УДК 575.21 + 577.2 https://doi.org/10.7124/FEEO.v39.1847

ПІДПАЛА О. В.✉
ЛУКАШ Л. Л.

Інститут молекулярної біології та генетики НАН України,
Україна, 03680, м. Київ, вул. Акад. Заболотного, 150
✉ *pidpala@ukr.net*

МОЗАЇЧНА СТРУКТУРА НУКЛЕОТИДНИХ ПОСЛІДОВНОСТЕЙ КАУЛІМОВІРУСІВ (*CAULIMOVIRUS*)

Мета. Проаналізувати нуклеотидні послідовності каулімовірусів роду *Caulimovirus* на наявність гомології до інтегрованих послідовностей параретровірусів та представників різних класів еукаріотних МГЕ. ***Методи***. Результати пошуку та ідентифікації МГЕ і ендогенних параретровірусних послідовностей здійснено за допомогою програми *CENSOR*. ***Результати***. Отримано дані біоінформатичного аналізу нуклеотидних послідовностей каулімовірусів роду *Caulimovirus* показав, що вони мають мозаїчну будову і переважно складаються із фрагментів ендогенних параретровірусів *Caulimovirus*-LSa, *Caulimovirus*-TOf, *Caulimovirus*-2_NS, SmelV, AeurV, *Caulimovirus*-MSa2, TABARE, VvinAV_sc2, а також фрагментів еукаріотних МГЕ. ***Висновки***. На основі одержаних результатів можна припустити, що еволюційний шлях Caulimoviridae може бути різнонаправленим, і їхні ендогенні послідовності можуть бути джерелом нових екзогенних епісомальних форм.

Ключові слова: каулімовіруси, ендогенні параретровіруси, еукаріотні МГЕ.

Ретроелементи охоплюють широке коло різноманітних нуклеотидних послідовностей від ретронів і ретроінтронів до ретротранспозонів, ретровірусів і параретровірусів (Hull, 1999; Hansen and Heslop-Harrison, 2004; Kunakh, 2013). Незважаючи на відмінність у їхній структурі та механізмах функціонування, найдревнішим компонентом у їхніх структурі є зворотня транскриптаза, яка і стала відправною точкою в еволюції цих елементів (Xiong and Eickbush, 1990). Найстарішою групою ретроелементів вважають бактеріальні ретрони, а наймолодшою групою є параретровіруси. Особливу увагу при дослідженні еволюційного шляху ретроелементів привертають ретроінтрони (Lambowitz and Belfort, 2015). Ймовірно, що перші ретроінтрони виникли у геномі бактерій, а у процесі утворення еукаріотних клітин вони увійшли до їхнього складу разом із мітохондріями і хлоропластами (Zimmerly et al., 2001). Є припущення, що саме бактеріальні ретроінтрони є ключовими елементами в еволюції складових елементів геному еукаріот і предками *non-LTR* ретротранспозонів (Lambowitz and Belfort, 2015; Garavís et al., 2013; Novikova and Belfort, 2017). На основі філогенетичного аналізу ретроелементів виявлено, що *non-LTR* ретротранспозони є прабатьками еукаріотних *LTR* ретротранспозонів, які у свою чергу дали початок кільком класам вірусів, включаючи ретровіруси і параретровіруси (Malik and Eickbush, 2001). Параретровіруси є дволанцюговими ДНК-вірусами, які використовують зворотну транскрипцію для реплікації, але, на відміну від ретровірусів, рідко інтегрують свій геном у геноми клітин-хазяїв. Вони об'єднують представників двох родин вірусів: *Hepadnaviridae* (у тварин) і *Caulimoviridae* (у рослин). Аналізуючи транспозонний ландшафт послідовностей бактеріальних ретроінтронів, поряд із різноманітними фрагментами еукаріотних мобільних генетичних елементів (МГЕ) нами виявлено фрагменти ендогенних каулімовірусів (Pidpala and Lukash, 2024). Метою даної роботи було проаналізувати нуклеотидні послідовності каулімовірусів роду *Caulimovirus* на наявність гомології до інтегрованих послідовностей параретровірусів та представників різних класів еукаріотних МГЕ.

Матеріали і методи

Інформацію про нуклеотидні послідовності екзогенних каулімовірусів одержано із Gen-

© 2026 Підпала О. В., Лукаш Л. Л. This work is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Bank і наведено у табл. 1. Пошук та ідентифікацію МГЕ і ендогенних параретровірусних послідовностей здійснено за допомогою програми *CENSOR* (https://www.girinst.org/censor/) (Kohany et al., 2006).

Результати та обговорення

Caulimoviridae – одна з п'яти родин вірусоподібних ретротранспозонів, що зустрічаються в еукаріотів, і єдина родина вірусів з дволанцюговим ДНК-геномом, яка інфікує рослини. На сьогодні за даними Міжнародного комітету з таксономії вірусів (ICTV) на основі форми віріону, організації геному, групи векторів і незначних варіацій у циклі реплікації родини *Caulimoviridae* виокремлюють 11 родів: *Badnavirus*, *Caulimovirus*, *Cavemovirus*, *Dioscovirus*, *Petuvirus*, *Rosadnavirus*, *Ruflodivirus*, *Solendovirus*, *Soymovirus*, *Tungrovirus* та *Vaccinivirus* (Teycheney et al., 2020). Віріони членів даної родини існують без оболонки та мають ізометричну або бацилоподібну форму. Вони складаються з нековалентно замкнутого кільцевого геному дволанцюгової ДНК (dsDNA) розміром приблизно від 6 до 9,8 т.п.н., який, подібно до ретровірусів, реплікується через проміжні РНК. Залежно від роду, до якого вони належать, геноми Caulimoviridae можуть містити від однієї (у *Petuvirus*) до дев'яти відкритих рамок зчитування (у *Soymovirus*). На відміну від ретровірусів, інтеграція в геном хазяїна не є обов'язковою фазою їхнього життєвого циклу. Тим не менш, їхня повна або часткова геномна ДНК може іноді захоплюватися геномом хазяїна під час так званого явища «нелегітимної рекомбінації» або негомологічного з'єднання кінців ДНК, яке може відбуватися під час соматичної репарації ДНК або мейотичної рекомбінації (Nasim and Dey, 2022) і, як і у випадку ретровірусів, вона стає стабільною та може успадковуватися вертикально, якщо вона інтегрується в зародкову лінію хазяїна.

Представники родини *Caulimoviridae* заражають широкий спектр однодольних і дводольних рослин, хоча члени одного виду мають, як правило, вузькі діапазони господарів. Деякі віруси передаються комахами-переносниками, а також віруси можуть передаватися при вегетативному розмноженні рослин. Географічний ареал багатьох *Caulimoviridae* широкий, вони можуть зустрічатися від тропіків до субарктичних регіонів. Симптоми, спричинені цими вірусами, різноманітні та залежать від вірусу, хазяїна та кліматичних умов. Загалом, переважають мозаїки, міжжилкові хлоротичні плями, відставання в рості, жовто-оранжеве знебарвлення листя та смуги.

Цікаво, що члени родів *Badnavirus*, *Caulimovirus*, *Cavemovirus*, *Petuvirus* і *Solendovirus* можуть мати як екзогенні епісомальні форми, так і ендогенні (вірусна ДНК інтегрована в ядерний геном господаря). Ендогенні вірусні елементи, схожі на існуючі *Caulimoviridae*, широко поширені в геномах судинних рослин (Kunakh, 2013), що свідчить про існування в минулому додаткових вірусів у родині, які зараз існують лише як викопні послідовності в геномі свого хазяїна. Інтеграція вірусної ДНК в геном не є обов'язковим кроком у циклах реплікації цих вірусів, а скоріше ДНК захоплюється ядерним геномом хазяїна шляхом негомологічного з'єднання кінців (також відомого як нелегітимна рекомбінація) (Vassilieff et al., 2023). Очевидно, що вірусні послідовності можуть стати резервуаром

Таблиця 1. Основна інформація про досліджувані каулімовіруси роду *Caulimovirus*

Повна назва екзогенного каулімовірусу	Скорочена назва	Господар	Номер у GenBank	Довжина послідовності п.н.	Кількість кодованих білків
Angelica bushy stunt virus	AnBSV	*Angelica dahurica*	NC_043523	8300	7
Atractylodes mild mottle virus	AMMV	*Atractylodes macrocephala*	NC_027924	8105	6
Carnation etched ring virus	CERV	*Dianthus caryophylinus*	NC_003498	7932	6
Cauliflower mosaic virus	CaMV	*Brassica oleracea*	NC_001497	8024	7
Dahlia mosaic virus	DMV	*Dahlia variabilis*	NC_018616	7916	8
Figwort mosaic virus	FMV	*Scrophularia nodosa*	NC_003554	7743	7
Horseradish latent virus	HrLV	*Cochlearia armoracia*	NC_018858	7954	8
Lamium leaf distortion virus	LLDV	*Lamium maculatum*	NC_010737	7713	6
Mirabilis mosaic virus	MiMV	*Mirabilis jalapa*	NC_004036	7857	7
Soybean Putnam virus	SPuV	*Glycine max*	NC_018505	7757	6
Strawberry vein banding virus	SVBV	*Fragaria vesca*	NC_001725	7876	6

генетичної мінливості екзогенного вірусу, що буде реалізовуватися шляхом рекомбінації чи активації (Staginnus and Richert-Pöggeler, 2006).

У даній роботі ми сконцентрувались на аналізі нуклеотидних послідовностей вірусів роду *Caulimovirus.* Програма CENSOR дозволяє не лише ідентифікувати фрагменти еукаріотних МГЕ, але і послідовності ендогенних каулімовірусів (рис.).

Представники роду *Caulimovirus* мають кільцеві та несегментовані геноми, що кодують від 6 до 8 білків, які необхідні для життєвого циклу та передачі вірусу. Серед основних білків: ORF1 – транспортний білок (Mov); ORF2 – фактор передачі попелиць (Atf); ORF3 – асоційований з віріоном білок (Vap); ORF4 – капсидний білок (CP); ORF5 – полімеразний поліпротеїн, що містить білки аспарагінової протеази (AP) і зворотної транскриптази (RT) з доменом РНКази H1 (RH1) та ORF6 – білок матриці тілець включення (Tav). Також деякі каулімовіруси кодують білки, функція яких ще невідома.

У досліджуваних послідовностях каулімовірусів виявлено від одного до чотирьох фрагментів восьми різних ендогенних послідовностей параретровірусів (EPRV): AeurV; *Caulimovirus*-2_NS; *Caulimovirus*-LSa; *Caulimovirus*-MSa2; *Caulimovirus*-TOf; SmelV; TABARE; VvinAV_sc2. Довжина одиночних фрагментів коливається від 30 п.н. до 6201 т.п.н., а загальний відсоток їхніх послідовностей становить від 32,36 до 78,50% довжини відповідних каулімовірусів (табл. 2). Рівні подібності коливаються від 67,00 до 93,55% (середнє значення – 74,68%).

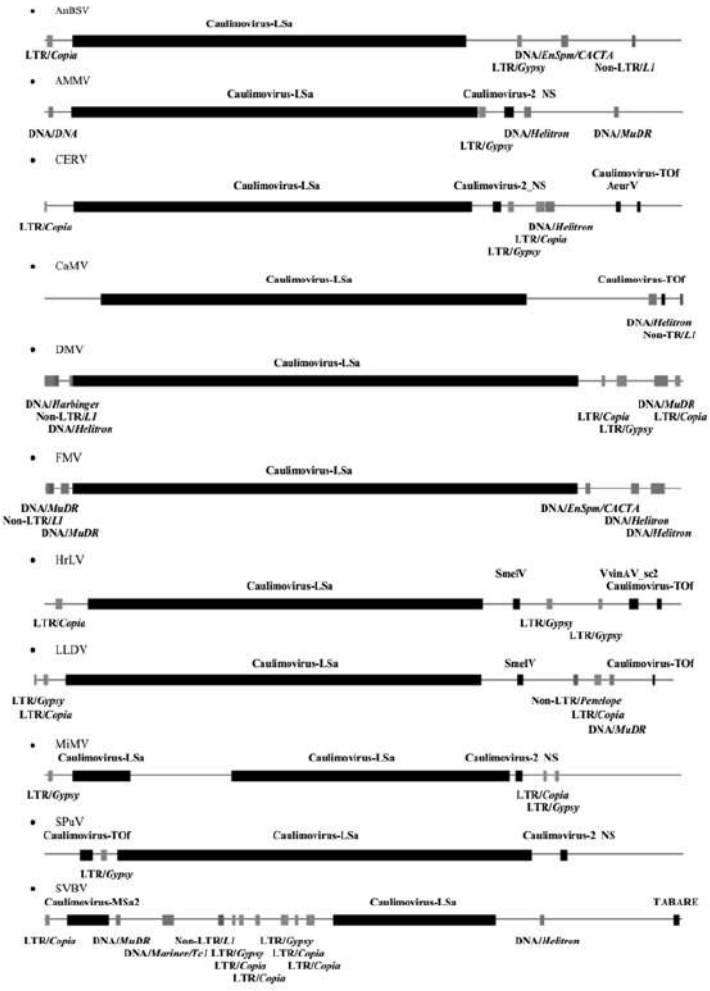


Рис. Фрагментарний профіль досліджуваних каулімовірусів.

Таблиця 2. Основна інформація про фрагменти інтегрованих вірусних послідовностей у досліджуваних каулімовірусів

Досліджувані каулімовіруси	Кількість фрагментів інтегрованих вірусних послідовностей	Комбінація послідовностей інтегрованих вірусних послідовностей	Їхній відсоток до загальної довжини вихідної послідовності каулімовірусу
AnBSV	1	LSa	61,37
AMMV	2	LSa + 2_NS	64,23
CERV	4	LSa + 2_NS+AeurV+TOf	64,02
CaMV	2	LSa + TOf	66,51
DMV	1	LSa	78,34
FMV	1	LSa	78,50
HrLV	4	LSa + SmelV + VvinAV_sc2 + TOf	64,07
LLDV	3	LSa + SmelV + TOf	65,87
MiMV	3	2 LSa + 2_NS	53,01
SPuV	3	TOf + LSa + 2_NS	67,05
SVBV	3	MSa2 + LSa + TABARE	32,36

Примітки: повні назви ендогенних вірусних послідовностей: LSa – *Caulimovirus*-LSa; 2_NS – *Caulimovirus*-2_NS; TOf – *Caulimovirus*-TOf; MSa2 – *Caulimovirus*-MSa2.

Найпоширенішими серед фрагментів EPRV в усіх проаналізованих каулімовірусів є послідовності *Caulimovirus*-LSa (LSa) – інтегрованого каулімовірусу *Lactuca sativa* (рис. 1, табл. 2). Ідентифіковано 1 або 2 фрагменти *Caulimovirus*-LSa у досліджуваних послідовностях каулімовірусів (табл. 2). Довжина одиночних фрагментів коливається від 690 п.н. до 6201 т.п.н., а загальний відсоток їхніх послідовностей становить від 25,20 до 78,50% довжини відповідних каулімовірусів. Варто зазначити, що фрагменти інтегрованого каулімовірусу *Caulimovirus*-LSa виявлено у послідовностях усіх 6 основних генів *Caulimovirus*. Частка послідовностей, гомологічних до *Caulimovirus*-LSa у різних генах коливалася від 0,19 до 100%. Найдовші фрагменти виявлено у послідовності гена, який кодує полімеразний поліпротеїн (від 92,66 до 100%). Цікаво, що у каулімовірусу *Figwort mosaic virus* фрагмент послідовності *Caulimovirus*-LSa ідентифіковано також у промоторній ділянці (553 п.н. – 54,92%). Якщо порівнювати довжину і гомологію фрагментів *Caulimovirus*-LSa у досліджуваних каулімовірусах відповідно до повнорозмірної послідовності інтегрованого каулімовірусу *Caulimovirus*-LSa (7519 п.н.), то варто зауважити, що вони охоплюють від 698 до 6173 п.н., що становить від 9,28 до 82,1% її повнорозмірної послідовності.

У послідовностях п’яти каулімовірусів (CERV; CaMV; HrLV; LLDV; SPuV) поряд із фрагментами інтегрованого каулімовірусу *Caulimovirus*-LSa ідентифіковано по одному фрагменту Caulimovirus-TOf-інтегрованого каулімовірусу *Taraxacum officinale* (рис. 1, табл. 2). Довжина даних фрагментів коливається від 30 п.н. до 147 п.н., а відсоток їхніх послідовностей варіює від 0,39 до 1,89% довжини відповідних каулімовірусів. Варто зауважити, що фрагменти послідовностей *Caulimovirus*-TOf ідентифіковано у некодованих ділянках трьох каулімовірусів CERV, CaMV і LLDV. У послідовності каулімовірусу HrLV фрагмент *Caulimovirus*-TOf виявлено у промоторній ділянці, а у послідовності каулімовірусу SPuV даний фрагмент *Caulimovirus*-TOf охоплював частину некодованої ділянки і промотор.

Поряд із іншими фрагментами інтегрованих каулімовірусів, у послідовностях чотирьох каулімовірусів (AMMV; CERV; MiMV; SPuV) ідентифіковано по одному фрагменту *Caulimovirus*-2_NS-інтегрованого каулімовірусу *Nicotiana sylvestris* (рис. 1, табл. 2). Довжина даних фрагментів коливається від 70 п.н. до 110 п.н., а відсоток їхніх послідовностей варіює від 0,90 до 1,36% довжини відповідних каулімовірусів. Цікаво, що фрагменти усіх чотирьох послідовностей *Caulimovirus*-2_NS ідентифіковано у гені, що кодує білок матриці тілець включення (Tav). Також у гені, що кодує білок матриці тілець включення у каулімовірусах HrLV і LLDV виявлено по одному фрагменту SmelV – інтегрованого флорендовірусу *Solanum melongena.* Їхня довжина становила відповідно 73 п.н. (0,92% довжини каулімовірусу HrLV) і 68 п.н. (0,88% довжини каулімовірусу LLDV).

У послідовності каулімовірусу *Carnation etched ring virus* (CERV), крім фрагментів інтегрованих каулімовірусів *Caulimovirus*-LSa, *Caulimovirus*-2_NS і *Caulimovirus*-TOf (табл. 2) ідентифіковано фрагмент флорендовірус-подібної послідовності *Aubrieta erubescens* AeurV довжиною 42 п.н.. Його ідентифіковано у некодуючій ділянці даного каулімовірусу, що складає 0,53% від його послідовності.

Фрагмент ендогенного флорендовірусу *Vitis vinifera* – VvinAV_sc2 довжиною 101 п.н. виявлено у промоторній ділянці послідовності каулімовірусу Horseradish latent virus (HrLV). Варто зазначити, що це четвертий фрагмент інтегрованого каулімовірусу у послідовності HrLV (табл. 2)

Ще два фрагменти інтегрованих каулімовірусів, окрім фрагмента *Caulimovirus*-LSa (табл. 2), зокрема інтегрованих каулімовірусів *Medicago sativa* – *Caulimovirus*-MSa2 та *Nicotiana tabacum* – TABARE ідентифіковано у послідовності каулімовірусу *Strawberry vein banding virus* (SVBV). Фрагмент *Caulimovirus*-MSa2 довжиною 508 п.н виявлено у послідовності гена, що кодує транспортний білок (Mov). Його довжина становить 51,47% даного гена і 6,45% повнорозмірної послідовності каулімовірусу SVBV. Що стосується фрагмента інтегрованого каулімовірусу TABARE довжиною 56 п.н., його ідентифіковано у некодуючій ділянці каулімовірусу SVBV і він охоплює 0,71% довжини даного каулімовірусу.

Отже, фрагментовані послідовності ендогенних параретровірусів присутні у генах усіх основних білків каулімовірусів (ORF1 + ORF2 + ORF3 + ORF4 + ORF5 + ORF6). Домінують фрагменти інтегрованих каулімовірусів *Caulimovirus*-LSa. У послідовностях шести досліджуваних каулімовірусів (AnBSV; AMMV; CERV; HrLV; LLDV; SPuV) вони охоплюють послідовності ORF1 + ORF2 + ORF3 + ORF4 + ORF5, а у трьох каулімовірусах (CaMV; DMV; FMV) вони гомологічні послідовностям генів усіх основних білків ORF1 + ORF2 + ORF3 + ORF4 + ORF5 + ORF6. Цікаво, що у каулімовірусу FMV фрагмент *Caulimovirus*-LSa зачіпає і промоторну ділянку. У двох випадках виявили фрагменти, які поєднують кілька генів. Зокрема, у випадку каулімовірусу MiMV це послідовності ORF1 + ORF2 і ORF4 + ORF5 + ORF6, а у каулімовірусу SVBV – ORF5 + ORF6.

У гені, що кодує білок матриці тілець включення (Tav), окрім фрагментів *Caulimovirus*-LSa, у чотирьох послідовностях каулімовірусів (AMMV; CERV; MiMV; SPuV) також виявлено фрагменти інтегрованого каулімовірусу *Caulimovirus*-2_NS та у двох послідовностях каулімовірусів (HrLV; LLDV) ідентифіковано фрагменти інтегрованого флорендовірусу SmelV. Також, у гені, що кодує транспортний білок (Mov), окрім фрагментів *Caulimovirus*-LSa, у двох послідовностях каулімовірусів (SPuV і SVBV) виявлено фрагменти інтегрованих каулімовірусів *Caulimovirus*-TOf і *Caulimovirus*-MSa2.

У промоторних ділянках досліджуваних каулімовірусів виявлено фрагменти інтегрованих каулімовірусів *Caulimovirus*-LSa (у FMV) і *Caulimovirus*-TOf та фрагмент ендогенного флорендовірусу VvinAV_sc2 (у HrLV). А у некодованих ділянках низки каулімовірусів ідентифіковано фрагменти флорендовірус-подібної послідовності AeurV (у CERV), фрагменти інтегрованого каулімовірусу *Caulimovirus*-TOf (у CERV, CaMV, LLDV і SPuV) та фрагмент інтегрованого каулімовірусу TABARE (у SVBV).

Поряд із фрагментами ендогенних параретровірусів, у послідовностях досліджуваних каулімовірусів, виявлено фрагменти мобільних генетичних елементів (МГЕ) (рис. 1, табл. 3). Вони присутні, переважно, у двох генах, що кодують транспортний білок (ORF1) і білок матриці тілець включення (ORF6) та у некодованих ділянках. Ідентифіковано фрагменти МГЕ різних класів/підкласів у різних комбінаціях: non-LTR + LTR + DNA (n = 4); non-LTR + DNA (n = 1); LTR + DNA (n = 3) або лише LTR (n = 3). Кількість фрагментів МГЕ коливається від одного до одинадцяти, а довжина – від 30 п.н. до 161 п.н., що становить від 0,63 до 7,67% загальної довжини відповідного каулімовірусу. Найбільше виявлено фрагментів послідовностей LTR ретроелементів (n = 26, із них 14 фрагментів надродини *Copia* і 12 фрагментів надродини *Gypsy*). Найменше ідентифіковано фрагментів non-LTR ретроелементів (n = 6). Вони належать до 2 надродин – *L1* і *Penelope* із перевагою *L1* ретроелементів. Фрагменти ДНК-транспозонів (n = 17), які виявлено у досліджуваних каулімовірусів, належать до 5 надродин ДНК-транспозонів: *En-Spm/CACTA, Harbinger, Helitron, Mariner/Tc1 і MuDR*. Переважають фрагменти *Helitron* (5 випадків) і *MuDR* (4 випадки).

Таблиця 3. Частка МГЕ у досліджуваних послідовностях

Каулімовірус	Кількість МГЕ, %			
	Загальна (n*)	ДНК-транспозони	LTR ретро-транспозони	Non-LTR ретро-транспозони
Angelica bushy stunt virus	2,48 (n = 4)	0,77	1,32	0,39
Atractylodes mild mottle virus	2,72 (n = 4)	1,99	0,73	–
Carnation etched ring virus	3,44 (n = 4)	1,09	2,34	–
Cauliflower mosaic virus	1,67 (n = 2)	1,15	0,52	–
Dahlia mosaic virus	7,06 (n = 7)	3,79	2,45	0,82
Figwort mosaic virus	5,51 (n = 6)	5,02	-	0,49
Horseradish latent virus	2,23 (n = 3)	–	2,23	–
Lamium leaf distortion virus	3,76 (n = 5)	0,75	2,42	0,58
Mirabilis mosaic virus	1,64 (n = 3)	–	1,64	–
Soybean Putnam virus	0,63 (n = 1)	–	0,63	–
Strawberry vein banding virus	7,67 (n = 11)	2,47	4,29	0,90

Отже, біоінформатичний аналіз нуклеотидних послідовностей каулімовірусів роду *Caulimovirus* показав, що вони мають мозаїчну будову і переважно складаються із фрагментів ендогенних параретровірусів а також фрагментів еукаріотних МГЕ. Стаття присвячена світлій пам'яті вірусологу і шановній людині Краєву В. Г.

Висновки

На основі одержаних результатів можна припустити, що еволюційний шлях *Caulimoviridae* може бути різнонаправленим, і їхні ендогенні послідовності можуть бути джерелом нових екзогенних епісомальних форм.

Інформація про фінансову підтримку, подяки: відсутня.

Конфлікт інтересів: автор Лукаш Л. Л. є членом редакційної колегії збірника наукових праць «Фактори експериментальної еволюції організмів». Інший автор заявляє про відсутність конфлікту інтересів.

References

1. Garavís M., González C., Villasante A. On the origin of the eukaryotic chromosome: The role of noncanonical DNA structures in telomere evolution. *Genome Biology and Evolution*. 2013. Vol. 5, № 6. P. 1142–1150. https://doi.org/10.1093/gbe/evt079.
2. Hansen C., Heslop-Harrison J. S. Sequences and phylogenies of plant pararetroviruses, viruses, and transposable elements. *Advances in Botanical Research*. 2004. Vol. 41. P. 165–193. https://doi.org/10.1016/S0065-2296(04)41004-0.
3. Hull R. Classification of reverse transcribing elements: A discussion document. *Archives of Virology*. 1999. Vol. 144, № 1. P. 209–213. https://doi.org/10.1007/s007050050498.
4. Kohany O., Gentles A. J., Hankus L., Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*. 2006. Vol. 7. P. 474. https://doi.org/10.1186/1471-2105-7-474.
5. Kunakh V. A. Mobile genetic elements and plant genome plasticity. Kyiv : Logos, 2013. 288 p.
6. Lambowitz A. M., Belfort M. Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiology Spectrum*. 2015. Vol. 3, № 1. https://doi.org/10.1128/microbiolspec.
7. Malik H. S., Eickbush T. H. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Research*. 2001. Vol. 11, № 7. P. 1187–1197. https://doi.org/10.1101/gr.185101
8. Nasim N., Dey N. Pararetroviruses: Plant infecting dsDNA viruses. *Plant Molecular Biology Reporter*. 2022. Vol. 40, № 1. P. 106–118. https://doi.org/10.1007/s11105-021-01294-7.
9. Novikova O., Belfort M. Mobile group II introns as ancestral eukaryotic elements. *Trends in Genetics*. 2017. Vol. 33. P. 773–783. https://doi.org/10.1016/j.tig.2017.07.009.
10. Pidpala O. V., Lukash L. L. Analysis of nucleotide sequences of bacterial retrointrons for the presence of homology to eukaryotic MGE. *Bulletin of the Ukrainian Society of Geneticists and Breeders*. 2024. Vol. 22, № 1–2. P. 4–9. https://doi.org/10.7124/visnyk.utgis.22.1-2.1683 [in Ukrainian]
11. Staginnus C., Richert-Pöggeler K. R. Endogenous pararetroviruses: two-faced travelers in the plant genome. *Trends in Plant Science*. 2006. Vol. 11 (10). P. 485–491. https://doi.org/10.1016/j.tplants.2006.08.008.
12. Teycheney P.-Y., Geering A. D. W., Dasgupta I., Hull R., Kreuze J. F., Lockhart B., Muller E., Olszewski N., Pappu H., Pooggin M. M., Richert-Pöggeler K. R., Schoelz J. E., Seal S., Stavolone L., Umber M. and ICTV Report Consortium. ICTV Virus Taxonomy Profile: Caulimoviridae. *Journal of General Virology*. 2020. Vol. 101. P. 1025–1026. https://doi.org/10.1099/jgv.0.001497.
13. Vassilieff H., Geering A. D. W., Choisne N., Teycheney P.-Y., Maumus F. Endogenous caulimovirids: fossils, zombies, and living in plant genomes. *Biomolecules*. 2023. Vol. 13 (7). P. 1069. https://doi.org/10.3390/biom13071069.

14. Xiong Y., Eickbush T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO Journal*. 1990. Vol. 9, № 10. P. 3353–3362. https://doi.org/10.1002/j.1460-2075.1990.tb07536.x.
15. Zimmerly S., Hausner G., Wu X. Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Research*. 2001. Vol. 29, № 5. P. 1238–1250. https://doi.org/10.1093/nar/29.5.1238.

PIDPALA O. V., LUKASH L. L.
Institute of Molecular Biology and Genetics of Natl. Acad. Sci. of Ukraine,
Ukraine, 03680, Kyiv, Akad. Zabolotnogo str., 150

MOSAIC STRUCTURE OF NUCLEOTIDE SEQUENCES OF CAULIMOVIRUSES (CAULIMOVIRUS)

Aim. To analyze the nucleotide sequences of *Caulimoviruses* of the genus *Caulimovirus* for homology to integrated sequences of pararetroviruses and representatives of different classes of eukaryotic MGEs. ***Methods.*** The results of the search and identification of MGE and endogenous pararetroviral sequences were performed using the CENSOR program. ***Results.*** It was obtained bioinformatic analysis data of the nucleotide sequences of *Caulimoviruses* of the genus *Caulimovirus* showed that they have a mosaic structure and mainly consist of fragments of endogenous pararetroviruses *Caulimovirus*-LSa, *Caulimovirus*-TOf, *Caulimovirus*-2_NS, SmelV, AeurV, *Caulimovirus*-MSa2, TABARE, VvinAV_sc2 as well as fragments of eukaryotic MGEs. ***Conclusions.*** Based on the results obtained, it can be assumed that the evolutionary path of Caulimoviridae may be multidirectional, and their endogenous sequences may be a source of new exogenous episomal forms.

Keywords: *Caulimoviruses*, endogenous pararetroviruses, eukaryotic MGEs.

Надходження статті: 30 квітня 2026 р.
Прийняття до друку після рецензування: 30 червня 2026 р.
Публікація: 30 листопада 2026 р.