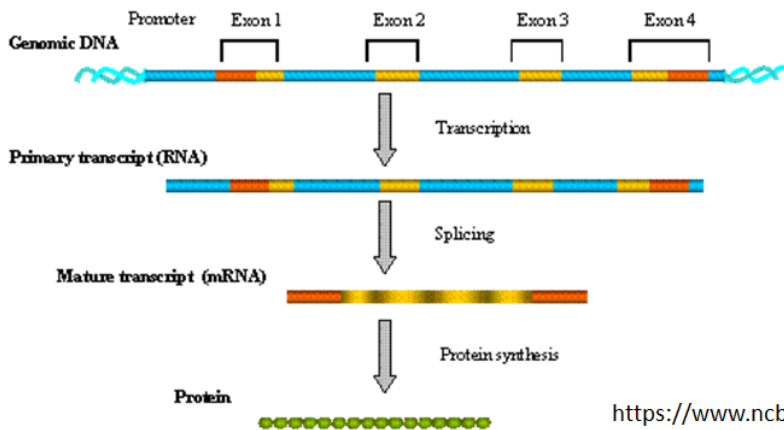
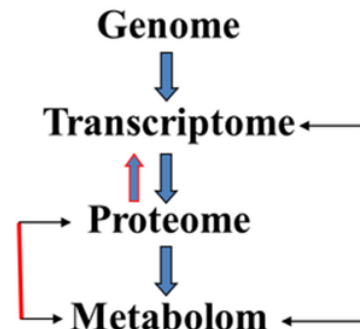


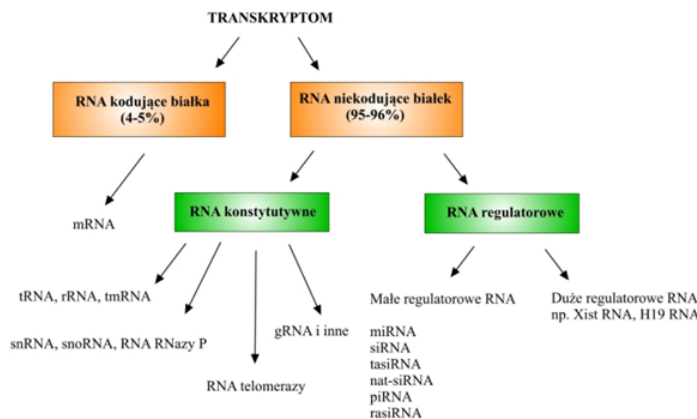
# Analizy transkryptomiczne

Analiza **transkryptomów** roślin pszenicy i rzepaku objęty rośliny rosnące w kontrastujących ze sobą systemach uprawy.



<https://www.ncbi.nlm.nih.gov/probe/docs/applexpression/>

**Transkryptom** to zestaw cząsteczek mRNA (lub ogólniej: transkryptów) obecny w określonym momencie w komórce, grupie komórek lub organizmie.



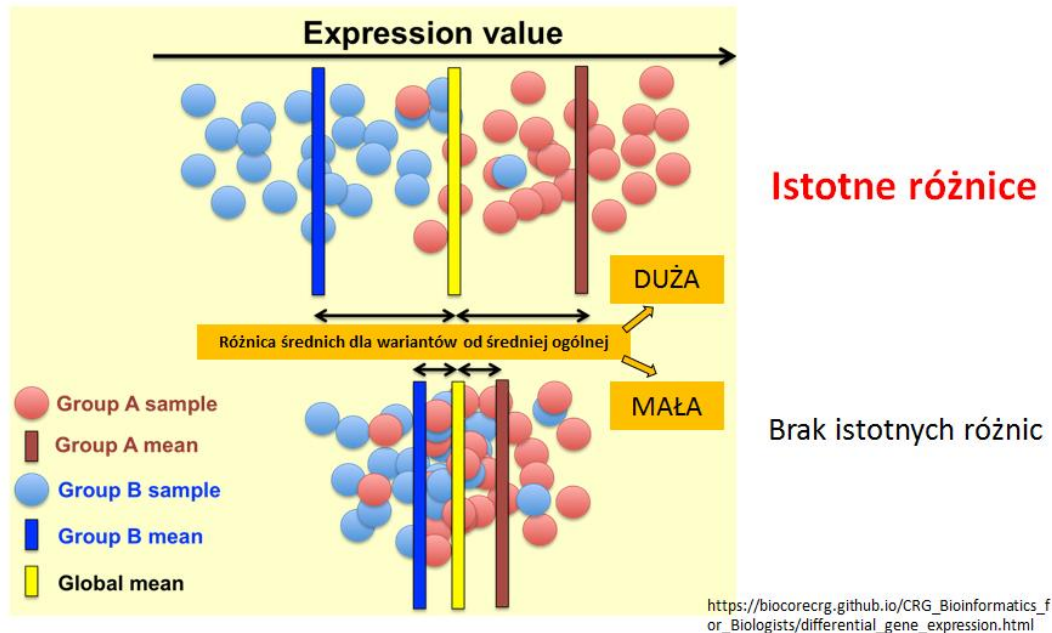
Kodujące RNA to **matryce do syntezy białek**, rzadko stanowią więcej niż 5% RNA komórkowego. U eukariontów syntetyzowane są jako **transkrypty pierwotne** a następnie **ulegają obróbce posttranskrypcyjnej**, podczas której dodawana jest czapeczka na końcu 5', ogon poli-A na końcu 3' oraz wycinane są introny. Dojrzała cząsteczka mRNA poza centralnie usytuowaną sekwencją kodującą zawiera również regiony nie ulegające translacji (3'- i 5'-UTR, ang. *untranslated region*).

Kodujące **mRNA** ulega **ciągłym modyfikacjom**, czas półtrwania **kilka minut, godzin**.

**Transkryptom w przeciwieństwie do genomu jest tworem bardzo dynamicznym.**

Komórki w odpowiedzi na różne czynniki uruchamiają i wyłączają transkrypcję genów, zmieniając w ten sposób swój transkryptom.

Często już kilka minut po zadziałaniu jakiegoś czynnika (np. stresu) na komórki można obserwować powstawanie transkryptów genów reakcji na ten czynnik.



## Cel badań

1. Zidentyfikowanie genów o zróżnicowanej ekspresji związanych z uprawą czystą lub z pokryciem
2. Interakcja między roślinami, wpływ roślin strączkowych na sąsiadujące rośliny poprzez cykle składników odżywczych, sygnalizację chemiczną i interakcje z mikroorganizmami.
3. Poznanie genów i szlaków metabolicznych uaktywnianych podczas odpowiedzi na uprawę czystą lub mieszaną

## Materiał i metody

1. Wybrane linie, w warunkach kontrolowanych, ocena liczby wschodów.
2. Analiza RNAseq dla linii w siewie czystym i mieszanym.  
Analiza różnicowej ekspresji genów.
3. Analiza metodą qPCR genów o zróżnicowanej ekspresji w odpowiedzi na sposób uprawy, projektowanie markerów na sekwencje wskazane po analizie danych z RNAseq.

# Materiał i metody



Czysty



Czysty



Z koniczyną

Z koniczyną

Z koniczyną



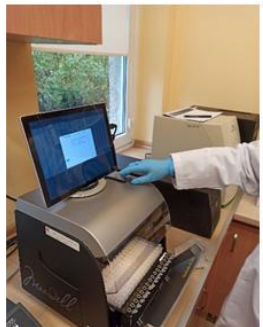
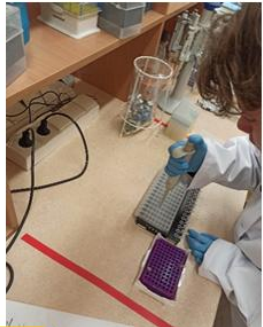
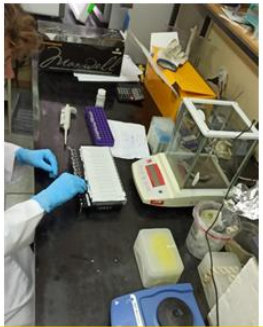
Z koniczyną



Homogenizacja prób



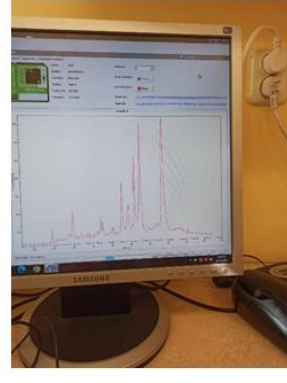
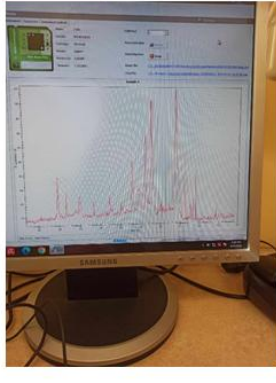
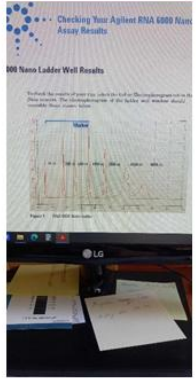
Izolacja RNA



Przechowywanie prób



Ocena jakości RNA



**Profilowanie ekspresji genów pozwala na poznanie genów i szlaków metabolicznych uaktywnianych podczas odpowiedzi na zróżnicowaną uprawę. Celem było zidentyfikowanie genów o zróżnicowanej ekspresji, związanych z uprawą rzepaku w siewie czystym i z wsiewką.**

### **Materiały i metody:**

Materiał do RNAseq: listki z 6 linii rzepaku z uprawy czystej, 8 linii z uprawy w komonicy, 7 linii z uprawy w koniczynie. Doświadczenie zostało przeprowadzone w szklarni. Całkowity RNA z liści został wyizolowany za pomocą zestawu do izolacji RNA. Stężenie i jakość matryc została oznaczona z wykorzystaniem Bioanalyzer RNA 6000 Nano Assay (Agilent Technologies).

Jakość materiału została potwierdzona po przesłaniu przez usługobiorcę. Również przez usługobiorcę została przygotowana biblioteka TruSeq mRNA stranded. Sekwencjonowanie wykonane zostało na NovaSeq6000 w konfiguracji 2x150bp, 40M odczytów. Dane zostały zapisane na dysku przenośnym i przekazane wykonawcy.

Grupa bioinformatyczna przeprowadziła analizę uzyskanych danych. Surowe dane zostały odniesione do sekwencji referencyjnej (sekwencja genomu rzepaku *Brassica napus* ssp. *oleifera*). Złożone sekwencje zostały scharakteryzowane funkcjonalnie. Wyznaczone zostały geny o ekspresji zróżnicowanej w odpowiedzi na sposób uprawy (DEGs). Została wykonana analiza typu KEGG (the Kyoto Encyclopedia of Genes and Genomes), która pozwoliła na identyfikację poszczególnych szlaków metabolicznych, do których należą wytypowane geny (charakterystyka funkcjonalna).

### **Wyniki:**

Wybrano 6 linii rzepaku z uprawy czystej, 8 linii rzepaku z uprawy w komonicy i 7 linii rzepaku z uprawy w koniczynie. Ich dobrą jakość po izolacji potwierdzono parametrem RIN (**Tab. 1**).

**Tabela 1.**

Linie, wybrane do doświadczenia, z warunków szklarniowych wraz z parametrem jakości matrycy.

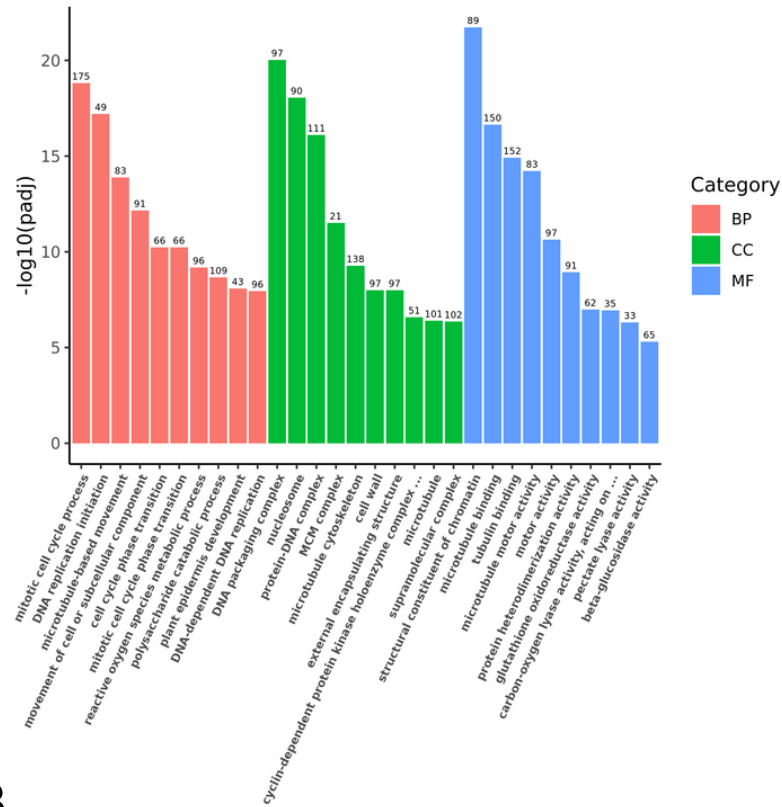
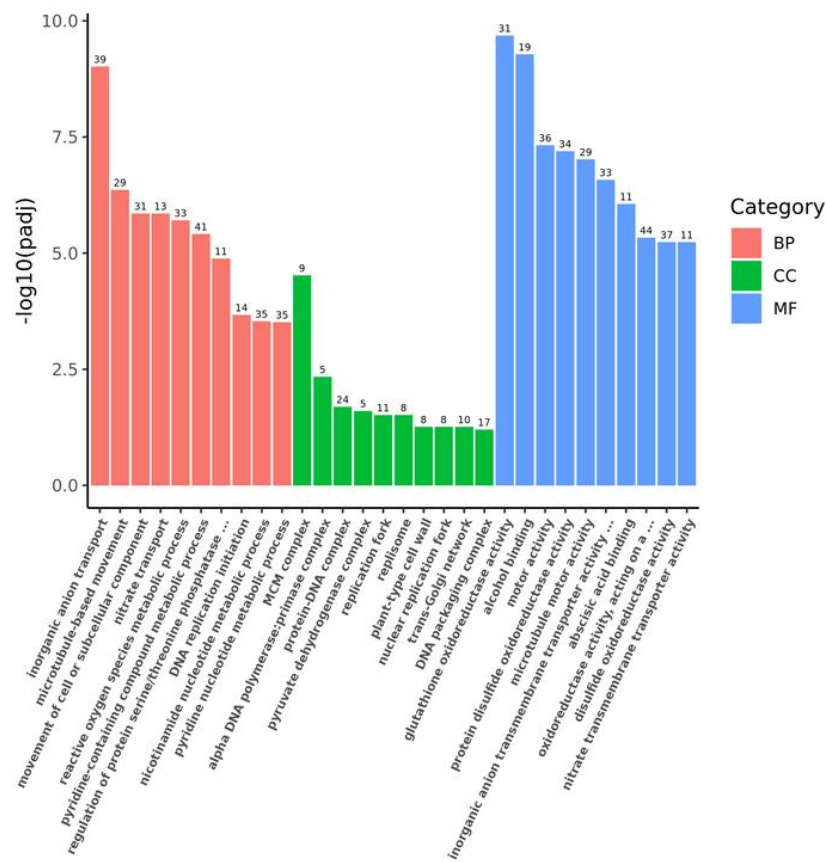
Próba	Name on the list	ng/ul	RIN	Próba	Name on the list	ng/ul	RIN
1	Rz2_2	624	8.1	15	Rz_kon1_21	377	6.8
2	Rz3_3	89	7.5	16	Rz_kon2_22	127	6.4
3	Rz4_4	1028	9.4	17	Rz_kon3_23	279	6.9
4	Rz5_25	810	9.1	18	Rz_kon4_24	102	6.5
5	Rz6_26	577	8.7	19	Rz_kon5_42	120	6.4
6	Rz7_27	914	9	20	Rz_kon6_43	199	8
7	Rz_kom1_17	218	6.8	21	Rz_kon7_44	127	8.8
8	Rz_kom2_18	396	7.4				
9	Rz_kom3_19	188	4.4				
10	Rz_kom4_20	133	6.7				
11	Rz_kom5_38	948	8.6				
12	Rz_kom6_39	661	8.5				
13	Rz_kom7_40	67	7.5				
14	Rz_kom8_41	187	7.5				

Odczytane sekwencje mogą zawierać odczyty o niskiej jakości lub odczyty z adapterami, które wpływają na jakość i wiarygodność biblioteki. Aby tego uniknąć niezbędne jest „przefiltrowanie” odczytów i uzyskanie odczytów czystych. Usuwane są odczyty z adapterami, odczyty, w których zawartość nukleotydów z odczytem niepewnym wynosi >10% i odczyty o niskiej jakości (**Tab. 2**). Duża liczba uzyskanych odczytów (min. 74 mln) gwarantuje wysoką jakość analiz transkryptomicznych.

**Tabela 2.** Analiza statystyczna wyników sekwencjonowania.

Próba	Biblioteka	Raw_reads	Raw_bases	Clean_reads	Clean_bases	Error_rate
Rz2_2	MRRA240019647-1a	79015256	11.85G	76137590	11.42G	0.01
Rz3_3	MRRA240019648-1a	89669950	13.45G	81338594	12.2G	0.01
Rz4_4	MRRA240019649-1a	99952450	14.99G	93561452	14.03G	0.01
Rz5_25	MRRA240019654-1a	105644380	15.85G	100166232	15.02G	0.01
Rz6_26	MRRA240019655-1a	82145460	12.32G	74954758	11.24G	0.01
Rz7_27	MRRA240019656-1a	95788230	14.37G	91610602	13.74G	0.01
Rz_kom1_17	MRRA240019682-1a	89922308	13.49G	83223450	12.48G	0.01
Rz_kom2_18	MRRA240019683-1a	79951628	11.99G	79924050	11.99G	0.01
Rz_kom3_19	MRRA240019684-1a	78701664	11.81G	78661596	11.8G	0.01
Rz_kom4_20	MRRA240019685-1a	80965358	12.14G	73098586	10.96G	0.01
Rz_kon1_21	MRRA240019690-1a	81597780	12.24G	74766902	11.22G	0.01

- Raw\_bases – liczba odczytów\* długość sekwencji
- Clean reads – odczyty sekwencji po odrzuceniu odczytów o niskiej jakości.
- Zmapowano uzyskane sekwencje do genomu referencyjnego *Brassica napus ssp oleifera*.
- Zmapowano >85% sekwencji, z czego ~81% sekwencji to sekwencje unikalne.
- Wykonano kwalifikację zidentyfikowanych genów do poszczególnych kategorii funkcjonalnych (GO).
- Rzepak w czystej uprawie vs rzepak w koniczynie: najczęściej genów należało do kategorii „podziały mitotyczne komórki” i „inicjacja replikacji DNA” (**Rys.1A**).
- Rzepak w czystej uprawie vs rzepak w koniczynie najczęściej genów należało do kategorii „transport anionów” i „ruch w oparciu o mikrotubule” (**Rys. 1B**).

**A****B**

**Ryc.1.** Przyporządkowanie zidentyfikowanych genów do poszczególnych kategorii funkcjonalnych (GO). **A**) rzepak w czystej uprawie vs rzepak w koniczynie, **B**) rzepak w czystej uprawie vs rzepak w koniczy

Sekwencje odniesiono do genomu referencyjnego rzepaku. Posłużono się programem HISAT2 (Tab. 3).

**Tabela 3.** Podsumowanie statystyczne mapowania do genomu referencyjnego.

Próba	Liczba odczytów	Liczba i %przypisanych odczytów	Unique_map	Multi_map
Rz2_2	76137590	69425851 (91.18%)	66205697 (86.96%)	3220154 (4.23%)
Rz3_3	81338594	75656689 (93.01%)	71343652 (87.71%)	4313037 (5.3%)
Rz4_4	93561452	86854429 (92.83%)	82666761 (88.36%)	4187668 (4.48%)
Rz5_25	100166232	93064240 (92.91%)	88525237 (88.38%)	4539003 (4.53%)
Rz6_26	74954758	69085739 (92.17%)	65737015 (87.7%)	3348724 (4.47%)
Rz7_27	91610602	85358025 (93.17%)	81145771 (88.58%)	4212254 (4.6%)
Rz_kom1_17	83223450	76456682 (91.87%)	71754043 (86.22%)	4702639 (5.65%)
Rz_kom2_18	79924050	73347615 (91.77%)	68725612 (85.99%)	4622003 (5.78%)
Rz_kom3_19	78661596	71895599 (91.4%)	67536718 (85.86%)	4358881 (5.54%)
Rz_kom4_20	73098586	66620649 (91.14%)	63072080 (86.28%)	3548569 (4.85%)
Rz_kon1_21	74766902	69241201 (92.61%)	65048971 (87.0%)	4192230 (5.61%)

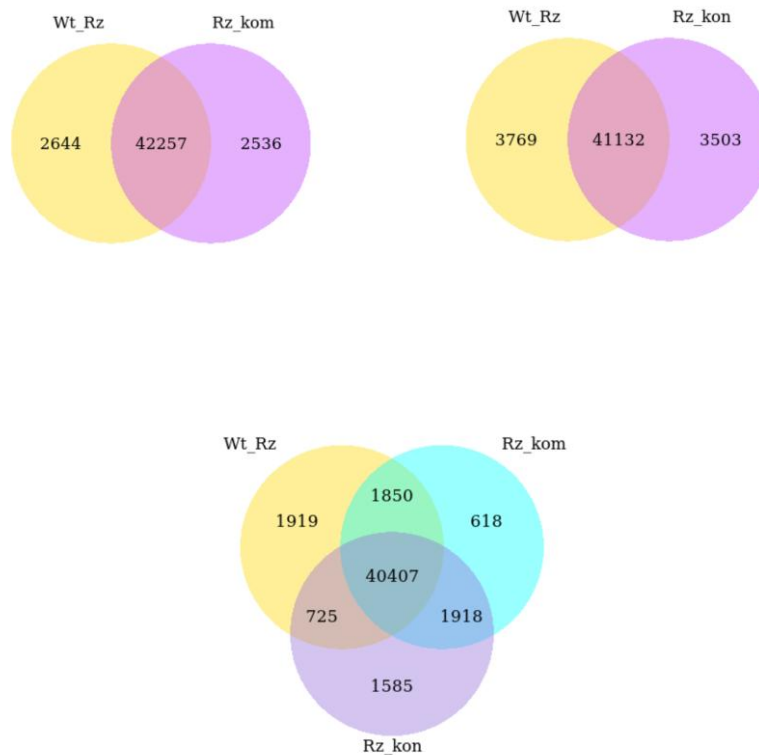
Total reads – wszystkie odczyty,

Total map – liczba odczytów, odniesiona do genomu referencyjnego,

Unique map – liczba odczytów, zmapowanych w miejsca unikalne w genomie,

Multi-map – liczba odczytów, zmapowanych w liczne miejsca w genomie,

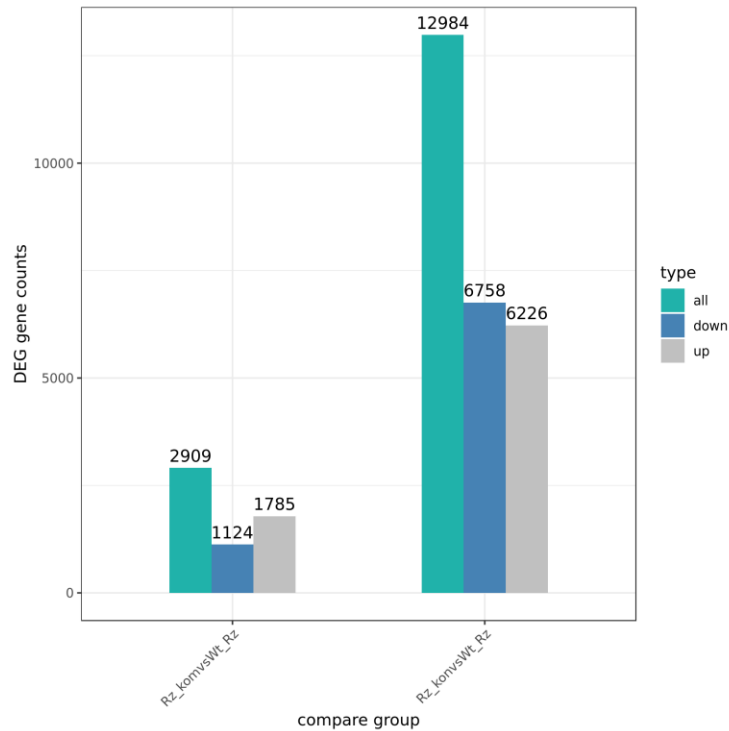
Dokonano porównania liczby sekwencji aktywowanych w uprawie czystej i mieszanej, rzepak i rzepak w komonicy oraz rzepak i rzepak w koniczynie. W porównaniu pierwszym, sekwencji specyficznych tylko dla rzepaku w siewie czystym było 2644, dla rzepaku w komonicy 2536. W porównaniu drugim zidentyfikowano dla rzepaku w siewie czystym 3769 sekwencji *versus* 3503 w siewie z koniczyną. Porównanie trzecie – rzepak w siewie czystym vs rzepak w komonicy i rzepak w koniczynie wykazała 4122 sekwencji specyficznych wobec 40407 wspólnych (**Rys 2**).



**Ryc 2.** Diagram Venna, porównanie liczby sekwencji wspólnych i oddzielnych dla rzepaku w siewie czystym, siewie w komonicy i siewie w koniczynie

Zidentyfikowano sekwencje o zróżnicowanej ekspresji pomiędzy rzepakiem w siewie czystym i mieszanym. Wskazano sekwencje, których ekspresja wzrosła lub zmalała oraz podsumowano wszystkie (**Rys.3**).

compare	all	up	down	threshold
Rz_konvsWt_Rz	12984	6226	6758	DESeq2 padj<=0.05  log2FoldChange >=1.0
Rz_komvsWt_Rz	2909	1785	1124	DESeq2 padj<=0.05  log2FoldChange >=1.0



**Ryc.3.** Sekwencje o różnicowanej ekspresji pomiędzy rzepakiem w siewie czystym i mieszanym.

Oddziaływanie licznych genów zebrano w poszczególne szlaki metaboliczne (analiza KEGG, Kyoto Encyclopedia of Genes and Genomes). W przypadku porównania rzepak w siewie czystym vs rzepak w koniczynie wskazano pięć najważniejszych szlaków metabolicznych: replikacja DNA, biosynteza fenylopropanoidów, białka motoryczne, metabolizm cysteiny i metioniny, wydłużanie kwasów tłuszczowych. Dla porównania rzepak w siewie czystym vs rzepak w komonicy również najważniejszym szlakiem okazała się replikacja DNA, białka motoryczne, ale oprócz tego metabolizm azotu, siarki, synteza alkaloidów piperidynowych i pyridynowych.

Wybrano pięć sekwencji o największej różnicy ekspresji pomiędzy rzepakiem w uprawie czystej, a rzepakiem w uprawie z koniczyną (**Tab. 4**).

**Tabela 4.** Sekwencje o największej różnicy ekspresji pomiędzy rzepakiem w uprawie czystej a rzepakiem w uprawie z koniczyną

1	BnaA01g21560D	530.9	530.9	6.3	ATP phosphoribosyltransferase 1
2	BnaA09g06210D	36.5	1664.4	5.5	BnaA09g06210D protein
3	BnaC03g49300D	112.3	2135.0	4.2	Asparagine synthetase
4	BnaA01g14450D	33.1	7294.0	7.7	Superoxide dismutase [Fe] 1
5	BnaC05g16070D	1.8	231.7	6.9	Cysteine protease XCP2

Log2FoldChange – wartość logarytmu z parametru określającego różnicę pomiędzy poziomem ekspresji jednej porównywanej grupy wobec drugiej

## **Podsumowanie analiz transkryptomicznych dla prób rzepaku:**

Do sekwencjonowania RNAseq wybrano linie rzepaku z uprawy czystej, uprawy z komonicy i koniczyną. Po izolacji potwierdzono jakość i integralność uzyskanego materiału parametrem RIN>4.4. W wyniku sekwencjonowania uzyskano 74 mln odczytów, co gwarantuje wysoką jakość analiz transkryptomu. Odniesiono pozyskane sekwencje do genomu referencyjnego grochu. Udało się zmapować ok. 85% sekwencji. 81% z nich zostało ulokowanych w miejscach unikalnych w genomie. Wskazano sekwencje o największej różnicy ekspresji pomiędzy liniami w uprawie czystej i mieszanej. Scharakteryzowano funkcjonalnie wskazane sekwencje (Tabela 4). Określono ich potencjalny udział w różnych szlakach metabolicznych: replikacja DNA, biosynteza fenyl-propanoidów, białka motoryczne, metabolizm cysteiny i metioniny, wydłużanie kwasów tłuszczowych. Dla porównania rzepak w siewie czystym vs rzepak w komonicy również najważniejszym szlakiem okazała się replikacja DNA, białka motoryczne, ale oprócz tego metabolizm azotu, siarki, synteza alkaloidów piperidynowych i pyridynowych.