

СТРУКТУРНА ОРГАНІЗАЦІЯ ДЕЯКИХ МІКРОСАТЕЛІТНИХ ЛОКУСІВ У ПРЕДСТАВНИКІВ РОДУ *CORYLUS*

А. М. МІЩЕНКО  0009-0003-1310-3703

І. О. АНДРЕЄВ  0000-0002-3706-8514

В. М. МЕЛЬНИК  0000-0002-0084-3142

В. А. КУНАХ  0000-0002-9418-3172

Інститут молекулярної біології і генетики НАН України,
вул. Акад. Заболотного, 150, м. Київ, 03143, Україна
e-mail: i.o.andreev@imbg.org.ua

Мета. Метою даної роботи було дослідження структурної організації дев'яти мікросателітних локусів, використаних для характеристики українських сортів фундука, на основі аналізу даних сиквенування ПЛП-продуктів та послідовностей з бази даних NCBI GenBank. **Методи.** ПЛП-ампліфікація зі специфічними праймерами, клонування і сиквенування ПЛП-продуктів, біоінформатичний аналіз. **Результати.** Проведено ампліфікацію чотирьох SSR-локусів з подальшим клонуванням і сиквенуванням отриманих ПЛП-продуктів, також пошук алельних варіантів дев'яти SSR-локусів для п'яти видів *Corylus*. Визначено структурну організацію повторюваних ділянок цих локусів та характер їхньої мінливості, проаналізовано частотний розподіл алельних варіантів. Більшість локусів має консервативну послідовність і містить прості тринуклеотидні повтори, кількість одиниць в яких зумовлює розмір відповідних варіантів. У складі трьох локусів знайдено комбіновані ди- та тринуклеотидні повтори або перервані повтори, і виявлено делеції та інсерції в окремих варіантах. Такі структурні особливості зумовлюють складний характер розподілу варіантів цих локусів за розміром. **Висновки.** Визначено структурну організацію дев'яти SSR-локусів, які пропонуються для ідентифікації українських сортів фундука. Виявлені структурні особливості слід враховувати при інтерпретації результатів ПЛП-аналізу цих локусів у сортів фундука.

Ключові слова: *Corylus*, SSR-маркери, нуклеотидна послідовність, генетична мінливість, молекулярно-генетичний аналіз.

Вступ. Ліщина (*Corylus L.*) — рід листопадних чагарників і дерев родини Betulaceae, який включає понад 15 видів, поширених у Євразії та Північній Америці. Найбільше промислове значення має ліщина європейська (*Corylus avellana L.*), яка є основним джерелом культурного фундука. Її горіхи цінуються за високий вміст жирів, білків, мікроелементів та біологічно активних речовин, що робить їх одним із ключових компонентів світової кондитерської промисловості. Стійке зростання попиту на фундук стимулює розвиток селекційних програм, спрямованих на підвищення врожайності, поліпшення якості плодів, адаптацію до різних кліматичних умов та формування стійкості до основних хвороб, зокрема фітофторозу, спричиненого *Anisogramma anomala* (Gökirmak et al., 2009; Botta et al., 2019).

Попри значні успіхи у селекції фундука в Туреччині, Італії, США та Китаї, генетичне різноманіття роду *Corylus*, а також генетичні зв'язки між окремими сортами залишаються дослідженими неповною мірою. Традиційні методи морфологічної класифікації не завжди забезпечують достатню точність, оскільки фенотипові ознаки можуть варіювати під впливом умов середовища. У цьому контексті особливої ваги набувають молекулярно-генетичні підходи, що дозволяють здійснювати об'єктивну ідентифікацію генотипів, простежувати родовід, виявляти ступінь спорідненості між сортами й оцінювати внутрішньовидове генетичне різноманіття (Botta et al., 2019).

Упродовж останніх десятиліть для досліджень генетичної структури *C. avellana* було запропоновано широкий спектр молекулярних маркерів, серед яких RAPD, AFLP, ISSR, SNP, проте найширше застосування отримали мікросателітні маркери (SSR).

Вони характеризуються високою частотою мутацій, кодомінантним успадкуванням, високою відтворюваністю результатів та рівномірним розподілом у геномі. SSR-маркери продемонстрували високу ефективність у дослідженнях генетичного різноманіття як культурних сортів фундука (Ellegren, 2000; Eckert, Hile, 2009; Gökirmak et al., 2009; Öztürk et al., 2018), так і диких популяцій ліщини (Gürçan et al., 2018; Tanhuanpää et al., 2019; Ershadi, Toolir, 2022), у визначенні географічного походження сортів, перевірки спорідненості, а також у виявленні генетичних копій сортів у складі колекцій. Крім того, вони стали основою для побудови карт зчеплення, локалізації QTL, асоційованих із господарсько-цінними ознаками, ідентифікації маркерів, пов'язаних із генами стійкості, та створення ефективних систем молекулярної паспортизації сортів фундука (Botta et al., 2019).

Мікросателітна ДНК — це короткі тандемні повтори довжиною від одного до шести нуклеотидів, які можуть бути розташовані як у кодувальних, так і некодувальних ділянках геному. Вони часто характеризуються високим рівнем поліморфізму навіть у близькоспоріднених генотипів, що робить їх універсальним інструментом аналізу у популяційній та селекційній генетиці. Останніми роками значного розвитку набули дослідження, спрямовані на розробку та характеристику SSR-маркерів для *C. avellana*. Завдяки прогресу у технологіях секвенування нового покоління (NGS) було ідентифіковано понад 700 поліморфних SSR-маркерів, отриманих із геномних бібліотек, збагачених повторами, а також близько 130 поліморфних EST-SSR, ідентифікованих із транскриптомних бібліотек (Bhattarai, Mehlenbacher, 2017; Botta et al., 2019).

Проте більшість досліджень зосереджена переважно на вивченні поліморфізму SSR-локусів за довжиною ампліконів без детального аналізу їхньої молекулярної структури. Водночас структурна організація цих локусів, зокрема, тип та кількість повторів, будова фланкуючих ділянок і наявність внутрішніх варіацій, може мати істотний вплив на їхню стабільність, ефективність ампліфікації і рівень інформативності. Дослідження цих характеристик поглиблює розуміння механізмів формування алейного різноманіття та поліморфізму SSR-локусів.

Проведене раніше дослідження (Mishchenko et al., 2026) показало високу ефективність дев'яти SSR-маркерів для ідентифікації та диференціації

сортів фундука української селекції. Проте структурні особливості цих локусів і природа виявлених алейних відмінностей залишалися нерозкритими.

Метою даної роботи було дослідження структурної організації дев'яти мікросателітних локусів, використаних раніше для характеристики сортів фундука української селекції, на основі аналізу даних секвенування ПЛР-продуктів та послідовностей з бази даних NCBI GenBank.

Матеріали і методи

Рослинний матеріал та виділення ДНК. Використано зразки рослинного матеріалу чотирьох сортів фундука (Свічковий, Пиріжок, Софіївський 1, Корончастий) з помологічної колекції Національного дендрологічного парку «Софіївка» НАН України, люб'язно надані членом-кореспондентом НАН України [І. С. Косенком].

Для виділення геномної ДНК використовували молоде листя, попередньо висушене з силікагелем. Екстракцію здійснювали за модифікованим СТАВ-методом (Mishchenko, Andreev, 2023), а якість та концентрацію ДНК оцінювали електрофоретичним розділенням зразків шляхом порівняння зі стандартною ДНК фага λ відомої концентрації, а також спектрофотометричним аналізом на приладі NanoDrop 1000 (Thermo Fisher Scientific).

Полімеразна ланцюгова реакція та фрагментний аналіз. Для ПЛР-ампліфікації використовували стандартну ПЛР-суміш з Taq-полімеразою, яка містила близько 30 нг геномної ДНК. Детальні послідовності праймерів та умови проведення аналізу описані раніше (Mishchenko et al., 2026).

Клонування та секвенування ПЛР-продуктів. Після розділення продуктів ампліфікації в агарозному гелі відібрані смуги вирізали з гелю. ДНК очищували за допомогою набору Silica Bead DNA Gel Extraction Kit (Thermo Scientific) відповідно до інструкції виробника.

Лігування очищених фрагментів у вектор pGEM-T проводили в реакційній суміші загальним об'ємом 15 мкл, яка містила: 4 мкл очищеного фрагмента ДНК, 4 мкл вектора pGEM-T, 1,5 мкл 10 × буфера для ДНК-лігази, 1,5 мкл PEG, 0,8 мкл T4 ДНК-лігази. Реакційну суміш інкубували протягом 14 год при температурі 4–7 °С.

Для трансформації використали компетентні клітини *Escherichia coli* штаму XL1-Blue, підготовлені за класичним CaCl₂-методом (Sambrook et al., 1989), з подальшою детекцією колоній за синім або білим забарвленням в присутності X-Gal та IPTG.

Біоінформатичний аналіз. Для пошуку і подальшого аналізу структурної організації мікро-

сателітних локусів використали загальну базу нуклеотидних послідовностей NCBI GenBank, а також геномні збірки *C. avellana* (GCA_901000735.2; GCA_965113065.2; GCA_965113165.2), *C. americana* Walter (GCA_036287905.1; GCA_036287885.1), *C. chinensis* Franch. (GCA_048127625.1); *C. colurna* L. (GCA_965233425.1; GCA_965233385.1), *C. heterophylla* Fisch. ex Trautv. (GCA_016403345.1; GCA_019343435.1).

Пошук проводили за допомогою веб-інструменту BLAST (Altschup et al., 1990).

Результати і обговорення

Для вивчення було відібрано чотири мікросателітних локуси (у дужках вказана назва сорту): BR414 (Софіївський 1), GB950 (Пиріжок), BR259 (Свічковий) та CAC-C008 (Корончастий), які у зазначених сортів знаходилися в гомозиготному стані і при ампліфікації зі специфічними праймерами формували одиночні чіткі смуги. Завдяки вибору генотипів з локусами у гомозиготному стані було одержано гомогенні продукти ПЛР, які після електрофоретичного розділення вирізували і після очистки використали для подальшого клонування і сиквенування. Для під-

твердження специфічності досліджених ділянок визначені нуклеотидні послідовності аналізували з використанням бази даних NCBI GenBank і після цього депонували в цій же базі під номерами PV926123-PV926126.

Для вивчення структурної організації мікросателітних локусів окрім послідовностей, сиквенування в ході дослідження придатності SSR-маркерів для ідентифікації українських сортів фундука (Mishchenko et al., 2026), ми використали також послідовності, знайдені в базі даних NCBI GenBank для *C. avellana* та інших видів роду — *C. americana*, *C. chinensis*, *C. colurna*, *C. heterophylla*. Зокрема, ми проаналізували прості повтори, наявні у складі мікросателітних локусів, їх мінливість, а також наявність інших мутацій (делецій або інсерцій), які впливають на довжину очікуваних ПЛР-продуктів. Повний перелік проаналізованих послідовностей, що містять зазначені мікросателітні локуси, разом з характеристиками простих повторів у їхньому складі та розміром виявлених алейних варіантів наведено в табл. 1. Також було проведено порівняльний аналіз розподілів частот варіантів, знайдених у цьому та попередньому дослідженнях, з огляду на визначену структуру локусів.

Таблиця 1. Виявлені у різних видів *Corylus* алейні варіанти мікросателітних локусів з визначеним мотивом простого повтору та номером відповідної послідовності в базі даних NCBI GenBank

Локус	Розмір алейя, п.н.	Мотив	Вид / сорт	Запис в базі даних NCBI GenBank
BR259	237	(TCA) ₈	<i>C. avellana</i> cv. Свічковий	PV926123
	234	(TCA) ₇	<i>C. avellana</i>	KJ193889
	249	(TCA) ₁₂	<i>C. avellana</i> cv. Tombul	XM_059593920
	234	(TCA) ₇	<i>C. avellana</i>	OZ219478
	228	(TCA) ₅	<i>C. avellana</i>	CBCOTP020000001
	237	(TCA) ₈	<i>C. chinensis</i>	CM106774
	237	(TCA) ₈	<i>C. colurna</i>	OZ247467
	237	(TCA) ₈	<i>C. colurna</i>	CBDDSP010000005
	234	(TCA) ₇	<i>C. americana</i> cv. Rush	CP125818
	234	(TCA) ₇	<i>C. americana</i> cv. Winkler	CP125829
	234	(TCA) ₇	<i>C. heterophylla</i>	CM028143
	240	(TCA) ₉	<i>C. heterophylla</i>	JADFUG010000175

Структурна організація деяких мікросателітних локусів у представників роду *Corylus*...

Продовження Таблиці 1.

Локус	Розмір алеля, п.н.	Мотив	Вид / сорт	Запис в базі даних NCBI GenBank
BR414	125	(AAT) ₇	<i>C. avellana</i> cv. Софіївський 1	PV926124
	122	(AAT) ₆	<i>C. avellana</i>	KJ193944
	122	(AAT) ₆	<i>C. avellana</i> cv. Tombul	XM_059591969
	125	(AAT) ₇	<i>C. avellana</i>	OZ219477
	125	(AAT) ₇	<i>C. avellana</i>	CBCOTP020000001
	128	(AAT) ₈	<i>C. chinensis</i>	CM106772
	128	(AAT) ₈	<i>C. colurna</i>	OZ247466
	137	(AAT) ₁₁	<i>C. colurna</i>	CBDDSP010000004
	112 ^a	(AAT) ₄	<i>C. americana</i> cv. Rush	CP125817
	135	(AAT) ₁₂	<i>C. americana</i> cv. Winkler	CP125828
	131	(AAT) ₉	<i>C. heterophylla</i>	CM028142
	131	(AAT) ₉	<i>C. heterophylla</i>	JADFUG010000330
CAC008	205	(AAT) ₅ (AAG) ₈	<i>C. avellana</i> cv. Корончастий	PV926125
	226 ^b	(AAT) ₅ (AAG) ₁₆	<i>C. avellana</i>	DQ151612
	202	(AAT) ₆ (AAG) ₆	<i>C. avellana</i> cv. Tombul	NC_081543
	244 ^b	(AAT) ₄ (AAG) ₁₀ (AAG) ₈	<i>C. avellana</i>	OZ219476
	238 ^b	(AAT) ₅ (AAG) ₃ (AAG) ₁₃	<i>C. avellana</i>	CBCOTP020000004
	238	(AAT) ₉ (AAG) ₁₅	<i>C. chinensis</i>	CM106773
	223	(AAT) ₈ (AAG) ₁₁	<i>C. colurna</i>	OZ247465
	221 ^{b,r}	(AAT) ₅ (AAG) ₁₀ (AAG) ₆	<i>C. colurna</i>	CBDDSP010000003
	244 ^b	(AAT) ₄ (AAG) ₁₀ (AAG) ₈	<i>C. americana</i> cv. Rush	CP125816
	229 ^b	(AAT) ₄ (AAG) ₈ (AAG) ₅	<i>C. americana</i> cv. Winkler	CP125827
	232 ^b	(AAT) ₅ (AAG) ₈ (AAG) ₅	<i>C. heterophylla</i>	CM028141
	230 ^b	(AAT) ₅ (AAG) ₇ (AAG) ₅	<i>C. heterophylla</i>	JADFUG010000040
GB950	157	(TGG) ₂ TCG(TGG) ₃ (GA) ₄	<i>C. avellana</i> cv. Пиріжок	PV926126
	160	(TGG) ₇ (GA) ₄	<i>C. avellana</i>	KT943908
	157	(TGG) ₄ TCGTGG(GA) ₄	<i>C. avellana</i>	XM_059598699
	157	(TGG) ₄ TCGTGG(GA) ₄	<i>C. avellana</i> cv. Tombul	NC_081546
	157	(TGG) ₄ TCGTGG(GA) ₄	<i>C. avellana</i>	OZ219480
	157	(TGG) ₄ TCGTGG(GA) ₄	<i>C. avellana</i>	CBCOTP020000006
	162	(TGG) ₂ TGCTG(TGC) ₃ TGG(GA) ₄ ^t	<i>C. chinensis</i>	CM106776
	157	(TGG) ₆ (GA) ₄	<i>C. colurna</i>	OZ247469
	153	(TGG) ₄ (GA) ₅	<i>C. colurna</i>	CBDDSP010000007
	157	TAG(TGGTGC) ₂ TGG(GA) ₄	<i>C. americana</i> cv. Rush	CP125819
	157	TAG(TGGTGC) ₂ TGG(GA) ₄	<i>C. americana</i> cv. Winkler	CP125830
	157	(TGG) ₆ (GA) ₄	<i>C. heterophylla</i>	CM028145
	160	TCG(TGG) ₆ (GA) ₄	<i>C. heterophylla</i>	JADFUG010000269

Продовження Таблиці 1.

Локус	Розмір алеля, п.н.	Мотив	Вид / сорт	Запис в базі даних NCBI GenBank
BR270	92	(CTG) ₆	<i>C. avellana</i>	KJ193894
	98	(CTG) ₈	<i>C. avellana</i>	XM_059578301
	98	(CTG) ₈	<i>C. avellana</i> cv. Tombul	NC_081541
	98	(CTG) ₈	<i>C. avellana</i>	OZ219475
	92	(CTG) ₆	<i>C. avellana</i>	CBCOTP020000003
	98	(CTG) ₈	<i>C. chinensis</i>	CM106770
	95	(CTG) ₇	<i>C. columa</i>	OZ247463
	95	(CTG) ₇	<i>C. columa</i>	CBDDSP010000001
	92	(CTG) ₆	<i>C. americana</i> cv. Rush	CP125814
	92	(CTG) ₆	<i>C. americana</i> cv. Winkler	CP125825
	92	(CTG) ₆	<i>C. heterophylla</i>	CM028139
	95	(CTG) ₇	<i>C. heterophylla</i>	JADFUG010000126
BR322	105	(ACT) ₇	<i>C. avellana</i>	KJ193907
	108	(ACT) ₈	<i>C. avellana</i>	XM_059576315
	108	(ACT) ₈	<i>C. avellana</i>	XM_059576314
	108	(ACT) ₈	<i>C. avellana</i> cv. Tombul	NC_081551
	105	(ACT) ₇	<i>C. avellana</i>	OZ219484
	105	(ACT) ₇	<i>C. avellana</i>	CBCOTP020000009
	105	(ACT) ₇	<i>C. chinensis</i>	CM106780
	105	(ACT) ₇	<i>C. columa</i>	OZ247472
	105	(ACT) ₇	<i>C. columa</i>	CBDDSP010000010
	111	(ACT) ₉	<i>C. americana</i> cv. Rush	CP125824
	105	(ACT) ₇	<i>C. americana</i> cv. Winkler	CP125835
	105	(ACT) ₇	<i>C. heterophylla</i>	CM028149
107 ^a	(ACT) ₇	<i>C. heterophylla</i>	JADFUG010000317	
GB949	156	(TGG) ₇ TGA(TGG) ₄	<i>C. avellana</i>	KT943907
	153	(TGG) ₆ TGA(TGG) ₄	<i>C. avellana</i>	XM_059606598
	153	(TGG) ₆ TGA(TGG) ₄	<i>C. avellana</i>	XM_059606599
	153	(TGG) ₆ TGA(TGG) ₄	<i>C. avellana</i> cv. Tombul	NC_081549
	147	(TGG) ₄ TGA(TGG) ₄	<i>C. avellana</i>	OZ219483
	156	(TGG) ₇ TGA(TGG) ₄	<i>C. avellana</i>	CBCOTP020000007
	156	(TGG) ₇ TGA(TGG) ₄	<i>C. chinensis</i>	CM106779
	150	(TGG) ₅ TGA(TGG) ₄	<i>C. columa</i>	OZ247470
	150	(TGG) ₁₀	<i>C. columa</i>	CBDDSP010000008
	159	(TGG) ₈ TGA(TGG) ₄	<i>C. americana</i> cv. Rush	CP125822
	156	(TGG) ₂ TGAGGG (TGG) ₃ (TGA) ₂ (TGG) ₃	<i>C. americana</i> cv. Winkler	CP125833
	153	(TGG) ₆ TGA(TGG) ₄	<i>C. heterophylla</i>	CM028147
	153	(TGG) ₆ TGA(TGG) ₄	<i>C. heterophylla</i>	JADFUG010000054

Продовження Таблиці 1.

Локус	Розмір алеля, п.н.	Мотив	Вид / сорт	Запис в базі даних NCBI GenBank
BR438	193	(CT) ₁₁ (GTCT) ₂ (TCA) ₈	<i>C. avellana</i>	KJ193954
	195	(CT) ₁₂ (GTCT) ₂ (TCA) ₈	<i>C. avellana</i>	XM_059601891
	195	(CT) ₁₂ (GTCT) ₂ (TCA) ₈	<i>C. avellana</i>	XM_059601892
	195	(CT) ₁₂ (GTCT) ₂ (TCA) ₈	<i>C. avellana</i> cv. Tombul	NC_081547
	199	(CT) ₁₁ (GTCT) ₂ (TCA) ₁₀	<i>C. avellana</i>	OZ219479
	191	(CT) ₁₀ (GTCT) ₂ (TCA) ₈	<i>C. avellana</i>	CBCOTP020000005
	186	(CT) ₈ (GTCT)(TCA) ₉	<i>C. chinensis</i>	CM106775
	181	(CT) ₈ (GT)(GTCT) ₂ (TCA) ₆	<i>C. colurna</i>	OZ247468
	184	(CT) ₈ (GT)(GTCT) ₂ (TCA) ₇	<i>C. colurna</i>	CBDDSP010000006
	205	(CT) ₁₅ (GT) ₃ CA(GTCT)(TCA) ₈	<i>C. americana</i> cv. Rush	CP125820
	195	(CT) ₁₆ (TCA) ₈	<i>C. americana</i> cv. Winkler	CP125831
	194	(CT) ₁₃ (GTCT)(TCA) ₈	<i>C. heterophylla</i>	JADFUG010000309
BR464	280	(ATC) ₇	<i>C. avellana</i>	KJ193960
	280	(ATC) ₇	<i>C. avellana</i>	XM_059604722
	280	(ATC) ₇	<i>C. avellana</i> cv. Tombul	NC_081548
	286	(ATC) ₉	<i>C. avellana</i>	OZ219482
	280	(ATC) ₇	<i>C. avellana</i>	CBCOTP020000002
	295	(ATC) ₁₂	<i>C. chinensis</i>	CM106777.1
	283	(ATC) ₈	<i>C. colurna</i>	OZ247473
	283	(ATC) ₈	<i>C. colurna</i>	CBDDSP010000011
	289	(ATC) ₁₀	<i>C. americana</i> cv. Rush	CP125821
	286	(ATC) ₉	<i>C. americana</i> cv. Winkler	CP125832
	283	(ATC) ₈	<i>C. heterophylla</i>	CM028146
	295	(ATC) ₁₂	<i>C. heterophylla</i>	JADFUG010000222

Примітка. Жирним виділено номери послідовностей, визначених в ході цього дослідження. ^a — локус містить делецію 4 п.н.; ^b — локус містить делецію 3 п.н.; ^c — локус містить інсерцію 12 п.н.; ^d — локус містить делецію 20 п.н.; ^e — локус містить інсерцію 2 п.н.

Аналіз послідовностей мікросателітних локусів з різних геномів показав, що більшість з них мають сталу структурну організацію, а мінливість їхньої довжини зумовлена лише варіаціями у кількості простих повторів. Водночас для деяких SSR-локусів виявлено й інші відмінності, зумовлені мутаціями, які можуть бути пов'язані як із видовими чи сортовими відмінностями, так і з мікромутаційними процесами у межах ділянок простих повторів.

Локус BR259 у представників роду *Corylus* містить тринуклеотидний повтор (TCA)_n, для якого спостерігається варіація кількості базових одиниць від п'яти до дванадцяти, що відповідає варіантам розміром від 229 до 249 п.н. (табл. 1). Така мінливість добре пояснює відмінності в розмірі алелів, кратні трьом нуклеотидам, виявлені в попередньому дослідженні (Mishchenko et al.,

2026). У сорту «Свічковий» (PV926123) до складу локусу входить повтор (TCA)₈, що відповідає алелю довжиною 237 п.н., який вписується у загальний ряд варіантів цього локусу, знайдених у інших видів. Частотний розподіл алелів BR259 у цьому і у попередньому дослідженнях наближений до нормального і це добре узгоджується з простою структурною організацією локусу без перерваних або складних повторів (рис. 1).

Локус BR414 містить тринуклеотидний мотив (AAT)_n, для якого спостерігається варіація кількості повторів у межах від чотирьох до дванадцяти, що відповідає алелям з розміром 112–137 п.н. (табл. 1). У сорту «Софіївський 1» (PV926124) ділянка простих повторів представлена (AAT)₇, що відповідає алелю довжиною 125 п.н., який знаходиться в центрі розподілу варіантів цього локусу (рис. 1). Порівняння з послідов-

ностями з GenBank свідчить про наявність як коротших (AAT)₄₋₆, так і довших (AAT)₈₋₁₂ варіантів у різних видів і сортів, що вказує на зміни числа повторів упродовж еволюції локусу. Частотний розподіл алелів BR414 у попередньому дослідженні характеризувався домінуванням одно-

го варіанта 125 п.н., що пояснюється простою, однорідною структурою повтору та відсутністю перерваних або комбінованих мотивів, які могли б сприяти формуванню більш розсіяного набору алелів.

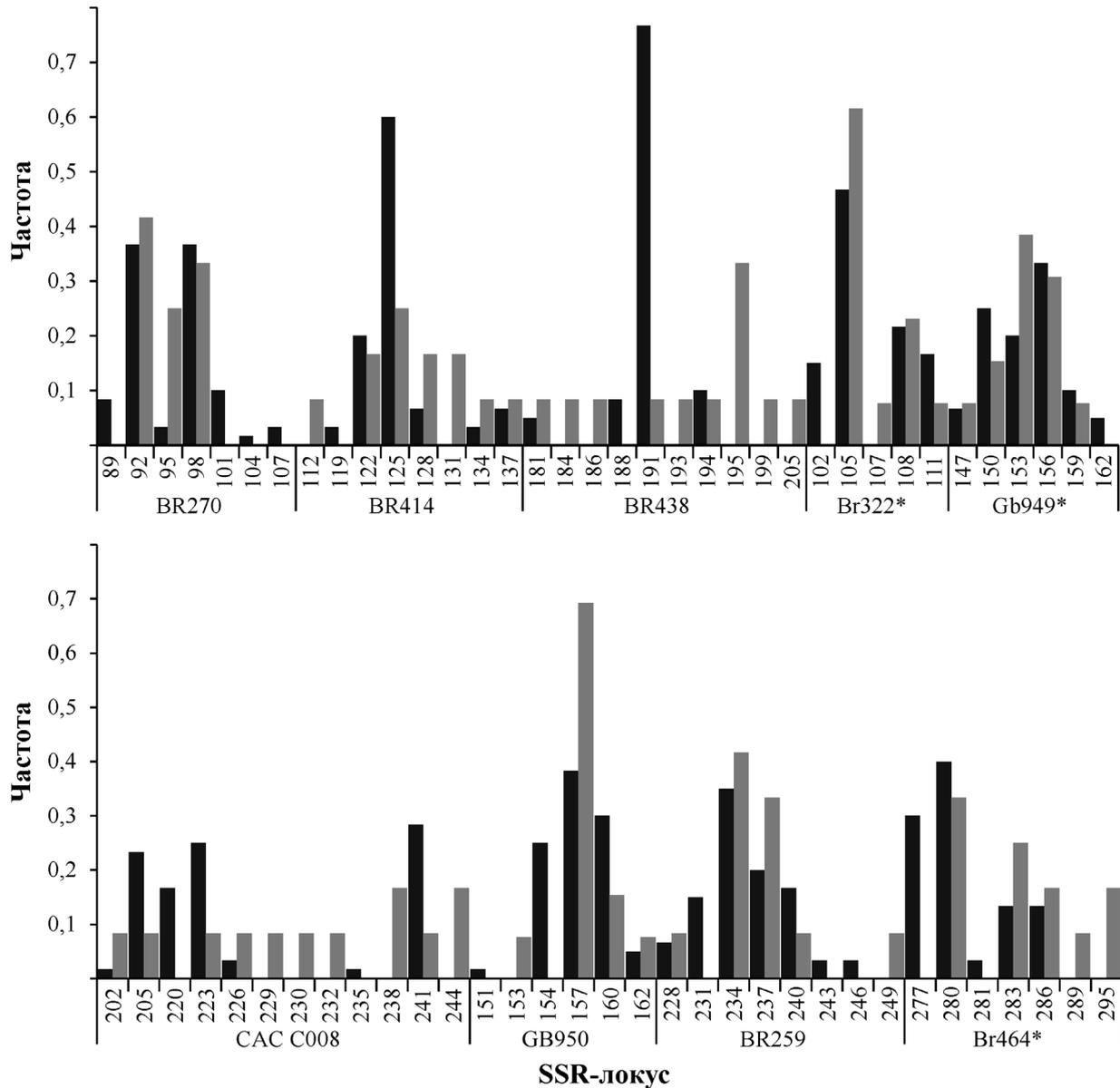


Рис. 1. Розподіл частот алельних варіантів досліджених мікросателітних локусів: ■ — за даними генетичного аналізу сортів фундука української селекції (Mishchenko et al., 2026); ■ — за результатами аналізу нуклеотидних послідовностей з бази даних NCBI GenBank. Під стовпчиками вказано розміри відповідних алелів в п.н. Розміри алелів для локусів, помічених зірочкою, скориговано за результатами їхньої оцінки по нуклеотидним послідовностям.

Структурна організація деяких мікросателітних локусів у представників роду *Corylus*...

Локус SAC-C008 має складну, комбіновану будову повторюваної ділянки, що включає два різні тринуклеотидні мотиви типу (AAT)_n та (AAG)_m, кількість яких істотно варіює у видів *Corylus* (табл. 1). У доступних послідовностях з бази даних GenBank знайдено алелі з розмірами в широкому діапазоні від 202 до 244 п.н., різниця між якими не завжди кратна розміру одиниці простого повтору, що характерно для SSR-локусів з однорідними повторами. Комбіновані мотиви можуть виникати внаслідок кількох механізмів: формування вторинних структур ДНК, локальних рекомбінацій між гомологічними повторюваними ділянками або послідовних одонуклеотидних замінів, які перетворюють один тип тринуклеотиду на інший. Подібне чергування добре узгоджується з моделями еволюції простих повторів, які припускають, що зміна одного тринуклеоти-

ду може передувати подальшій експансії нового повтору впритул до старого з подальшим формуванням комбінованих повторів в межах одного локусу (Buschiazzo, Gemmell, 2006). Для локусу SAC-C008 виявлено перебудови внутрішньої організації локусу у вигляді інсерції довжиною 12 п.н., яка присутня у геномах кількох різних видів, а також делеція розміром 20 п.н. в геномі одного зі зразків *C. colurna*. У сорту «Корончастий» (PV926125) локус представлений алелем розміром 205 п.н., який містить комбінацію повторів (AAT)₅(AAG)₈. Цей алельний варіант подібний за організацією до виявлених у деяких інших представників *C. avellana*, *C. chinensis* та *C. colurna*, проте є простішим порівняно з більш довгими та перерваними варіантами повторів, характерними для *C. colurna*, *C. americana* та *C. heterophylla* (рис. 2а).

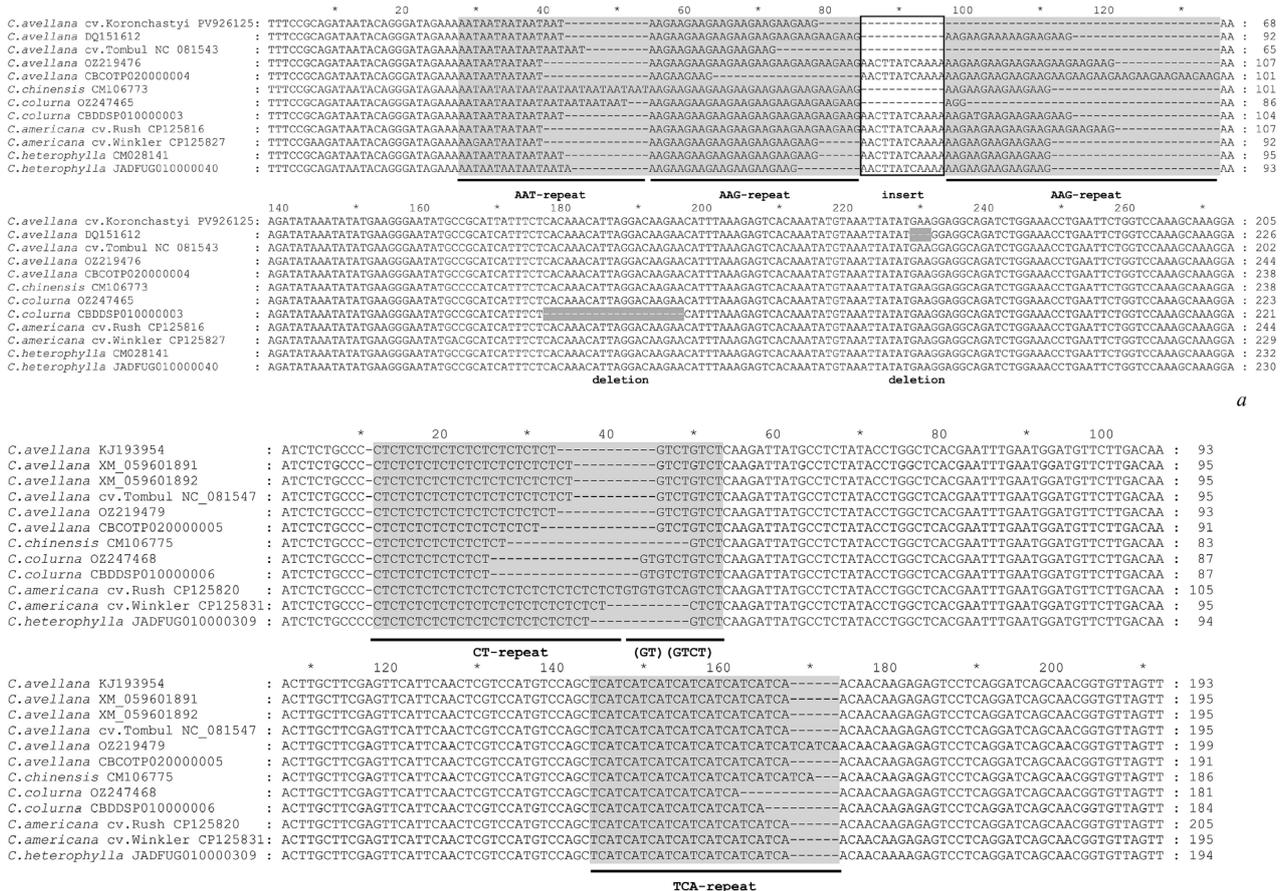


Рис. 2. Структурна організація мікросателітних локусів SAC-C008 (а) і BR438 (б). Сірим кольором виділено ділянки простих повторів, рамкою показано інсерцію, також позначено місця делецій у окремих варіантах.

Вказана в роботі (Akin et al., 2016) структурна організація локусу SAC-C008, яка містить перерваний однорідний повтор (AAG)₁₁(AAG)₃, відрізняється від визначеної нами. У виявлених нами варіантах локусу повтор (AAG)_n закономірно зберігається та представлений більшою кількістю копій, але водночас спостерігається варіабельність за числом копій і у повтору (AAT)_m, про який не згадують вказані автори. Така складна організація разом із наявністю в геномах окремих зразків інсерції довжиною 12 п.н., яка ділить повтор (AAG)_n на дві частини, впливає на характер мінливості довжини різних варіантів локусу SAC-C008, оскільки внесок в неї вносять зміни в кількості обох типів повторів. Крім того, у двох випадках було виявлено делеції, які також впливають на розмір окремих варіантів (рис. 2а).

Визначена нами структурна організація локусу SAC-C008 дозволяє пояснити асиметричний розподіл алельних варіантів з широким розмахом розмірів, знайдених у попередньому дослідженні (Mishchenko et al., 2026) (рис. 1). Такий розподіл зазвичай свідчить про підвищену внутрішню мінливість локусу, яка характерна саме для мікросателітів із комбінованими повторами. Доведено, що наявність двох ділянок з різними повторами у межах одного SSR-локусу значно підвищує його мінливість і призводить до формування більшої кількості варіантів (Bull et al., 1999). Це підтверджують також роботи інших авторів (Akin et al., 2016; Freixas-Coutin et al., 2019), за даними яких SAC-C008 належить до числа локусів з найбільшою кількістю алелів.

Локус GB950 вирізняється складною та варіабельною структурою ділянки повторів, сформованої тринуклеотидним мотивом (TGG)_n, який можуть переривати вставки тринуклеотиду TCG та інших похідних TGG (табл. 1), які виникали ймовірно в результаті точкових замін (Ananda et al., 2014). В геномах представників роду *Corylus* для локусу GB950 виявлено варіанти розміром 153–162 п.н., відмінності яких за розміром загалом пояснюються мінливістю кількості тринуклеотидних повторів, однак в окремих випадках спостерігаються відхилення від цієї закономірності, як от у *C. chinensis*, де в одному з повторів присутня одонуклеотидна делеція. На противагу цьому, в геномах *C. colurna*, *C. heterophylla* та у послідовності *C. avellana* KT943908 цей повтор є гомогенним. Крім того, ближче до 5'-кінця локусу розташований динуклеотидний повтор (GA)₄, який в геномі одного із зразків *C. colurna* по-

довжений до (GA)₅, і такі відмінності можуть накладатися на мінливість довжини локусу, зумовлену варіаціями у кількості одиниць основного повтору (TGG)_n.

У *C. chinensis* виявлено незвичайний варіант ділянки повторів локусу GB950, а саме (TGG)₂TGCTG(TGC)₃. Цей варіант суттєво відрізняється від типових для цього локусу, які містять однорідний або перерваний повтор (TGG)_n, і утворився ймовірно в результаті трансверсії G→C з подальшою експансією похідного мотиву TGC і одонуклеотидної делеції в одному з тринуклеотидів, що вказує на давність цієї еволюційної події. Такі одонуклеотидні заміни відбуваються з меншою частотою, ніж зміна кількості повторів, та можуть бути інформативними для досліджень мікроеволюційних процесів. Зокрема, переривання простого повтору підвищує його стабільність та зменшує частоту подальших мутацій, завдяки чому змінена послідовність є кращим популяційним маркером порівняно із звичайними повторами (Ananda et al., 2014). У *C. americana* спостерігається істотна зміна організації повторюваної ділянки, зумовлена трансверсією G→C з формуванням шестинуклеотидного повтору (TGGTGC)₂, якому передують змінений тринуклеотид TAG. У сорту «Пиріжок» (PV926126) локус представлений алелем розміром 157 п.н. з ділянкою повторів (TGG)₂TCG(TGG)₃, яка відрізняє його від решти інших зразків *C. avellana*, для більшості яких властивий варіант локусу тієї ж довжини, але з структурою повтору типу (TGG)₄TCGTGG.

Частотний розподіл алелів GB950 в обох випадках наближений до нормального, без виразного домінування одного варіанта (рис. 1). Водночас, локус GB950 в нашому дослідженні (Mishchenko et al., 2026) мав помірну інформативність, хоча за результатами інших вчених (Akin et al., 2016) він був одним із найінформативніших локусів, що можна пояснити відмінностями у генетичній композиції зразків, взятих для аналізу в цих роботах.

Локус BR270 містить простий тринуклеотидний повтор (CTG)_n (табл. 1). У доступних послідовностях з бази даних GenBank кількість одиниць повтору варіює від 6 до 8, що відповідає алелям розміром 92–98 п.н. Відмінності в розмірі варіантів локусу зумовлені лише варіаціями у кількості одиниць повтору. У *C. avellana*, *C. chinensis* та *C. americana* переважають алелі з шістьма або вісьмома повторами, тоді як у

C. colurna та *C. heterophylla* частіше фіксуються проміжні варіанти з сімома повторами.

Довжина алелів BR270, знайдених у попередньому дослідженні (Mishchenko et al., 2026), варіювала у ширшому діапазоні 89–107 п.н., а розподіл їхніх частот характеризувався асиметрією з переважанням алелів 92 та 98 у рівному співвідношенні (рис. 1). Водночас алель 95 слабо представлений, що і зумовлює відхилення розподілу від нормального. Слід зазначити, що алель 95 відповідає повтору (CTG)₇, який відсутній в наявних у GenBank геномах *C. avellana* та *C. chinensis*.

Локус BR322 містить простий повтор (ACT)_n, для якого в геномах представників роду *Corylus* спостерігається відносно обмежена варіабельність кількості повторів (табл. 1). За даними GenBank, кількість мотивів варіює переважно від семи до восьми, рідше до дев'яти, що відповідає алельним варіантам з розміром 105–111 п.н. Повтори (ACT)₇ та (ACT)₈ зустрічаються переважно в зразках *C. avellana*, *C. chinensis* та *C. colurna*, тоді як у *C. americana* виявлено варіант (ACT)₉ з довжиною 111 п.н. В геномі одного зі зразків *C. heterophylla* знайдено варіант з нетиповою довжиною 107 п.н., поява якого зумовлена інсерцією двох нуклеотидів.

Алельний розподіл BR322 характеризується наявністю одного переважаючого алеля 105, тоді як інші варіанти зустрічаються рідше (рис. 1). Водночас, у вибірці українських сортів фундука було знайдено алель з довжиною 102 п.н., відсутній в GenBank. Характер частотного розподілу алелів добре узгоджується з простою структурою локусу та обмеженою кількістю можливих алельних станів, де більшість мутаційних подій зводиться до зміни числа одиниць повтору. Це зумовлює формування компактного алельного набору з домінуванням найбільш стабільного варіанта (Li et al., 2004; Ananda et al., 2014).

Локус GB949 містить перерваний повтор, сформований тринуклеотидним мотивом (TGG)_n, з вставкою тринуклеотиду TGA або більш складних фрагментів (табл. 1). Кількість повторів TGG у першому повторі варіює від чотирьох до семи, після чого розташований тринуклеотид TGA і додатковий повтор (TGG)₄. У одного зі зразків *C. colurna* ця ділянка представлена (TGG)₁₀, а у *C. americana* cv. Winkler, навпаки, зазнала значних змін у вигляді точкових мутацій в окремих одиницях першої частини повтору, що призвели до його переривання в кількох місцях.

Розмір знайдених варіантів варіює у діапазоні 147–162 п.н. Алельний розподіл GB949 у попередньому дослідженні наближений до нормального, без вираженого домінування одного варіанта, однак має нерівну верхівку з двох близьких за частотою алелів. Це можна пояснити складною перерваною організацією локусу.

Локус BR438 містить дві ділянки простих повторів, одна з яких сформувалася на основі динуклеотидного повтору (CT)_n, а друга містить повтор (TCA)_m (рис. 26). В геномах видів *Corylus* кількість повторів CT в цьому локусі становить 8–16, а кількість повторів TCA — 6–10. Розмір варіантів варіює, відповідно, в діапазоні 181–205 п.н. Внаслідок такої організації відмінності між варіантами локусу за довжиною не завжди кратні трьом. Основними мутаційними подіями для BR438 є зміна кількості копій тринуклеотиду TCA і динуклеотиду CT, які призводять до зміни співвідношення цих двох мотивів. Водночас на 5'-кінці повтору (CT)_n розташована ділянка з мінливою послідовністю, довжина якої варіює від 4 у *C. chinensis* (GTCT) до 12 нуклеотидів у *C. americana* cv. Rush ((GT)₃CA(GTCT)) (табл. 1). У більшості зразків тут розташований повтор (GTCT)₂. Другий простий повтор найчастіше представлений (TCA)₈. Виключення складають по одному зразку *C. avellana* та *C. chinensis*, у яких повтор подовжений до (TCA)₁₀ та (TCA)₉, а також зразки *C. colurna*, в яких цей локус навпаки містить скорочений варіант повтору (TCA)₆ або (TCA)₇. Для алельного розподілу BR438 у попередньому дослідженні (Mishchenko et al., 2026) характерне виражене домінування одного алеля (191) з частотою 0,767, що в кілька разів перевершує частоти всіх інших (рис. 1). Натомість, за даними аналізу доступних в GenBank нуклеотидних послідовностей домінуючим виявився алельний варіант 195, який знайдений в геномах *C. avellana* та *C. americana* cv. Winkler (табл. 1).

До складу локусу BR464 входить ділянка повторів, сформована тринуклеотидним мотивом (ATC)_n (табл. 1). У варіантах локусу, знайдених в базі даних GenBank, кількість повторів варіює від 7 до 12, відповідні варіанти локусу мають розмір 280–295 п.н. Спостерігається чітка видова диференціація варіантів локусу за кількістю повторів. У *C. avellana* переважають варіанти (ATC)₇ або (ATC)₉. У *C. colurna* виявлено проміжний варіант (ATC)₈, у *C. chinensis* — (ATC)₁₀, а у двох інших видів кількість копій тринуклеотиду ATC варіює від 8 до 12. Усі знайдені варі-

анти формують безперервний ряд фрагментів, довжина яких змінюється кратно трьома. Водночас, у досліджених українських сортів фундука (Mishchenko et al., 2026) один з алелів BR464 не вкладається у зазначений ряд, що ймовірно спричинено делецією або інсерцією, не пов'язаною з ділянкою повтору.

Загалом, більша частина досліджених локусів (BR259, BR270, BR322, BR414, BR464) містить один простий тринуклеотидний повтор і має консервативні прилеглі ділянки, в яких не виявлено делецій або інсерцій. Всі вони представлені безперервним рядом варіантів з регулярно зростаючим розміром, які за даними цього і попереднього досліджень мають наблизений до нормального розподіл. Сюди ж можна віднести також локус GB949, який хоча і містить перерваний повтор, але також представлений серією варіантів, різниця між розмірами яких кратна трьома. Розмір ПЛР-продуктів цих локусів зазвичай можна чітко пов'язати з кількістю простих повторів у їхньому складі. Інші три локуси (BR438, GB950, CAC-C008) містять комбіновані та перервані повтори. Крім того, в окремих варіантах цих локусів знайдено делеції або інсерції. Все це значно ускладнює інтерпретацію результатів їхнього ПЛР-аналізу, оскільки може призводити до формування ПЛР-продуктів однакового розміру, але з різною нуклеотидною послідовністю. Водночас внаслідок комбінованої структури, ці локуси мають більший набір варіантів, а отже й більшу інформативність для ідентифікації сортів.

Розподіл частот алельних варіантів за результатами аналізу нуклеотидних послідовностей, які було відсиквеновано нами, а також знайдених у базі даних GenBank, для більшості локусів подібний до розподілу частот алелів у вибірці українських сортів та трьох видів *Corylus*, досліджених раніше (Mishchenko et al., 2026). До їх числа зокрема увійшли ті, що містять один простий повтор. Для частини локусів кількість варіантів була подібною або більшою у вибірці українських сортів, хоча в цій роботі було проаналізовано додатково ще два види. В інших випадках, а саме для локусів з комбінованими повторами, таких як CAC-C008 та BR438, кількість виявлених алельних варіантів була більшою в цьому дослідженні.

Висновки

На основі аналізу даних сиквенування ПЛР-продуктів та послідовностей з бази даних NCBI GenBank визначено структурну організацію дев'яти мікросателітних локусів, які пропонуються як молекулярні маркери для ідентифікації українських сортів фундука. Показано, що більшість з цих локусів містять один простий тринуклеотидний повтор, мають консервативні прилеглі ділянки і представлені в геномах видів *Corylus* безперервним рядом алельних варіантів з розміром, який зростає відповідно до збільшення кількості повторюваних одиниць всередині локусу. Водночас три локуси (BR438, GB950, CAC-C008) містять комбіновані динуклеотидні та тринуклеотидні або ж перервані повтори, в них також знайдено делеції або інсерції, що впливають на розмір відповідних алельних варіантів. Вказані особливості структурної організації слід приймати до уваги при інтерпретації результатів ПЛР-аналізу алельних варіантів досліджених SSR-локусів у сортів фундука.

Фінансування. Дослідження виконано за часткової фінансової підтримки Фонду Сімонса (IMBG Simons Foundation Grant for Ukrainian institutions No. FI-PD-Ukraine-00017453; Andreev I. O., Mishchenko A. M.).

Перелік літератури

1. Akin M., Nyberg A., Postman J., Mehlenbacher S., Bassil N. V. A multiplexed microsatellite fingerprinting set for hazelnut cultivar identification. *European Journal of Horticultural Science*. 2016. Vol. 81(6). P. 327–338. doi: 10.17660/eJHS.2016/81.6.6.
2. Altschup S. F., Gish W., Miller W., Myers E. W., Lipman D. J. Basic Local Alignment Search Tool. *J Mol Biol*. 1990. Vol. 215(3). P. 403–410. doi: 10.1016/S0022-2836(05)80360-2.
3. Ananda G., Hile S. E., Breski A., Wang Y., Kelkar Y., Makova K. D., Eckert K. A. Microsatellite interruptions stabilize primate genomes and exist as population-specific single nucleotide polymorphisms within individual human genomes. *PLOS Genetics*. 2014. Vol. 10(7). P. 1–20. doi: 10.1371/journal.pgen.1004498.
4. Bhattarai G., Mehlenbacher S. A. *In silico* development & characterization of tri-nucleotide simple sequence repeat markers in hazelnut (*Corylus avellana* L.). *PLoS ONE*. 2017. Vol. 12(5). doi: 10.1371/journal.pone.0178061.
5. Botta R., Molnar T. J., Erdogan V., Valentini N., Torello Marinoni D., Mehlenbacher S. A. Hazelnut (*Corylus* spp.) breeding In: *Advances in Plant Breeding Strategies: Nut and Beverage Crops*. Cham, Springer International Publishing, 2019. pp. 157–219.
6. Bull L. N., Pabon-Pena C. R., Freimer N. B. Compound microsatellite repeats: Practical and theoretical features. *Cold Spring Harbor Laboratory Press*. 1999. Vol. 9. P. 830–838.
7. Buschiazzo E., Gemmell N. J. The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays*. 2006. Vol. 28. P. 1040–1050. doi: 10.1002/bies.20470.

8. Eckert K. A., Hile S. E. Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Molecular Carcinogenesis*. 2009. Vol. 48. P. 379–388. doi: 10.1002/mc.20499.
9. Ellegren H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 2000. Vol. 16(12). P. 551–558.
10. Ershadi A., Toolir J. F. Evaluation of genetic diversity of Iranian populations of hazelnut (*Corylus avellana* L.) using SSR markers. *Journal of Nuts*. 2022. Vol. 13(4). P. 273–288. doi: 10.22034/jon.2022.1955752.1162.
11. Freixas-Coutin J. A., An S., Postman J., Bassil N. V., Yates B., Shukla M., Saxena P. K. Development of a reliable *Corylus* sp. reference database through the implementation of a DNA fingerprinting test. *Planta*. 2019. Vol. 249(6). P. 1863–1874. doi: 10.1007/s00425-019-03131-4.
12. Gökirmak T., Mehlenbacher S. A., Bassil N. V. Characterization of European hazelnut (*Corylus avellana*) cultivars using SSR markers. *Genetic Resources and Crop Evolution*. 2009. Vol. 56(2). P. 147–172. doi: 10.1007/s10722-008-9352-8.
13. Gürçan K., Mehlenbacher S. A., Köse M. A., Balık H. I. Population structure analysis of European hazelnut (*Corylus avellana*). *Acta Horticulturae*. 2018. Vol. 1226. P. 87–92. doi: 10.17660/ActaHortic.2018.1226.12.
14. Li Y.-C., Korol A. B., Fahima T., Nevo E. Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution*. 2004. Vol. 21(6). P. 991–1007. doi: 10.1093/molbev/msh073.
15. Mishchenko A. M., Andreev I. O. Selection and optimization of the method for DNA isolation and purification from *Corylus* species for PCR analysis. *Faktori eksperimental'noi evolucii organizmiv*. 2023. Vol. 32(1). P. 53–58. doi: 10.7124/FEE0.v32.1535.
16. Mishchenko A. M., Andreev I. O., Hryshchenko N. V., Kravchenko S. A., Kunakh V. A. Assessment of the informativeness of SSR markers for the identification and differentiation of Ukrainian hazelnut varieties. *Cytology and Genetics*. 2026. Vol. 60(2). P. in press.
17. Öztürk S. C., Göktay M., Allmer J., Doğanlar S., Frary A. Development of simple sequence repeat markers in hazelnut (*Corylus avellana* L.) by next-generation sequencing and discrimination of Turkish hazelnut cultivars. *Plant Molecular Biology Reporter*. 2018. Vol. 36(5–6). P. 800–811. doi: 10.1007/s11105-018-1120-0.
18. Sambrook J., Fritsch E. R., Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2nd ed. Ed. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press. 1989.
19. Tanhuanpää P., Heinonen M., Bitz L., Rokka V. M. Genetic diversity and structure in the northern populations of European hazelnut (*Corylus avellana* L.). *Genome*. 2019. Vol. 62(8). P. 537–548. doi: 10.1139/gen-2018-0193.

Стаття надійшла до редакції 11.09.2025
прийнята до друку 27.09.2025

STRUCTURAL ORGANIZATION OF SOME MICROSATELLITE LOCI IN *CORYLUS* SPECIES

A. M. Mishchenko, I. O. Andreev, V. M. Mel'nyk,
V. A. Kunakh

Institute of Molecular Biology and Genetics,
National Academy of Sciences of Ukraine,
150 Akad. Zabolotnoho St., Kyiv, 03143, Ukraine
e-mail: i.o.andreev@imbg.org.ua

The **aim** of this study was to investigate the structural organization of nine microsatellite loci used for the characterization of Ukrainian hazelnut varieties, based on the analysis of PCR product sequencing data and sequences deposited in the NCBI GenBank database. **Methods.** PCR amplification with specific primers, cloning and sequencing of PCR products, bioinformatic analysis. **Results.** Four SSR loci were amplified and the obtained PCR products were cloned and sequenced. A search for allelic variants of nine SSR loci was also performed for five *Corylus* species. The structural organization and variability of the repetitive regions of these loci were investigated, and the frequency distribution of allelic variants was analyzed. Most loci have a conservative sequence and contain simple trinucleotide repeats; the number of repeat units determines the size of the corresponding variants. Three loci were found to have combined dinucleotide and trinucleotide repeats or interrupted repeats, and some of the variants were found to have deletions and insertions. These structural features determine the complex nature of the size distribution of these loci variants. **Conclusions.** The structural organization of nine simple sequence repeat (SSR) loci, which have been proposed for identifying Ukrainian hazelnut varieties, was identified. These features should be considered when interpreting the PCR analysis results.

Keywords: *Corylus*, SSR-markers, nucleotide sequence, genetic variation, molecular-genetic analysis.