760763	0.96	-	0	gene 15.3
chr1	AUGUSTUS	CDS	134191348	134191490	0.65	-	2	gene 16.1
chr1	AUGUSTUS	CDS	134194561	134195068	1	-	0	gene 16.1
chr1	AUGUSTUS	CDS	134188665	134189534	0.49	-	0	gene 16.2
chr1	AUGUSTUS	CDS	134191432	134191490	0.33	-	2	gene 16.2
chr1	AUGUSTUS	CDS	134194561	134195068	1	-	0	gene 16.2
chr1	AUGUSTUS	CDS	134188665	134189357	0.51	-	0	gene 16.3
chr1	AUGUSTUS	CDS	171096767	171096996	0.98	+	0	gene 17.1
chr1	AUGUSTUS	CDS	171097212	171097277	0.85	+	1	gene 17.1
chr1	AUGUSTUS	CDS	171097482	171097602	0.5	+	1	gene 17.1
chr1	AUGUSTUS	CDS	171100017	171100194	0.31	+	0	gene 17.2
chr1	AUGUSTUS	CDS	171101101	171101678	1	+	2	gene 17.2
chr1	AUGUSTUS	CDS	171096767	171096996	0.98	+	0	gene 17.3
chr1	AUGUSTUS	CDS	171097212	171097277	0.85	+	1	gene 17.3
chr1	AUGUSTUS	CDS	171099086	171099153	0.21	+	1	gene 17.3
chr1	AUGUSTUS	CDS	171101101	171101678	1	+	2	gene 17.3
chr1	AUGUSTUS	CDS	171403636	171403730	0.4	+	0	gene 18.1
chr1	AUGUSTUS	CDS	171403806	171403904	0.88	+	1	gene 18.1
chr1	AUGUSTUS	CDS	171404623	171404797	0.92	+	1	gene 18.1
chr1	AUGUSTUS	CDS	171404900	171405034	0.99	+	0	gene 18.1
chr1	AUGUSTUS	CDS	171405186	171405392	1	+	0	gene 18.1
chr1	AUGUSTUS	CDS	171401526	171401790	1	+	0	gene 18.2
chr1	AUGUSTUS	CDS	171402572	171402669	0.38	+	2	gene 18.2
chr1	AUGUSTUS	CDS	171402859	171403067	0.55	+	0	gene 18.2
chr1	AUGUSTUS	CDS	171403806	171403904	0.88	+	1	gene 18.2
chr1	AUGUSTUS	CDS	171404623	171404797	0.92	+	1	gene 18.2
chr1	AUGUSTUS	CDS	171404900	171405034	0.99	+	0	gene 18.2
chr1	AUGUSTUS	CDS	171405186	171405392	1	+	0	gene 18.2
chr1	AUGUSTUS	CDS	171401526	171401790	1	+	0	gene 18.3
chr1	AUGUSTUS	CDS	171402253	171402266	0.41	+	2	gene 18.3
chr1	AUGUSTUS	CDS	180524695	180524923	0.77	+	0	gene 19.1
chr1	AUGUSTUS	CDS	180525075	180525216	1	+	2	gene 19.1

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	180525357	180525477	0.9	+	1	gene 19.1
chr1	AUGUSTUS	CDS	180525575	180525637	0.99	+	0	gene 19.1
chr1	AUGUSTUS	CDS	180525791	180525937	1	+	0	gene 19.1
chr1	AUGUSTUS	CDS	180521296	180521842	1	+	0	gene 19.2
chr1	AUGUSTUS	CDS	180521923	180521998	1	+	2	gene 19.2
chr1	AUGUSTUS	CDS	180522112	180522227	0.94	+	1	gene 19.2
chr1	AUGUSTUS	CDS	180523526	180524040	0.75	+	2	gene 19.2
chr1	AUGUSTUS	CDS	180521296	180521842	1	+	0	gene 19.3
chr1	AUGUSTUS	CDS	180521923	180521998	1	+	2	gene 19.3
chr1	AUGUSTUS	CDS	180522112	180522227	0.94	+	1	gene 19.3
chr1	AUGUSTUS	CDS	180523526	180524036	0.22	+	2	gene 19.3
chr1	AUGUSTUS	CDS	180524691	180524923	0.22	+	1	gene 19.3
chr1	AUGUSTUS	CDS	180525075	180525216	1	+	2	gene 19.3
chr1	AUGUSTUS	CDS	180525357	180525477	0.9	+	1	gene 19.3
chr1	AUGUSTUS	CDS	180525575	180525637	0.99	+	0	gene 19.3
chr1	AUGUSTUS	CDS	180525791	180525937	1	+	0	gene 19.3
chr1	AUGUSTUS	CDS	189472154	189472512	1	-	2	gene 20.1
chr1	AUGUSTUS	CDS	189472584	189472746	1	-	0	gene 20.1
chr1	AUGUSTUS	CDS	189472821	189473071	0.47	-	2	gene 20.1
chr1	AUGUSTUS	CDS	189473399	189473519	0.77	-	0	gene 20.1
chr1	AUGUSTUS	CDS	189472154	189472512	1	-	2	gene 20.2
chr1	AUGUSTUS	CDS	189472584	189472746	1	-	0	gene 20.2
chr1	AUGUSTUS	CDS	189472821	189473009	0.27	-	0	gene 20.2
chr1	AUGUSTUS	CDS	189473100	189473275	0.47	-	2	gene 20.3
chr1	AUGUSTUS	CDS	189473399	189473519	0.77	-	0	gene 20.3
chr1	AUGUSTUS	CDS	191520283	191521137	0.47	-	0	gene 21.1
chr1	AUGUSTUS	CDS	191521232	191521387	0.91	-	0	gene 21.1
chr1	AUGUSTUS	CDS	191518223	191520187	0.45	-	0	gene 21.2
chr1	AUGUSTUS	CDS	191520322	191521137	0.44	-	0	gene 21.2
chr1	AUGUSTUS	CDS	191521232	191521387	0.91	-	0	gene 21.2
chr1	AUGUSTUS	CDS	191518223	191520133	0.53	-	0	gene 21.3
chr1	AUGUSTUS	CDS	200602835	200602875	0.99	-	2	gene 22.1
chr1	AUGUSTUS	CDS	200602979	200603039	0.81	-	0	gene 22.1
chr1	AUGUSTUS	CDS	200603941	200604505	0.9	-	1	gene 22.1
chr1	AUGUSTUS	CDS	200604619	200604913	0.97	-	2	gene 22.1

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	200605090	200605411	0.57	-	0	gene 22.1
chr1	AUGUSTUS	CDS	200602835	200602875	0.99	-	2	gene 22.2
chr1	AUGUSTUS	CDS	200602979	200603039	0.81	-	0	gene 22.2
chr1	AUGUSTUS	CDS	200603941	200604505	0.9	-	1	gene 22.2
chr1	AUGUSTUS	CDS	200604619	200604913	0.97	-	2	gene 22.2
chr1	AUGUSTUS	CDS	200605090	200605390	0.39	-	0	gene 22.2
chr1	AUGUSTUS	CDS	200605862	200606275	0.96	-	0	gene 22.2
chr1	AUGUSTUS	CDS	200606365	200606404	0.97	-	1	gene 22.2
chr1	AUGUSTUS	CDS	200606700	200606715	0.91	-	2	gene 22.2
chr1	AUGUSTUS	CDS	200606954	200607221	0.88	-	0	gene 22.2
chr1	AUGUSTUS	CDS	200605521	200605526	0.46	-	0	gene 22.3
chr1	AUGUSTUS	CDS	200605862	200606275	0.96	-	0	gene 22.3
chr1	AUGUSTUS	CDS	200606365	200606404	0.97	-	1	gene 22.3
chr1	AUGUSTUS	CDS	200606700	200606715	0.91	-	2	gene 22.3
chr1	AUGUSTUS	CDS	200606954	200607221	0.88	-	0	gene 22.3
chr1	AUGUSTUS	CDS	207315723	207315867	0.35	-	1	gene 23.1
chr1	AUGUSTUS	CDS	207315969	207316072	1	-	0	gene 23.1
chr1	AUGUSTUS	CDS	207312686	207313326	1	-	2	gene 23.2
chr1	AUGUSTUS	CDS	207313456	207313567	0.36	-	0	gene 23.2
chr1	AUGUSTUS	CDS	207312686	207313326	1	-	2	gene 23.3
chr1	AUGUSTUS	CDS	207315248	207315318	0.26	-	1	gene 23.3
chr1	AUGUSTUS	CDS	207315727	207315867	0.65	-	1	gene 23.3
chr1	AUGUSTUS	CDS	207315969	207316072	1	-	0	gene 23.3
chr1	AUGUSTUS	CDS	207387371	207387409	0.58	+	0	gene 24.1
chr1	AUGUSTUS	CDS	207387515	207387716	0.88	+	0	gene 24.1
chr1	AUGUSTUS	CDS	207387820	207387935	0.9	+	2	gene 24.1
chr1	AUGUSTUS	CDS	207388184	207388501	1	+	0	gene 24.1
chr1	AUGUSTUS	CDS	207386685	207386974	0.48	+	0	gene 24.2
chr1	AUGUSTUS	CDS	207387268	207387409	0.28	+	1	gene 24.2
chr1	AUGUSTUS	CDS	207387515	207387716	0.88	+	0	gene 24.2
chr1	AUGUSTUS	CDS	207387820	207387935	0.9	+	2	gene 24.2
chr1	AUGUSTUS	CDS	207388184	207388501	1	+	0	gene 24.2
chr1	AUGUSTUS	CDS	207386685	207386974	0.48	+	0	gene 24.3
chr1	AUGUSTUS	CDS	207387063	207387159	0.33	+	1	gene 24.3
chr1	AUGUSTUS	CDS	214783781	214783951	0.72	+	0	gene 25.1

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	214784221	214785195	0.86	+	0	gene 25.1
chr1	AUGUSTUS	CDS	214787493	214787758	0.62	+	0	gene 25.1
chr1	AUGUSTUS	CDS	214787881	214788445	0.93	+	1	gene 25.1
chr1	AUGUSTUS	CDS	214790428	214790577	0.73	+	0	gene 25.1
chr1	AUGUSTUS	CDS	214783781	214783951	0.72	+	0	gene 25.2
chr1	AUGUSTUS	CDS	214784221	214785195	0.86	+	0	gene 25.2
chr1	AUGUSTUS	CDS	214785569	214785595	0.48	+	0	gene 25.2
chr1	AUGUSTUS	CDS	214787034	214787295	0.36	+	0	gene 25.3
chr1	AUGUSTUS	CDS	214787596	214787758	0.29	+	2	gene 25.3
chr1	AUGUSTUS	CDS	214787881	214788445	0.93	+	1	gene 25.3
chr1	AUGUSTUS	CDS	214790428	214790577	0.73	+	0	gene 25.3
chr1	AUGUSTUS	CDS	216923340	216923480	1	+	0	gene 26.1
chr1	AUGUSTUS	CDS	216923605	216923772	1	+	0	gene 26.1
chr1	AUGUSTUS	CDS	216932888	216932935	0.85	+	0	gene 26.1
chr1	AUGUSTUS	CDS	216933032	216933136	0.73	+	0	gene 26.1
chr1	AUGUSTUS	CDS	216933228	216933319	0.95	+	0	gene 26.1
chr1	AUGUSTUS	CDS	216933402	216933507	0.93	+	1	gene 26.1
chr1	AUGUSTUS	CDS	216933588	216933674	1	+	0	gene 26.1
chr1	AUGUSTUS	CDS	216933777	216934190	0.92	+	0	gene 26.1
chr1	AUGUSTUS	CDS	216931711	216931758	0.28	+	0	gene 26.2
chr1	AUGUSTUS	CDS	216932888	216932935	0.85	+	0	gene 26.2
chr1	AUGUSTUS	CDS	216933032	216933136	0.73	+	0	gene 26.2
chr1	AUGUSTUS	CDS	216933228	216933319	0.95	+	0	gene 26.2
chr1	AUGUSTUS	CDS	216933402	216933507	0.93	+	1	gene 26.2
chr1	AUGUSTUS	CDS	216933588	216933674	1	+	0	gene 26.2
chr1	AUGUSTUS	CDS	216933777	216934190	0.92	+	0	gene 26.2
chr1	AUGUSTUS	CDS	216923340	216923480	1	+	0	gene 26.3
chr1	AUGUSTUS	CDS	216923605	216923772	1	+	0	gene 26.3
chr1	AUGUSTUS	CDS	216923875	216923889	0.2	+	0	gene 26.3
chr1	AUGUSTUS	CDS	217299271	217300004	0.92	+	0	gene 27.1
chr1	AUGUSTUS	CDS	217300101	217300521	0.83	+	1	gene 27.1
chr1	AUGUSTUS	CDS	217298679	217298984	1	+	0	gene 27.2
chr1	AUGUSTUS	CDS	217299271	217300004	0.92	+	0	gene 27.2
chr1	AUGUSTUS	CDS	217300101	217300521	0.83	+	1	gene 27.2
chr1	AUGUSTUS	CDS	217298679	217298984	1	+	0	gene 27.3
chr1	AUGUSTUS	CDS	217299076	217299120	0.63	+	0	gene 27.3

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	217606364	217609450	0.67	-	0	gene 28.1
chr1	AUGUSTUS	CDS	217594550	217594607	0.45	-	1	gene 28.2
chr1	AUGUSTUS	CDS	217598785	217598918	0.58	-	0	gene 28.2
chr1	AUGUSTUS	CDS	217599022	217599159	0.51	-	0	gene 28.2
chr1	AUGUSTUS	CDS	217594550	217594607	0.45	-	1	gene 28.3
chr1	AUGUSTUS	CDS	217598785	217598918	0.58	-	0	gene 28.3
chr1	AUGUSTUS	CDS	217599022	217599252	0.15	-	0	gene 28.3
chr1	AUGUSTUS	CDS	217604758	217604940	0.17	-	0	gene 28.3
chr1	AUGUSTUS	CDS	217606430	217609450	0.17	-	0	gene 28.3
chr1	AUGUSTUS	CDS	226118919	226119412	0.93	-	2	gene 29.1
chr1	AUGUSTUS	CDS	226119728	226119796	0.78	-	2	gene 29.1
chr1	AUGUSTUS	CDS	226121738	226121779	0.89	-	2	gene 29.1
chr1	AUGUSTUS	CDS	226122580	226122733	0.59	-	0	gene 29.1
chr1	AUGUSTUS	CDS	226125785	226126007	0.47	-	1	gene 29.1
chr1	AUGUSTUS	CDS	226126110	226126402	0.65	-	0	gene 29.1
chr1	AUGUSTUS	CDS	226128487	226128588	0.35	-	0	gene 29.1
chr1	AUGUSTUS	CDS	226118919	226119412	0.93	-	2	gene 29.2
chr1	AUGUSTUS	CDS	226119728	226119796	0.78	-	2	gene 29.2
chr1	AUGUSTUS	CDS	226121738	226121779	0.89	-	2	gene 29.2
chr1	AUGUSTUS	CDS	226122580	226122733	0.59	-	0	gene 29.2
chr1	AUGUSTUS	CDS	226125785	226126007	0.47	-	1	gene 29.2
chr1	AUGUSTUS	CDS	226126110	226126402	0.65	-	0	gene 29.2
chr1	AUGUSTUS	CDS	226128487	226128582	0.32	-	0	gene 29.2
chr1	AUGUSTUS	CDS	226129517	226129589	0.57	-	1	gene 29.2
chr1	AUGUSTUS	CDS	226129712	226130010	0.87	-	0	gene 29.2
chr1	AUGUSTUS	CDS	226129349	226129448	0.35	-	1	gene 29.3
chr1	AUGUSTUS	CDS	226129542	226129589	0.43	-	1	gene 29.3
chr1	AUGUSTUS	CDS	226129712	226130010	0.87	-	0	gene 29.3
chr1	AUGUSTUS	CDS	234752186	234752661	0.86	-	2	gene 30.1
chr1	AUGUSTUS	CDS	234752737	234753040	0.38	-	0	gene 30.1
chr1	AUGUSTUS	CDS	234752186	234752661	0.86	-	2	gene 30.2
chr1	AUGUSTUS	CDS	234752737	234753030	0.47	-	2	gene 30.2
chr1	AUGUSTUS	CDS	234754938	234755295	0.46	-	0	gene 30.2
chr1	AUGUSTUS	CDS	234757253	234757753	0.97	-	0	gene 30.2
chr1	AUGUSTUS	CDS	234757884	234758225	0.77	-	0	gene 30.2
chr1	AUGUSTUS	CDS	234758436	234758723	0.55	-	0	gene 30.2

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	234756879	234756884	0.42	-	0	gene 30.3
chr1	AUGUSTUS	CDS	234757253	234757753	0.97	-	0	gene 30.3
chr1	AUGUSTUS	CDS	234757884	234758225	0.77	-	0	gene 30.3
chr1	AUGUSTUS	CDS	234758436	234758723	0.55	-	0	gene 30.3
chr1	AUGUSTUS	CDS	244003087	244003221	0.66	+	0	gene 31.1
chr1	AUGUSTUS	CDS	244003311	244003820	1	+	0	gene 31.1
chr1	AUGUSTUS	CDS	244002394	244002676	1	+	0	gene 31.2
chr1	AUGUSTUS	CDS	244002765	244002991	0.66	+	2	gene 31.2
chr1	AUGUSTUS	CDS	244002394	244002676	1	+	0	gene 31.3
chr1	AUGUSTUS	CDS	244002765	244002983	0.34	+	2	gene 31.3
chr1	AUGUSTUS	CDS	244003073	244003221	0.34	+	2	gene 31.3
chr1	AUGUSTUS	CDS	244003311	244003820	1	+	0	gene 31.3
chr1	AUGUSTUS	CDS	257235438	257235920	0.97	-	0	gene 32.1
chr1	AUGUSTUS	CDS	257236009	257236587	1	-	0	gene 32.1
chr1	AUGUSTUS	CDS	257236694	257237140	0.79	-	0	gene 32.1
chr1	AUGUSTUS	CDS	257237227	257237949	0.66	-	0	gene 32.1
chr1	AUGUSTUS	CDS	257238011	257238688	0.69	-	0	gene 32.2
chr1	AUGUSTUS	CDS	257235438	257235920	0.97	-	0	gene 32.3
chr1	AUGUSTUS	CDS	257236009	257236587	1	-	0	gene 32.3
chr1	AUGUSTUS	CDS	257236694	257237140	0.79	-	0	gene 32.3
chr1	AUGUSTUS	CDS	257237227	257237965	0.24	-	1	gene 32.3
chr1	AUGUSTUS	CDS	257238087	257238688	0.28	-	0	gene 32.3
chr1	AUGUSTUS	CDS	259188166	259188337	1	+	0	gene 33.1
chr1	AUGUSTUS	CDS	259188443	259188545	0.93	+	2	gene 33.1
chr1	AUGUSTUS	CDS	259188815	259189290	0.55	+	1	gene 33.1
chr1	AUGUSTUS	CDS	259189899	259190353	0.49	+	2	gene 33.1
chr1	AUGUSTUS	CDS	259188166	259188337	1	+	0	gene 33.2
chr1	AUGUSTUS	CDS	259188443	259188545	0.93	+	2	gene 33.2
chr1	AUGUSTUS	CDS	259188815	259189447	0.25	+	1	gene 33.2
chr1	AUGUSTUS	CDS	259189560	259189614	0.27	+	1	gene 33.2
chr1	AUGUSTUS	CDS	259189901	259190353	0.51	+	0	gene 33.3
chr1	AUGUSTUS	CDS	266079971	266080055	1	-	1	gene 34.1
chr1	AUGUSTUS	CDS	266081521	266081693	1	-	0	gene 34.1

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	266081811	266081886	1	-	1	gene 34.1
chr1	AUGUSTUS	CDS	266082004	266082116	0.53	-	0	gene 34.1
chr1	AUGUSTUS	CDS	266079971	266080055	1	-	1	gene 34.2
chr1	AUGUSTUS	CDS	266081521	266081693	1	-	0	gene 34.2
chr1	AUGUSTUS	CDS	266081811	266081886	1	-	1	gene 34.2
chr1	AUGUSTUS	CDS	266082004	266082118	0.24	-	2	gene 34.2
chr1	AUGUSTUS	CDS	266084056	266084179	0.95	-	0	gene 34.2
chr1	AUGUSTUS	CDS	266084272	266084334	0.58	-	0	gene 34.2
chr1	AUGUSTUS	CDS	266084576	266084626	0.6	-	0	gene 34.2
chr1	AUGUSTUS	CDS	266086525	266086596	1	-	0	gene 34.2
chr1	AUGUSTUS	CDS	266086699	266086767	1	-	0	gene 34.2
chr1	AUGUSTUS	CDS	266086904	266087049	0.99	-	2	gene 34.2
chr1	AUGUSTUS	CDS	266087174	266087237	1	-	0	gene 34.2
chr1	AUGUSTUS	CDS	266083734	266083826	0.47	-	0	gene 34.3
chr1	AUGUSTUS	CDS	266083907	266083968	0.53	-	2	gene 34.3
chr1	AUGUSTUS	CDS	266084056	266084179	0.95	-	0	gene 34.3
chr1	AUGUSTUS	CDS	266084272	266084334	0.58	-	0	gene 34.3
chr1	AUGUSTUS	CDS	266084576	266084626	0.6	-	0	gene 34.3
chr1	AUGUSTUS	CDS	266086525	266086596	1	-	0	gene 34.3
chr1	AUGUSTUS	CDS	266086699	266086767	1	-	0	gene 34.3
chr1	AUGUSTUS	CDS	266086904	266087049	0.99	-	2	gene 34.3
chr1	AUGUSTUS	CDS	266087174	266087237	1	-	0	gene 34.3
chr1	AUGUSTUS	CDS	268180407	268180555	0.99	-	2	gene 35.1
chr1	AUGUSTUS	CDS	268180660	268180747	1	-	0	gene 35.1
chr1	AUGUSTUS	CDS	268180839	268180990	1	-	2	gene 35.1
chr1	AUGUSTUS	CDS	268181073	268181232	1	-	0	gene 35.1
chr1	AUGUSTUS	CDS	268181345	268181500	1	-	0	gene 35.1
chr1	AUGUSTUS	CDS	268181610	268181828	1	-	0	gene 35.1
chr1	AUGUSTUS	CDS	268182335	268182631	1	-	0	gene 35.1
chr1	AUGUSTUS	CDS	268182740	268182946	1	-	0	gene 35.1
chr1	AUGUSTUS	CDS	268183097	268183222	1	-	0	gene 35.1
chr1	AUGUSTUS	CDS	268177883	268178230	1	-	0	gene 35.2
chr1	AUGUSTUS	CDS	268178313	268178461	1	-	2	gene 35.2
chr1	AUGUSTUS	CDS	268178566	268178653	1	-	0	gene 35.2
chr1	AUGUSTUS	CDS	268178756	268178907	1	-	2	gene 35.2
chr1	AUGUSTUS	CDS	268178991	268179105	0.39	-	0	gene 35.2
chr1	AUGUSTUS	CDS	268177883	268178230	1	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268178313	268178461	1	-	2	gene 35.3
chr1	AUGUSTUS	CDS	268178566	268178653	1	-	0	gene 35.3

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	268178756	268178907	1	-	2	gene 35.3
chr1	AUGUSTUS	CDS	268178991	268179150	0.61	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268179427	268179500	0.58	-	2	gene 35.3
chr1	AUGUSTUS	CDS	268179652	268179755	0.58	-	1	gene 35.3
chr1	AUGUSTUS	CDS	268179981	268180324	0.49	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268180407	268180555	0.99	-	2	gene 35.3
chr1	AUGUSTUS	CDS	268180660	268180747	1	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268180839	268180990	1	-	2	gene 35.3
chr1	AUGUSTUS	CDS	268181073	268181232	1	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268181345	268181500	1	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268181610	268181828	1	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268182335	268182631	1	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268182740	268182946	1	-	0	gene 35.3
chr1	AUGUSTUS	CDS	268183097	268183222	1	-	0	gene 35.3
chr1	AUGUSTUS	CDS	280826651	280826908	0.2	-	0	gene 36.1
chr1	AUGUSTUS	CDS	280827449	280827611	1	-	1	gene 36.1
chr1	AUGUSTUS	CDS	280829256	280829351	0.98	-	1	gene 36.1
chr1	AUGUSTUS	CDS	280829467	280829576	1	-	0	gene 36.1
chr1	AUGUSTUS	CDS	280823262	280823343	0.69	-	1	gene 36.2
chr1	AUGUSTUS	CDS	280824053	280824173	0.6	-	2	gene 36.2
chr1	AUGUSTUS	CDS	280825153	280825307	0.51	-	1	gene 36.2
chr1	AUGUSTUS	CDS	280825403	280825881	0.86	-	0	gene 36.2
chr1	AUGUSTUS	CDS	280825951	280826391	0.52	-	0	gene 36.2
chr1	AUGUSTUS	CDS	280827449	280827611	1	-	1	gene 36.2
chr1	AUGUSTUS	CDS	280829256	280829351	0.98	-	1	gene 36.2
chr1	AUGUSTUS	CDS	280829467	280829576	1	-	0	gene 36.2
chr1	AUGUSTUS	CDS	280823262	280823343	0.69	-	1	gene 36.3
chr1	AUGUSTUS	CDS	280824053	280824173	0.6	-	2	gene 36.3
chr1	AUGUSTUS	CDS	280825153	280825307	0.51	-	1	gene 36.3
chr1	AUGUSTUS	CDS	280825403	280825881	0.86	-	0	gene 36.3
chr1	AUGUSTUS	CDS	280825951	280826289	0.29	-	0	gene 36.3
chr1	AUGUSTUS	CDS	282672702	282673160	1	+	0	gene 37.1
chr1	AUGUSTUS	CDS	282677336	282678117	0.74	+	0	gene 37.1
chr1	AUGUSTUS	CDS	282678282	282678316	0.69	+	1	gene 37.1
chr1	AUGUSTUS	CDS	282679461	282679522	0.76	+	2	gene 37.1
chr1	AUGUSTUS	CDS	282680909	282680952	0.42	+	0	gene 37.1
chr1	AUGUSTUS	CDS	282682405	282682525	0.77	+	1	gene 37.1
chr1	AUGUSTUS	CDS	282683694	282683981	0.83	+	0	gene 37.1
chr1	AUGUSTUS	CDS	282672702	282673160	1	+	0	gene 37.2

Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	282677336	282678121	0.26	+	0	gene 37.2
chr1	AUGUSTUS	CDS	282682301	282682326	0.32	+	0	gene 37.3
chr1	AUGUSTUS	CDS	282682405	282682525	0.77	+	1	gene 37.3
chr1	AUGUSTUS	CDS	282683694	282683981	0.83	+	0	gene 37.3
chr1	AUGUSTUS	CDS	284056282	284056540	1	-	1	gene 38.1
chr1	AUGUSTUS	CDS	284056725	284056909	0.27	-	0	gene 38.1
chr1	AUGUSTUS	CDS	284058043	284058072	0.21	-	0	gene 38.2
chr1	AUGUSTUS	CDS	284058255	284058593	0.84	-	0	gene 38.2
chr1	AUGUSTUS	CDS	284059037	284059096	0.99	-	0	gene 38.2
chr1	AUGUSTUS	CDS	284059196	284059222	1	-	0	gene 38.2
chr1	AUGUSTUS	CDS	284056282	284056540	1	-	1	gene 38.3
chr1	AUGUSTUS	CDS	284056725	284056965	0.36	-	2	gene 38.3
chr1	AUGUSTUS	CDS	284057635	284057707	0.25	-	0	gene 38.3
chr1	AUGUSTUS	CDS	284058255	284058593	0.84	-	0	gene 38.3
chr1	AUGUSTUS	CDS	284059037	284059096	0.99	-	0	gene 38.3
chr1	AUGUSTUS	CDS	284059196	284059222	1	-	0	gene 38.3
chr1	AUGUSTUS	CDS	286290988	286291175	0.48	+	0	gene 39.1
chr1	AUGUSTUS	CDS	286292735	286292869	0.56	+	1	gene 39.1
chr1	AUGUSTUS	CDS	286293493	286294054	0.65	+	1	gene 39.1
chr1	AUGUSTUS	CDS	286296013	286296370	0.98	+	0	gene 39.1
chr1	AUGUSTUS	CDS	286296660	286296916	0.38	+	2	gene 39.1
chr1	AUGUSTUS	CDS	286282502	286283965	0.36	+	0	gene 39.2
chr1	AUGUSTUS	CDS	286285920	286286277	0.83	+	0	gene 39.2
chr1	AUGUSTUS	CDS	286290998	286291175	0.23	+	2	gene 39.2
chr1	AUGUSTUS	CDS	286292735	286292869	0.56	+	1	gene 39.2
chr1	AUGUSTUS	CDS	286293493	286294054	0.65	+	1	gene 39.2
chr1	AUGUSTUS	CDS	286296013	286296370	0.98	+	0	gene 39.2
chr1	AUGUSTUS	CDS	286296660	286296916	0.38	+	2	gene 39.2
chr1	AUGUSTUS	CDS	286282502	286283965	0.36	+	0	gene 39.3
chr1	AUGUSTUS	CDS	286285920	286286277	0.83	+	0	gene 39.3
chr1	AUGUSTUS	CDS	286286924	286286928	0.41	+	2	gene 39.3
chr1	AUGUSTUS	CDS	292270620	292271786	0.88	-	0	gene 40.1
chr1	AUGUSTUS	CDS	292274660	292274856	0.61	-	2	gene 40.1
chr1	AUGUSTUS	CDS	292274964	292275576	0.59	-	0	gene 40.1
chr1	AUGUSTUS	CDS	292275801	292276010	0.88	-	0	gene 40.1

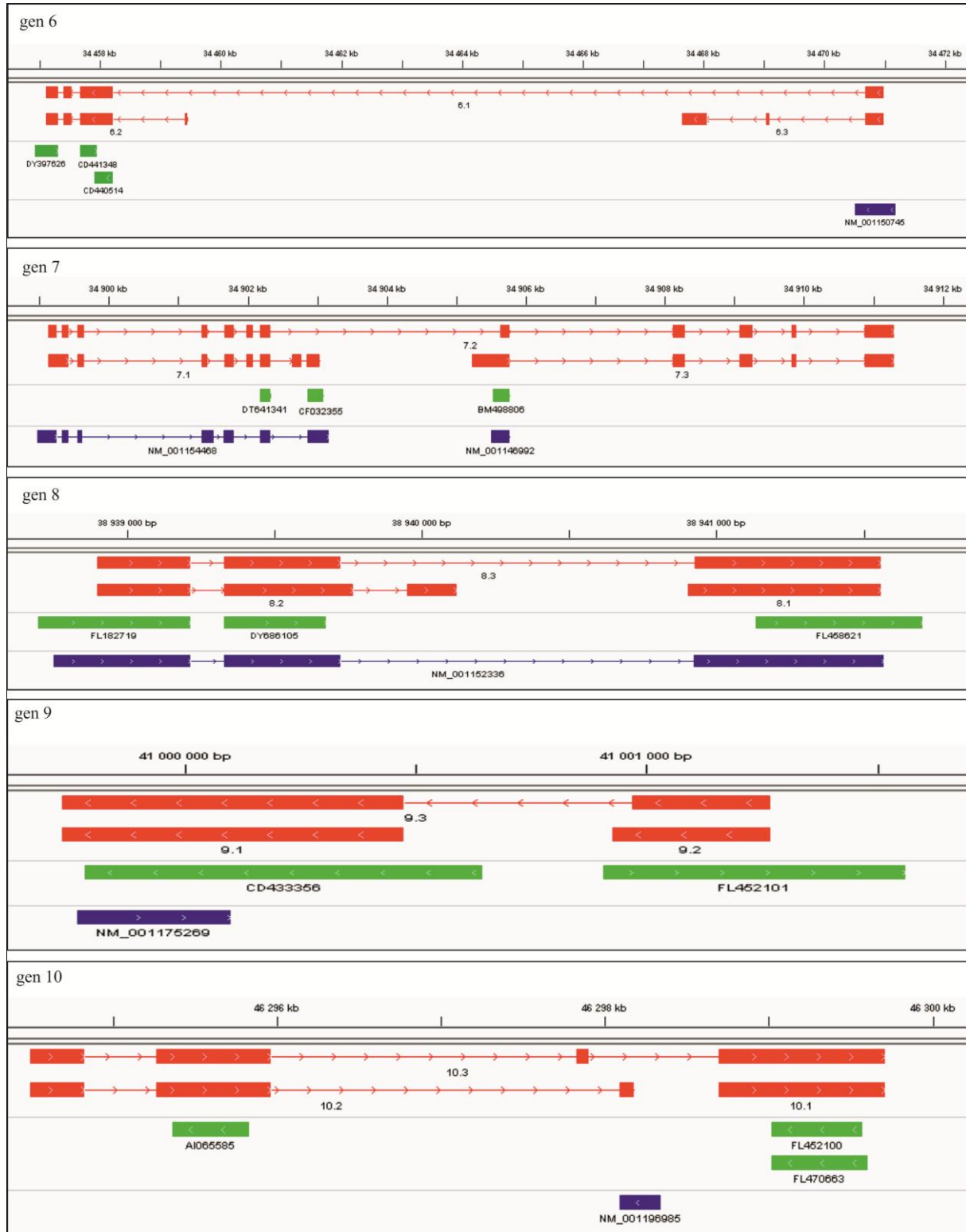
Aneks 1 Koordynaty puli 42 genów przewidzianych przez program AUGUSTUS

chr1	AUGUSTUS	CDS	292274953	292275576	0.33	-	0	gene 40.2
chr1	AUGUSTUS	CDS	292275801	292276010	0.88	-	0	gene 40.2
chr1	AUGUSTUS	CDS	292270620	292271786	0.88	-	0	gene 40.3
chr1	AUGUSTUS	CDS	292274660	292274677	0.22	-	0	gene 40.3
chr1	AUGUSTUS	CDS	297665649	297665661	0.64	-	1	gene 41.1
chr1	AUGUSTUS	CDS	297665777	297665910	0.86	-	0	gene 41.1
chr1	AUGUSTUS	CDS	297665999	297666135	0.94	-	2	gene 41.1
chr1	AUGUSTUS	CDS	297666636	297666786	0.43	-	0	gene 41.1
chr1	AUGUSTUS	CDS	297665649	297665661	0.64	-	1	gene 41.2
chr1	AUGUSTUS	CDS	297665777	297665910	0.86	-	0	gene 41.2
chr1	AUGUSTUS	CDS	297665999	297666135	0.94	-	2	gene 41.2
chr1	AUGUSTUS	CDS	297666636	297666913	0.4	-	1	gene 41.2
chr1	AUGUSTUS	CDS	297667362	297667414	0.4	-	0	gene 41.2
chr1	AUGUSTUS	CDS	297667531	297667746	0.54	-	0	gene 41.2
chr1	AUGUSTUS	CDS	297667241	297667414	0.43	-	0	gene 41.3
chr1	AUGUSTUS	CDS	297667531	297667746	0.54	-	0	gene 41.3
chr1	AUGUSTUS	CDS	299599133	299599222	0.53	+	0	gene 42.1
chr1	AUGUSTUS	CDS	299599802	299600083	0.73	+	0	gene 42.1
chr1	AUGUSTUS	CDS	299600405	299600535	0.2	+	0	gene 42.2
chr1	AUGUSTUS	CDS	299600757	299600879	0.71	+	1	gene 42.2
chr1	AUGUSTUS	CDS	299601057	299601215	0.37	+	1	gene 42.2
chr1	AUGUSTUS	CDS	299601384	299601586	0.87	+	1	gene 42.2
chr1	AUGUSTUS	CDS	299610755	299612148	0.85	+	2	gene 42.2
chr1	AUGUSTUS	CDS	299614107	299614464	0.75	+	0	gene 42.2
chr1	AUGUSTUS	CDS	299614754	299615010	0.91	+	2	gene 42.2
chr1	AUGUSTUS	CDS	299599151	299599222	0.43	+	0	gene 42.3
chr1	AUGUSTUS	CDS	299599802	299599956	0.26	+	0	gene 42.3
chr1	AUGUSTUS	CDS	299600757	299600879	0.71	+	1	gene 42.3
chr1	AUGUSTUS	CDS	299601057	299601149	0.3	+	1	gene 42.3
chr1	AUGUSTUS	CDS	299601384	299601586	0.87	+	1	gene 42.3
chr1	AUGUSTUS	CDS	299610755	299612148	0.85	+	2	gene 42.3
chr1	AUGUSTUS	CDS	299614107	299614464	0.75	+	0	gene 42.3
chr1	AUGUSTUS	CDS	299614754	299615010	0.91	+	2	gene 42.3

Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



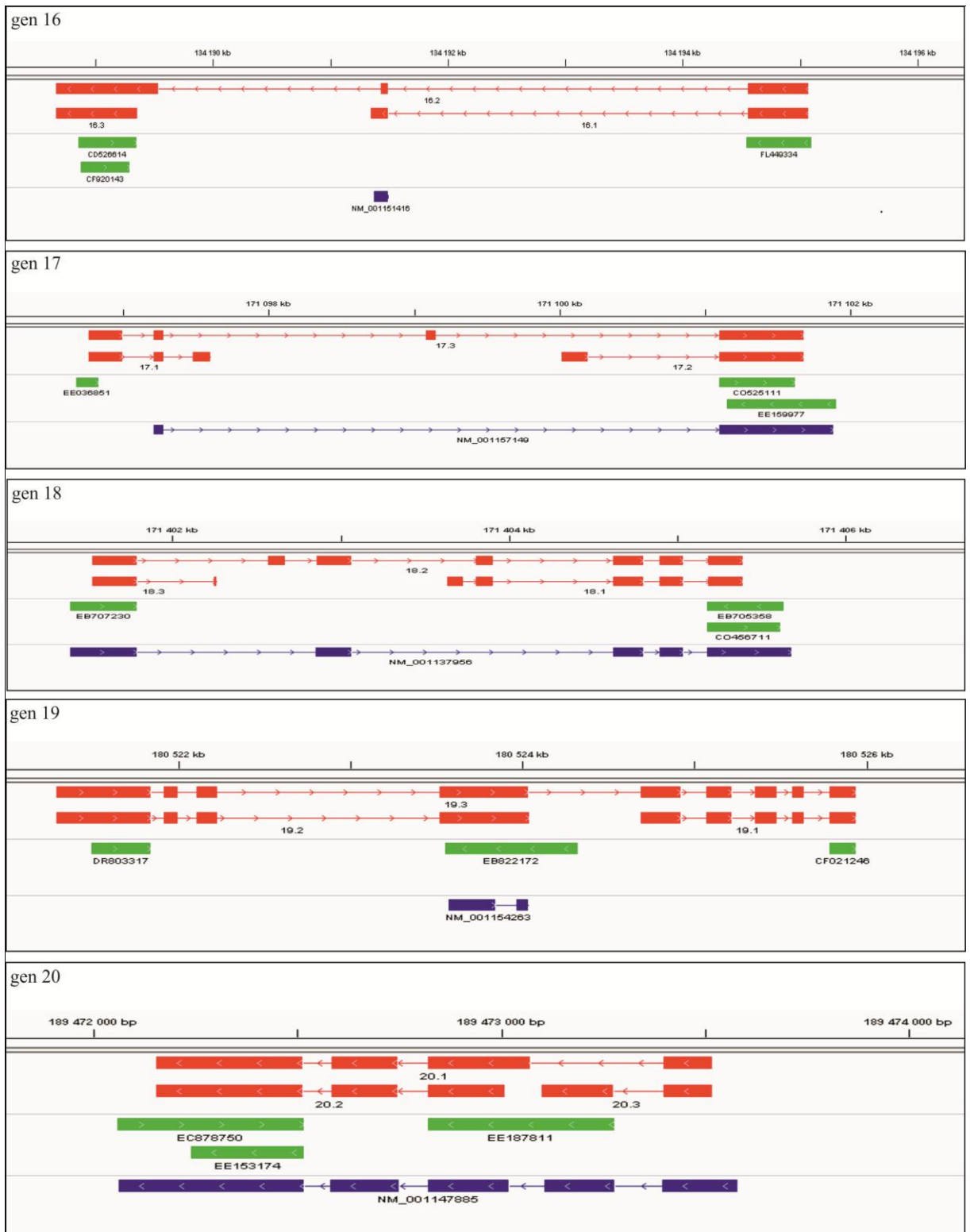
Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



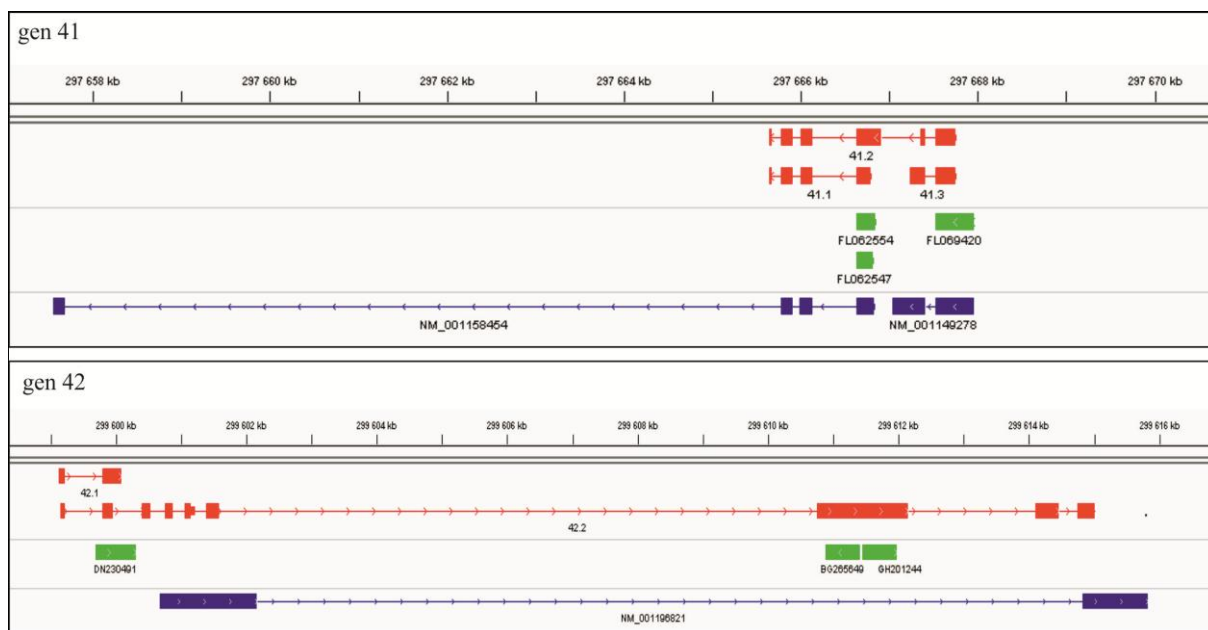
Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



Aneks 2 Graficzne przedstawienie transkryptów 42 genów przewidzianych przez program AUGUSTUS



Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

Numer identyfikacyjny genu	Locus	Kodowana domena białkowa (wraz z numerem identyfikacyjnym)	Zlogarytmowana zmiana poziomu ekspresji SL1	Zlogarytmowana zmiana poziomu ekspresji SL7
GRMZM2G700208	10:33339390-33404487	PF00657: Lipaza GDSL/Acyłhydrolaza	2.12769	2.25479
GRMZM5G808836/ GRMZM5G833207	3:205769416-205781129	PF00005: ABC transporter, PF09818: Domena ATPazowa związana z ABC transporterami	1.04.1766	7.28844
GRMZM2G010044	4:238877211-238881188	PF00330: Rodzina akonitaz	5.22086	5.04959
GRMZM2G133988	4:238487229-238559507	PF00330: Rodzina akonitaz	4.81164	5.20674
GRMZM2G363052	7:85636679-85639723	PF00847: Domena Apetala 2	2.51939	4.54232
GRMZM2G045239	8:168021434-168025531	Nie zidentyfikowano	3.13642	5.05772

Zestawienie informacji na temat 6 genów o ekspresji zmienionej w linii wrażliwej S79757 w obu badanych punktach czasowych pod wpływem oprysku herbicydem Roundup®.

Numer identyfikacyjny genu	Locus	Kodowana domena białkowa (wraz z numerem identyfikacyjnym)	Zlogarytmowana zmiana poziomu ekspresji TL1	Zlogarytmowana zmiana poziomu ekspresji TL7
GRMZM2G008053, zma-MIR528a	1:6414992-6420909	Brak	-6.18078	-6.17429
GRMZM2G062724	1:10033093-10037543	PF05495: Domena palca cynkowego CHY	3.92592	3.25049
asn3	1:45114257-45126350	B6ETR5: Syntetaza asparaginy 3	7.87405	5.03012
GRMZM2G038893	1:166799683-166810410	PF00069: Domena kinazy proteinowej, PF0771 : Białkowa kinaza tyrozynowa	1.02.7396	2.97234
GRMZM2G169535	1:180728677-180729284	Nie zidentyfikowano	5.28277	4.81907
GRMZM2G477872	1:197070826-197080079	PF00005: ABC transporter, PF0106: trenasporter typu, PF08370: Roślinna domena związana plejotropową odpornością na leki (transporter ABC)	3.30938	4.07062
GRMZM2G445343	1:217831271-217831586	Nie zidentyfikowano	8.30912	3.82925
AC207347.3_FG005, MFSP	1:252793889-252818048	PF07690: Nadrodzina głównych koordynatorów, PF00083:	1.59475	1.49689

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

		Transporter cukru		
AC177897.2_FG002, GRMZM2G134973	1:260133894 -260139207	PF03105: Domena SPX	2.01095	2.03236
XLOC_009842	1:43979812- 43982812	PF01679: Modulator potencjału membran białkowo-lipidowych	1.07.2732	6.94243
asn3 (locus) - syntetaza asparaginy 3	1:45114257- 45126350	Nie zidentyfikowano	8.02211	6.40493
GRMZM2G011240 GRMZM2G311182 GRMZM2G318780	1:56746420- 56792250	PF00862: Syntetaza sacharozy, PF00534: Grupa 1 transferaz glikozylowych	2.98886	2.90497
GRMZM2G040441	1:84537954- 84543910	PF03094: Rodzina białek MLO	-3.93697	-3.54929
GRMZM2G105189	1:246200684 -246205255	PF00657: Lipaza GDSL/Acyłhydrolaza	-5.74732	-5.39935
GRMZM2G042895	10:76586622 -76588863	PF00010: Domena wiążąca DNA typu helisa-pętla-helisa	-4.24724	-1.04.1306
GRMZM2G001853	10:10934545 9-109346707	Nie zidentyfikowano	8.12921	1.07.646
GRMZM2G092427	10:14789056 1-147892066	PF00504: Białko wiążące chlorofil A-B	-3.17699	-3.52564
GRMZM2G094375	10:12166678 -12169103	PF07732, PF00394: oksydaza miedziowa	6.84204	-4.49162
TPS6	10:55603291 -55756434	PFAM ID: PF01397: N-końcowa domena syntetazy terpenów,	5.71294	13.1122

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

		PF03936: Rodzina syntetaz terpenów, domena wiążąca metale		
GRMZM2G051921 GRMZM2G051943	2:33506947-33535449	PF00187: Białko rozpoznające chitynę, PF00182: Chitynaza klasy I	5.92719	7.75292
GRMZM2G166776	2:224064455-224065350	Nie zidentyfikowano	6.48662	8.45772
GRMZM2G032977	2:4847230-4850042	PF02265: Nukleaza S1/P1	4.45273	4.11087
GRMZM2G130008 GRMZM2G575328 GRMZM2G584110 GRMZM5G840111	2:7842898-8223839	Nie zidentyfikowano	7.02916	4.00358
GRMZM2G086430	2:16348455-16355671	PF03105: Domena SPX, PF07690: Nadrodzina głównych koordynatorów,	1.03.5641	3.41901
GRMZM2G053206	2:7842898-8223839	Nie zidentyfikowano	4.61133	5.29987
GRMZM2G339866 GRMZM2G506430	3:32463053-32534797	PF11820: Domena o nieznanych funkcjach (DUF3339)	-6.74024	-6.03651
GRMZM2G094712	3:195312989-195319148	PF00155: Aminotransferaza klasy I i II	2.96368	3.37166
AC208915.3_FG010 CAB48	3:223353862-223534761	PF00504: Białko wiążące chlorofil A-B	-4.34747	-4.84249

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

GRMZM2G104549				
GRMZM2G120661				
GRMZM2G429955				
GRMZM2G546097				
L-FNRII	3:147285154 -147293158	BAA88237.1: Ferredoksyna	-2.24668	-2.31248
GRMZM2G138258 GRMZM2G138338	3:147397872 -147407494	PF08263: N-końcowa domena leucynowa, PF00560: Domena bogata w powtórzenia leucynowe, PF07714: Białkowa kinaza tyrozynowa, PF00069: Domena kinazy białkowej	-5.41979	-1.04.2021
GRMZM2G092474	3:186793463 -186794596	PF00314: Rodzina taumatyn	5.26858	7.56614
GRMZM2G121878 GRMZM2G348512	3:215424678 -215477991	PF00484: Anhydraza węglanowa	-2.76167	-2.32402
GRMZM2G310755 GRMZM2G406438 GRMZM2G423366 GRMZM2G429938	3:223353862 -223534761	Nie zidentyfikowano	-4.73574	-4.46644
GRMZM2G020801	4:28116851- 28125443	PF00330: Rodzina akonitaz, PF00694: C-końcowa domena akonitazy	1.56943	1.68237
GRMZM2G168365	4:96310950- 96312854	PF03083: Transporter błonowy biorący udział w międzykomórkowej wymianie cukru,	1.06.9164	6.77425

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

		PF04193: Domena z powtórzoną pętlą PQ		
GRMZM2G109830 GRMZM2G165387	4:138361886 -138661895	PF00069: Domena kinazy proteinowej, PF07714: Białkowa kinaza tyrozynowa	-5.5257	-7.20603
GRMZM2G070508	4:229843442 -229849297	PF00067: Cytochrom P450	-3.44887	-5.31204
A2	5:66107720- 66115045	PF03171: Nadrodzina oksygenaz zależnych od żelaza (II) I 2-oksylglutaranu	5.61581	5.76345
GRMZM2G107544	5:5332470- 5333562	Nie zidentyfikowano	7.27891	1.05.9737
GRMZM2G025833	5:30766577- 30777461	PF01373: Rodzina hydrolaz glikozylowych (14)	5.79891	1.05.1676
TPS2	5:71262954- 71266273	PF01397: N-końcowa domena syntetazy terpenów, PF03936: Rodzina syntetaz terpenów, domena wiążąca metale	-1.79769e +308	-3.69437
GRMZM2G007466	5:183121076 -183127363	PF00023: Powtórzenie ankirynowe, PF07714: Białkowa kinaza tyrozynowa, PF00069: Domena kinazy białkowej	-2.98819	-2.26478
GRMZM2G180328	6:147868150 -147870321	PF02365: Białko NAM	3.44637	2.92039

GRMZM2G096475	6:161966267 -161967646	PF02987: Białka późnej embriogenezy	8.60217	8.92328
FDX1 FDX5 pFD2	6:1339636- 1374388	PF00111: Domena wiążąca Fe ₂ S ₂ (rodzina ferredoksyn)	-2.9041	-2.69957
glp1 (locus)	6:123777181 -123778357	Nie zidentyfikowano	-5.20798	-5.18888
GRMZM2G061469	6:134648482 -134651586	PF03595: Kanał anionowy zależny od napięcia	4.31008	5.07223
DHN1	6:137142552 -137144384	PF00257: Dehydryna	5.01235	8.55118
GRMZM2G162359	7:10651270- 10653006	PF00704: Rodzina hydrolaz glikozylowych (18)	2.92787	2.67243
GRMZM2G056369	7:154185082 -154191068	PF00463: Rodzina liaz izocytrynionowych	5.08804	-4.49008
Lhcb4*1/Lhcb48- BE10/Lhcb48-BE10	7:157276116 -157581165	chlorophyll a/b binding protein precursor	-2.615	-3.35214
GRMZM2G174192	7:172655998 -172665181	PF00201: UDP- glukuronylotransfera za i UDP- glukozylotransferaza	8.86182	6.30934
GRMZM2G132547	8:147641920 -147652575	PF06418: N-końcowa domena syntetazy CTP, PF00117: Amidotransferaza glutaminy klasa I	2.66592	2.16304

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

GRMZM2G058491	8:27636813-27643204	PF01695: wiążąca ATP zbliżona do IstB, PF00004: Rodzina ATPaz związanych z różnymi aktywnościami komórki (AAA), PF05496: N-końcowa domena helikazy ruvB DNA związanej ze strukturą Holliday'a, PF07728: Domena AAA (zbliżona do nadrodziny dyneiny)	3.35137	2.10259
AC197705.4_FG003	8:118103039-118103810	PF01161: Białko wiążące fosfatydyloetanoleinę	4.25609	4.57508
GRMZM2G118770	8:175058637-175063179	PF00390: N-końcowa domena dehydrogenazy jabłczanu, PF03949: Domena wiążąca NAD dehydrogenazy jabłczanu	5.66932	6.14265
PEP1	9:61295925-61301850	PF00311: Karoboksylaza fosfoenolopirogronianu	-2.91311	-3.39408
GRMZM2G366532	9:139510669-139519756	PF00012: Białko Hsp70, PF06723: Białko MreB/Mbl	6.12962	7.22259
GRMZM2G173685	9:20459073-20460609	Nie zidentyfikowano	8.97107	3.33827
GRMZM2G006973	9:24596811-24600134	PF00023: Powtórzenie ankirowane, PF07714: Białkowa	3.86658	3.63806

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

		kinaza tyrozynowa, PF00069: Domena kinazy białkowej		
asn4 (locus) – syntetaza asparaginy 4	9:137784209-137791159	Nie zidentyfikowano	6.53178	4.03181
Numer identyfikacyjny genu	Locus	Kodowana domena białkowa (wraz z numerem identyfikacyjnym)	Zlogarytmowana zmiana poziomu ekspresji TL1	Zlogarytmowana zmiana poziomu ekspresji TL7
GRMZM2G008053, zma-MIR528a	1:6414992-6420909	Brak	-6.18078	-6.17429
GRMZM2G062724	1:10033093-10037543	PF05495: Domena palca cynkowego CHY	3.92592	3.25049
asn3	1:45114257-45126350	B6ETR5: Syntetaza asparaginy 3	7.87405	5.03012
GRMZM2G038893	1:166799683-166810410	PF00069: Domena kinazy proteinowej, PF0771 : Białkowa kinaza tyrozynowa	1.02.7396	2.97234
GRMZM2G169535	1:180728677-180729284	Nie zidentyfikowano	5.28277	4.81907
GRMZM2G477872	1:197070826-197080079	PF00005: ABC transporter, PF0106: trenasporter typu, PF08370: Roślinna domena związana plejotropową odpornością na leki (transporter ABC)	3.30938	4.07062

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

GRMZM2G445343	1:217831271 -217831586	Nie zidentyfikowano	8.30912	3.82925
AC207347.3_FG005, MFSP	1:252793889 -252818048	PF07690: Nadrodzina głównych koordynatorów, PF00083: Transporter cukru	1.59475	1.49689
AC177897.2_FG002, GRMZM2G134973	1:260133894 -260139207	PF03105: Domena SPX	2.01095	2.03236
XLOC_009842	1:43979812- 43982812	PF01679: Modulator potencjału membran białkowo-lipidowych	1.07.2732	6.94243
asn3 (locus) - syntetaza asparaginy 3	1:45114257- 45126350	Nie zidentyfikowano	8.02211	6.40493
GRMZM2G011240 GRMZM2G311182 GRMZM2G318780	1:56746420- 56792250	PF00862: Syntetaza sacharozy, PF00534: Grupa 1 transferaz glykozylowych	2.98886	2.90497
GRMZM2G040441	1:84537954- 84543910	PF03094: Rodzina białek MLO	-3.93697	-3.54929
GRMZM2G105189	1:246200684 -246205255	PF00657: Lipaza GDSL/Acyłhydrolaza	-5.74732	-5.39935
GRMZM2G042895	10:76586622 -76588863	PF00010: Domena wiążąca DNA typu helisa-pętla-helisa	-4.24724	-1.04.1306
GRMZM2G001853	10:10934545 9-109346707	Nie zidentyfikowano	8.12921	1.07.646
GRMZM2G092427	10:14789056 1-147892066	PF00504: Białko wiążące chlorofil A-B	-3.17699	-3.52564

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

GRMZM2G094375	10:12166678 -12169103	PF07732, PF00394: oksydaza miedziowa	6.84204	-4.49162
TPS6	10:55603291 -55756434	PFAM ID: PF01397: N-końcowa domena syntetazy terpenów, PF03936: Rodzina syntetaz terpenów, domena wiążąca metale	5.71294	13.1122
GRMZM2G051921 GRMZM2G051943	2:33506947- 33535449	PF00187: Białko rozpoznające chitynę, PF00182: Chitynaza klasy I	5.92719	7.75292
GRMZM2G166776	2:224064455 -224065350	Nie zidentyfikowano	6.48662	8.45772
GRMZM2G032977	2:4847230- 4850042	PF02265: Nukleaza S1/P1	4.45273	4.11087
GRMZM2G130008 GRMZM2G575328 GRMZM2G584110 GRMZM5G840111	2:7842898- 8223839	Nie zidentyfikowano	7.02916	4.00358
GRMZM2G086430	2:16348455- 16355671	PF03105: Domena SPX, PF07690: Nadrodzina głównych koordynatorów,	1.03.5641	3.41901
GRMZM2G053206	2:7842898- 8223839	Nie zidentyfikowano	4.61133	5.29987
GRMZM2G339866 GRMZM2G506430	3:32463053- 32534797	PF11820: Domena o nieznanych funkcjach (DUF3339)	-6.74024	-6.03651

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

GRMZM2G094712	3:195312989 -195319148	PF00155: Aminotransferaza klasy I i II	2.96368	3.37166
AC208915.3_FG010 CAB48 GRMZM2G104549 GRMZM2G120661 GRMZM2G429955 GRMZM2G546097	3:223353862 -223534761	PF00504: Białko wiążące chlorofil A-B	-4.34747	-4.84249
L-FNR II	3:147285154 -147293158	BAA88237.1: Ferredoksyna	-2.24668	-2.31248
GRMZM2G138258 GRMZM2G138338	3:147397872 -147407494	PF08263: N-końcowa domena leucynowa, PF00560: Domena bogata w powtórzenia leucynowe, PF07714: Białkowa kinaza tyrozynowa, PF00069: Domena kinazy białkowej	-5.41979	-1.04.2021
GRMZM2G092474	3:186793463 -186794596	PF00314: Rodzina taumatyn	5.26858	7.56614
GRMZM2G121878 GRMZM2G348512	3:215424678 -215477991	PF00484: Anhydraza węglanowa	-2.76167	-2.32402
GRMZM2G310755 GRMZM2G406438 GRMZM2G423366 GRMZM2G429938	3:223353862 -223534761	Nie zidentyfikowano	-4.73574	-4.46644
GRMZM2G020801	4:28116851- 28125443	PF00330: Rodzina akonitaz, PF00694:	1.56943	1.68237

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

		C-końcowa domena akonitazy		
GRMZM2G168365	4:96310950-96312854	PF03083: Transporter błonowy biorący udział w międzykomórkowej wymianie cukru, PF04193: Domena z powtórzoną pętlą PQ	1.06.9164	6.77425
GRMZM2G109830 GRMZM2G165387	4:138361886-138661895	PF00069: Domena kinazy proteinowej, PF07714: Białkowa kinaza tyrozynowa	-5.5257	-7.20603
GRMZM2G070508	4:229843442-229849297	PF00067: Cytochrom P450	-3.44887	-5.31204
A2	5:66107720-66115045	PF03171: Nadrodzina oksygenaz zależnych od żelaza (II) I 2-oksylglutaranu	5.61581	5.76345
GRMZM2G107544	5:5332470-5333562	Nie zidentyfikowano	7.27891	1.05.9737
GRMZM2G025833	5:30766577-30777461	PF01373: Rodzina hydrolaz glikozylowych (14)	5.79891	1.05.1676
TPS2	5:71262954-71266273	PF01397: N-końcowa domena syntetazy terpenów, PF03936: Rodzina syntetaz terpenów, domena wiążąca metale	-1.79769e +308	-3.69437
GRMZM2G007466	5:183121076-183127363	PF00023: Powtórzenie ankirynowe, PF07714: Białkowa	-2.98819	-2.26478

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

		kinaza tyrozynowa, PF00069: Domena kinazy białkowej		
GRMZM2G180328	6:147868150 -147870321	PF02365: Białko NAM	3.44637	2.92039
GRMZM2G096475	6:161966267 -161967646	PF02987: Białka późnej embriogenezy	8.60217	8.92328
FDX1 FDX5 pFD2	6:1339636- 1374388	PF00111: Domena wiążąca Fe ₂ S ₂ (rodzina ferredoksyn)	-2.9041	-2.69957
glp1 (locus)	6:123777181 -123778357	Nie zidentyfikowano	-5.20798	-5.18888
GRMZM2G061469	6:134648482 -134651586	PF03595: Kanał anionowy zależny od napięcia	4.31008	5.07223
DHN1	6:137142552 -137144384	PF00257: Dehydryna	5.01235	8.55118
GRMZM2G162359	7:10651270- 10653006	PF00704: Rodzina hydrolaz glikozylowych (18)	2.92787	2.67243
GRMZM2G056369	7:154185082 -154191068	PF00463: Rodzina liaz izocytrynionowych	5.08804	-4.49008
Lhcb4*1/Lhcb48- BE10/Lhcb48-BE10	7:157276116 -157581165	chlorophyll a/b binding protein precursor	-2.615	-3.35214
GRMZM2G174192	7:172655998 -172665181	PF00201: UDP- glukuronylotransfera za i UDP-	8.86182	6.30934

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

		glukozylotransferaza		
GRMZM2G132547	8:147641920 -147652575	PF06418: N-końcowa domena syntetazy CTP, PF00117: Amidotransferaza glutaminy klasa I	2.66592	2.16304
GRMZM2G058491	8:27636813- 27643204	PF01695: wiążąca ATP zbliżona do IstB, PF00004: Rodzina ATPaz związanych z różnymi aktywnościami komórki (AAA), PF05496: N-końcowa domena helikazy ruvB DNA związanej ze strukturą Holliday'a, PF07728: Domena AAA (zbliżona do nadrodziny dyneiny)	3.35137	2.10259
AC197705.4_FG003	8:118103039 -118103810	PF01161: Białko wiążące fosfatydyloetanolanę	4.25609	4.57508
GRMZM2G118770	8:175058637 -175063179	PF00390: N-końcowa domena dehydrogenazy jabłczanu, PF03949: Domena wiążąca NAD dehydrogenazy jabłczanu	5.66932	6.14265
PEP1	9:61295925- 61301850	PF00311: Karoboksylaza fosfoenolopirogronianu	-2.91311	-3.39408
GRMZM2G366532	9:139510669 -139519756	PF00012: Białko Hsp70, PF06723: Białko MreB/Mbl	6.12962	7.22259

Aneks 3 Zestawienie genów o zmienionej ekspresji w obu badanych punktach czasowych

GRMZM2G173685	9:20459073-20460609	Nie zidentyfikowano	8.97107	3.33827
GRMZM2G006973	9:24596811-24600134	PF00023: Powtórzenie ankirynowe, PF07714: Białkowa kinaza tyrozynowa, PF00069: Domena kinazy białkowej	3.86658	3.63806
asn4 (locus) – syntetaza asparaginy 4	9:137784209-137791159	Nie zidentyfikowano	6.53178	4.03181

Zestawienie informacji na temat 61 genów o ekspresji zmienionej w linii tolerancyjnej S245 w obu badanych punktach czasowych pod wpływem oprysku herbicydem Roundup®.

Aneks 4 Dorobek naukowy

PUBLIKACJE

1. Grześ M., Nowacka-Woszek J., Szczerbal I., Czerwińska J., Gracz J., Świtoński M., *A Comparison of Coding Sequence and Cytogenetic Localization of the Myostatin Gene in the Dog, Red Fox, Arctic Fox and Chinese Raccoon Dog*

Cytogenet Genome Res 2009;126:173–179

2. Watkin E., McInnes A., Gurda D., Gracz J., Steenkamp E.T., Stępkowski T., *Distinct Bradyrhizbium communities nodulate legumes native to temperate and tropical monsoon Australia*

Molecular Phylogenetics and Evolution 2013, 63 (2):265-77

3. Gracz J., Świercz A., Twardowski T., *Computational analysis of alternative spliced gene on maize chromosome 1*

BioTechnologia, 2012 , 93(1):41-47

UDZIAŁ W KONFERENCJACH NAUKOWYCH

1. Gracz J., Twardowski T., *Estimation of alternative splicing events on maize chromosome 1*, Plant Genome Evolution Conference, 4-6.09.2011 r., Holandia, Amsterdam (konferencja międzynarodowa, poster)

2. Gracz J., Twardowski T., *Alternative splicing events on maize chromosome 1*, mRNA Fate – Life and Death of the mRNA in the Cytoplasm, 23-26.05.2012 r., Riva del Garda, Włochy (konferencja międzynarodowa, poster)

3. Gracz J., Twardowski T., *Alternative splicing events on maize chromosome 1*, “RNA regulators of gene expression” The 72nd Harden Conference, Robinson College, 22 – 25.07.2012 r., Cambridge, Wielka Brytania (konferencja międzynarodowa, poster)

4. Gracz J., Tyczewska A., Hoffa A., Żmieńko A., Świercz A., Błażewicz J. Twardowski T., *Alternative splicing events in two maize lines under herbicide stress conditions*, Polish-German Biochemical Societies Joint Meeting, 11-14.09.2012 r., Poznań (konferencja międzynarodowa, poster)

5. Gracz J., Tyczewska A., Hoffa A., Żmieńko A., Świercz A., Błażewicz J., Twardowski T., *The role of alternative splicing in two maize lines under herbicide stress conditions*, 48th Congress of the Polish Biochemical Society, 2-5.09.2013 r., Toruń (konferencja krajowa, referat)
6. Gracz J., Tyczewska A., Hoffa A., Żmieńko A., Świercz A., Błażewicz J., Twardowski T., *The role of alternative splicing in two maize lines under herbicide stress conditions*, 2nd International Conference on Agricultural & Horticultural Sciences, 3-5.02.2014 r., Hyderabad, Indie (konferencja międzynarodowa, poster)
7. Gracz J., Tyczewska A., Hoffa A., Żmieńko A., Świercz A., Błażewicz J., Twardowski T., *Alternative splicing events in two maize lines under herbicide stress conditions*, 56th Annual Maize Genetics Conference, 13-16.03.2014 r., Pekin, Chiny (konferencja międzynarodowa, poster)
8. Gracz J., Tyczewska A., Hoffa A., Żmieńko A., Świercz A., Błażewicz J., Twardowski T., *Alternative splicing events in two maize lines under herbicide stress conditions*, Post-transcriptional Gene Expression Regulation in Plants, 30.06.-2.07.2014 r., Poznań (konferencja międzynarodowa, poster)

UDZIAŁ W PROJEKTACH BADAWCZYCH

1. Grant Polskiego Towarzystwa Genetycznego dla Kół Naukowych na rok 2008 „Poszukiwanie polimorfizmu w eksonie 3 genu miostatyny u psów zróżnicowanych pod względem umięśnienia”, wykonawca
2. Grant Ministerstwa Nauki i Szkolnictwa Wyższego pt. „Alternatywny splicing u kukurydzy zwyczajnej indukowany w warunkach stresu herbicydowego”, nr N N310 309839, okres realizacji: 22.10.2010r. - 21.10.2013r., wykonawca
3. Projekt badawczy NCN realizowany przez osobę rozpoczynającą karierę naukową nieposiadającą stopnia naukowego doktora pt. „Poszukiwanie sekwencji regulujących proces splicingu w warunkach stresu herbicydowego u kukurydzy zwyczajnej”, nr 2011/01/N/NZ9/02900, okres realizacji: 16.12.2011 - 15.06.2014 r., kierownik projektu
4. Projekt badawczy NCN realizowany przez osoby rozpoczynające karierę naukową posiadające stopień naukowy doktora (konkurs „ze stopniem doktora”) pt. Analiza epigenomu i proteomu kukurydzy zwyczajnej (*Zea mays*), linii odpornej i wrażliwej na herbicyd”, nr 2011/01/D/NZ9/03631, okres realizacji: 12.12.2011 - 11.12.2014 r., wykonawca

5. Projekt badawczy NCN MAESTRO pt. "Kompleksowe podejście w celu zdefiniowania molekularnych podstaw odpowiedzi kukurydzy na stres herbicydowy", nr 2012/06/A/NZ9/00125, okres realizacji 2013 - 2018, wykonawca

DOŚWIADCZENIE NAUKOWE

- 19 – 25.06.2011 kurs praktyczny pt. „Bioinformatics.Computer methods in molecular biology” organizowany przez International Center for Genetic Engineering and Biotechnology, Triest
- 2005 – 2008 działalność w Kole Naukowym Studentów Biotechnologii „Operon” przy Uniwersytecie Przyrodniczym w Poznaniu
- 09 – 10. 2007 Praktyki zawodowe w Pracowni Biologii Molekularnej Instytutu Dendrologii Polskiej Akademii Nauk w Kórniku
- 09 – 10. 2008 Praktyki zawodowe w Dziale Technologii i Kontroli Jakości Zakładów Chemicznych „Unia”

WYRÓŻNIENIA

- 06. 2008 – I miejsce w konkursie prac naukowych Sesji Studenckich Kół Naukowych
- 2009 – Medal rektora Uniwersytetu Przyrodniczego w Poznaniu za osiągnięcia w studiach
- 2012 – Stypendium w ramach projektu pt. „Wsparcie stypendialne dla doktorantów na kierunkach uznanych za strategiczne z punktu widzenia rozwoju Wielkopolski”, Podziałania 8.2.2 „Regionalne Strategie Innowacji”, Działania 8.2, Priorytetu VIII, Programu Operacyjnego Kapitał Ludzki finansowanego ze środków Europejskiego Funduszu Społecznego
- 02. 2014 – wyróżnienie w konkursie na najlepszą prezentację posterową podczas 2nd International Conference on Agricultural and Horticultural Sciences, która odbyła się w dniach 3-5.02.2014 w Hyderabadzie