

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](mailto:d.grzebelus@urk.edu.pl)**

**Dr hab. Marek Szklarczyk, prof. URK**

**Dr inż. Alicja Macko-Podgórn, prof. URK**

# Cele projektu w 2022 r.

## **W temacie badawczym 1 (Identyfikacja wariantów strukturalnych):**

- identyfikacja polimorficznych insercji wybranych rodzin retrotranspozonów LTR (ang. *Long Terminal Repeat*) oraz miniaturowych ruchomych elementów genetycznych (ang. *Miniature Inverted-repeat Transposable Elements*; MITE) w oparciu o resekwencjonowane genomy buraka cukrowego i pastewnego

## **W temacie badawczym 2 (Identyfikacja rejonu genomu obejmującego gen jednonasienności):**

- wyprowadzenie populacji roślin różniących się pod względem fenotypu jedno/wielonasienności, fenotypowanie cechy jedno/wielonasienności
- przygotowanie preparatów genomowego DNA roślin o określonym fenotypie do wysokowydajnego genotypowania

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

- wskazanie różnic w organizacji sekwencji przeciwstawnych alleli locus X/x (Rf1/rf1);
- określenie, czy obecność restorera X (Rf1) jest skorelowana z poziomem mRNA dla otwartych ramek odczytu tego locus;
- optymalizacja produkcji rekombinantowego białka RF1 oraz jego oczyszczenie;
- zaprojektowanie sond do wzbogacania bibliotek GBS w sekwencje z chromosomu IV.

***Wszystkie założone cele zostały zrealizowane***

# Materiały i metody

## W temacie badawczym 1 (Identyfikacja wariantów strukturalnych):

### Materiały roślinne

- Złożenie referencyjne genomu buraka cukrowego (RefBeet), biblioteki odczytów NGS 10 genomów buraka cukrowego i dwóch genomów buraka pastewnego

### Metody

- PopoolationTE2 (narzędzie bioinformatyczne do identyfikacji *de novo* ruchomych elementów genetycznych na podstawie surowych odczytów NGS)
- Genotypowanie PCR

## W temacie badawczym 2 (Identyfikacja rejonu genomu obejmującego gen

### Materiały roślinne

### jednonasienności):

- 88 populacji F2 uzyskanych w wyniku przekrzyżowania rośliny jednonasiennej z wielonasieną (materiały hodowlane KHBC)

### Metody

- Uprawa w polu, fenotypowanie w fazie kwitnienia, izolacja genomowego DNA

## W temacie badawczym 3 (Identyfikacja rejonu genomu obejmującego

### Materiały roślinne

### geny restorerowe):

- Populacje otrzymane w KHBC w Straszku w wyniku krzyżowań typu: obiekt męskosterylny (MS) × kandydat na linię dopełniającą (linię O): MS KRA 15 37, 3S 13 175 oraz 5S 1125.

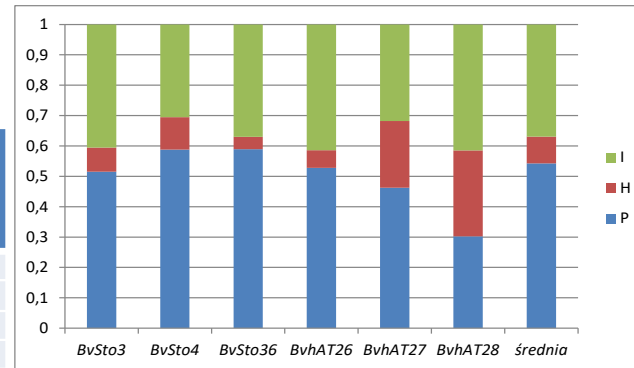
### Metody

- Analiza białek: SDS-PAGE, oczyszczanie na kolumnach Ni-NTA
- Real-time RT-PCR: system QuantStudio 3, reporter – SYBR Green I, oznaczenie względne
- Sekwencjonowanie DNA: system PacBio – Sequel II

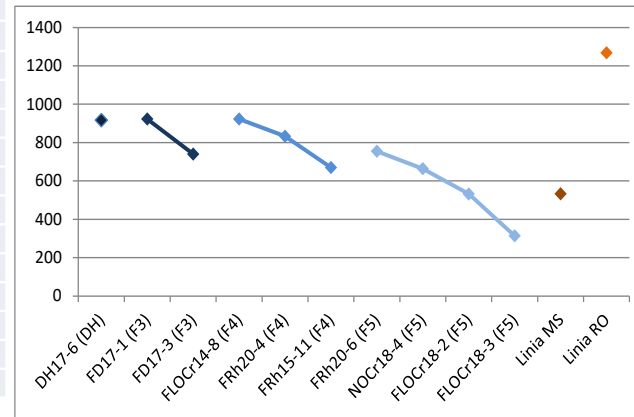
# Identyfikacja polimorficznych insercji wybranych rodzin miniaturowych ruchomych elementów genetycznych (MITE)

Liczba insercji sześciu rodzin MITE zidentyfikowanych w 12 genomach buraka, w tym homozygotycznych (I) i heterozygotycznych (H)

	DH17-6	FD17-1	FD17-3	FRh15-11	FRh20-4	FRh20-6	FLOCr18-2	FLOCr18-3	NOCr18-4	FLOCr14-8	Linia MS	Linia RO	Ogółem
<b>BvSto3</b>	<b>1115</b>	<b>1136</b>	<b>1105</b>	<b>1029</b>	<b>1084</b>	<b>1114</b>	<b>1084</b>	<b>780</b>	<b>1060</b>	<b>1110</b>	<b>808</b>	<b>1157</b>	<b>2164</b>
I	888	914	929	889	897	937	956	734	926	902	705	851	
H	227	222	176	140	187	177	128	46	134	208	103	306	
<b>BvSto4</b>	<b>835</b>	<b>794</b>	<b>794</b>	<b>765</b>	<b>834</b>	<b>832</b>	<b>756</b>	<b>503</b>	<b>770</b>	<b>793</b>	<b>470</b>	<b>766</b>	<b>1803</b>
I	545	559	600	601	614	626	599	462	594	540	382	478	
H	290	235	194	164	220	206	157	41	176	253	88	288	
<b>BvSto36</b>	<b>1409</b>	<b>1394</b>	<b>1320</b>	<b>1316</b>	<b>1414</b>	<b>1340</b>	<b>1246</b>	<b>1093</b>	<b>1282</b>	<b>1362</b>	<b>1196</b>	<b>1457</b>	<b>3213</b>
I	1261	1186	1188	1227	1272	1197	1182	1059	1201	1239	1101	1160	
H	148	208	132	89	142	143	64	34	81	123	95	297	
<b>BvhAT26</b>	<b>252</b>	<b>276</b>	<b>255</b>	<b>252</b>	<b>264</b>	<b>241</b>	<b>243</b>	<b>247</b>	<b>260</b>	<b>260</b>	<b>281</b>	<b>286</b>	<b>550</b>
I	218	233	229	239	229	214	230	234	240	229	224	215	
H	34	43	26	13	35	27	13	13	20	31	57	71	
<b>BvhAT27</b>	<b>161</b>	<b>157</b>	<b>165</b>	<b>166</b>	<b>165</b>	<b>171</b>	<b>141</b>	<b>165</b>	<b>159</b>	<b>151</b>	<b>169</b>	<b>171</b>	<b>301</b>
I	100	100	104	96	97	108	91	108	83	77	111	73	
H	61	57	61	70	68	63	50	57	76	74	58	98	
<b>BvhAT28</b>	<b>403</b>	<b>395</b>	<b>407</b>	<b>409</b>	<b>413</b>	<b>375</b>	<b>397</b>	<b>413</b>	<b>415</b>	<b>413</b>	<b>416</b>	<b>407</b>	<b>581</b>
I	246	237	256	215	232	236	277	290	238	179	284	199	
H	157	158	151	194	181	139	120	123	177	234	132	208	



← Frekwencja wariantów obserwowanych w miejscach insercji MITE z sześciu rodzin; I – homozygotyczna insercja, H – heterozygota, P – homozygota bez insercji



← Liczba heterozygotycznych miejsc insercji kopii elementów MITE w genomach buraka w zależności od stopnia uwsobnienia

Nazwa kopii	Metoda	DH17-6	FD17-1	FD17-3	FRh 15-11	FRh 20-4	FRh 20-6	FLOCr 18-2	FLOCr 18-3	NOCr 18-4	FLOCr 14-8	Linia MS	Linia RO
BvSto3.124	PCR	I	P	P	P	P	P	I	P	P	P	P	H
	PopoolationTE2	0,9	0	0	0	0	0	0,8	0	0	0	0	0,6
BvSto3.165	PCR	I	I	I	I	I	P	I	I	I	I	H	H
	PopoolationTE2	0,66 (H)	0,7	0,61 (H)	0,9	0,8	0	0,8	0,9	0,7	0,9	0 (P)	0 (P)
BvSto4.008	PCR	I	I	I	P	I	I	I	I	I	I	I	I
	PopoolationTE2	1	0,9	0 (P)	0,89 (I)	0,9	0,9	1	1	0,9	0,8	1	0,9
BvSto4.043	PCR	P	I	I	P	P	P	P	P	P	I	H	H
	PopoolationTE2	0	0,9	1	0	0	0	0	0	0	0,8	0,4	0,5
BvSto36.134	PCR	H	I	I	P	I	H	P	P	P	I	P	P
	PopoolationTE2	0 (P)	1	0,9	0	0,9	0,6	0	0	0	0,8	0	0
BvSto36.154	PCR	I	H	P	I	I	I	I	I	I	P	P	P
	PopoolationTE2	0,9	0 (P)	0	0,8	1	0,7	0,9	1	0,8	0	0	0
BvhAT27.002	PCR	P	H	P	P	P	P	P	P	P	P	H	I
	PopoolationTE2	0	0 (P)	0	0	0	0	0	0	0	0	0,6	0 (P)
BvhAT27.028	PCR	H	P	H	P	P	P	P	P	P	P	H	H
	PopoolationTE2	0,4	0	0,85 (I)	0	0	0	0	0	0	0	0,7	0,5

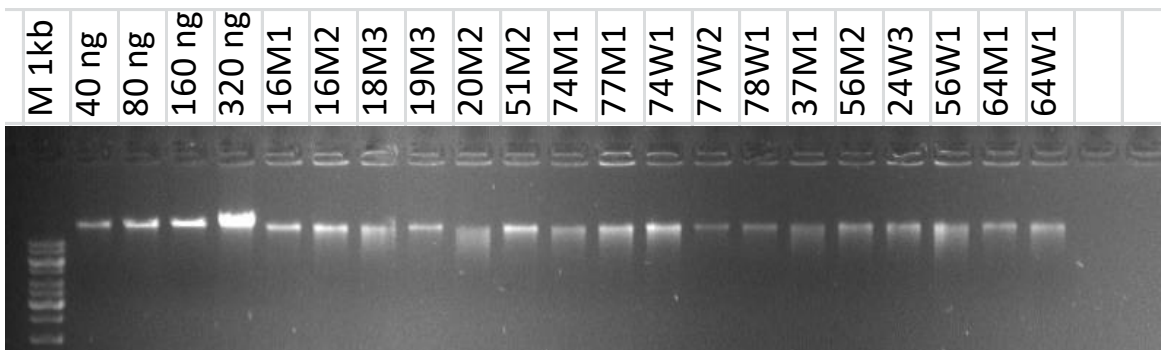
← Porównanie wyników bioinformatycznej identyfikacji insercji elementów MITE i weryfikacji PCR.

## Wyrowadzenie populacji roślin różniących się pod względem fenotypu jedno/wielonasienności, fenotypowanie cechy jedno/wielonasienności

Fenotypowanie cechy jedno/wielonasienności przeprowadzono dla 88 populacji F2. Tylko w obrębie 35 z nich zaobserwowano oczekiwaną segregację na rośliny jedno- i wielonasienne, w pozostałych populacjach nie stwierdzono występowania roślin jednonasiennych, z wyjątkiem populacji NASKRAK F2 21 45 [(FLOCr18 3 x FD17 3)F2], dla której zaobserwowano wyłącznie rośliny jednonasienne.

Częstość występowania roślin jednonasiennych w populacjach segregujących była na ogół niższa niż oczekiwano dla cechy jednogenowej warunkowanej recesywnym allelem. Z każdej segregującej populacji do dalszych prac wytypowano rośliny różniące się fenotypem. Pobrano od jednej do trzech roślin z każdego fenotypu reprezentujących każdą populację segregującą. Ogółem wybrano 210 roślin, z których 67 stanowiły rośliny jednonasienne, a pozostałe 143 rośliny były wielonasienne.

## Przygotowanie preparatów genomowego DNA roślin o określonym fenotypie do wysokowydajnego genotypowania



Przykładowy elektroforegram obrazujący jakość otrzymanych preparatów DNA. M 1kb – marker wielkości; 40, 80, 160, 320 ng – ścieżki referencyjne z DNA faga  $\lambda$  o podanych stężeniach, pozostałe ścieżki – DNA roślin reprezentujących segregujące populacje F2

## Analiza sekwencyjna locus *X/x (Rf1/rf1)*

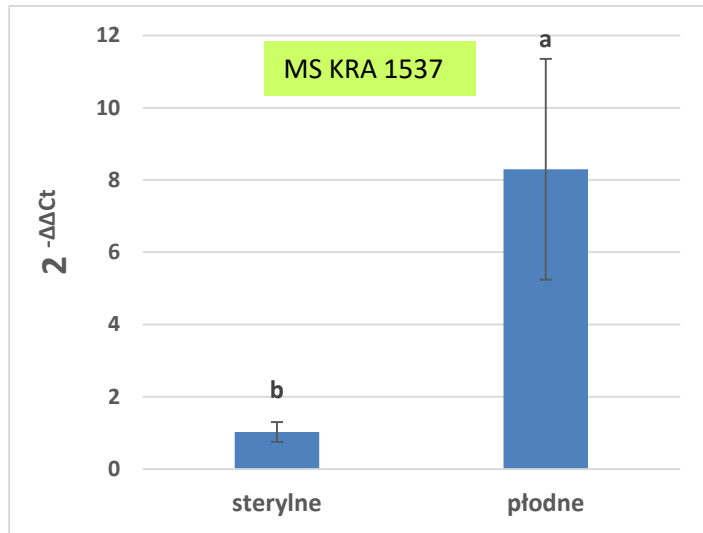
W ramach przeprowadzonych analiz zsekwencjonowano produkty amplifikacji z roślin męskopłodnych i męskosterylnych, które następnie zmapowano do sekwencji nukleotydowej allelu restorerowego *Rf1*. Analizy te wykazały 62 polimorfizmy sekwencyjne, które różnicowały odczyty z roślin męskosterylnych w stosunku do odczytów z roślin męskopłodnych oraz referencji *Rf1*, które cechowała obecność tego samego wariantu.

**Tabela wynikowa (fragment)**

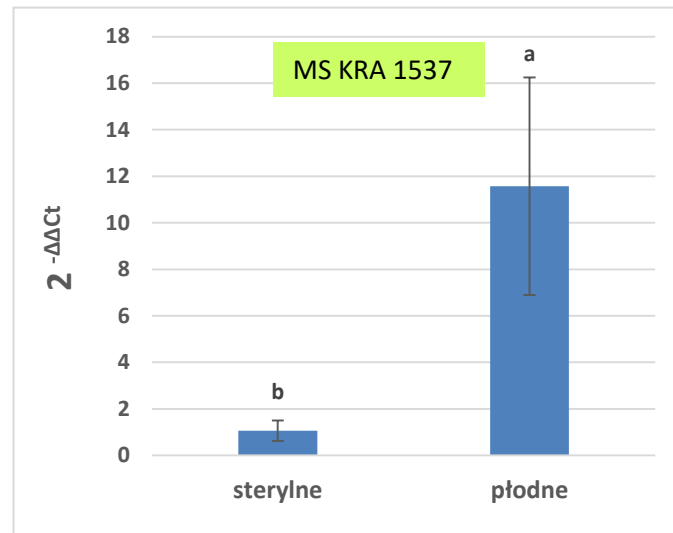
Pozycja nukleotydowa referencji	Wariant: referencja/ odczyty z roślin płodnych	Wariant: odczyty z roślin sterylnych	Pozycja nukleotydowa referencji	Wariant: referencja/ odczyty z roślin płodnych	Wariant: odczyty z roślin sterylnych
1512	t	a	12661	g	t
1515	g	a	15846	t	c
1524	t	c	15916	a	t
1527	a	-	15922	c	a
1570-1572	gtg	-	15930	t	c
1580-1584	ggaca	-	15961	g	a
1626	g	a	15968	t	a
2628	g	-	16151	g	a
2674	a	t	16154	g	a
3494	g	a	16158	g	a

## Analiza ekspresji locus X/x (*Rf1/rf1*) - analiza akumulacji mRNA

W przypadku populacji MS KRA 1537 analizy real-time RT-PCR wykazały, iż u roślin z przywróconą płodnością (męskopłodnych) poziom akumulacji transkryptów locus *Rf1/rf1* jest o ok. rząd wielkości wyższy niż u roślin męskosterylnych. Z kolei rośliny męskosterylne i męskopłodne z populacji 3S 13 175 nie różniły się istotnie pod względem poziomu akumulacji badanych transkryptów.



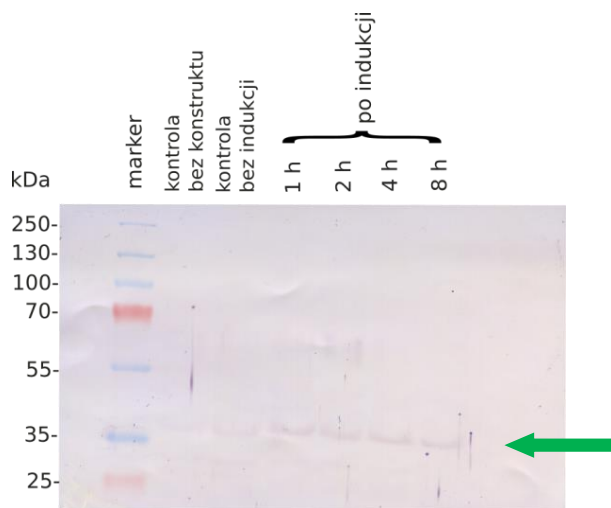
Rf1-RT-3f1/Rf1-Lorf-r



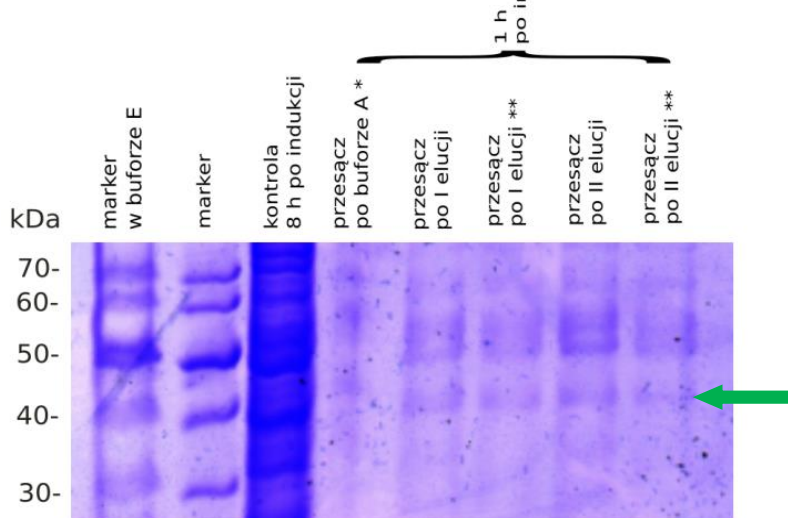
Rf1-Lorf-f/Rf1-Sorf-frev

**Porównanie ekspresji genu *Rf1* u męskosterylnych i męskopłodnych roślin z populacji MS KRA 1537**

# Analiza ekspresji locus *X/x (Rf1/rf1)* - optymalizacja produkcji i oczyszczanie rekombinantowego białka RF1



Western blot



SDS PAGE preparatów po oczyszczeniu

W wyniku analizy western blot słaby sygnał dla docelowego białka (ok. 40 kDa) otrzymano tylko dla klonu 63 hodowanego w temp. 30 °C. Po oczyszczaniu na kolumnach N-NTA zarówno w preparatach po pierwszej jak i po drugiej elucji wykryto białko o oczekiwanej wielkości 40 kDa. Preparaty te były zanieczyszczone innymi białkami.

## Analizy genu *Z (Rf2)*

W ramach tegorocznych prac zaprojektowano serię 300 sond do wzbogacania bibliotek NGS w sekwencje z chromosomu 4, w którym jest zlokalizowany gen *Rf2*.

>Probe 47

```
ttagccatacaatacaaatgcaatgtactaaaggatagtccataggaagaaaaggctactgtacacagtcacccacaa  
gaatgtgcatccagatgtaattgtgcagtcgaactaatggt
```

Przykładowa sonda



## **W temacie badawczym 1 (Identyfikacja wariantów strukturalnych):**

1. Kopie ruchomych elementów genetycznych z grupy MITE stanowią wydajne źródło łatwo identyfikowalnych polimorfizmów insercyjnych.
2. MITE z nadrodzin *hAT*-like i *Stowaway* różnią się liczbą kopii i poziomem polimorfizmu insercyjnego w genomach buraka.
3. Genotypowanie populacji buraka pod kątem TEASV może posłużyć do określenia poziomu uwsobnienia tych populacji, np. na etapie wyprowadzania linii DH z heterozygotycznych roślin donorowych.

## **W temacie badawczym 2 (Identyfikacja rejonu genomu obejmującego gen jednonasienności):**

1. W większości badanych populacji F2 (53) nie zaobserwowano roślin jednonasiennych, pozostałe populacje (35) segregowały pod względem jedno/wielonasienności w warunkach polowych.
2. Recesywna cecha jednonasienności zwykle obserwowana była dla mniejszej liczby roślin w obrębie populacji segregujących, niż wynikałoby to z teoretycznego stosunku rozszczepień 3:1.

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

1. Zidentyfikowano 62 polimorfizmy sekwencyjne różnicujące allele *Rf1* i *rf1*. Są one rozłożone na długości ok. 15 kb allelu restorerowego (*Rf1*).
2. W jednej z analizowanych populacji rośliny męskopłodne z allelem *Rf1* wykazują ok. 10-krotnie wyższą akumulację transkryptów tego locus w porównaniu do roślin męskosterylnych. Wynik ten wskazuje na udział efektów ilościowych w przywracaniu płodności przez gen *Rf1*.
3. Sygnały dla białka rekombinowanego udało się uzyskać dla jednego z badanych klonów przy zastosowaniu immunodetekcji z przeciwciałami Anty-6xHIS Tag. Niewielką ilość tego białka wyizolowano przy zastosowaniu chromatografii powinowactwa. Pozyskany preparat jest jednak zanieczyszczony białkami towarzyszącymi, co wskazuje na konieczność pozyskiwania docelowego białka z żelu po elektroforezie.
4. Zaprojektowano zaplanowany zestaw sond do wzbogacania bibliotek NGS w odczyty z chromosomu 4.

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
<b>temat badawczy 1</b>				
1.1	Liczba resekwencjonowanych genomów wykorzystanych do identyfikacji polimorficznych miejsc insercji ruchomych elementów genetycznych	12	12	1,00
1.2	Liczba rodzin ruchomych elementów genetycznych, dla których zidentyfikowano polimorficzne miejsca insercji	6	6	1,00
<b>temat badawczy 2</b>				
2.1	Liczba populacji F2, segregujących pod względem cechy jedno/wielonasienności poddanych ocenie fenotypowej	88	88	1,00
2.2	Liczba roślin segregujących pod względem cechy jedno/wielonasienności, z których wyizolowano DNA	210	210	1,00
<b>temat badawczy 3</b>				
3.1	Liczba sekwencjonowanych bibliotek	2	2	1,00
3.2	Liczba populacji objętych analizą ekspresji	2	2	1,00
3.3	Liczba klonów bakteryjnych użytych do produkcji białka rekombinantowego	4	4	1,00
3.4	Liczba zaprojektowanych sond do wzbogacania bibliotek GBS	300	300	1,00
<b>ŚREDNIA</b>				1,00
<b>% REALIZACJI ZADANIA</b>				100%

## Wykaz publikacji 2022:

### Doniesienia konferencyjne:

Macko-Podgórn A., Piestrzyńska W., Grzebelus D., „Transposable elements in the genome of sugar beet”; poster; Polski Kongres Genetyki, Kraków, 27-30 czerwca 2022

Szklarczyk M., Wesołowski W., Szlachtowska A., Domnicz B., “Expression analysis of the *Rf1* restorer gene in sugar beet”; poster; Polski Kongres Genetyki, Kraków, 27-30 czerwca 2022