

Zadanie nr 23:

Globalna analiza wariantów strukturalnych w genomach buraka oraz identyfikacja rejonów powiązanych z jednonasiennością i męską sterylnością

Okres realizacji: **72 miesiące (lata 2021-2026)**

Zespół wykonawców projektu:

Prof. dr hab. Dariusz Grzebelus – kierownik projektu; email: d.grzebelus@urk.edu.pl

Dr hab. Marek Szklarczyk, prof. URK

Dr inż. Alicja Macko-Podgórn, prof. URK

Cele projektu w 2021 r.

W temacie badawczym 1 (Identyfikacja wariantów strukturalnych):

- identyfikacja rodzin ruchomych elementów genetycznych obejmujących retrotranspozony LTR (ang. *Long Terminal Repeat*) oraz miniaturowe ruchome elementy genetyczne (ang. *Miniature Inverted-repeat Transposable Elements*; MITE) w oparciu o złożenia genomów buraka cukrowego dostępne w domenie publicznej
- resekwencjonowanie genomów 10 roślin buraka cukrowego i dwóch roślin buraka pastewnego o różnym pochodzeniu

W temacie badawczym 3 (Identyfikacja rejonu genomu obejmującego geny restorerowe):

- wskazanie różnic pomiędzy sekwencjami przeciwstawnych alleli locus X/x ($Rf1/rf1$)
- określenie, czy działanie restorera X ($Rf1$) jest związane z poziomem akumulacji jego mRNA
- uzyskanie klonów bakteryjnych z wklonowaną do wektora ekspresyjnego otwartą ramką odczytu genu X ($Rf1$)

Wszystkie założone cele zostały zrealizowane

Materiały i metody

W temacie badawczym 1 (Identyfikacja wariantów strukturalnych):

Materiały roślinne

- Dwa złożenia referencyjne genomów buraka cukrowego (RefBeet i EL10)
- 10 roślin reprezentujących materiały hodowlane KHBC o różnej charakterystyce
- Dwie rośliny reprezentujące materiały hodowlane MHR (linia O i linia MS)

Metody

- Narzędzia bioinformatyczne do identyfikacji *de novo* ruchomych elementów genetycznych: retrotranspozonów LTR (LTRharvest) i MITE (MITE-hunter)
- Izolacja genomowego DNA, ocena jakości, sekwencjonowanie NGS na platformie Illumina
- Narzędzia bioinformatyczne do obróbki i mapowania odczytów NGS oraz identyfikacji wariantów strukturalnych

W temacie badawczym 3 (Identyfikacja rejonu genomu obejmującego geny restorerowe):

Materiały roślinne

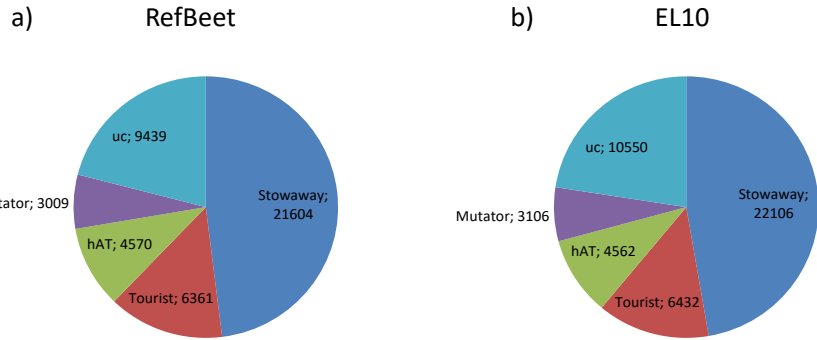
- 2S1329, 2S1387 oraz 3S1123 – populacje otrzymane po krzyżowaniach typu: obiekt męskosterylny (MS) × kandydat na linię dopełniającą (linię O), segregowały na rośliny męskopłodne i męskosterylne (KHBC w Straszkanie Sp. z o.o.)

Metody

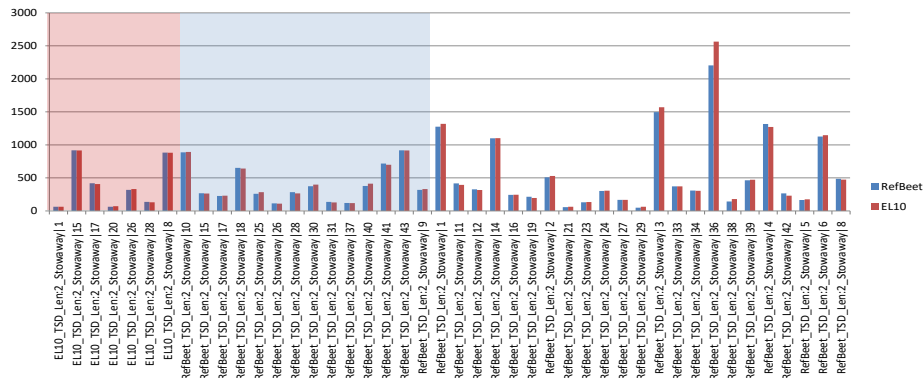
- Sekwencjonowanie DNA w serwisie zewnętrznym
- Hybrydyzacja northern z użyciem sondy znakowanej digoksygeniną (DIG)
- Klonowanie molekularne w wektorze plazmidowym typu ekspresyjnego (pQE32, Qiagen)

Identyfikacja rodzin ruchomych elementów genetycznych obejmujących retrotranspozony LTR oraz miniaturowe ruchome elementy genetyczne MITE

MITE

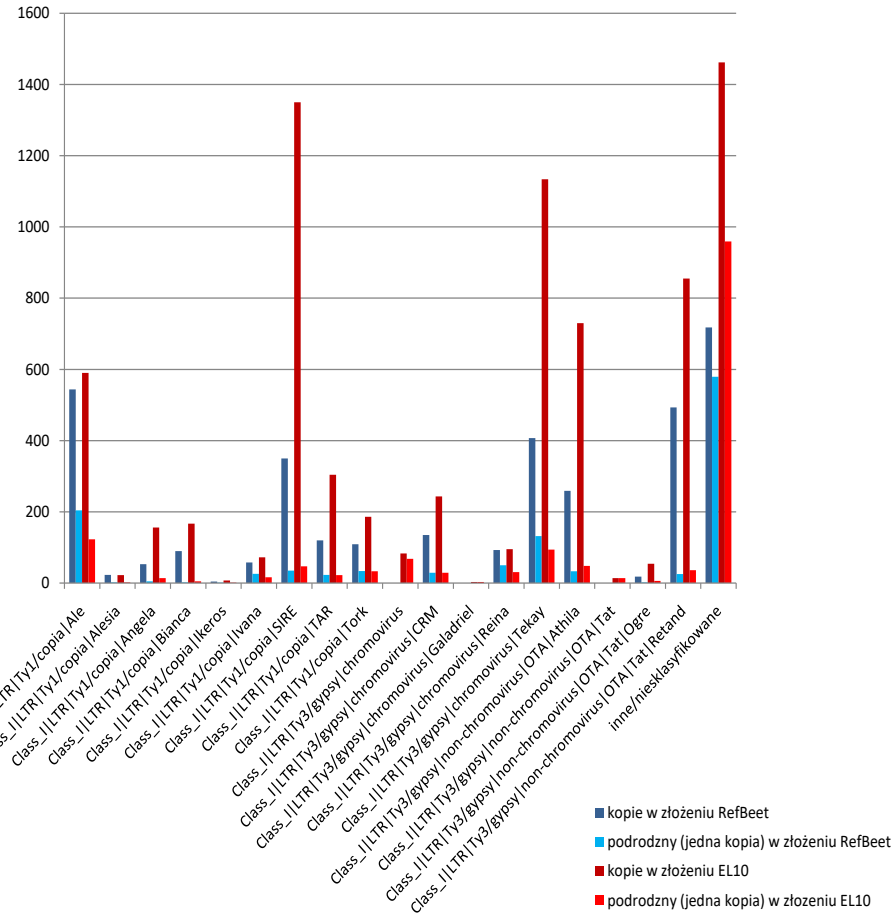


Udział każdej ze zidentyfikowanych grup elementów MITE w całkowitej liczbie zidentyfikowanych kopii w złożeniu RefBeet (a) i w złożeniu EL10 (b).



Liczebność zidentyfikowanych rodzin elementów *Stowaway* MITE.

LTR-RT



Liczba retrotranspozonów LTR w złożeniach RefBeet i EL10 buraka cukrowego. Osobny słupek dla każdej grupy ilustruje liczbę rodzin reprezentowanych przez jedną kopię w genomie.

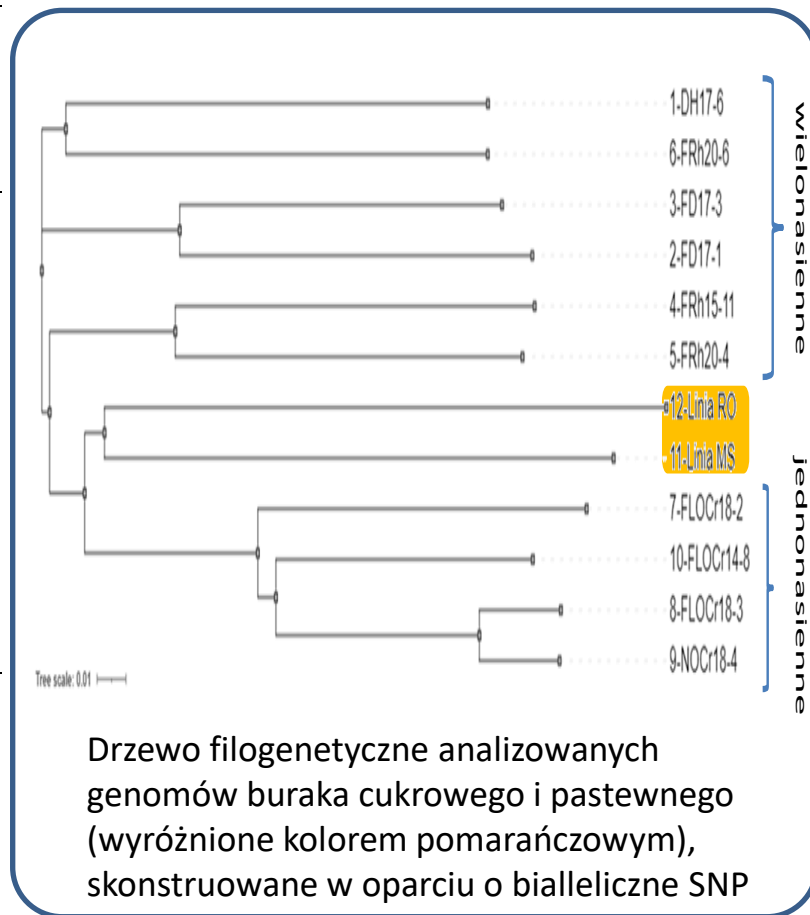
Resekwencjonowanie genomów 10 roślin buraka cukrowego i dwóch roślin buraka pastewnego o różnym pochodzeniu

Wyniki sekwencjonowania, filtrowania i mapowania odczytów sekwencjonowania genomów buraka cukrowego i pastewnego

Nr	Nazwa	Liczba par odczytów otrzymanych w wyniku sekwencjonowania	Procent par odczytów, które przeszły kontrolę jakości	Procent par odczytów, które zostały prawidłowo zmapowane do genomu
1	DH17-6	99798179	96,57%	99,28%
2	FD17-1	85896225	96,91%	99,18%
3	FD17-3	93211190	96,71%	99,32%
4	FRh15-11	98655808	96,46%	98,73%
5	FRh20-4	129188633	96,67%	98,98%
6	FRh20-6	93816654	96,76%	99,12%
7	FLOCr18-2	83835319	97,40%	99,36%
8	FLOCr18-3	92053346	96,70%	99,36%
9	NOCr18-4	102419920	96,36%	99,28%
10	FLOCr14-8	115610794	96,03%	99,33%
11	Linia MS	95574805	96,83%	99,13%
12	Linia RO	106728377	96,45%	99,13%

Charakterystyka zidentyfikowanych polimorfizmów

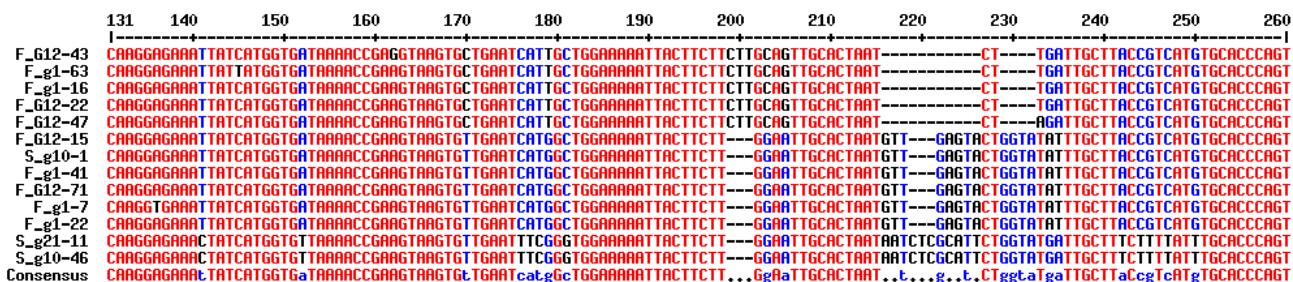
liczba polimorfizmów	11 917 236
liczba SNPów	9 570 990
liczba MNPów (np.CC>TT)	1 658 508
liczba indeli	1 319 180
liczba innych (np. kompleksowa substytucja ACT>TCGA)	372 581
liczba pozycji z wieloma alternatywnymi allelami	1 289 871
liczba SNPów z wieloma alternatywnymi allelami	112 884



Wskazanie różnic pomiędzy sekwencjami przeciwstawnych alleli locus *X/x (Rf1/rf1)*

Analizie poddano trzy męskosterylne (S) i trzy męskopłodne (F) rośliny z populacji 3S1123.

Uliniowienie sekwencji genu *X/Rf1* z roślin sterylnych wykazało, iż u sekwencje te grupują się w trzy typy. Jeden z tych typów odnaleziono także w sekwencjach uzyskanych dla roślin płodnych. W roślinach płodnych wystąpił jeszcze jeden typ, którego nie znaleziono u roślin sterylnych (wyróżniając typy sekwencji nie brano pod uwagę wariantów unikalnych tzn. odnotowanych wyłącznie w jednej sekwencji – mogły one wynikać z błędów polimerazy DNA w toku amplifikacji).



Uliniowienie sekwencji genu *X/Rf1* z roślin płodnych (F) i reprezentatywnych sekwencji z roślin sterylnych (S) (fragment).

Sekwencje reprezentatywne dla wyróżnionych typów ↓

	247	250	276	295	298	300	310-311	324	331	346	368	375	387	403	407	434
F_G12-43	C	G	A	A	A	G	CG	T	G	-	A	G	T	A	G	G
S_G10-1	C	G	A	C	A	A	GA	G	G	A	A	T	A	C	A	A
F_G1-41	C	G	G	C	A	A	GA	G	G	A	A	T	A	C	A	A
S_G21-11	T	T	A	A	T	G	CA	T	T	A	T	T	A	C	A	A
S_G10-46	T	T	A	A	T	G	CA	T	T	A	A	G	T	C	A	A

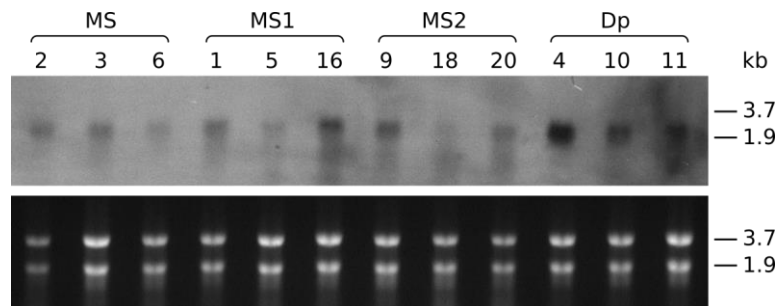
← Pozycje sekwencji konsensusowej

Polimorfizmy sekwencyjne uwidocznione w analizowanych sekwencjach genu *X/Rf1* z roślin płodnych (F) i sterylnych (S) (szare cieniowanie - polimorfizmy specyficzne dla roślin płodnych) (fragment).

Ogółem w obrębie uliniowanego regionu pomiędzy wyróżnionymi typami sekwencji odnotowano 64 polimorfizmy pojedynczego nukleotydu (SNP), 12 polimorfizmów wielonukleotydowych (MNP), 8 indeli oraz 3 polimorfizmy o charakterze mieszanym. Spośród odnotowanych polimorfizmów 18 było specyficznych dla unikalnego typu sekwencji z roślin płodnych.

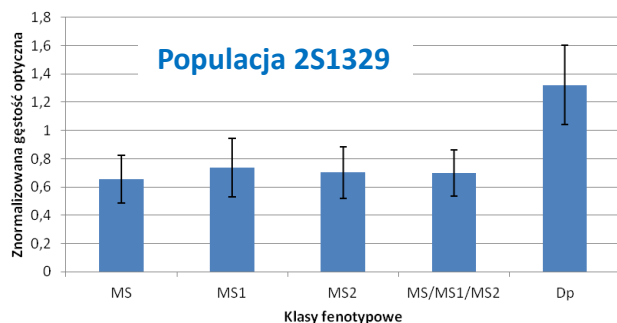
Określenie, czy działanie restorera *X (Rf1)* jest związane z poziomem akumulacji jego mRNA

Analizie poddano 12 roślin z populacji 2S1329 oraz 18 roślin z populacji 2S1387. Badane rośliny reprezentowały cztery klasy fenotypowe wyróżnione na bazie analizy żywotności pyłku: MS – tylko sterylne ziarna pyłku, MS1 – mniej niż 50% żywotnych ziaren pyłku, MS2 – 50% żywotnych ziaren pyłku, Dp – więcej niż 50% żywotnych ziaren pyłku.



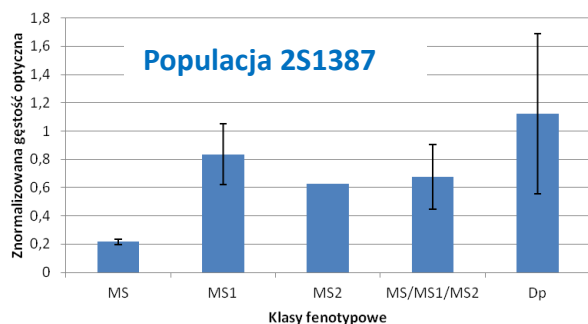
Sygnaly hybrydizacyjne uzyskane dla RNA z pąków kwiatowych po hybrydizacji z sondą dla genu *Rf1* u roślin przynależących do 4 klas fenotypowych populacji 2S1329 (górze). Odpowiedni żel z preparatami RNA (dół).

W wyniku hybrydizacji u wszystkich badanych roślin z obydwu populacji wykryto transkrypt o wielkości 2,4 kb.



Populacja 2S1329

Analiza densytometryczna otrzymanych sygnałów hybrydizacyjnych wykazała istotnie wyższy poziom ekspresji genu *Rf1* u roślin należących do klasy Dp aniżeli u roślin należących do pozostałych klas fenotypowych (dwustronny test t-Studenta; $p = 0,019$; $\alpha = 0,05$), pomiędzy którymi nie zaobserwowano istotnych różnic w akumulacji badanego mRNA.



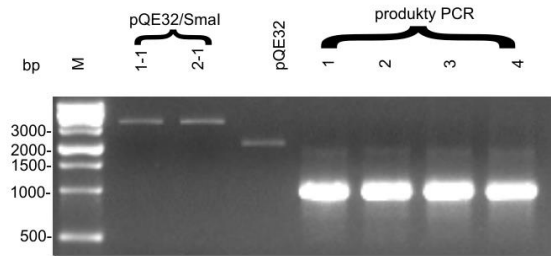
Populacja 2S1387

Analiza densytometryczna otrzymanych sygnałów hybrydizacyjnych wykazała istotnie niższy poziom ekspresji genu *Rf1* u roślin należących do klasy MS aniżeli u roślin należących do pozostałych klas fenotypowych, pomiędzy którymi nie zaobserwowano istotnych różnic w ilości transkryptu badanego genu. Analiza statystyczna (dwustronny test t-Studenta, $\alpha = 0,05$) nie wykazała istotnych różnic znormalizowanych wartości OD pomiędzy roślinami dobrze pyłącymi i męskosterylnymi ($p = 0,232$).

Średnie arytmetyczne znormalizowanej gęstości optycznej odpowiadającej akumulacji mRNA genu *Rf1* u roślin przynależących do różnych klas fenotypowych (zaznaczono słupki błędów standardowych).

Uzyskanie klonów bakteryjnych z wklonowaną do wektora ekspresyjnego otwartą ramką odczytu genu *X (Rf1)*

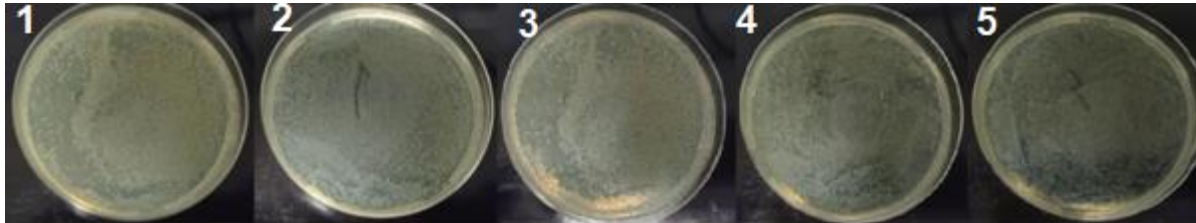
Przeznaczony do klonowania odcinek DNA zawierał fragment otwartej ramki odczytu (ORF) genu *X/Rf1*. Fragment ten uzyskano metodą PCR z użyciem starterów Rf1-Sorf-f oraz Rf1-Sorf-r, matrycę stanowił sklonowany w wektorze pGEM-T (Promega) fragment cDNA genu *X/Rf1* – konkretnie był to klon 5/37 uzyskany w toku wcześniejszych prac. Elektroforeza mieszanin po reakcji amplifikacji ujawniła obecność pojedynczego, bardzo wydajnie powielonego produktu PCR o spodziewanej długości wynoszącej ok. 1 kb.



Wykorzystanym do klonowania wektorem ekspresyjnym był plazmid pQE32 (Qiagen), który przygotowano do ligacji poprzez trawienie restryktazą *SmaI*. Uzyskany produkt restrykcji miał oczekiwaną wielkość wynoszącą 3,4 kb.

Wynik elektroforezy: strawionego enzymem *SmaI* plazmidu pQE32 wyizolowanego z klonów 1-1 i 2-1, nietrawionego plazmidu pQE32 wyizolowanego z klonu 1-1 (kontrola dla trawienia restrykcyjnego) oraz produktów PCR uzyskanych na bazie klonu 5/37 przy użyciu starterów Rf1-Sorf-r / Rf1-Sorf-r.

Po uprzednim oczyszczeniu z żelu produkt PCR i strawiony plazmid poddano ligacji na tępe końce, po czym mieszaninę ligacyjną wprowadzono do szczepu *E. coli* DH10B uzyskując bardzo liczne kolonie transformantów.



Wynik transformacji szczepu *E. coli* DH10B mieszaniną ligacyjną (szalki zawierają agar LB z ampicyliną).

DNA 52 transformantów wykorzystano jako matrycę do PCR ze starterami Rf1-Sorf-f/Rf1-Sorf-r (PCR w wariacie kolonijny). Dla 13 kolonii uzyskano produkt amplifikacji o spodziewanej wielkości ok. 1 kb. Z klonów odpowiadających tym koloniom wyizolowano plazmidowy DNA, który poddawano elektroforezie wraz z kontrolą w postaci niezrekombinowanego plazmidu pQE32. W preparatach plazmidowego DNA uzyskanych dla 9 badanych klonów forma CCC migrowała w strefie wskazującej na obecność insertu o wielkości ok. 1 kb. Klony te uznano za rekombinanty z pożądanym insertem i poddano sekwencjonowaniu Sangera.

Wnioski

W temacie badawczym 1 (Identyfikacja wariantów strukturalnych):

1. W genomach buraka cukrowego i pastewnego rezydują liczne rodziny ruchomych elementów genetycznych, w tym retrotranspozonów LTR oraz miniaturowych elementów klasy II (MITE).
2. Jakość złożenia genomowego wpłynęła na skuteczność identyfikacji rodzin retrotranspozonów LTR. Ponad dwa razy więcej kopii retrotranspozonów LTR zidentyfikowano z złożeniu EL10 o zdecydowanie wyższej ciągłości (3474 kopie w złożeniu RefBeet versus 7527 kopii w złożeniu EL10). Ciągłość złożenia nie wpłynęła natomiast istotnie na liczbę zidentyfikowanych kopii elementów MITE.
3. Resekwencjonowane genomy buraka cukrowego i pastewnego cechują się różnymi poziomami homozygotyzowania oraz wysokim poziomem polimorfizmu (w porównaniu do złożenia referencyjnego RefBeet), obejmującym substytucje nukleotydowe oraz indele.

W temacie badawczym 3 (Identyfikacja rejonu genomu obejmującego geny restorerowe):

1. W wyniku analizy sekwencyjnej locus *X/Rf1* zidentyfikowano 18 polimorfizmów specyficznych dla roślin z przywróconą płodnością. Polimorfizmy te będzie można użyć do konwersji w markery PCR służące diagnozowaniu obecności restorera *X/Rf1* w materiałach hodowlanych.
2. W eksperymentach hybrydacyjnych najwyższą akumulację mRNA genu *X/Rf1* odnotowywano wśród roślin z przywróconą płodnością (męskopłodnych, dobrze pylących). W grupie tej zdarzały się jednak także rośliny o niskiej akumulacji (korespondującej z obserwowaną dla roślin męskosterylnych), co może wynikać z niepełnej korelacji pomiędzy poziomem mRNA *Rf1* i białka RF1.
3. Otrzymano klony *Escherichia coli*, które w obrębie wektora ekspresyjnego pQE32 (Qiagen) mają insert zawierający fragment otwartej ramki odczytu (ORF) restorera *X/Rf1*. Trwają prace nad określeniem, które klony mają ten insert w orientacji gwarantującej otrzymanie rekombinantowego białka RF1.

6. Miernik zadania – stopień realizacji

l.p.	miernik	wartość miernika podana w opisie zadania	wartość miernika zrealizowana	stopień realizacji zadania
1	2	3	4	5
temat badawczy 1				
1.1	Liczba złożzeń genomu wykorzystanych do identyfikacji rodzin ruchomych elementów genetycznych	2	2	1,00
1.2	Liczba baz danych zawierających reprezentatywne sekwencje elementów ruchomych	2	2	1,00
1.3	Liczba resekwencjonowanych genomów	12	12	1,00
temat badawczy 3				
3.1	Liczba sekwencjonowanych bibliotek	2	2	1,00
3.2	Liczba populacji objętych analizą ekspresji	2	2	1,00
3.3	Liczba pozytywnie zweryfikowanych klonów bakteryjnych do produkcji białka rekombinantowego	4	4	1,00
			ŚREDNIA	1,00
			% REALIZACJI ZADANIA	100%

*W roku 2021 (pierwszy rok realizacji projektu)
nie planowano prezentacji doniesień konferencyjnych i publikacji*