

Застосування кластерного аналізу для групування сортів *Brassica oleracea* var. *italica* за тесту на відмінність

О. Й. Дидів¹, В. В. Хареба², О. В. Хареба², Н. В. Лещук^{3*}, Н. С. Орленко³, О. Б. Орленко³

¹Львівський національний університет природокористування, вул. Володимира Великого, 1, м. Дубляни, Львівський р-н, Львівська обл., 80381, Україна

²Національна академія аграрних наук України, вул. Омеляновича-Павленка, 9, м. Київ, 01010, Україна

³Український інститут експертизи сортів рослин, вул. Генерала Родимцева, 15, м. Київ, 03041, Україна,
*e-mail: nadiya1511@ukr.net

Мета. Використати кластерний аналіз морфологічних ознак для спрощення ідентифікації сортів *Brassica oleracea* var. *italica* та сформувати групи схожих сортів для тесту на відмінність. **Методи.** У процесі роботи послуговувалися аналітичним, математичним і статистичним методами. Як вхідну інформацію для статистичного опрацювання отриманих результатів застосовували відомості про результати експертизи на відмінність, однорідність і стабільність (ВОС) із бази даних автоматизованої інформаційної системи Українського інституту експертизи сортів рослин. Моделювання кластерів здійснювали за допомогою статистичного пакета IBM SPSS Statistics «Statistical Package for the Social Sciences». **Результати.** За 32 ознаками для тесту на відмінність, однорідність і стабільність проведено морфологічний опис сортів капусти броколі. Морфологічні кодові формули останніх, складені з відповідних кодів прояву ідентифікаційних ознак вегетативних і генеративних органів рослин, слугували джерелом вихідних даних. Серед 41 сорту, описаного за 32 морфологічними характеристиками, вдалося виокремити лише дві групи подібних за ідентифікаційними ознаками сортів. Як параметри моделі застосовували два типи змінних: цільова ознака «голівка: антоціанове забарвлення», фокусна – «голівка: забарвлення». Повний перелік характеристик був таким: «рослина: за висотою (за збиральної стиглості)», «листок: положення (на початку формування голівки)», «листова пластинка: хвилястість краю», «листова пластинка: пухирчастість», «черешок: за довжиною», «голівка: забарвлення», «голівка: антоціанове забарвлення», «голівка: за щільністю», «квітка: забарвлення», «квітка: інтенсивність жовтого забарвлення», «чоловіча стерильність». Способом комп'ютерного моделювання було сформовано кластери з 17 подібних сортів капусти броколі та 9 контрольних об'єктів (сортів), ідентифікація яких передбачала 11 морфологічних ознак. **Висновки.** Для пошуку відмітних ознак у процесі тесту на відмінність сорти капусти броколі було згруповано в кластери за такими морфологічними характеристиками, як положення листка на початку формування голівки; хвилястість краю листової пластинки; пухирчастість листової пластинки; довжина черешка; забарвлення голівки; наявність антоціанового та інтенсивність жовтого забарвлення.

Ключові слова: капуста броколі; статистичний аналіз; класифікація; сорт; код; відмінність; кластер; ознака; колекція.

Вступ

Сучасні інформаційні технології дають змогу накопичувати, акумулювати та зберігати досить велику кількість даних [1], що є характерним також і для Українського інституту експертизи сортів рослин (далі – УІЕСР).

Результати кваліфікаційної експертизи на відмінність, однорідність і стабільність (ВОС-тест) сортів капусти броколі як науково-технічні дані накопичуються роками та досить часто є різномірними (невпорядкованими). Вони можуть бути представлені як в числовому, так і категорійному форматі.

Olha Dydiv

<https://orcid.org/0000-0003-4155-5945>

Volodymyr Khareba

<https://orcid.org/0000-0001-9947-2689>

Olena Khareba

<https://orcid.org/0000-0002-6763-1988>

Nadiia Leshchuk

<https://orcid.org/0000-0001-6025-3702>

Natalia Orlenko

<https://orcid.org/0000-0003-0494-2065>

Oleksandr Orlenko

<https://orcid.org/0009-0001-3309-0757>

Різний набір вхідних параметрів великої кількості алгоритмів кластеризації може спричинити відмінність, а відповідно і складність оцінювання її результатів. Одним зі способів оцінювання результатів кластеризації є розташування об'єктів і центрів кластерів на координатній площині, що можливо лише якщо дані не багаторозмірні. В іншому разі необхідно використовувати методику [2].

Проблема вибору серед алгоритмів для максимально правильної кластеризації полягає в тому, що більшість з них очікують вхідним параметром зазвичай заздалегідь невідому кількість кластерів [2, 3]. Саме тому для встановлення оптимальної чисельності останніх потрібно послуговуватися деякими емпіричними правилами. Завдання кластеризації стають значно вагомішими, якщо кластери застосовують не лише для наочного представлення об'єктів, але й для розпізнавання нових.

Кожен новий сорт належить до тієї групи, приєднання до якої найкраще задовольняє критерії якості кластеризації та з об'єктами в якій він найбільше схожий за характеристикою [4].

Броколі або спаржева капуста (*Brassica oleracea* var. *italica*) – однорічна овочева рослина родини капустяних, підвид цвітної капусти. Її споживча стиглість настає за максимального розвитку суцвіття (головки), яке є продуктивним органом. Час досягання головок визначають за умови, що 50% рослин вступили в фенологічну фазу збиральної стиглості. Залежно від сорту він може бути дуже раннім (до 35 діб); раннім (35–40); середнім (41–50); пізнім (51–60); дуже пізнім (понад 60 діб) [5].

Відповідність критеріям відмінності, однорідності та стабільності, а також господарсько-цінним характеристикам для задоволення потреб споживачів, відсутність загрози довіллю та здоров'ю людини є обов'язковими вимогами для поширення на території України нових сортів капусти броколі. Їхній морфологічний опис за ознаками вегетативних і генеративних органів рослин здійснюють у відповідні фенологічні фази росту та розвитку [6].

Науково-технічна експертиза сортів рослин групи овочевих – це комплекс польових і лабораторних досліджень, результати яких формують джерелознавчу базу даних. Кінцевого інформаційного продукту для практичного використання досягають через алгоритм статистичного опрацювання. Саме тому досить актуальним є пошук сучасних статистичних методів оброблення результатів

(кодів прояву ознак) з тесту на відмінність у процесі ідентифікації сортів [3, 7].

Моніторинг іноземної та вітчизняної літератури продемонстрував, що кластерний аналіз (*англ.* Data clustering) забезпечує структурування вибірки об'єктів (ознак) на підмножини (кластери) так, щоб об'єкти, з яких вони складаються, були схожими між собою й відрізнялися від об'єктів з інших [1, 3, 8]. Отже, кластеризацію можна вважати задачею багатокритеріальної оптимізації, алгоритм і вибір параметрів якої (функція відстані, порогове значення щільності або кількість очікуваних кластерів) залежать від конкретного набору даних та мети використання результатів [8].

На думку низки дослідників, технологічну процедуру з визначення відмінності сортів, за відсутності статистичного методу, можна спростити, здійснивши автоматизоване групування найподібніших з них за кодами прояву ідентифікаційної ознаки (QN, QL, PQ). Відмінні сорти в межах групи слід використовувати не тільки для тесту на відмінність, але й для тесту на стабільність [9]. Оцінювання груп можна здійснювати, застосовуючи різні статистичні методи інтелектуального аналізу кодів і ступенів прояву характеристик нових сортів капусти броколі.

Оскільки кількість ознак, використовуваних для групування під час тесту на відмінність, обмежена («рослина: кількість стебел», «головка: забарвлення», «рослина: час збиральної стиглості», «рослина: чоловіча стерильність»), для статистично обґрунтованого розширення їхнього діапазону виникає потреба працювати з кодами прояву у відповідні фенологічні фази росту та розвитку рослин [10–12]. Спростити цей процес можна за допомогою комп'ютерного оброблення даних, використовуючи метод кластерного аналізу, що дасть змогу зменшити розмірність кодів і ступенів їхнього прояву, які підлягають аналізу, та згрупувати сорти з морфологічними ознаками для оперативного та прозорого пошуку серед них хоча б однієї відмінної [6, 8].

Мета досліджень – використати кластерний аналіз морфологічних ознак для спрощення ідентифікації сортів *Brassica oleracea* var. *italica* та сформувати групи схожих сортів для тесту на відмінність.

Матеріали та методика досліджень

Як вхідну інформацію для статистичного опрацювання отриманих результатів використовували відомості про результати експертизи на відмінність, однорідність і стабільність (ВОС) з бази даних автомати-

зованої інформаційної системи УІЕСР. Поширений на території України сорт капусти броколі має офіційно оприлюднений опис із 32 морфологічних ознак вегетативних і генеративних органів, представлений кодами та ступенем їхнього прояву в морфологічній кодовій формулі. Саме коди ознак вегетативних і генеративних органів рослин сортів капусти броколі, що перебувають у базі УІЕСР, формують масив даних для тесту на відмінність [13]. Перелік сортів за країнами походження подано у таблиці 1.

Таблиця 1

**Перелік сортів капусти броколі
за країнами походження**

Назва сорту	Країна походження
'Libra'	Болгарія
'Klieopatra'	Канада
'TM-04/FL', 'Limba', 'Apolena'	Чехія
'Naxos', 'Rumba', 'Kaptan', 'Tambora', 'Green Magic', 'Marathon', 'Parthenon', 'Koros', 'Eos', 'Cigno', 'Babilon', 'Abilon', 'Stirling', 'Cusco'	Франція
'Kostal', 'Agassi F1'	Німеччина
'Moysan', 'Samoa'	Італія
'Empieror'	Японія
'Batavia F1', 'Bielinda', 'Korvat F', 'Korvet', 'Milady F1', 'Skiff', 'Monaco F1', 'Ironman F1', 'Quinta F1', 'Beaumont F1', 'Corato', 'Monrello', 'Belstar F1', 'Bay Meadows Cemes F1', 'Reggi', 'Besty', 'Orantes', 'SV1002BL', 'Steel', 'Edberg', 'Batory', 'Stromboli', 'Larsson', 'Titanium', 'Tesmari', 'Vicario'	Голландія
'Wiarus', 'Cezar'	Польща
'Baro Star'	Південна Корея
'Atlantik', 'Monrello', 'McLaren', 'Gongga'	Швейцарія
'Lednitska', 'Vitamina', 'Cezar', 'TM-04/FL', 'Trubadur', 'Muline', 'Cazzy F1'	Україна
'Yahuar', 'SSTs 5186 R', 'Spartan Erli'	США

Предметом проведених досліджень був вихідний масив даних кодів прояву морфологічних ознак сортів капусти броколі вітчизняної та іноземної селекції.

Як цільову змінну обрано ознаку «головка: антоціанове забарвлення», фокусну – «головка: забарвлення». Повний перелік характеристик був таким: «рослина: за висотою (за збиральної стиглості)», «листок: положення (на початку формування головки)», «листова пластинка: хвилястість краю», «листова пластинка: пухирчастість», «черешок: за довжиною», «головка: забарвлення», «головка: антоціанове забарвлення», «головка: за щільністю», «квітка: забарвлення», «квітка: інтенсивність жовтого забарвлення», «чоловіча стерильність» [11, 14].

У процесі комп'ютерного опрацювання даних (кодів морфологічних ознак) використо-

ували дві шкали. А саме: номінальну – для групування подібних сортів за такими морфологічними характеристиками, як «листок: положення (на початку формування головки)», «квітка: забарвлення», «листова пластинка: хвилястість краю», «листова пластинка: пухирчастість», «головка: забарвлення»; порядкову – для ознак «рослина: за висотою (за збиральної стиглості)», «черешок: за довжиною», «головка: за щільністю», «квітка: інтенсивність жовтого забарвлення» [6].

Результати експертизи опрацьовували статистичним методом кластерного аналізу через пошук алгоритму подібності ознак для формування груп-кластерів. Моделювання останніх, а також розрахунки виконували у середовищі статистичного пакета IBM SPSS Statistics «Statistical Package for the Social Sciences», застосовуючи параметри класифікаційного методу Nearest Neighbor Analysis для генетично маркованих ознак [4, 7, 8].

Вищевказаним методом передбачено, що початково є деяка кількість об'єктів з точною класифікацією подібних сортів капусти броколі, тому необхідно виробити правило, яке дасть змогу зараховувати новий сорт до одного з можливих класів (набору подібних за морфологічними ознаками сортів рослин).

Алгоритм кластеризації складається з добору коефіцієнта (k) та атрибутів (n), що визначають міру подібності, для нових сортів капусти броколі [6]. Де k – це кількість записів, які будуть вважатися близькими за використання таких правил:

$(x,y) \geq 0$, $d(x,y) = 0$ тоді й лише тоді, коли $x = y$;

$$d(x,y) = d(y,x);$$

$d(x,z) \leq d(x,y) + d(y,z)$ за умови, що точки x , y , z не лежать на одній прямій.

x , y , z – це вектори ознак об'єктів, які порівнюють.

Значення атрибутів упорядковують, використовуючи як міру схожості спостережень відстань Евкліда, і розраховують за формулою:

$$D_E = \sqrt{\sum_1^n (x_i - y_i)^2},$$

де n – кількість атрибутів.

Відповідно до міжнародних вимог «Методики проведення експертизи сортів капусти на відмінність, однорідність і стабільність (Guidelines for the conduct of tests for distinctness, uniformity and stability UPOV)», тестування на ВОС проводили в польових умовах протягом двох послідовних вегетаційних сезонів.

Результати морфологічного опису сортів за роки досліджень автоматично акумулюються

в базі даних з інстальованою таблицею морфологічних ознак і підлягають розподілу за ознаками, роками та пунктом досліджень.

Якщо відстань $d(in, ip)$ менша за деяке значення σ , то можна зробити висновок, що елементи розташовані близько один до одного та перебувають в єдиному кластері; в іншому разі елементи розміщені в різних кластерах. Розбиття елементів на групи повинно відповідати деякому критерію – певному функціоналу, який ще називають цільовою функцією. Тому для полегшення оцінювання відмінності найкраще обирати цільову змінну «головка: забарвлення» як альтернативу заданих ознак, рекомендованих методикою

для групування сортів-кандидатів із подібними загальновідомими сортами. Для класифікації використовують ознаки, окремо або в комбінаціях з іншими, які не варіюють або дуже слабо варіюють у межах сорту [6].

Результати досліджень

Загальновідомі сорти капусти броколі вітчизняної та іноземної селекції, заявки на які подавали починаючи з 2002 року і дотепер, є складниками національних сортових рослинних ресурсів. Нині інформаційна система УІЕСР налічує 41 їхню морфологічну формулу та щорічно поповнюється ще трьома-чотирма сортами (рис. 1).

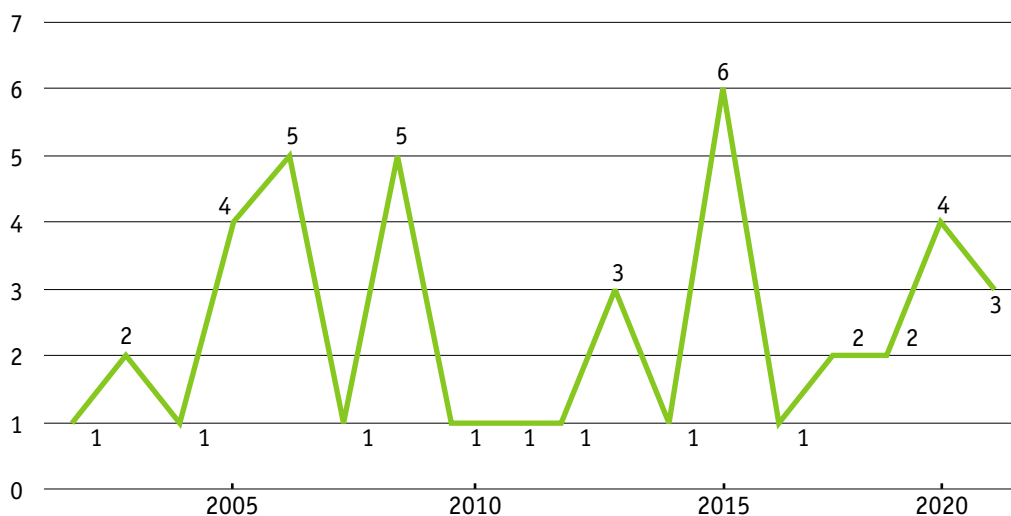


Рис. 1. Динаміка формування національних сортових ресурсів капусти броколі за період 2002–2022 рр. [13]

Повний перелік груп сортів капусти броколі за результатами моделювання низки (11) морфологічних ознак, коди яких утворюють кодові формули, наведено в таблиці 2.

Із сукупності досліджуваних сортів було виокремлено чотири групи. До першої увійшли пари 'Batory' й 'Monrello' та 'Titanium' і 'Sv1002bl'. Відстань між ними нульова за алгоритмом найближчих сусідів, що свідчить про збіг 11 з 32 морфологічних ознак. А саме: «рослина: за висотою (за збиральної стиглості)», «листок: положення (на початку формування головки)», «листочка пластинка: хвилястість краю», «листочка пластинка: пухирчастість», «черешок: за довжиною», «головка: забарвлення», «головка: антоціанове забарвлення», «головка: за щільністю», «квітка: забарвлення», «квітка: інтенсивність жовтого забарвлення», «чоловіча стерильність». Утім за деякими ознаками, що не потрапили до вибірки для класифікації, вищевказані сорти відрізняються. Так, 'Batory' має блакитно-зелене забарвлення листкової пластинки, середній за шириною та довжиною (включно

з черешком) листок та велику кількість його часток. Листкова пластинка 'Monrello' сіро-зелена, листок широкий і довгий, кількість його часток середня.

Пара 'Sv1002bl' і 'Titanium' різняться за часом збиральної стиглості – пізній і середній відповідно. Головка за збиральної стиглості у сорту 'Sv1002bl' зелена з помірними інтенсивністю забарвлення та структурою тканини; у 'Titanium' – блакитно-зелена зі слабкими інтенсивністю забарвлення та структурою тканини.

Унаслідок аналізу масиву даних результатів морфологічного опису виявлено, що подібні за ідентифікаційними характеристиками сорти, базуючись лише на інформації про них та їхні взаємозв'язки, можна сформувати у групи (кластери) методом кластеризації. Найкращою кластеризація є тоді, коли об'єкти з однієї групи максимально схожі між собою та сильно відрізняються від об'єктів з інших.

У процесі досліджень ідентифіковано 41 сорт капусти броколі за 11 морфологічними

Таблиця 2

Групування подібних сортів капусти броколі за відстанню до найближчих сусідів
(k – Nearest Neighbor and Distance)

Focal Record	Nearest Neighbors			Nearest Distances		
	1	2	3	1	2	3
Agassi F1	Apolena	Stirling	Batory	1,414	2,000	2,000
Apolena	Agassi F1	Stirling	Steel	1,414	2,000	2,000
Babilon	Koros	Limba	Cigno	2,000	2,449	2,449
Baro Star	Steel	Apolena	Samoa	2,000	2,449	2,449
Batavia F1	Green Magic	Beaumont F1	Marathon	1,414	2,000	2,000
Batory	Monrello	Orantes	Corato	0,000	2,000	2,000
Bay Meadows Cemes F1	Parthenon	Beaumont F1	Batavia F1	2,449	2,449	2,449
Beaumont F1	Parthenon	Batavia F1	Belstar F1	1,414	2,000	2,000
Belstar F1	Beaumont F1	Batavia F1	Trubadur	2,000	2,000	2,449
Besty	Larsson	Samoa	Cusco	1,414	2,000	2,000
Cazy F1	Trubadur	Green Magic	Larsson	2,000	2,000	2,449
Cigno	Reggi	Cusco	Limba	1,414	2,000	2,000
Corato	Orantes	Batory	Monrello	2,000	2,000	2,000
Cusco	Cigno	Besty	Monaco F1	2,000	2,000	2,449
Green Magic	Marathon	Batavia F1	Trubadur	1,414	1,414	2,000
Ironman F1	Agassi F1	Limba	Monaco F1	2,000	2,449	2,449
Koros	Babilon	Limba	Cigno	2,000	2,449	2,449
Larsson	Besty	Orantes	Corato	1,414	2,000	2,000
Limba	Cigno	Apolena	Ironman F1	2,000	2,000	2,449
Marathon	Green Magic	Batavia F1	Parthenon	1,414	2,000	2,449
Milady F1	Trubadur	Ironman F1	Marathon	2,828	3,162	3,162
Monaco F1	Cusco	Cigno	Ironman F1	2,449	2,449	2,449
Orantes	Corato	Batory	Monrello	2,000	2,000	2,000
Parthenon	Beaumont F1	Bay Meadows Cemes F1	Batavia F1	1,414	2,449	2,449
Reggi	Cigno	Cusco	Limba	1,414	2,449	2,449
Samoa	Stromboli	Besty	Larsson	2,000	2,000	2,449
Steel	Baro Star	Apolena	Sv1002bl	2,000	2,000	2,449
Stirling	Apolena	Agassi F1	Batory	2,000	2,000	2,000
Stromboli	Samoa	Larsson	Orantes	2,000	2,449	2,449
	Sv1002bl	Steel	Apolena	0,000	2,449	2,449
Trubadur	Green Magic	Cazy F1	Belstar F1	2,000	2,000	2,449
Cezar	Batavia F1	Marathon	Parthenon	1,414	1,414	2,000
Lednitska	Milady F1	Trubadur	Marathon	2,000	2,000	2,449
Moycan	Ironman F1	Monaco F1	Limba	2,000	2,449	2,828
Muline	Larsson	Besty	Cazy F1	1,414	2,000	2,000
Naxos	Stromboli	Samoa	Baro Star	3,162	3,162	3,162
Quinta F1	Agassi F1	Apolena	Cigno	1,414	2,000	2,000
Rumba	Stirling	Corato	Babilon	1,414	2,000	2,000
Tm-04/Fl	Cazy F1	Bay Meadows Cemes F1	Batavia F1	2,449	2,449	2,449

ознаками. Сформована модель включала 17 тренувальних та 9 контрольних об'єктів (сортів). Для останніх за допомогою комп'ютерного

моделювання утворено кластери, в граничних межах яких найпростіше шукати відмітні за однією чи групою ознак сорти.

'Bay Meadows Cemes F1'	0000005010005000209500700000000	11148003
'Batavia F1'	03000000010305000309500500000000	11148001
'Cezar'	03000000010505000509500500000000	03148001
'Marathon'	03000000030505000509300500000000	08148003
'Parthenon'	03000003010507000309500700000000	08148004
'Green Magic'	03000003030305000309500500000000	08148001
'Beaumont F1'	03000005010307000509500700000000	08148005
'Belstar F1'	03300000010305000309500700000000	11148002
'TM-04/Fl'	05500000010105000309300500000000	05148001

Результати частотного аналізу прояву морфологічних ознак капусти броколі подано на рисунку 2. Модель подібних сортів капусти броколі, сформовану методом найближчих сусідів, подано на інтерактивній діаграмі простору показників «Predictor Space» (рис. 3).

Кожна вісь координат моделі відображає один із показників у ній [«рослина: за висотою (за збиральної стиглості)», «листова

пластинка: хвилястість краю», «листок: положення (на початку формування головки)»].

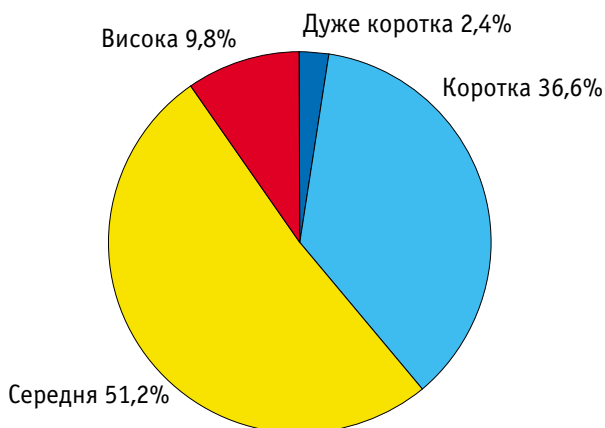
Статистичним методом на діаграмі встановлено такі фокусні сорти: 'Green Magic', 'Parthenon', 'Marathon', 'Cazy', 'Belstar F1', 'Cusco', 'Orantes', 'Besty', 'Monrello', 'Batory', 'Corato', 'Stromboli', 'Steel', 'Baro Star', 'Cigno', 'Babilon', 'Titanium'. Також визначено, що сорт 'Babilon' є унікальним за ознакою «че-

решок: за довжиною» (дуже короткий), а 'Bay Meadows Cemes F1' – за характеристикою «рослина: за висотою» (дуже коротка). Антоціанове забарвлення головки відсутнє в сортів 'Agassi F1', 'Apolena', 'Besty', 'Corato', 'Larsson', 'Naxos', 'Quinta F1', 'Samoa', 'Stromboli', 'Lednitska', 'Milady F1', 'Moysan', 'Cusco', 'Monaco F1'; наявне – в 'Reggi', 'Baro Star', 'Batory', 'Cazzy F1', 'Monrello',

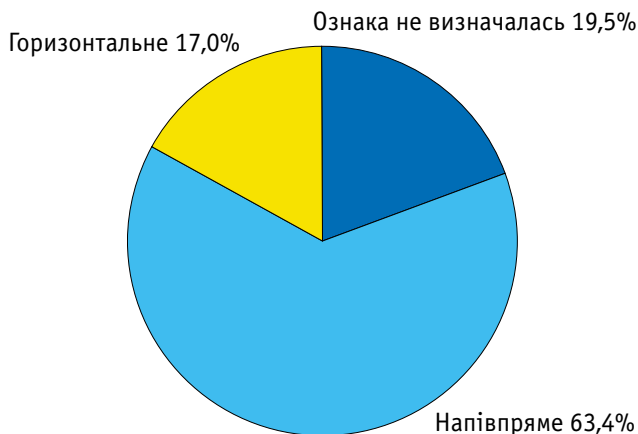
'Muline', 'Orantes', 'Rumba', 'Steel', 'Stirling', 'SV1002BL', 'Titanium', 'TM-04/FL', 'Babilon', 'Batavia F1', 'Beaumont F1', 'Belstar F1', 'Cezar', 'Green Magic', 'Ironman F1', 'Koros', 'Limba', 'Marathon', 'Parthenon', 'Trubadur', 'Cigno', 'Bay Meadows Cemes F1'.

Короткими під час збиральної стиглості є сорти 'Agassi F1', 'Apolena', 'Baro Star', 'Batory', 'Besty', 'Cazzy F1', 'Corato', 'Larsson',

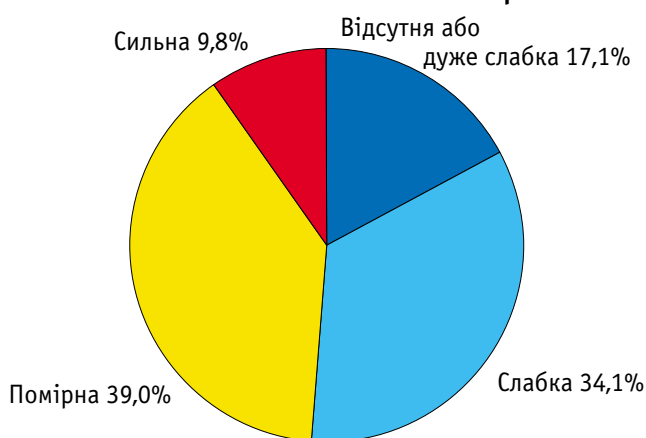
**Рослина: за висотою
(за збиральної стиглості)**



**Листок: положення
(на початку формування головки)**



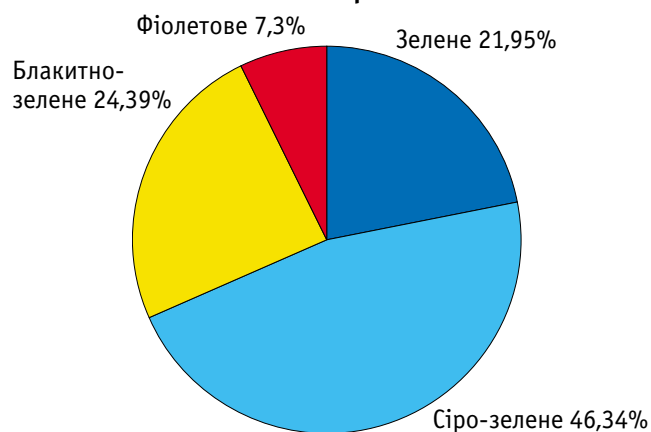
Листкова пластинка: хвилястість краю



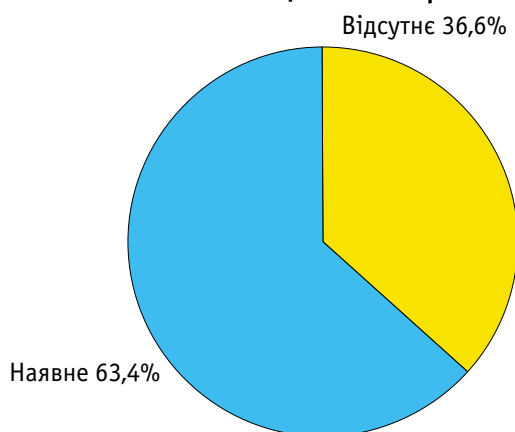
Черешок: за довжиною



Головка: забарвлення



Головка: антоціанове забарвлення



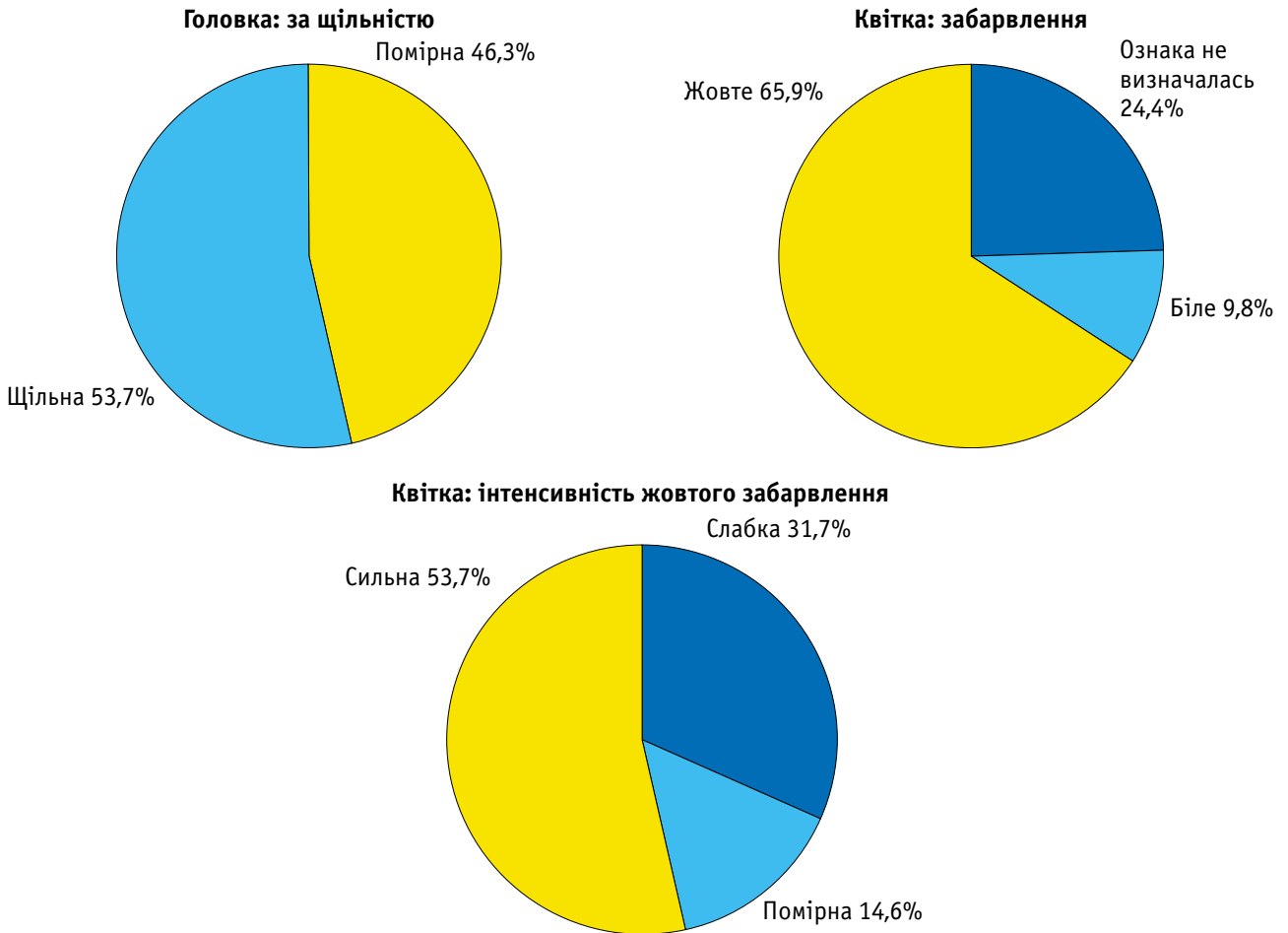


Рис. 2. Групування колекції загальновідомих сортів капусти броколі за морфологічними ознаками

'Monrello', 'Muline', 'Naxos', 'Orantes', 'Quinta F1', 'Rumba', 'Samoa', 'Steel', 'Stirling', 'Stromboli', 'SV1002BL', 'Titanium TM-04/FL'; високими – 'Babilon', 'Batavia F1', 'Beaumont F1', 'Belstar F1', 'Cezar', 'Green Magic', 'Ironman F1', 'Koros', 'Lednitska', 'Limba', 'Marathon', 'Milady F1', 'Moycan', 'Parthenon', 'Trubadur', 'Cigno', 'CUSCO', 'Monaco F1', 'Reggi'.

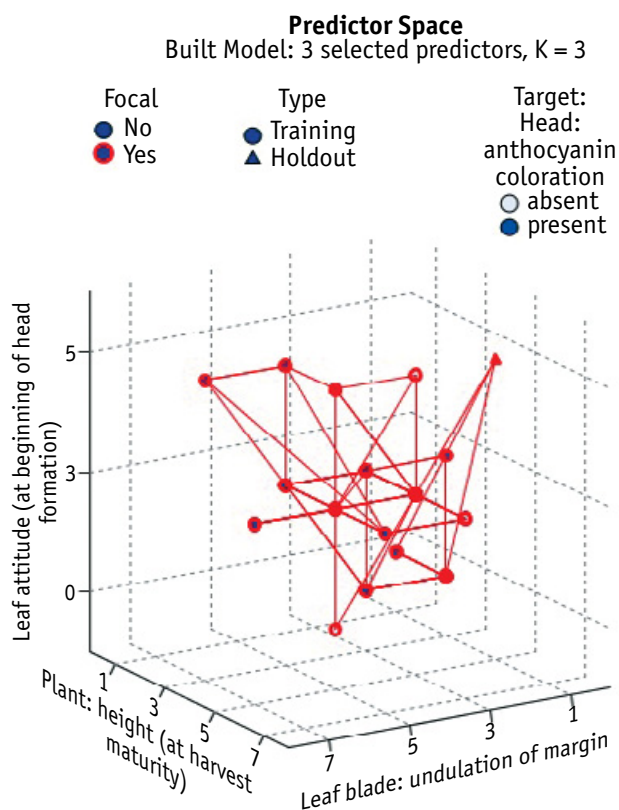
Положення листка на початку формування головки є горизонтальним у 'Naxos', 'Stromboli', 'Rumba', 'Stirling', 'TM-04/FL', 'Babilon', 'Koros'; напівпрямим – в 'Agassi F1', 'Apolena', 'Besty', 'Larsson', 'Quinta F1', 'Samoa', 'Lednitska', 'Milady F1', 'Moycan', 'CUSCO', 'Monaco F1', 'Reggi', 'Baro Star', 'Batory', 'Cazzy F1', 'Monrello', 'Muline', 'Orantes', 'Steel', 'SV1002BL', 'Titanium', 'Belstar F1', 'Ironman F1', 'Limba', 'Trubadur', 'Cigno'; напівпохилим – у 'Corato', 'Batavia F1', 'Beaumont F1', 'Cezar', 'Green Magic', 'Marathon', 'Parthenon', 'Bay Meadows Cemes F1'.

Хвилястість краю листкової пластинки відсутня або дуже слабка у сортів 'Batavia F1', 'Beaumont F1', 'Cezar', 'Parthenon', 'Bay Meadows', 'Cemes F1', 'TM-04/FL', 'Belstar

F1'; помірна – в 'Batavia F1', 'Beaumont F1', 'Cezar', 'Parthenon', 'Bay Meadows Cemes F1', 'TM-04/FL', 'Belstar F1'; сильна – в 'Koros', 'Steel', 'SV1002BL', 'Titanium'; дуже сильна – в 'Green Magic', 'Marathon', 'Stromboli', 'Besty', 'Larsson', 'Samoa', 'Moycan', 'CUSCO', 'Monaco F1', 'Baro Star', 'Cazzy F1', 'Muline', 'Ironman F1', 'Trubadur'. У сортів 'Bay Meadows Cemes F1', 'TM-04/FL', 'Naxos', 'Reggi', 'Baro Star' пухирчастість листкової пластинки відсутня або слабка.

Черешок у 'Bay Meadows Cemes F1', 'TM-04/FL', 'Reggi', 'Cezar', 'Stirling', 'Agassi F1', 'Apolena', 'Lednitska', 'Bator', 'Monrello', 'Marathon', 'Moycan', 'Ironman F1', 'SV1002BL', 'Titanium', 'Batavia F1', 'Belstar F1', 'Corato', 'Rumba', 'Quinta F1', 'Orantes', 'Limba', 'Cigno', 'Koros', 'Green Magic', 'Stromboli', 'Larsson', 'Monaco F1', 'Cazzy F1', 'Muline', 'Trubadur' середньої довжини, а в 'Naxos', 'Baro Star', 'Parthenon', 'Milady F1', 'Steel', 'Beaumont F1', 'Besty', 'Samoa', 'CUSCO', 'Babilon' – довгий.

Відстань між сортами 'Agassi F1' та 'Apolena', 'Batavia F1' та 'Green Magic', 'Beaumont F1' та 'Parthenon', 'Besty' та 'Larsson', 'Cigno' та 'Reggi', 'Green Magic' та



This chart is a lower-dimensional projection of the predictor space, which contains a total of 9 predictors.

Рис. 3. Головна діаграма моделі найбільшої подібності

'Limba'	1335535153315552359355719525551	15136001
'Ironman F1'	1335532513351555225955719525550	05148003
'Milady F1'	13355513153517533251155591313550	82148001
'Trubadur'	13355535933395552259055591010370	06148001
'Moycan'	13355537133515371371073719323579	16136001
'Lednitska'	13355715155515552251555591010550	04148001
'Koros'	13535517975395373379555719725339	13136003
'Babilon'	13535523955393352339753719525579	19136001
'Corato'	15055535153315352451055519325559	15136003
'Larsson'	15355325133315552451055509325779	18136004
'Vicario'	15355325153315353351053519325559	21136003
'Muline'	15355523933315353259133599325550	07148001
'Quinta F1'	15355525153315351251055719325770	06148003
'Orantes'	15355525153315352359153509323579	15136002
'Titanium'	15355525175715352459153799325579	20136002
'Steel'	15355727173517551379355799325971	16136003
'Batory'	15355735155515552379353511025779	16136006
'Samoa'	15357527133317373451035511023351	20136001
'Agassi F1'	15370325953515571351135711025570	06148002
'Apolena'	15375725153515352351035719525551	14136001
'SV1002BL'	15375737175715352379355799325779	16136002
'Besty'	15377525133317353351035511025579	15136005
'Monrello'	15377525155515552359353511025779	16136005
'Stromboli'	15553335133315353451035599323351	17136001
'Rumba'	15555525955395371459353711025559	13136002
'Stirling'	15555535155515352479355711025779	19136002
'Naxos'	15577337153117531451055519310551	12028009
'Monaco F1'	17335527133315351351155711323770	05148002
'Reggi'	17355535155115352251155711525559	13136001
'Cigno'	17355535955315373479553711325539	18136003
'CUSCO'	17375335133317352451075719525559	20136003
'Cazzy F1'	25355313133315334259153519300500	09148001
'Baro Star'	253753271333117551379135799525331	18136001

'Marathon', 'Larsson' та 'Besty', 'Parthenon' та 'Beaumont F1', 'Cezar' та 'Batavia F1' становила 1,414; між 'Muline' та 'Larsson', 'Quinta F1' та 'Agassi F1', 'Rumba' та 'Stirling', 'Babilon' та 'Koros', 'Baro Star' та 'Steel', 'Belstar F1' та 'Beaumont F1', 'Cazzy F1' та 'Trubadur', 'Corato' та 'Orantes', 'CUSCO' та 'Cigno', 'Ironman F1' та 'Agassi F1', 'Koros' та 'Babilon', 'Limba' та 'Cigno', 'Orantes' та 'Corato', 'Samoa' та 'Stromboli', 'Steel' та 'Baro Star', 'Stirling' та 'Apolena', 'Stromboli' та 'Samoa', 'Trubadur' та 'Green Magic', 'Lednitska' та 'Milady F1', 'Moycan' та 'Ironman F1', 'Bay Meadows Cemes F1' та 'Parthenon', 'Monaco F1' та 'CUSCO', 'Tm-04/F1' та 'Cazzy F1', 'Milady F1' та 'Trubadur', 'Naxos' та 'Stromboli' – більше ніж 2.

Як параметри моделі використовували два типи змінних: цільова – ознака «головка: антоціанове забарвлення», фокусна – «головка: забарвлення». За допомогою комп'ютерного моделювання було сформовано кластери 17 подібних сортів, ідентифікація яких передбачала 11 морфологічних ознак.

Кластер для 33 контрольних сортів капусти броколі має таку структуру:

Отже, загальновідомі сорти капусти броколі, внесені до Державного Реєстру сортів рослин, придатних для поширення в Україні, є переважно середніми за довжиною (вимірювання проводили під час збиральної стиглості) – 51,2%, чисельність коротких становить

36,6%, дуже коротких – 2,4%, довгих – 9,8%. Частка ознак, за статистичними даними яких характеризуються сорти, є такою: 19,5% сортів мають наполовину пряме положення листка на початку формування головки, 63,4% – горизонтальне, 17,1% – поникле. Хвилястість краю листової пластинки у 51% слабка, а у 12% – відсутня. Пухирчастість листової пластинки слабка у 51% сортів, середня – у 31%, відсутня – у 12%, сильна – в 5%. 75% сортів мають середню довжину черешка; 46% властивий зелений колір головки, 7% – фіолетовий. Для 63% характерне антоціанове забарвлення, а інтенсивність жовтого кольору середня у 53%.

Висновки

Проведення тесту на відмінність сортів капусти броколі через морфологічний опис ознак – це трудомісткий процес, що потребує автоматизованого опрацювання результатів статистичним методом кластерного аналізу масиву даних.

За допомогою кластерного аналізу вдалося зменшити розмірність вихідних даних для пошуку відмітних ознак між сортами. Водночас за подібними ознаками сформовано групи, об'єкти в яких схожі між собою та відрізняються від об'єктів з інших.

Завдяки ідентифікації 41 сорту капусти броколі за 32 морфологічними характеристиками виокремлено 2 пари подібних сортів за 11 ознаками. Саме кластерний аналіз забезпечив оперативність, об'єктивність і вірогідність отриманих результатів (кодів прояву ознак) для групування ознак за тесту на відмінність.

Сформована за групами кластеризації колекція загальновідомих сортів капусти броколі характеризується такими статистичними даними: 19,5% сортів мають наполовину пряме положення листка на початку формування головки, 63,4% – горизонтальне, 17,1% – поникле. Хвилястість краю листової пластинки у 51% слабка, а у 12% – відсутня. Пухирчастість листової пластинки слабка у 51% сортів, середня – у 31%, відсутня – у 12%, сильна – в 5%. 75% сортів мають середню довжину черешка; 46% властивий зелений колір головки, 7% – фіолетовий. Для 63% характерне антоціанове забарвлення, а інтенсивність жовтого кольору середня у 53% сортів.

Використана література

1. Zhang P., Guan J.-J., Huang Q.-M. et al. Phenotypic diversity of phalaenopsis based on statistic analysis and data mining. *Frontiers in Artificial Intelligence and Applications*. 2016. Vol. 281. P. 486–493.

2. Тищенко В. М., Панченко П. М., Чернишева О. П. Ідентифікація сортів та селекційних ліній пшениці озимої за збалансованими кількісними ознаками з використанням кластерного аналізу. *Вісник Полтавської державної аграрної академії*. 2013. № 3. С. 28–35. doi: 10.31210/visnyk2013.03.04
3. Compton M. E. Statistical methods suitable for the analysis of plant tissue culture data. *Plant Cell Tissue and Organ Culture*. 1994. Vol. 37, Iss. 3. P. 217–242. doi: 10.1007/BF00042336
4. Leschuk N., Orlenko N., Khareba O., Dydiv O. The use of grouping morphological characteristics of *Lettuce varieties* L. var. *capitata* for the difference test in Ukraine. *International Journal of Botany Studies*. 2020. Vol. 5, Iss. 6. P. 516–522.
5. Хареба В. В., Дидів О. Й., Дидів І. В., Лещук Н. В. Агробіологічна оцінка гібридів капусти броколі в умовах Західного Лісостепу України. *Plant Varieties Studying and Protection*. 2018. Т. 14, № 2. С. 240–244. doi: 10.21498/2518-1017.14.2.2018.134776
6. Test Guidelines for the conduct of tests for distinctness, uniformity and stability of Buckwheat (*Fagopyrum esculentum* Moench) (TG/278/1) / UPOV. Geneva, 2012. 27 p.
7. Чеботарь С. В., Сиволап Ю. М. Дифференциация, идентификация и создание базы данных сортов *T. aestivum* L. украинской селекции на основе STMS-анализа. *Цитология і генетика*. 2001. Т. 35, № 6. С. 8–27. URL: <https://cytgen.com/ru/2001/18-27N6V35.htm>
8. Тищенко В. М., Дінець О. М. Використання кластерного аналізу в селекції пшениці озимої для пошуку генотипів збалансованих за господарсько-корисними ознаками. *Збірник наукових праць науково-практичної конференції професорсько-викладацького складу* (м. Полтава, 17–18 травня 2017 р.). Полтава : РВВ ПДАА, 2017. С. 215–217.
9. Пузік Л. М., Бондаренко В. А. Екологічна стабільність гібридів капусти броколі. *Вісник Харківського національного аграрного університету. Серія : Рослинництво, селекція і насінництво, плодовоовочівництво і зберігання*. 2015. № 1. С. 15–20.
10. Anami B. S., Naveen N. M., Surendra P. Automated Paddy Variety Recognition from Color-Related Plant Agro-Morphological Characteristics. *International Journal of Image, Graphics and Signal Processing*. 2019. Vol. 11, Iss. 1. P. 12–22. doi: 10.5815/ijigsp.2019.01.02
11. Sajedur R. Genetic Analysis of Leaf and Sprout Traits of Cabbage and Brussels Sprout : MSc Thesis Plant Breeding. Wageningen, 2019. 62 p. URL: <https://edepot.wur.nl/474618>
12. Turbin V. A., Sokolov A. S., Kosterna E., Rosa R. Effect of plant density on the growth, development and yield of brussels sprouts (*Brassica oleracea* L. var. *gemmifera* L.). *Acta Agrobotanica*. 2014. Vol. 67, Iss. 4. P. 51–58. doi: 10.5586/aa.2014.049
13. Державний реєстр сортів рослин, придатних для поширення в Україні на 2023 рік / Мін-во аграр. політики та прод-ва України. Київ, 2023. URL: <https://minagro.gov.ua/file-storage/reestr-sortiv-roslin>
14. Pusik L., Pusik V., Lyubymova N., Bondarenko V. Study into formation of nutritional value of cauliflower depending on the agrobiological factors. *Eastern-European Journal of Enterprise Technologies*. 2018. Vol. 6, Iss. 11. P. 29–35. doi: 10.15587/1729-4061.2018.147748

References

1. Zhang, P., Guan, J.-J., Huang, Q.-M., Liu, Y.-F., & Zhang, J.-H. (2016). Phenotypic diversity of phalaenopsis based on statistic analysis and data mining. *Frontiers in Artificial Intelligence and Applications*, 281, 486–493.
2. Tyschenko, V. M., Panchenko, P. M., & Chernyshova, O. P. (2013). Identification of the sorts of breeding lines of winter wheat with respect to balanceness of quantitative characteristics using cluster analysis. *Bulletin of Poltava State Agrarian Academy*, 3, 28–35. doi: 10.31210/visnyk2013.03.04 [In Ukrainian]

3. Compton, M. E. (1994). Statistical methods suitable for the analysis of plant tissue culture data. *Plant Cell Tissue and Organ Culture*, 37(3), 217–242. doi: 10.1007/BF00042336
4. Leschuk, N., Orlenko, N., Khareba, O., & Dydiv, O. (2020). The use of grouping morphological characteristics of *Lettuce varieties* L. var. *capitata* for the difference test in Ukraine. *International Journal of Botany Studies*, 5(6), 516–522.
5. Khareba, V. V., Dydiv, O. Y., Dydiv, I. V., & Leschuk, N. V. (2018). Agrobiological assessment of broccoli hybrids under the conditions of the Western Forest-Steppe of Ukraine. *Plant Varieties Studying and Protection*, 14(2), 240–244. doi: 10.21498/2518-1017.14.2.2018.134776 [In Ukrainian]
6. UPOV. (2012). *Test Guidelines for the conduct of tests for distinctness, uniformity and stability of Buckwheat (Fagopyrum esculentum Moench) (TG/278/1)*. Geneva: UPOV.
7. Chebotar, S. V., & Syvolap, Yu. M. (2001). Differentiation, identification and creation of a database of varieties of *T. aestivum* L. of Ukrainian selection based on STMS-analysis. *Cytology and Genetics*, 6, 18–27.
8. Tyshchenko, V. M., & Dinets, O. M. (2017). The use of cluster analysis in the selection of winter wheat to search for genotypes balanced by economically useful traits. In *Collection of scientific works of the scientific and practical conference of professors and teachers*. Poltava: Editorial and publishing department of the Poltava State Agrarian Academy. [In Ukrainian]
9. Puzik, L. M., & Bondarenko, V. A. (2015). Ecological stability of broccoli hybrids. *The Bulletin of Kharkiv National Agrarian University. Crop production, breeding and seed production, horticulture*, 1, 15–20. [In Ukrainian]
10. Anami, B. S., Naveen, N. M., & Surendra, P. (2019). Automated Paddy Variety Recognition from Color-Related Plant Agro-Morphological Characteristics. *International Journal of Image, Graphics and Signal Processing*, 11(1), 12–22. doi: 10.5815/ijigsp.2019.01.02
11. Sajedur, R. (2019). *Genetic Analysis of Leaf and Sprout Traits of Cabbage and Brussels Sprout* (MSc Thesis Plant Breeding). Wageningen. Retrieved from <https://edepot.wur.nl/474618>
12. Turbin, V. A., Sokolov, A. S., Kosterna, E., & Rosa, R. (2014). Effect of plant density on the growth, development and yield of brussels sprouts (*Brassica oleracea* L. var. *gemmifera* L.). *Acta Agrobotanica*, 67(4), 51–58. doi: 10.5586/aa.2014.049
13. Ministry of Agrarian Policy and Food of Ukraine. (2023). *State register of plant varieties suitable for distribution in Ukraine as of May, 2023*. Retrieved from <https://minagro.gov.ua/file-storage/reyster-sortiv-roslin> [In Ukrainian]
14. Pusik, L., Pusik, V., Lyubymova, N., & Bondarenko, V. (2018). Study into formation of nutritional value of cauliflower depending on the agrobiological factors. *Eastern-European Journal of Enterprise Technologies*, 6(11), 29–35. doi: 10.15587/1729-4061.2018.147748

UDC 635.35:519.2

Dydiv, O. Y.¹, Khareba, V. V.², Khareba, O. V.², Leshchuk, N. V.^{3*}, Orlenko, N. S.³, & Orlenko, O. B.³ (2023). Application of cluster analysis for grouping *Brassica oleracea* var. *italica* varieties for the difference test. *Plant Varieties Studying and Protection*, 19(4), 207–216. <https://doi.org/10.21498/2518-1017.19.4.2023.291221>

¹Lviv National Environmental University, 1 Volodymyra Velykoho St., Dubliany, Lviv district, Lviv region, 80381, Ukraine

²National Academy of Agrarian Sciences of Ukraine, 9 Omelianovycha-Pavlenka St., Kyiv, 01010, Ukraine

³Ukrainian Institute of Plant Varieties Examination, 15 Henerala Rodymtseva St., Kyiv, 03041, Ukraine, *e-mail: nadiya1511@ukr.net

Purpose. To use cluster analysis of morphological characters to simplify the identification of *Brassica oleracea* var. *italica* and form groups of similar varieties for the test of difference. **Methods.** Analytical, mathematical and statistical methods were used in the work. As input information for the statistical processing of the obtained results, information on the results of the examination for distinctness, uniformity and stability (DUS) from the database of the Automated Information System of the Ukrainian Institute for Plant Varieties Examination was used. Cluster modelling was carried out using the IBM SPSS Statistics “Statistical Package for the Social Sciences”. **Results.** A morphological description of broccoli varieties was carried out on the basis of 32 characteristics for the examination of distinctness, uniformity and stability. The morphological code formulae of the latter, composed of the corresponding codes for the manifestation of identifying characteristics of vegetative and generative organs of plants, served as a source of initial data. Out of 41 varieties described by 32 morphological characteristics, only two groups were found to be similar in terms of the identifying characteristics of the varieties. Two

types of variables were used as parameters of the model: target – characteristic “Head: anthocyanin colour”, focal – “Head: colour”. The full list of characteristics was as follows “plant: by height (at harvest maturity)”, “leaf: position (at beginning of head formation)”, “leaf blade: wavy edge”, “leaf blade: blistering”, “petiole: by length”, “head: colour”, “head: anthocyanin colour”, “head: by density”, “flower: colour”, “flower: intensity of yellow colour”, “male sterility”. Using computer modelling, clusters of 17 similar broccoli varieties and 9 control objects (varieties) were formed, the identification of which involved eleven morphological characteristics. **Conclusions.** In order to search for distinguishing characteristics in the process of testing the difference of cabbage varieties, broccoli was grouped into clusters according to such morphological characteristics as the position of the leaf at the beginning of the formation of the head; waviness of the edge of the leaf blade; blistering of the leaf plate; petiole length; head colour; presence of anthocyanin and intensity of yellow colour.

Keywords: broccoli; statistical analysis; classification; variety; code; difference; cluster; sign; collection.

Надійшла / Received 13.09.2023

Погоджено до друку / Accepted 25.10.2023