

УДК: 577.218:577.29

**АНАЛІЗ ПРОЯВУ ДНК-КОНСТРУКЦІЙ ЗІ САЙТ-СПЕЦИФІЧНОЮ РЕКОМБІНАЗНОЮ СИСТЕМОЮ CRE/LOXP У ДЕКІЛЬКОХ ПОКОЛІННЯХ ТРАНСФОРМАНТІВ *ARABIDOPSIS THALIANA* ЗА ДОПОМОГОЮ МУЛЬТИВАРІАТИВНОГО МЕТОДУ**

**А. С. СЕКАН**, здобувач,

*ДУ «Інститут харчової біотехнології та геноміки» НАН України,*

**О. С. МИРОНИЧЕВА**, кандидат сільськогосподарських наук,

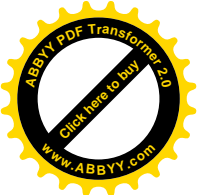
*Таврійський державний агротехнологічний університет*

*E-mail: ehirta3@gmail.com*

***Анотація.** Представлені результати мультиваріативного статистичного аналізу прояву двох векторних ДНК-конструкцій в декількох поколіннях трансформантів *Arabidopsis thaliana*. Розроблено статистичну модель класифікаційного типу, за допомогою котрої можливо достовірно розділити між собою досліджувані конструкції на дві окремих події.*

***Ключові слова.** Мультиваріантний аналіз, метод головних компонент, сайт-специфічна рекомбіназна система Cre/LoxP, генетична трансформація*

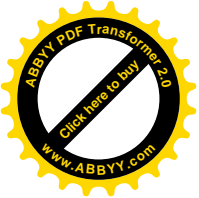
Останнім часом зростає потреба в комп'ютерному моделюванні й адаптації класичних статистичних технологій для універсалізації цих підходів для того, щоб забезпечити більш якісне вирішення відповідних завдань у біологічних дослідженнях [1]. Загалом, для розроблення нової методики в генетичній інженерії створюють різноманітні векторні конструкції з наступним тестуванням на певних рослинах. Проте інколи спостерігається варіація в ефективності використання розроблених конструкцій під час зміни рослинного об'єкту [2]. Це, в свою чергу, викликає труднощі в комерційному застосуванні розробленої методології. Окрім цього, під час отримання трансгенної рослини важливо контролювати вплив різного роду факторів на здійснення процесу трансформації. Таким чином, в кінці проведеної роботи накопичується значний масив даних, що в свою чергу ускладнює відтворення здійсненої роботи та отримання подібних результатів. Тому розроблення емпіричної моделі процесу для класифікації



даних з метою їх масштабування та прогнозування прояву синтезованого конструкту, є важливим завданням під час планування та здійснення експерименту [3]. Моделювання досліджуваного процесу з метою класифікації, або прогнозування отриманих емпіричних даних дає можливість виявити прихований фактор, який впливає на успішність проведення експерименту, а також дає можливість пояснити результати роботи на більш високому професійному рівні [4].

В рослинній біотехнології під час здійснення генетичної трансформації дослідник отримує величезну кількість результатів, що важко аналізувати. Це призводить до втрати інформації, яку можна отримати з емпіричних даних. Разом з тим, за допомогою новітніх статистичних програм, таких як SPSS (statistical package for social science), Statistica та ін. стає можливим структурувати отримані результати з їх наступним аналізом. Окрім побудови зручної статистичної моделі прояву досліджуваного об'єкту з можливістю передбачення та масштабування експерименту стає можливим також і побудова схематичної класифікаційної моделі результатів.

Так, за допомогою методу головних компонент (PCA, Principal Component Analysis) здійснюють розподілення результатів, отриманих в результаті здійснення експериментів, націлених на дослідження прояву обраного об'єкту. Як один із варіантів, вдається знайти можливий прихований взаємозв'язок між, здавалось, зовсім різними факторами, які на перший погляд не мають нічого спільного. Загалом, PCA є класичним методом пониження розмірності набору отриманих даних шляхом їх перетворення в новий набір змінних величин (головних компонент) з метою підсумовування основних ознак даних [5]. Ця технологія застосовується для аналізу величин, які не корелюють між собою, та може бути інтерпретована як напрямок, який максимізує варіацію проєкцій отриманих даних на площині. На сьогоднішній день немає жодної публікації, в якій би описувалось використання мультиваріативного статистичного аналізу

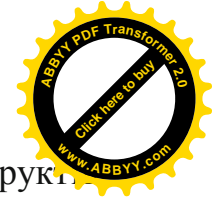
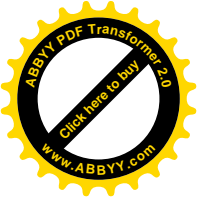


емпіричних даних, отриманих внаслідок проведення генетичної трансформації рослин.

Тому, **метою дослідження** було створення за допомогою методу головних компонент і дискримінантного аналізу статистичної моделі, яка б описала ефективність генетичної трансформації рослин з можливістю прогнозування подальшого розвитку подій для наступних поколінь трансгенів.

**Матеріали та методи дослідження.** Для побудови одноваріантної статистичної моделі використовувались данні досліджень прояву декількох ДНК-конструкцій, які містять сайт-специфічну рекомбіназну систему Cre/loxP. Досліджувані векторні конструкції були створені для трансформації арабідопсису з метою отримання трансгенів, вільних від маркерних або селективних послідовностей. Детально процес трансформації описаний у роботах [6, 7]. Як результат, було створено матрицю даних розміром 120×15 (досліджувалось всього 120 ліній трансформантів у двох поколіннях; протягом цього періоду було здійснено 15 досліджень). Отримані данні обробляли за допомогою мультиваріативного аналізу, а саме за допомогою методу головних компонент (PCA, Principal Component Analysis) та дискримінантного аналізу (PLS-DA, Projection to Latent Structures – Discriminant Analysis). Аналіз даних виконувався в середовищі SIMCA 13,0 (Soft Independent Modelling of Class Analogy, Umetrics AB, Sweden). для обчислення моделі з наступним розмежуванням спостережених груп.

**Результати та їх обговорення.** Як правило, побудова PCA-моделі базується на одночасному визначенні таких параметрів спостережених варіант, як  $R^2X$  (ступінь варіативності досліджуваних даних) і  $Q^2X$  (ступінь потенційного передбачення майбутніх результатів). Ці параметри характеризуються взаємозамінністю та є безрозмірними величинами, однак, без значення  $R^2X$  неможливо визначити  $Q^2X$  та навпаки. Під час побудови моделі важливо  $Q^2X > 0,5$  що дає право стверджувати про її прийнятність. Значення  $Q^2X > 0,9$  характеризує створену модель як відмінну. У випадку



побудови статистичної моделі для прояву двох досліджуваних конструкцій шляхом пророщування трансформантів на селективному середовищі було визначено загальне  $Q^2=0,4$  (табл. 1). Це є недостатнім для створення прийнятної класифікаційної моделі.

### 1. Класифікаційна статистична модель, побудована за допомогою методу PCA

Component	R <sup>2</sup> Cent.	R <sup>2</sup> (cum)	E.value	Q <sup>2</sup>	Limit	Q <sup>2</sup> (cum)
Cre/loxP 1	0,444	0,444	6,66	0,386	0,0703	<b>0,386</b>
Ce/loxP 2	0,136	0,58	2,04	0,0221	0,0745	<b>0,399</b>

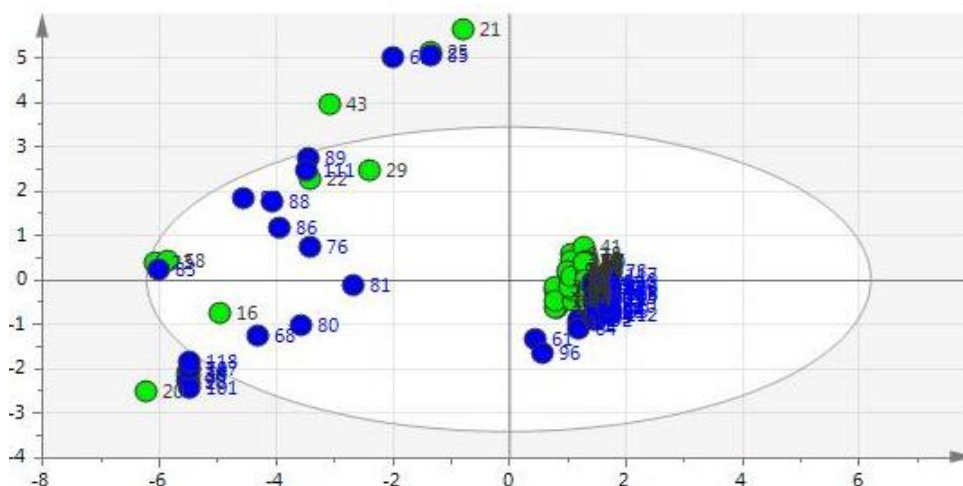
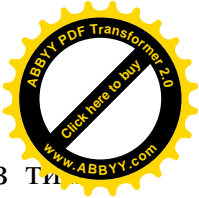
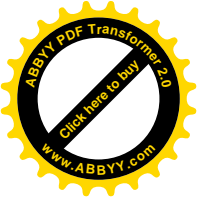


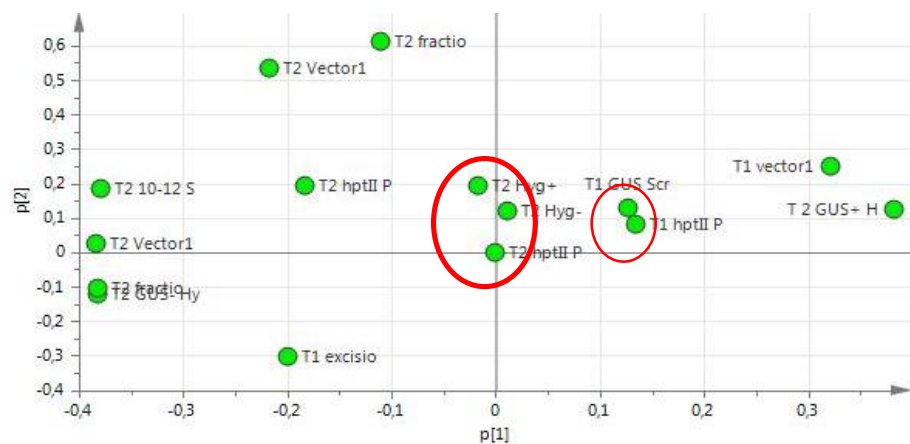
Рис. 1. Проекція результатів селективного відбору трансформантів (кількість резистентних рослин для кожної з трансформованих ліній) з ДНК-конструкціями Cre/loxP1 (точки зеленого кольору) і Cre/loxP2 (точки синього кольору) на селективному середовищі MS з антибіотиком гігроміцином (100 мг/л). Вісь ординат – R<sup>2</sup> значення; вісь абсцис – Q<sup>2</sup> значення.

За допомогою графічного зображення на рис. 1 показано досить чітке кластерне розмежування, що вказує на створення саме класифікаційної моделі. Використання методу PCA в даному випадку дає можливість

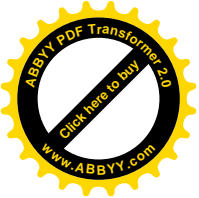


розмежувати досліджувані конструкції на два окремих класи. Разом з тим спостерігається значна частка викидань, що знижує точність побудови моделі, а отже й понижає значення параметрів  $R^2$  та  $Q^2$ . За спостереженням розсіюванням результатів можна зробити висновок, що прояв T-ДНК Cre/loxP2 в отриманих лініях трансформантів є менш стабільним. Це значно ускладнює подальше прогнозування прояву даного конструкту, як і створення відповідної статистичної моделі.

З метою отримання прийнятнішого варіанту класифікаційної моделі був застосований інший метод аналізу даних PLS-DA. Загалом, для побудови окремого графіку за методом головних компонент враховуються дані всієї матриці, а не лише окремої групи досліджень. Тому, за допомогою методу PLS-DA ми встановили найбільш значущу частину результатів проведених досліджень (перемінних) з аналізу двох синтезованих конструктів – Cre/loxP1 і Cre/loxP2 (рис. 2).



**Рис. 2.** Розподілення результатів досліджень, отриманих протягом двох поколінь трансформантів арабідопсису за двома ДНК-конструктами на площині координат першої та другої компонент, розраховане за допомогою методу PLS-DA. Вісь ординат – результати, отримані від використання ДНК-конструкту Cre/loxP2. Вісь абсцис – результати, отримані від використання ДНК-конструкту Cre/loxP2.

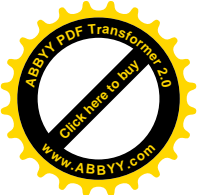


Основним принципом даної методології є виявлення значущих перемінних серед загальної маси отриманих даних. Таким чином, за допомогою PLS-DA вдається встановити, яка частина здійснених експериментів є найбільш значущою з наступним відокремленням її від основної маси латентних даних (ті, в яких спостерігається значна частина викидань) та проведення класифікаційного аналізу саме із цією частиною результатів.

T1 GUS scri – лінії з позитивним результатом GUS-тесту в поколінні T<sub>0</sub>; T1 hptII pos – лінії з позитивним результатом ПЛР-аналізу на присутність T-ДНК в геномі покоління T<sub>0</sub>; T1 ves1 – лінії з позитивним результатом ПЛР-аналізу на присутність *loxP*-обмежених ділянок T-ДНК в геномі покоління T<sub>0</sub>; T1 exci – лінії покоління T<sub>0</sub>, де зафіксовано події видалення *loxP*-обмежених ділянок ДНК за допомогою рекомбінази Cre; T2 Hvg+ – кількість трансформантів на кожен лінійку покоління T<sub>1</sub>, резистентних до антибіотику; T2 Hvg- – кількість трансформантів на кожен лінійку покоління T<sub>1</sub>, резистентних до антибіотику; T2 GUS+ H – лінії трансформантів покоління T<sub>1</sub>, позитивно забарвлені за GUS-тестом на селективному середовищі; T2 GUS- H – лінії трансформантів покоління T<sub>1</sub>, негативно забарвлені за GUS-тестом на селективному середовищі; T2 10-12 S – зразки лінійку трансформантів покоління T<sub>1</sub>, перенесених в ґрунт; T2 fractio – лінійку трансформантів покоління T<sub>1</sub> з позитивним результатом GUS-тесту за умов вирощування в ґрунті; T2 hptII pos – лінійку з позитивним результатом ПЛР-аналізу на присутність T-ДНК в геномі покоління T<sub>1</sub>; T2 hptII neg – лінійку з позитивним результатом ПЛР-аналізу на присутність T-ДНК в геномі покоління T<sub>1</sub>; T2 ves1 – лінійку з позитивним результатом ПЛР-аналізу на присутність *loxP*-обмежених ділянок T-ДНК в геномі покоління T<sub>1</sub>.

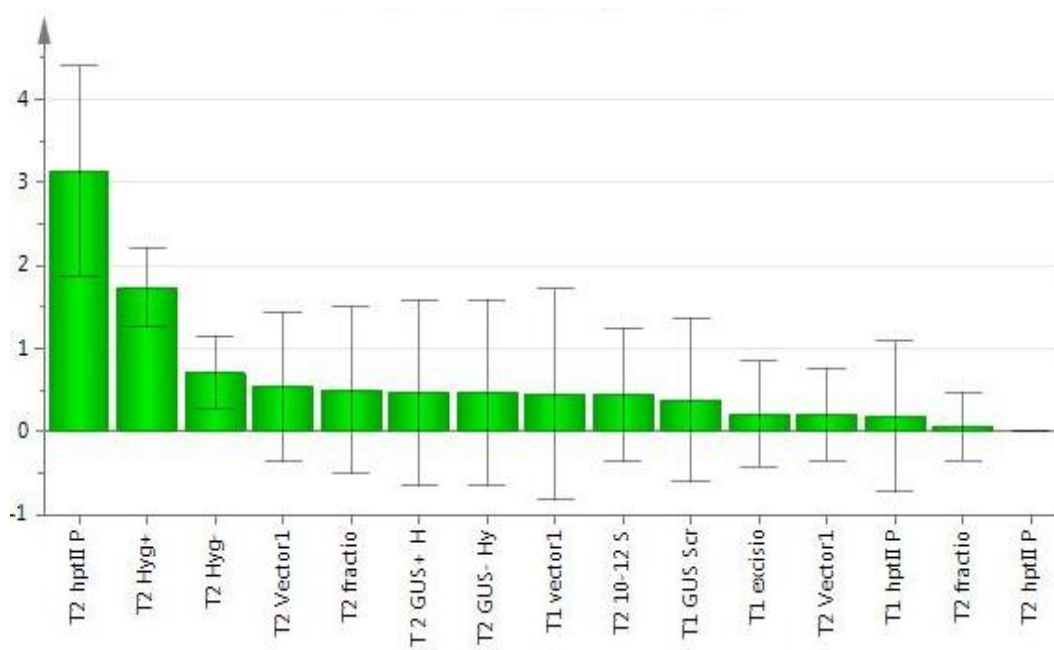
В результаті проектування даних із досліджень двох компонент (ДНК-конструктів) виявлено певне групування деяких результатів. Так, окремою групою із сильним факторним зв'язком виступають результати, отримані від досліджень на резистентність рослин до антибіотику на селективному





середовищі MS, а також результати ПЛР-аналізу на присутність Т-ДНК в геномі трансформанту. Саме за допомогою цих перемінних можливо розділити досліджувані векторні конструкти на окремі спостережені класи. Спостережені класи досліджень резистентності трансформантів та проведення ПЛР-аналізу мають пряму залежність і знаходяться в центрі схеми, чим пояснюють загальне розподілення дисперсії серед інших класів на даному графіку. Також, за допомогою аналізу PLS-DA було підтверджено, що в коефіцієнтному еквіваленті найбільш значущими результати селективного відбору трансформантів покоління  $T_1$  та ПЛР-аналізу на присутність Т-ДНК в геномі (рис. 3). Як результат, нам вдалось визначити найбільш значущу частину виконаних дослідів. Це, в свою чергу, дає нам можливість в подальшому проводити саме ці експерименти. Під час здійснення тестування подібної векторної конструкції, або ж продовження тестування даних конструктів в майбутньому буде достатньо проведення лише ПЛР-аналізу та селективного відбору трансформантів. Використання методології мультиваріативного аналізу (в даному випадку PLS-DA) дає можливість спростити дослідження обраного об'єкту, а отже мінімізувати затрати часу та зусиль для досягнення поставленої мети.

Для моделі PLS-DA (рис. 3) за допомогою SIMCA й відображення коефіцієнтного співвідношення результатів прописано рівняння регресійного характеру, як  $Y = Y_{avg} + xB + F$ , де  $Y_{avg}$  є константою (не відображено на рис. 3),  $B$  – коефіцієнт самої моделі, за яким побудовано рангування на рис. 3,  $F$  – окреме значення для кожного з аналізованих досліджень. За допомогою виведеного рівняння можливо описати взаємозв'язок між спостереженими факторами шляхом порівняння результатів між виконаними дослідженнями. За силою зв'язку всіх спостережених класів (здійснених досліджень), котрі можна описати даним рівнянням, побудовано схему рангування – VIP (Variable Importance for the Projection, рис. 3) [8, 9].



**Рис. 3. Рівень кореляції результатів досліджень з використання ДНК-конструкцій *Cre/loxP1* і *Cre/loxP2* протягом двох поколінь трансформантів.** Вісь ординат – коефіцієнт кореляції спостережених класів; вісь абсцис – проектування результатів прояву Т-ДНК *Cre/loxP1* та *Cre/loxP2* протягом двох поколінь в трансформантах арабідопсису.

Дана схема наглядно описує співвідношення у зв'язку між спостереженими класами і виявляє можливу приховану взаємозалежність між ними. На VIP-схемі демонструються довірчі інтервали для кожного із зображених значень, як правило, при 95 % рівні. Найсильніший взаємозв'язок, як і у випадку аналізу PLS-DA, продемонстрований між спостереженими класами селективного відбору трансформантів та ПЛР-аналізу. Рівень взаємозв'язку між іншими дослідженнями менший за величину 0,5 а отже є незначущим. Окрім цього, виходячи з аналізу PLS-DA та VIP-схеми, не спостерігається жодного зв'язку між успішністю отримання вільних від маркерних послідовностей трансформантів та експресією гену *gus*, хоча в конструкті *Cre/loxP2* маркерний ген *gus* розташований попереду гену рекомбінази.

Оскільки за допомогою методу PLS-DA було виявлено перемінні, які дають найбільший відсоток викидань і тим самим значно впливають на





побудову статистичної моделі, подальший аналіз здійснювали з урахуванням лише трьох перемінних (спостережені класи селективного відбору трансформантів та ПЛР-аналізу). В результаті повторної побудови статистичної моделі за допомогою цього методу, її основні параметри виглядали, як  $Q^2=0,6$  (Табл. 2) для обох ДНК-конструкцій, що дозволяє прийняти дану модель та розділити спостережені класи на два кластери (рис. 4).

## 2. Класифікаційна статистична модель, побудована за допомогою методу PLS-DA

Component	R <sup>2</sup> Cent.	R <sup>2</sup> (cum)	E.nvalue	R <sup>2</sup> Y	R <sup>2</sup> Y (cum)	Q <sup>2</sup>	Limit	Q2 (cum)
Cre/loxP1	0,431	0,431	1,29	0,593	0,593	0,583	0	<b>0,583</b>
Cre/loxP2	0,363	0,794	1,09	0,00841	0,601	0,00667	0	<b>0,586</b>

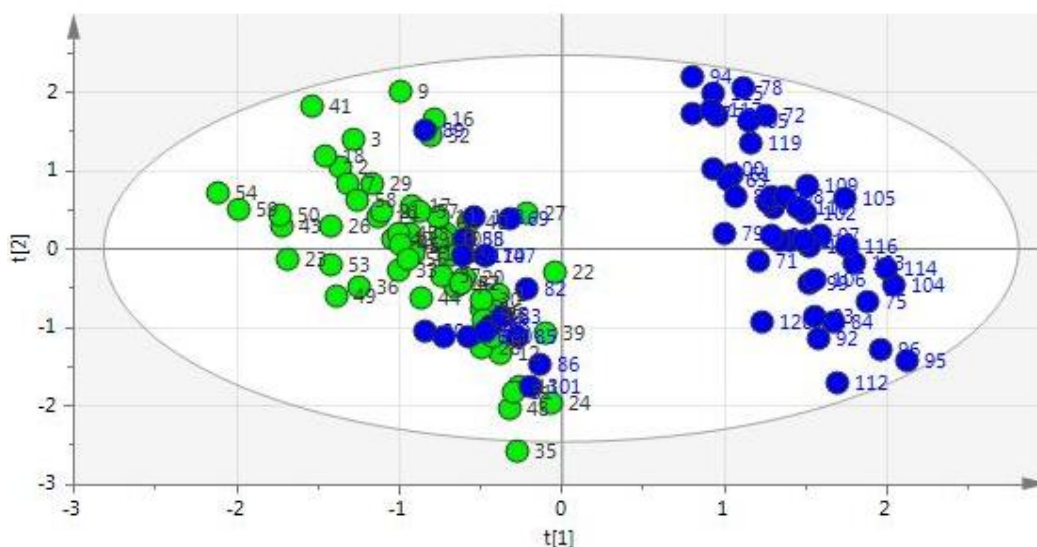
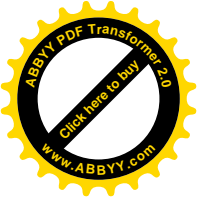


Рис. 4. Результуюча схема кластерного розподілення між двома компонентами *Cre/loxP1* та *Cre/loxP2* на селективному середовищі MS з гігроміцином (100 мг/л).

За допомогою PLS-DA побудовано нове схематичне зображення (рис. 4), що максимально відображає розподілення результатів на дві окремих групи, які й характеризують досліджувані ДНК-конструкції.



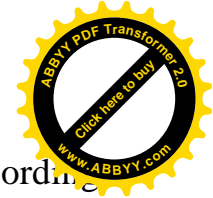
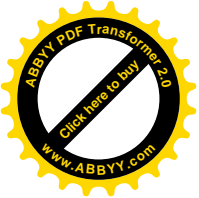
За результатами PCA-аналізу вдалось визначити, що результати досліджень піддаються класифікації. Однак статистична модель, побудована за допомогою PCA-аналізу, недостатньо гарно описує прояв двох ДНК-конструкцій та не має відповідного рівня прогнозування. Тому, з метою побудови класифікаційного типу моделі, яка описує прояв створених векторів, нами було здійснено ще один тип мультиваріативного аналізу – PLS-DA. В результаті вдалось визначити найбільш значущу частину досліджень та відокремити її від загальної кількості здійснених дослідів. Це, в свою чергу, дає нам можливість оптимізувати планування експериментів для подальшого аналізу конструкцій, а також передбачити їх прояв в майбутніх поколіннях без здійснення додаткової серії досліджень. Доведено, що побудована класифікаційна модель дає можливість передбачити розвиток подій на 60 %, що є достатнім для її використання.

### **Висновки**

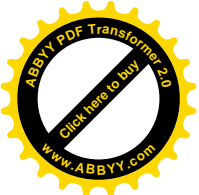
1. В результаті проведеної роботи розроблено статистичну модель класифікаційного типу, котра описує прояв двох ДНК-конструкцій в декількох поколіннях трансформантів арабідопсису.
2. За допомогою мультиваріативного аналізу встановили, що дві досліджувані векторні конструкції можливо достовірно розділити між собою на дві окремих події.
3. За допомогою створеної моделі можливо спрогнозувати майбутній прояв ДНК-конструкцій в наступних поколіннях із точністю 60 %.
4. В результаті здійснення VIP-аналізу встановили, які дослідження є найважливішими, таким чином оптимізувавши планування експериментів для подальшого аналізу ДНК-конструкцій.

### **СПИСОК ЛІТЕРАТУРИ**

1. Barnes D.J. Introduction to Modeling for Biosciences / Barnes D.J., Chu D. // Springer. – 2010. – pp. 334.



2. Yang S.-O. Metabolic discrimination of *Catharanthus roseus* calli according to their relative location using H-NMR and principal component analysis/ Yang S.-O., Kim S.-H., Kim Y., Kim H.-S., Chun Y.-J., Choi H. K. // Biosci. Biotechnol. Biochem. – 2009. – Vol.73, No 9. – P. 2032 – 2036.
3. Epstein J.M. Why model? / Epstein J.M. // Jasss. – 2008. – Vol.11, No 4. – P. 12 – 17.
4. Le H. Multivariate analysis of cell culture bioprocess data – lactate consumption as process indicator / Le H., Kabbur S., Pollastrini L., Sun Z., Mills K., Johnson K., Karypis G., Hu W.-Sh. // J. Biotechnol. – 2012. – Vol.162. – P. 210 – 223.
5. Yeng K.Y. Principal component analysis for clustering gene expression data / Yeng K.Y., Ruzzo W.L. // Bioinform. – 2001. – Vol.17, No.9. – P. 763 – 774.
6. Секан А. С. Створення конструкцій з сайт-специфічною рекомбіназною системою Cre/loxP під контролем 35S промотору та їх використання для отримання трансформантів *Arabidopsis thaliana*, вільних від маркерних послідовностей / А. С.Секан, С. В. Ісаєнков // Наукові доп. НУБіП. – 2014. - №5 (47). – Режим доступу: [http://nbuv.gov.ua/j-pdf/Nd\\_2014\\_5\\_3.pdf](http://nbuv.gov.ua/j-pdf/Nd_2014_5_3.pdf)
7. Секан А. С. Ефективність трансформації *Arabidopsis thaliana* ДНК-конструкціями з рекомбіназною системою Cre/loxP / А. С. Секан, С. В. Ісаєнков // Biotech Acta. – 2015. – V.8., №7 – P. 48 – 55.
8. Galindo-Prieto B. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS) / Galindo-Prieto B., Eriksson L., Trygg J. // J Chemom. – 2014. – Vol. 28. – P. 623–632.
9. Wei M. Multivariate modeling on biomass properties of cassava stems based on an experimental design / Wei M., Geladi P., Lestander T.A., Xie G., Xiong Sh. // Anal Bioanal Chem. – 2015. – V.207, No.18. – P. 5443 – 52.



**АНАЛИЗ ПРОЯВЛЕНИЯ ДНК-КОНСТРУКЦИЙ С  
САЙТ-СПЕЦИФИЧЕСКОЙ РЕКОМБИНАЗНОЙ СИСТЕМОЙ  
CRE/LOXP В НЕСКОЛЬКИХ ПОКОЛЕНИЯХ ТРАНСФОРМАНТОВ  
ARABIDOPSIS THALIANA С ПОМОЩЬЮ МУЛЬТИВАРИАТИВНОГО  
МЕТОДА**

**А. С. Секан, Е. С. Мироничева**

*Аннотация.* Представлены результаты мультивариативного статистического анализа проявления двух векторных ДНК-конструкций в нескольких поколениях трансформантов *Arabidopsis thaliana*. Разработана статистическую модель классификационного типа, с помощью которой можно достоверно разделить между собой исследуемые конструкции на две отдельные группы.

*Ключевые слова.* Мультивариативный метод, анализ главных компонент, сайт-специфическая система Cre/loxP, генетическая трансформация

**OF DEVELOPED DNA-CONSTRUCTIONS WITH SITE-SPECIFIC  
RECOMBINASE SYSTEM CRE/LOXP DURING THE SEVERAL  
GENERATIONS IN ARABIDOPSIS THALIANA TRANSFORMANTS  
USING THE MULTIVARIATE METHOD**

**A. Sekan, O. Myronycheva**

*Abstract.* Here are represented the results of the development of two vector DNA-constructions during several generations of the *Arabidopsis thaliana* transformants, analyzed by multivariate statistical method. The statistical classification model is developed and allows to significantly divide studied constructions on the two separated events.

*Key words:* multivariate analysis, Principal Component Analysis, site-specific recombinase system Cre/loxP, genetic transformation