

УДК: 633.34:631.527.5

## Класифікація сортів сої за господарськими ознаками з допомогою кластерного аналізу

Білявська Л.Г. , Рибальченко А.М. 

Полтавська державна аграрна академія

 stryzhak.am@gmail.com

Білявська Л.Г., Рибальченко А.М. Класифікація сортів сої за господарськими ознаками з допомогою кластерного аналізу. Збірник наукових праць «Агробіологія», 2020. № 2. С. 7–15.

Biljavs'ka L.G., Rybal'chenko A.M. Kласифікація сортів сої за господарськими ознаками з допомогою кластерного аналізу. Zбірник наукових праць «Агробіологія», 2020. no. 2, pp. 7–15.

Рукопис отримано: 30.09.2020 р.

Прийнято: 15.10.2020 р.

Затверджено до друку: 24.11.2020 р.

doi: 10.33245/2310-9270-2020-161-2-7-15

Метою дослідження було удосконалити метод оцінювання колекційного матеріалу сої з метою підбору пар для схрещування за допомогою кластерного аналізу, провести кластеризацію зразків за тривалістю вегетаційного періоду, висотою рослин, висотою прикріплення нижнього бобу, товщиною стебла в нижній частині, кількістю гілок на рослині, кількістю продуктивних вузлів на рослині, кількістю бобів на рослині, кількістю насіння з рослини, кількістю насінин в бобі, масою насіння з рослини, урожайністю (г/м<sup>2</sup>), стійкістю до вилягання, стійкістю до розтріскування бобів. На основі проведеного кластерного аналізу сто сорока п'яти колекційних зразків сої за визначеними ознаками встановили, що колекційний матеріал розподілюється в межах п'яти кластерів. Було виділено генотипи сої як носії високого рівня кількісних ознак, особливо ознак потенціалу врожайності, які можна використовувати як перспективний селекційний матеріал.

В окремий кластер об'єднано зразки з аналогічним набором ознак всередині кластера і достовірною відмінністю з іншими.

Зразки першого кластера відзначаються середніми і близькими до середніх у колекції значеннями всіх досліджуваних ознак. Зразки другого кластера відзначаються високими значеннями кількості гілок на рослині, кількості продуктивних вузлів на рослині та кількості бобів на рослині. Третій кластер складався із зразків з високими значеннями кількості насінин в бобі. Однак за іншими ознаками зразки цієї групи мають низькі значення. Зразки четвертого кластера більш пізньостиглі, ніж перших трьох. Вони виділяються за більшою висотою рослини і висотою прикріплення нижнього бобу. П'ятий кластер зразків відзначається найтривалішим вегетаційним періодом та високими значеннями більшості ознак, крім кількості насінин в бобі, стійкості до вилягання і стійкості до розтріскування бобів. Батьківські форми розмістилися у чотирьох кластерах із п'яти.

Проведений кластерний аналіз допоміг здійснити підбір батьківських пар для проведення гібридизації та створення нового вихідного матеріалу сої. Перевага кластерного аналізу полягає в тому, що він дає змогу ідентифікувати колекційні зразки збалансовано за комплексом цінних господарських ознак, а не виділити генотипи за окремими кількісними ознаками.

**Ключові слова:** соя, селекція, колекція, зразок, кластер, кластерний аналіз, вихідний матеріал.

**Постановка проблеми.** Успіх селекції переважно визначається підбором матеріалу, з яким буде проводитися робота, точніше підбором батьківських пар для схрещування, оскільки гібридизація – це основний спосіб отримання нових сортів. Якщо не підібрані відповідні батьки, гени яких мають бути рекомбіновані у новому

сорті, незважаючи на створену модель і бажаний тип сорту, значного успіху досягти неможливо. Під час створення нових сортів селекціонер має закласти в них якомога більше корисних ознак. Проблема полягає у відсутності надійного способу підбору батьківських пар за комплексом великої кількості різномірних ознак [12].

Для всебічного оцінювання селекційного матеріалу за комплексом ознак ряд учених застосовують методи багатовимірної статистики, а саме факторний кластерний аналіз. Універсальність кластерного аналізу, як методу, дає змогу застосовувати його для різних культур і в поєднанні з іншими статистичними процедурами [1, 2].

**Аналіз останніх досліджень.** Кластерний аналіз був створений для розбиття багатьох дослідних об'єктів та ознак на однорідні, певною мірою, групи, або кластери. Особливістю та перевагою цього аналізу є те, що в ньому виключається вибірка, що дає змогу робити розбиття (сортування) об'єктів не за одним параметром, а за рядом ознак [3, 4].

Кластерний аналіз (*англ. Data clustering*) – це задача розподілу заданої вибірки об'єктів (ситуацій) на підмножини, які називаються кластерами, так, щоб кожен кластер складався зі схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися [13]. Кластерний аналіз – це багатовимірний статистичний метод, що виконує збір даних, які містять інформацію про вибірку об'єктів, і потім впорядковує об'єкти в порівняно однорідні групи (кластери). Кластер – група елементів, які характеризуються загальною властивістю, головною метою кластерного аналізу – знаходження груп схожих об'єктів у вибірці [9].

На відміну від традиційної статистичної обробки результатів одномірними методами, багатомірний системний аналіз за своєю природою є гнучкою технологією математичного аналізу. Алгоритм системного аналізу в кожному випадку залежить від мети селекціонера щодо оцінювання специфічності відмінності вихідного і селекційного матеріалу за системними властивостями, яке необхідне для вирішення конкретної задачі на цьому етапі селекції [11].

Кластерний аналіз, як різновид системного, дає змогу формально проводити багатомірну класифікацію, а змістовність цієї класифікації і цінність у вирішенні конкретних селекційних задач визначається обсягом взятих в аналіз ознак. Використання цього методу в селекції – розгрупувати селекційний матеріал на відмінні типи (кластери). Цей метод класифікації є засобом кількісного представлення робочих гіпотез щодо відмінності селекційного матеріалу за певною стороною проявлення макроскопічної мінливості [10].

Генетична різноманітність є джерелом варіацій для селекції, зумовлюючи створення нових форм господарсько цінних рослин з поліпшеними властивостями. Вузька генетична основа знижує ефективність селекції, оскільки

не дає змоги подолати вразливість до несприятливих чинників і обмежує можливості комбінаторики спадкового матеріалу під час гібридизації. У зв'язку з цим важливого значення набуває збереження генетичних ресурсів і оцінювання їх різноманітності для подальшого використання в практичних цілях [6, 20].

Під час роботи селекціонер часто стикається з великим обсягом матеріалу з цілого набору різних за своєю природою ознак, який необхідно якимось чином систематизувати, щоб виділити кращі форми за комплексом господарсько цінних ознак [14, 15].

Однією з проблем для дослідника під час вивчення різноманітності генетичних колекцій рослин є аналіз великого масиву даних. Зазвичай основним статистичним методом класифікації даних є кластерний аналіз, що дає змогу структурувати колекційні зразки [19]. Кластерний аналіз у дослідженнях застосовується під час вивчення ступеня генетичної спорідненості [16, 18].

Ідентифікація зразків у кластерному аналізі дає змогу виділити генотипи з мінімальною евклідовою відстанню між кількісними ознаками, що підтверджує високий рівень збалансованості основних генеративних, вегетативних ознак та урожайності цих генотипів. Використання кластерного аналізу для ідентифікації генотипів наближає до створення теоретичної і практичної моделі сорту, в якого сприятливе співвідношення кількісних та якісних ознак дає змогу протистояти негативним впливам навколишнього середовища та формувати високу продуктивність [21].

На відміну від багатьох інших статистичних процедур, кластерний аналіз використовують тоді, коли дослідник не має яких-небудь апріорних гіпотез, і полягає у визначенні серед вихідної множини оптимального значення цільової функції. Більшість алгоритмів класифікації побудовано на використанні методу К-середніх або алгоритму побудови деревоподібної графіки (ієрархічне дерево) під час формування кластерів відмінності або відстані між об'єктами. Ці відстані можуть визначатися в одномірному або багатомірному просторах. Найпростіше відстані між об'єктами в одномірному або багатомірному просторах можна обчислити через евклідові відстані [17].

Використання багатовимірних методів аналізу в селекційному процесі допомагає скоротити його тривалість через більш повне і комплексне оцінювання чинників, що впливають на реалізацію генетичного потенціалу рослин.

**Метою дослідження** передбачено адаптувати метод оцінювання колекційного матеріалу сої

для підбору батьківських форм для схрещування за допомогою кластерного аналізу. Завданням дослідження було провести кластеризацію колекційних зразків сої за господарсько цінними ознаками: тривалістю вегетаційного періоду, висотою рослин, висотою прикріплення нижнього бобу, товщиною стебла в нижній частині, кількістю гілок на рослині, кількістю продуктивних вузлів на рослині, кількістю бобів на рослині, кількістю насіння з рослини, кількістю насіння в бобі, масою насіння з рослини, урожайністю ( $\text{г}/\text{м}^2$ ), стійкістю до вилягання, стійкістю до розтріскування бобів.

**Матеріал і методи дослідження.** Польові дослідження проводили в 2013–2015 рр. на дослідному полі Полтавської державної аграрної академії, що за зональним розподілом належить до центральної підзони Лісостепу України. Ґрунт дослідної ділянки – чорнозем опідзолений на лесі, вміст гумусу в орному шарі 0–20 см – 3,95–4,36 %. Кількість гідролізованого азоту в орному шарі становить 5,96 мг, доступного для рослин фосфору – 9,5 мг, калію – 14,2 мг на 100 г Ґрунту. Агротехніка вирощування сої – загальноприйнята для зони.

Об'єктом дослідження була колекція сої, яка налічувала 145 зразків різного еколого-географічного походження. Вивчали колекційні зразки, які походили з 14 країн світу: України, Росії, США, Канади, Китаю, Японії, Польщі, Франції, Чехії, Білорусі, Казахстану, Австрії, Молдови, Сербії. У період вегетації проводили фенологічні спостереження відповідно до загальноприйнятої методики [5, 8]. Біометричні

виміри та структурний аналіз урожаю проводили згідно з Широкиим уніфікованим класифікатором роду *Glycine max (L.) Merr* [7].

Для врахування одночасної дії головних елементів клімату – температури та опадів, застосовували інтегральний показник – гідротермічний коефіцієнт (ГТК) Г.Т. Селянинова. Метеорологічні умови періоду вегетації сої у роки досліджень відрізнялись (табл. 1).

Для статистичної обробки даних колекційних зразків сої кластерного аналізу застосовували комп'ютерну програму «Statistica 6.0» версії STATSOFT.

**Результати дослідження та обговорення.** Багатовимірна статистика дає змогу селекціонерів об'єктивно оцінити вихідний матеріал за набором ознак. Використання кластерного аналізу за найбільш важливими господарсько цінними ознаками дало змогу розбити зразки вивченої колекції на групи з різномірною селекційною значимістю (дендрограма кластеризації). Колекційні зразки сої методом кластерного аналізу було розподілено на 5 кластерів за комплексом досліджуваних ознак. Під час застосування кластерного аналізу (методом К-середніх) колекційні зразки сої було розділено на кластери за співвідношенням подібних ознак, у яких батьківські форми розмістилися у чотирьох кластерах із п'яти.

Найефективніше проводити оцінювання колекційного матеріалу на основі проведення кластерного аналізу з метою підбору батьківських форм для схрещування, які належать до різних кластерів.

Таблиця 1 – Метеорологічні дані за період травень–вересень (за даними Полтавської метеорологічної станції) за 2013–2015 рр. та середня багаторічна норма

Рік	Місяць	Середньомісячна температура повітря, °С	Відхилення від багаторічної (+,-), °С	Сума опадів, мм	Відхилення від багаторічної (+,-), мм	ГТК
2013	травень	20,1	+ 4,7	56,6	+ 5,6	0,90
	червень	21,5	+ 2,8	86,3	+ 26,3	1,42
	липень	21,0	+ 0,2	67,7	- 3,3	1,02
	серпень	20,9	+ 1,3	40,5	- 5,5	0,70
	вересень	12,5	- 1,9	103,9	+ 59,9	2,89
2014	травень	18,5	+ 3,4	58,4	+ 7,4	0,98
	червень	18,4	- 0,3	134,3	+ 74,3	2,42
	липень	22,0	+ 1,2	45,8	- 25,2	0,67
	серпень	22,5	+ 2,9	30,4	- 15,6	0,55
	вересень	15,4	+ 1,0	71,3	+ 27,3	2,10
2015	травень	16,3	+ 0,9	65,3	+ 14,3	1,33
	червень	20,5	+ 1,8	120,0	+ 60,0	1,98
	липень	21,1	+ 0,3	39,9	31,1	0,67
	серпень	21,7	+ 2,1	8,5	- 37,5	0,13
	вересень	18,7	+ 4,3	4,4	- 39,6	0,10

У межах колекції виділено п'ять типів колекційних зразків з різним проявом кількісних ознак (рис. 1).

Здебільшого належність колекційного зразка до кластера обумовлювалася ознаками «кількість гілок на рослині», «кількість продуктивних вузлів на рослині» і «кількість бобів на рослині».

Зразки першого кластера відзначаються середніми і близькими до середніх у колекції значеннями всіх досліджуваних ознак. Зразки другого кластера відзначаються високими значеннями кількості гілок на рослині, кількості продуктивних вузлів на рослині та кількості бобів на рослині. Третій кластер складався із зразків з високими значеннями кількості насінин у бобі. Однак за іншими ознаками зразки цієї групи мають низькі значення. Зразки четвертого кластера більш пізньостиглі, ніж

перших трьох. Вони виділяються за більшою висотою рослини і висотою прикріплення нижнього бобу. П'ятий кластер зразків відзначається найтривалішим вегетаційним періодом та високими значеннями більшості ознак: висотою рослин, висотою прикріплення нижнього бобу, товщиною стебла в нижній частині, кількістю гілок на рослині, кількістю продуктивних вузлів на рослині, кількістю бобів на рослині, кількістю насіння з рослини, кількістю насінин в бобі, масою насіння з рослини, урожайністю ( $\text{г}/\text{м}^2$ ), крім кількості насінин в бобі, стійкості до вилягання і стійкості до розтріскування бобів. Найбільш урожайні зразки потрапили до п'ятого типу, найменш урожайні – до третього.

Найбільше – 33 зразки – розмістилися в першому кластері: Аннушка, Gaillard, Березина, Єлена, АС Bravor, Краса Поділля, Аме-

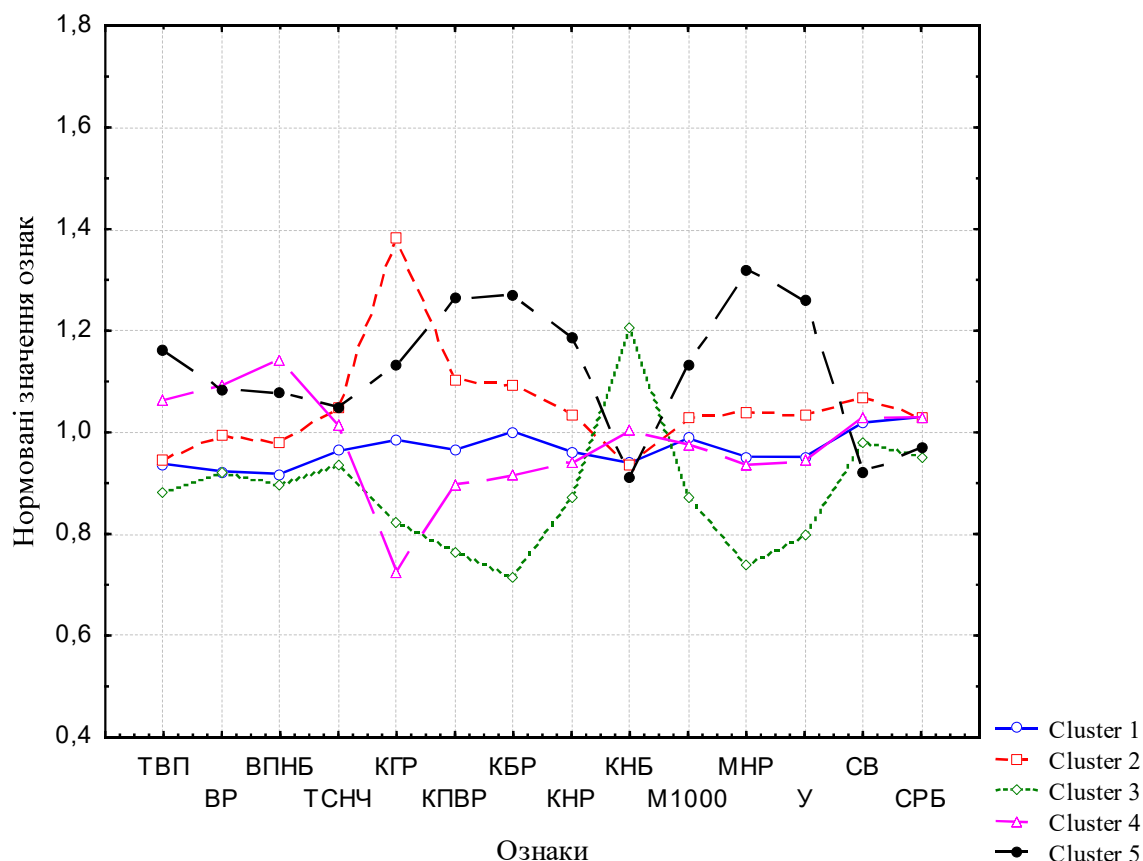


Рис. 1. Розподіл колекційних зразків сої на кластери за методом К-середніх, 2013–2015 рр.

**Примітка:** ТВП – тривалість періоду вегетації, ВР – висота рослини, ВПНБ – висота прикріплення нижнього бобу, ТСНЧ – товщина стебла в нижній частині, КГР – кількість гілок на рослині, КПВР – кількість продуктивних вузлів на рослині, КБР – кількість бобів на рослині, КНР – кількість насіння з рослини, КНБ – кількість насінин в бобі, М 1000 – маса 1000 насінин, МНР – маса насіння з рослини, У – урожайність ( $\text{г}/\text{м}^2$ ), СВ – стійкість до вилягання, СРБ – стійкість до розтріскування бобів.

тист, Медея, Хуторяночка, Антрацит, Ятрань, Роксолана, Особлива, Хорол, Glacier, Ювілейна, Анжеліка, Мальвіна, Подяка, Лара, Прикарпатська 96, Подільська 416, Омега вінницька, Срібна Рута, Стратегія, Аркадія одеська, Дені, Офелія, Княжна, Святкова, Артеміда, Феміда, Чернівецька 9. До другого кластера увійшли 23 зразки сої: ОАС Vision, LF-8, СН 32-15, Armour, Селекта 101, Лика, Атланта, Знахідка, Фаєтон, Merlin, Устя, Байка, Оріана, Л-101, Вільшанка, Скеля, Мрія, Юг-40, Поема, Фортуна, Чернівецька 8, Смуглянка, Анатоліївка.

У третьому кластері зосереджено 30 зразків: Анастасія, Легенда, Лада, Білявка, Танаїс, Білосніжка, Діона, Золотиста, Галі, Кобза, Юг-30, Злата, Сузір'я, Karikachi, Optimus, Toshi day, Nattawa, Либідь, Ксеня, Естафета, Київська-27, Ствига, Norpro, Dunajka, Харківська 80, Спринт, Говерла, Горлиця, Ніна, Sacura. Четвертий кластер налічує 27 зразків: Васильківська, Київська-98, Мавка, Сильвія, ВНИИОЗ-76, Смолянка, Валюта, Фея, MN 0901, Лариса, Деймос, Вінничанка, HeiNong 44, Корсак, Галина, Селекта 201, Монада, Донька, Kent, Марината, Ірина, Подільська 1, Антошка, L 94-1110, Таврія, Іна, Селекта 302. П'ятий

кластер об'єднав 32 зразки: Адамос, Алмаз, Хвиля, КиВін, Альба, Маша, Агат, Подолян-ка, Фарватер, Срібна, Чара, Славія, Ельдорадо, Іванка, KG-70, Blackjack 21, Hejiao 87-94-3, Sui 76-5191, Георгіна, Harosoy e1E2e3, L 71-920, Wayne L 72-1401, L 63-3117, L 65-540, Connor, KG-30, HeiNong 10, Вілана, Даная, Алмати, Дельта, Седмиця.

Групуванням господарських ознак колекційних зразків за абсолютними значеннями сої встановлено чіткий їх розподіл на два віддалених кластери, один з яких поєднував тривалість періоду вегетації, кількість насіння з рослини, висоту рослини, масу 1000 насінин, урожайність г/м<sup>2</sup> (рис. 2).

Другий кластер складався з блоку підкластерів, пов'язаних з елементами продуктивності (маса насіння з рослини, кількість бобів на рослині, кількість продуктивних вузлів на рослині) та морфо-біологічними ознаками (товщина стебла в нижній частині, висота прикріплення нижнього бобу).

В ультраскоростиглій та скоростиглій групах стиглості більшість колекційних зразків сої розподілилася в межах з першого до третього кластерів. У середньостиглій та пізньостиглій – з четвертого до п'ятого (табл. 2).

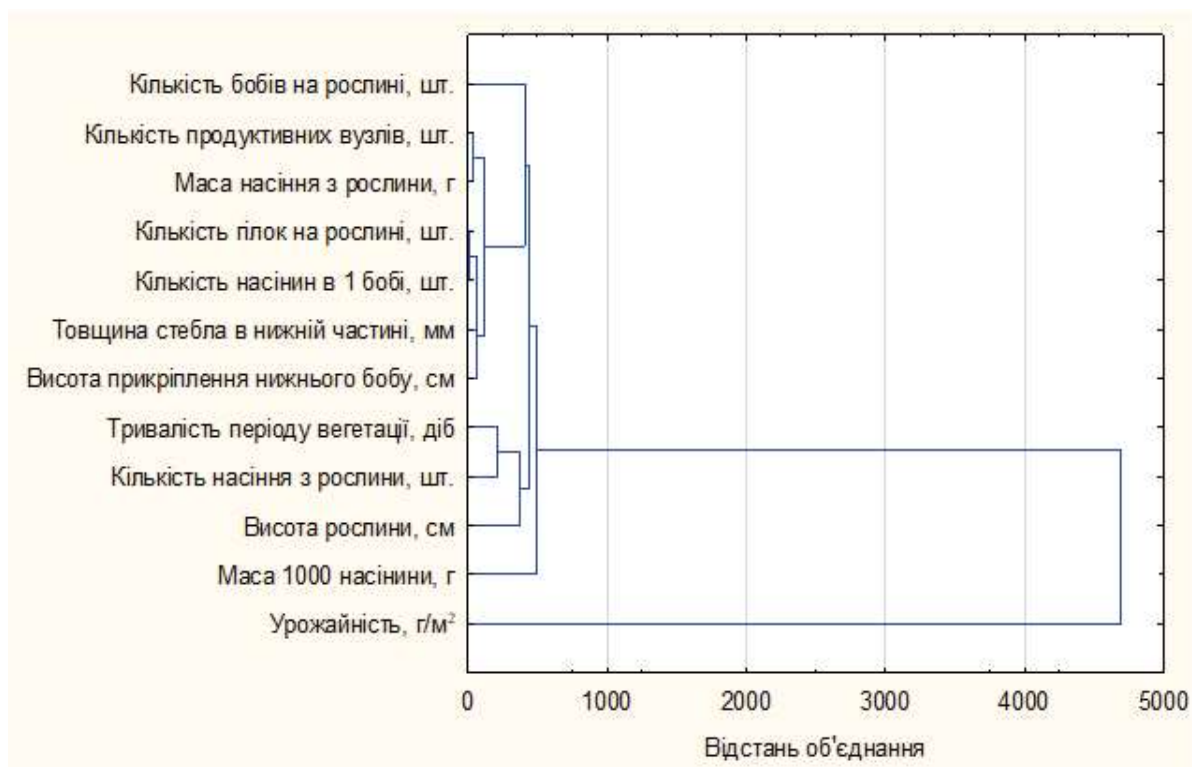


Рис. 2. Угрупування господарських ознак сої за абсолютними значеннями, 2013–2015 рр.

Таблиця 2 – Розподіл колекційних зразків сої за типами залежності між мінливістю вегетаційного періоду та належністю до кластера, 2013–2015 рр.

Група стиглості	Кількість зразків	Кластер	Кількість зразків	
			шт.	%
Ультраскоростиглі	16	1	2	12,5
		2	2	12,5
		3	12	75,0
		4	-	-
		5	-	-
Скоростиглі	80	1	29	36,2
		2	18	22,5
		3	17	21,3
		4	12	15,0
		5	4	5,0
Середньостиглі	27	1	2	7,4
		2	3	11,1
		3	1	3,7
		4	11	40,7
		5	10	37,1
Пізньостиглі	22	1	-	-
		2	-	-
		3	-	-
		4	4	18,2
		5	18	81,8

Формування п'яти кластерів у колекційних зразках сої свідчить про їх віддаленість один від одного. Дані досліджень та аналіз інформації, висвітленої у наукових літературних джерелах щодо кластеризації цінних господарських ознак сої доводять, що для добору батьківських форм для проведення гібридизації краще залучати зразки різних кластерів. Вивчення мінливості цінних господарських ознак сої забезпечує результативне ведення селекції. Значною перевагою кластерного аналізу є то, що він дає змогу ідентифікувати колекційні зразки збалансовано за комплексом цінних господарських ознак, а не виділити генотипи за окремими кількісними ознаками.

**Висновки.** У результаті кластерного аналізу колекційні зразки сої розподілено на п'ять кластерів за господарсько цінними ознаками: тривалістю вегетаційного періоду, висотою рослин, висотою прикріплення нижнього бобу, товщиною стебла в нижній частині, кількістю гілок на рослині, кількістю продуктивних вузлів на рослині, кількістю бобів на рослині, кількістю насіння з рослини, кількістю насіння в бобі, масою насіння з рослини, урожайністю ( $\text{г}/\text{м}^2$ ), стійкістю до вилягання, стійкістю до розтріскування бобів. Зразки в межах одного кластера подібні за співвідношенням ознак. В окремий кластер об'єднано зразки з аналогічним набором ознак всередині кластера і достовірною відмінністю з іншими. Батьківські

форми, виділені за комплексом цінних господарських ознак, для проведення гібридизації розмістилися у чотирьох кластерах із п'яти: Злата, Адамос, ОАС Vision, Gaillard, СН 32-15, Алмаз, Устя, Славія, LF-8, КиВін, Лада, Мрія, Хвиля, Ельдорадо, Поема, Фарватер. Отже, на основі детального вивчення результатів кластерного аналізу кількісних ознак у сої можливо більш ефективно проводити селекційну роботу зі створення нового вихідного матеріалу.

#### СПИСОК ЛІТЕРАТУРИ

1. Абуґалиєва А.И., Дидоренко С.В. Генетическое разнообразие сортов сои различных групп спелости по признакам продуктивности и качества. Вавиловский журнал генетики и селекции. 2016. Т. 20. № 3. С. 303–310. DOI: <https://doi.org/10.18699/VJ16.168>.
2. Белявская Л.Г., Рыбальченко А.М. Скрининг коллекции сои по скороспелости и продуктивности в условиях Левобережной Лесостепи Украины. Зернобобовые и крупяные культуры. 2019. № 1 (29). С. 63–69. DOI: <https://doi.org/10.24411/2309-348X-2019-11074>.
3. Кластерний аналіз в селекції гречки / Вільчинська Л.А. та ін. Вісник Українського товариства генетиків і селекціонерів. 2017. Том 15. № 2. С. 145–149.
4. Губа І.І., Стариченко В.М. Кластеризація колекційних зразків жита озимого за кількістю квіток та іншими господарськоцінними ознаками. Наукові праці Інституту біоенергетичних культур і цукрових буряків. 2018. Вип. 26. С. 113–119.
5. Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований). Москва: Агропромиздат, 1985. 351 с.

6. Отбор перспективных образцов для селекции фасоли с использованием кластерного анализа в условиях южной лесостепи Западной Сибири / Казыдуб Н.Г. и др. Вестник Омского государственного аграрного университета. 2014. № 4 (16). С. 8–14.

7. Широкий уніфікований класифікатор роду *Glycine max.* (L). Merr. / Кобизева Л.Н. та ін. Ін-т рослинництва ім. В.Я. Юр'єва. Х., 2004. 37 с.

8. Методические указания по изучению коллекции зерновых бобовых культур / Корсаков Н.И. и др. Л.: ВИР, 1975. 59 с.

9. Кластерний аналіз у селекції зернобобових культур / Коханюк Н. В. та ін. Корми і кормовиробництво. 2019. Вип. 87. С. 9–19. DOI: <https://doi.org/10.31073/kormovyrobnytstvo201987-02>.

10. Литун П.П. Идентификация генотипов в селекционных популяциях. Селекция и семеноводство. 1980. № 46. С. 27–34.

11. Системний аналіз в селекції польових культур: навч. посіб. / Літун П.П. та ін. Харків, 2009. 354 с.

12. Майданюк В.О., Холодняк О.Г. Метод підбору батьківських пар при створенні нових адаптивних сортів кабачка для відкритого ґрунту півдня України. Зрошуване землеробство. 2014. Вип. 61. С. 137–139.

13. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика. 1988. 176 с.

14. Маракаева Т.В. Кластерный анализ в селекции чечевицы. Успехи современного естествознания. 2018. № 12. Ч. 1. С. 75–80.

15. Мартынов С.П. Кластерный анализ саратовских сортов яровой пшеницы по коэффициентам родства. Цитология и генетика. Т. 23. № 4. 1989. С. 37–43.

16. Мельник А.В. Використання кластерного аналізу за підбору сортів і гібридів ріпаку ярого для вирощування в Лівобережному Степу України. Вісник Полтавської державної аграрної академії. 2013. № 4. С. 6–11.

17. Оцінка сортів гороху за комплексом господарсько цінних ознак / Присяжнюк О.І. та ін. Цукрові буряки. 2013. № 5. С. 16–17.

18. Стариченко В.М. Використання кластерного аналізу для оцінки F2 популяції сої. Збірник наукових праць ННЦ «Інститут землеробства НААН». 2012. Вип. 3–4. С. 124–134.

19. Систематизація сортів та селекційних ліній пшениці озимої за кількісними ознаками в умовах контрольованого середовища з використанням кластерного аналізу / Тищенко В.М. та ін. Вісник Полтавської державної аграрної академії. 2018. № 3. С. 56–65. DOI: <https://doi.org/10.31210/visnyk2018.03.09>

20. Тищенко В.Н., Чекалин Н.М., Зюков М.Е. Использование кластерного анализа для идентификации и отбора высокопродуктивных генотипов озимой пшеницы на ранних этапах селекции. Факторы экспериментальной эволюции организмов. 2004. Т. 2. С. 270–278.

21. Использование кластерного анализа как метода индивидуального отбора у проса (*Panicum miliaceum* L.) / Чекалин Н.М. та ін. Вісник Полтавської державної аграрної академії. 2009. № 2. С. 10–17.

## REFERENCES

1. Abugalieva, A.I., Didorenko S.V. (2016). Geneticheskoe raznoobrazie sortov soi razlichnyh grupp spelosti po priznakam produktivnosti i kachestva [Genetic diversity of soybean cultivars belonging to different ripeness groups with regard to performance and quality]. Vavilovskij zhurnal genetiki i selekcii [Vavilov journal of Genetics and Breeding]. Vol. 20, no. 3, pp. 303–310. Available at: <https://doi.org/10.18699/VJ16.168>.

2. Belyavskaya, L.G., Rybalchenko, A.M. (2019). Skrining kolekcii soi po skorospelosti i produktivnosti v usloviyah Levoberezhnoj Lesostepi Ukrainy [Screening soya collection as to early maturing and productivity in the conditions of the left bank Forrest-steppe of Ukraine]. Zernobobovye i krupyanye kultury [Legumes and groat crops], no. 1 (29), pp. 63–69. Available at: <https://doi.org/10.24411/2309-348X-2019-11074>.

3. Vilchinska, L.A., Gorodinska, O.P., Diyanchuk, M.V., Kaminna O.O. (2017). Klasternij analiz v selekciji grechki [Cluster analysis in the selection of buckwheat]. Visnyk Ukrayinskogo tovarystva genetykiv i selekcioneriv [The Bulletin of Ukrainian Society of Geneticists and Breeders]. Vol. 15, no. 2, pp. 145–149.

4. Guba, I.I., Starichenko, V.M. (2018). Klasterizaciya kolekcijnih zrazkiv zhita ozimogo za kilkisty kvitok ta inshimi gospodarskocinnimi oznakami [Clustering of winter rye collections on the number of flowers and other economically valuable traits]. Naukovi praci Institutu bioenergetichnih kultur i cukrovih buryakiv [Scientific Papers of the Institute of Bioenergy Crops and Sugar Beet], no. 26, pp. 113–119.

5. Dospekhov, B.A. (1985). Metodika polevogo opyta (s osnovami statisticheskoy obrabotki rezultatov issledovaniy) [Methods of field experiment (with fundamentals of statistical processing of research results)]. Moscow, Agropromizdat, 351 p.

6. Kazydub, N.G., Marakaeva, T.V., Korobejnikova, M.M., Epanchinev M.V. (2014). Otbor perspektivnyh obrazcov dlya selekcii fasoli s ispolzovaniem klasternogo analiza v usloviyah yuzhnoj lesostepi Zapadnoj Sibiri [Selection of perspective standards for selection of kidney bean with the use of cluster analysis in the conditions of south forest-steppe of Western Siberia]. Vestnik Omskogo gosudarstvennogo agrarnogo universiteta [Bulletin of Omsk State Agrarian University], no. 4 (16), pp. 8–14.

7. Kobizeva, L.N., Ryabchun, V.K., Bezugla, O.M. (2004). Shirokiy unifikovaniy klasifikator rodu Glycine max. (L). Merr [Great unified classifier kind of Glycine max. (L). Merr]. Kharkiv, Institute of Plant Breeding names of V.Y. Yur'yev, 37 p.

8. Korsakov, N.I., Adamova, O.A., Budakova, V.I. (1975). Metodicheskiye ukazaniya po izucheniyu kolekcii zernovykh bobovykh kultur [Methodical instructions to studies of leguminous and cereal crops collections]. Leningrad, VIR, 59 p.

9. Kohanyuk, N.V., Temchenko, I.V., Shtuc, T.M., Lehman, A.A., Barvinchenko, S.V. (2019). Klasternij analiz u selekciji zernobobovih kultur [Cluster analysis in the breeding of leguminous crops]. Kormy i kormovyrobnytstvo [Feeds and Feed Production], no. 87, pp. 9–19. Available at: <https://doi.org/10.31073/kormovyrobnytstvo201987-02>.

10. Litun, P.P. (1980). Identifikaciya genotipov v selekcionnyh populyacijah [Identification of genotypes in breeding populations]. *Selekciya i semenovodstvo* [Breeding and Seed Production], no. 46, pp. 27–34.

11. Litun, P.P., Kirichenko, V.V., Petrenkova, V.P., Kolomacka, V.P. (2009). Sistemnij analiz v selekciyi polovih kultur: navch. posib. [System analysis in field crop selection]. Kharkiv, 354 p.

12. Majdanyuk, V.O., Holodnyak, O.G. (2014). Metod pidboru batkivskih par pri stvorenni novih adaptivnih sortiv kabachka dlya vidkritogo gruntu pivdnya Ukrainy [Method of selection of parental pairs when creating new adaptive varieties zucchini for open ground of the Southern Ukraine]. *Zroshuvane zemlerobstvo* [Irrigated agriculture], no. 61, pp. 137–139.

13. Mandel, I.D. (1988). Klasternyj analiz [Cluster analysis]. Moscow, Finance and statistics, 176 p.

14. Marakaeva, T.V. (2018). Klasternyj analiz v selekci chechevicy [Cluster analysis in lentil breeding]. *Uspehi sovremenogo estestvoznaniya* [Advances in current in natural sciences], no. 12, Part 1, pp. 75–80.

15. Martynov, S.P. (1989). Klasternyj analiz saratovski h sortov yarovoj pshenicy po koefficientam rodstva [Cluster analysis of Saratov spring wheat varieties by kinship coefficients]. *Citologiya i genetika* [Cytology and genetics]. Vol. 23, no. 4, pp. 37–43.

16. Melnik, A.V. (2013). Viktoristannya klasternogo analizu za pidboru sortiv i gibridiv ripaku yarogo dlya viroshuvannya v livoberezhnomu stepu Ukrainy [Use of cluster analysis in the selection of varieties and hybrids of spring rape for cultivation in the Left-Bank Forest-Steppe Ukraine]. *Visnyk Poltavskoi derzhavnoi agrarnoi akademii* [News of Poltava State Agrarian Academy], no. 4, pp. 6–11.

17. Prisyazhnyuk, O.I., Kalyuzhna, E.A., Ukrainec, V.V., Shevchenko, O.P. (2013). Ocinka sortiv gorohu za kompleksom gospodarsko-cinnih oznak [Evaluation of pea varieties on a set of economically valuable features]. *Cukrovi buryaki* [Sugar Beet], no. 5, pp. 16–17.

18. Starichenko, V.M. (2012). Viktoristannya klasternogo analizu dlya ocinki F2 populyaciyi soyi [Using the cluster analysis for an estimation F2 population of soybean]. *Zbirnyk naukovykh prac NNC «Instytut zemlerobstva NAAN* [Collection of scientific papers of NSC “Institute of agriculture of NAAS”], no. 3–4, pp. 124–134.

19. Tishenko, V.M., Gusenkova, O.V., Dubenec, M.V., Kolisnik, A.V. (2018). Sistemizaciya sortiv ta selekciynih liniy pshenicy ozimoyi za kilkisnimi oznakami v umovah kontrolovanogo seredovisha z vikoristannyam klasternogo analizu [Systemization of sort and selection lines of wheat lines by numbership in conditions of controlled environment with using cluster analysis]. *Visnyk Poltavskoi derzhavnoi agrarnoi akademii* [Bulletin of Poltava State Agrarian Academy], no. 3, pp. 56–65. Available at: <https://doi.org/10.31210/visnyk2018.03.09>.

20. Tishenko, V.N., Chekalin, N.M., Zyukov, M.E. (2004). Ispolzovanie klasternogo analizu dlya identifikacii i otbora vysokoproduktivnyh genotipov ozimoi pshenicy na rannih etapah selekci [The use of cluster analysis for the identification and selection of highly productive winter wheat genotypes at the early stages of breeding]. *Faktery eksperimentalnoi evolyucii organizmiv* [Factors in experimental evolution of organisms]. Vol. 2, pp. 270–278.

21. Chekalin, N.M., Tishenko, V.N., Panchenko, P.M., Sidorenko, V.S. (2009). Ispolzovanie klasternogo analizu kak metoda individualnogo otbora u prosa (*Panicum miliaceum* L.) [Use cluster analysis as method of millet individual selection *Panicum miliaceum* L.]. *Visnyk Poltavskoi derzhavnoi agrarnoi akademii* [News of Poltava State Agrarian Academy], no. 2, pp. 10–17.

### **Классификация сортов сои по хозяйственным признакам с помощью кластерного анализа**

**Белявская Л.Г., Рыбальченко А.М.**

Целью исследования было усовершенствовать метод оценивания коллекционного материала сои с целью подбора пар для скрещивания с помощью кластерного анализа, провести кластеризацию образцов по продолжительности вегетационного периода, высоте растений, высоте прикрепления нижнего боба, толщине стебля в нижней части, количеству ветвей на растении, количеству продуктивных узлов на растении, количеству бобов на растении, количеству семян с растения, количеству семян в бобе, массе семян с растения, урожайности ( $г/м^2$ ), устойчивости к полеганию, стойкости к растрескиванию бобов. На основе проведенного кластерного анализа сто сорока пяти коллекционных образцов сои по определенным признакам установили, что коллекционный материал распределился в пределах пяти кластеров. Были выделены генотипы сои как носители высокого уровня количественных признаков, особенно признаков потенциала урожайности, которые можно использовать как перспективный селекционный материал.

В отдельный кластер объединены образцы с аналогичным набором признаков внутри кластера и достоверным отличием с другими.

Образцы первого кластера отмечаются средними и близкими к средним по коллекции значениями всех исследуемых признаков. Образцы второго кластера обладают высокими значениями количества ветвей на растении, количества продуктивных узлов на растении и количества бобов на растении. Третий кластер состоял из образцов с высокими значениями количества семян в бобе. Однако по другим признакам образцы этой группы имеют низкие значения. Образцы четвертого кластера более позднеспелые, чем первых трех. Они выделяются большей высотой растения и высотой прикрепления нижнего боба. Пятый кластер образцов отмечается самым продолжительным вегетационным периодом и высокими значениями большинства признаков, кроме количества семян в бобе, устойчивости к полеганию и устойчивости к растрескиванию бобов. Родительские формы разместились в четырех кластерах из пяти.

Проведенный кластерный анализ помог осуществить подбор родительских пар для проведения гибридизации и создания нового исходного материала сои. Преимущество кластерного анализа заключается в том, что он позволяет идентифицировать коллекционные образцы сбалансировано по комплексу ценных хозяйственных признаков, а не выделить генотипы по отдельным количественным признакам.

**Ключевые слова:** соя, селекция, коллекция, образец, кластер, кластерный анализ, исходный материал.



### Cluster analysis in soybean varieties classification by economic characteristics

Bilyavska L., Rybalchenko A.

The aim of the study was to improve the method of soybean collection material assessment with the purpose of choosing pairs for hybridization using cluster analysis, to cluster samples according to the duration of growing period, plant height, the height of lower pod attachment, stem thickness in the lower part, the number of branches and productive joints on a plant, the number of pods and seeds per plant, the number of seeds in a pod, seed weight per plant, yield (g/m<sup>2</sup>), and the resistance to lodging and bean cracking. Based on the conducted cluster analysis of one hundred and forty-five soybean collection samples concerning the determined signs, it has been established that the collection material was distributed within five clusters. Soybean genotypes have been singled out as carriers of a high level of quantitative characteristics, especially the signs of yield potential which can be used as a promising selection material.

The samples having analogous set of signs in the cluster and reliable distinction from others have been united in a separate cluster.

The samples of the first cluster in the collection are characterized by average and close to average values of all the studied characteristics. The samples of the second cluster are characterized by high values as to the number of branches and productive joints on a plant and also the number of beans per plant. The third cluster consisted of samples having high values of the number of seeds in a bean. However, as to other signs, the samples of this group have low values. The samples of the fourth cluster are later-maturing, than those in the first three groups. They are characterized by larger plant height and the height of lower bean attachment. The fifth sample cluster is distinguished by the longest growing period and high values of the majority of signs except the number of seeds in a bean and also resistance to lodging and bean splitting. The parent forms were in four clusters out of five ones.

The cluster analysis helped to conduct the choosing of parent pairs for hybridization and creation of soybean new initial material. The advantage of cluster analysis consists in being able to identify collection samples in balance according to the complex of valuable economic characteristics, rather than single out genotypes according to separate quantitative signs.

**Key words:** soybean, selection, collection, sample, cluster, cluster analysis, initial material.



Copyright: Білявська Л.Г., Рибальченко А.М. © This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



Білявська Л.Г.  
Рибальченко А.М.

ID: <https://orcid.org/0000-0003-3856-7718>  
ID: <https://orcid.org/0000-0002-2308-7853>