

СПОНТАННІ ЗАМІНИ НУКЛЕОТИДІВ У ГЕНАХ Х-ВІРУСУ  
КАРТОПЛІ

З'ясування факторів, що визначають типи і темпи заміни нуклеотидів у геномах має важливе значення для вдосконалення розповсюджених моделей молекулярної еволюції, дослідження еволюційного процесу, розпізнавання кодуєчих і некодуєчих РНК, ідентифікації біологічно важливих мотивів таких як реплікаційні сигнали, регуляторні ділянки тощо. Однак, багато питань стосовно факторів, що впливають на заміни нуклеотидів у геномах вірусів рослин, все ще залишаються нез'ясованими.

В нашій роботі проведено порівняльний комп'ютерний аналіз заміни нуклеотидів у геномних сиквенсах десяти пар близько споріднених штамів Х-вірусу картоплі (ХВК) з однаковою довжиною геномів і генів.

Виявлено значне варіювання частоти заміни залежно від пар нуклеотидів і штамів, довжини та локалізації ділянок геномів і генів, а також типу заміни та їх кодонних позицій. Найбільші відношення частот заміни знайдено в різних парах нуклеотидів ( $u \rightarrow c/c \rightarrow g$ , 1240), кодонних позиціях (третья/друга, 17,3) та типах (транзиції/трансверсії, 5,8); найменші – у прямих і зворотних заміни нуклеотидів однієї пари (1,03-1,6). Частота трансверсій набагато менша, але значно більш варіабельна, ніж транзицій і не має чіткої залежності від довжини гена (нуклеотидної ділянки).

**Ключові слова:** Х-вірус картоплі, геномні сиквенси, заміни нуклеотидів, транзиції, трансверсії, комп'ютерний аналіз.

Дискусійність положення про провідну роль мутаційної мінливості в біологічній еволюції спонукає до глибокого і різнобічного вивчення типів, механізмів і властивостей хромосомних, геномних та генних мутацій. Особливу увагу вчених привертають мало вивчені спонтанні точкові мутації, що спричиняють заміни поодиноких амінокислот у білках та/або поодиноких нуклеотидів у генах [8, 15, 17].

Спонтанні заміни нуклеотидів та амінокислот проявляють широкий спектр дії на життєздатність, вірулентність, та ефективність передачі вірусу, а також на прояв симптомів ураження та вірусостійкості рослин [1, 2, 4, 5, 9, 12]. Проте дані щодо частоти цих мутацій [3, 10, 11, 14, 17] а також їх залежності від функціональної активності і рівня експресії генів, нуклеотидних контекстів та позицій у генах і кодонах дуже фрагментарні і суперечливі [7, 9, 13, 15, 16, 19]. У той же час дані про нуклеотидні заміни в наборах сиквенсів споріднених біологічних об'єктів мають особливо важливе значення для моделювання процесів і темпів еволюції [Т, 6, 14, 16, 17, 19], оцінки ймовірності та достовірності філогенетичних дерев [Т, 19], встановлення кодуєчих ділянок геномів [18] тощо.

Зважаючи на це, наша робота присвячена порівняльним дослідженням спонтанних нуклеотидних заміни у генах п'яти близько споріднених штамів Х-вірусу картоплі (ХВК). Порівняльні аналізи були проведені за типами заміни (еквівалентні, нееквівалентні, транзиції, трансверсії, одно-, дво- та трьохнуклеотидні) і місцем їх локалізації (пари штамів, гени, а також геномні, генні та кодонні позиції).

**Матеріали і методи.** *Модельні об'єкти.* В роботі використано 5 близько споріднених штамів Х-вірусу картоплі з однаковою довжиною геномів і генів, які відрізняються між собою лише за поодинокими замінами нуклеотидів (від 25,4 до 41,7 на 1000). Геномні сиквенси дослідних штамів були отримані з генетичного банку даних GenBank за такими номерами доступу: штамп ks (Korean strain) - AF373782.1; штамп rb (resistance-breaking strain) - M95516; штамп s1 - X05198.1; штамп os - AB056718.1; штамп x3 - D00344.1. Для порівняльного аналізу транзицій і трансверсій ці нуклеотидні заміни розподілені нами на типи, підтипи і варіанти на основі розповсюджених назви типів нуклеотидів, запропонованих Індійським університетом США: А, Т, U, G, C – стандартні нуклеотиди; R - пуринові (puRine, А або G); Y - піримідинові (pYrimidine, C, U або T); S - сильні (Strong, G або C) та W - слабкі нуклеотиди (Week, А, U або T).

*Методи досліджень.* Заміни нуклеотидів визначали в десяти парах комбінацій з п'яти штамів вірусу, зіставляючи сиквенси генів першого штаму з чотирма іншими (ks / os, ks / rb, ks / s1, ks / x3), другого штаму з трьома (os / rb, os / s1, os / x3), третього з двома (rb / s1, rb / x3)

і четвертого з п'ятим (s1 / x3). В усіх п'яти генах (Re, Cp, Tgb1, Tgb2 та Tgb3) кожної пари підраховували кількість еквівалентних і нееквівалентних однонуклеотидних замін в 1-ій, 2-ій та 3-ій позиціях кодонів, двохнуклеотидних замін – в 1 і 2, 1 і 3 та 2 і 3 позиціях, а також кількість трьохнуклеотидних замін.

Комп'ютерний аналіз нуклеотидних замін у вірусних генах проводили за власними маленькими вузько спеціалізованими комп'ютерними програмами (утилітами), написаними мовою QBasic для вирішення кожної конкретної задачі досліджень.

Випадкову збіжність позицій нуклеотидних замін моделювали на 213-нуклеотидних ділянках вірусних генів за генератором випадкових чисел у діапазоні 1-213, використовуючи функцію QBasic INT(RND\*213+1). Ці числа являють собою позиції випадкових мутацій у випадкових кодонних позиціях (першій, другій або третій). Частина з отриманих чисел, які при діленні на 3 дають залишки 1, 2 або 0 представляють позиції випадкових нуклеотидних замін в одній кодонній позиції: першій, другій чи третій, відповідно. Позиції випадкових замін у двох кодонних позиціях представляють числа з залишками від ділення на три рівними 1 або 2, 1 або 0 чи 2 або 0.

Для кожного варіанта кодонних позицій нуклеотидних замін генерували 100 повторностей дійсної кількості замін на ділянках кожного гена і вираховували середнє значення модельованих позицій, які збігаються на ділянках двох, трьох, чотирьох та п'яти генів чи мають різні значення на всіх ділянках.

Статистичну обробку даних (обчислення середніх значень та стандартних відхилень) проводили за вбудованими статистичними функціями прикладного програмного пакету Microsoft Excel 2002.

**Результати досліджень.** Кількість нуклеотидних замін у десяти пар штамів Х-вірусу картоплі. Сумарна кількість замін нуклеотидів у генах десяти пар штамів ХВК прямо пропорційна довжині гена (табл. 1). Її величина варіює від 182,2 у генах реплікази (Re) до 5,4 в генах транспортного білка (Tgb3). Кінцеві некодуючі ділянки геномів 3'UTR та 5'UTR ідентичні у всіх штамів вірусу (не мають нуклеотидних замін), а ідентичність генів має обернено пропорційну залежність від їх довжини і становить від 97,7 % (ген Tgb2) до 95,8 % (ген Re). Кількість замін на 1000 нуклеотидів пропорційна довжині гена і варіює від 41,7 (ген Re) до 25,4 (ген Tgb2).

Таблиця 1

**Заміни нуклеотидів у генах та нетрансляційних ділянках десяти пар штамів Х-вірусу картоплі**

Пари штамів	Кількість замін в генах та нетрансляційних ділянках:							
	Re	Cp	Tgb1	Tgb2	Tgb3	3'UTR	5'UTR	Середнє
ks / os	183	23	21	6	5	0	0	47,6
ks / rb	162	28	21	10	3	0	0	44,8
ks / s1	184	25	23	7	7	0	0	49,2
ks / x3	164	24	20	7	6	0	0	44,2
os / rb	211	30	25	9	4	0	0	55,8
os / s1	217	26	28	5	8	0	0	56,8
os / x3	199	23	23	5	5	0	0	51,0
rb / s1	206	25	25	12	6	0	0	54,8
rb / x3	178	24	20	12	5	0	0	47,8
s1 / x3	118	9	13	6	5	0	0	30,2
Сума замін	1822	237	219	79	54	0	0	1326
Середнє значення	182,2	23,7	21,9	7,9	5,4	0	0	132,6
Замін на 1000 нуклеотидів	41,7	33,3	32,2	22,7	25,4	0,0	0,0	155,2
Довжина гена	4371	711	681	348	213	71	84	837,8
Ідентичність генів, %	95,8	96,6	96,8	97,7	97,5	100	100	88,7

Серед десяти можливих комбінацій пар із п'яти дослідних штамів ( $5 \times 4 : 2 = 10$ ) найменшу спорідненість мають пари  $os / s1$ ,  $os / gb$  та  $gb / s1$ , у яких середня кількість замін на геном становить, відповідно, 56,8; 55,8 і 54,8. До найбільш споріднених пар штамів відносяться  $s1 / x3$  та  $ks / x3$ , що мають 30,2 і 44,2 нуклеотидних замін на геном. У пари  $s1 / x3$  найменша кількість замін (найбільша спорідненість) проявляється також у довгих генах (Re, Cp і Tgb1), однак за замінами нуклеотидів у гені Tgb2 найбільшу спорідненість мають пари  $os / s1$  та  $os / x3$ , а за геном Tgb3 найбільш спорідненими є штами  $ks$  і  $gb$  (лише 3 заміни на ген).

*Порівняння транзицій і трансверсій у генах п'яти штамів ХВК.* Серед 12-ти можливих комбінацій замін нуклеотидів:  $a \leftrightarrow g$ ,  $a \leftrightarrow u$ ,  $a \leftrightarrow c$ ,  $g \leftrightarrow u$ ,  $g \leftrightarrow c$  та  $u \leftrightarrow c$  ( $4 \times 3 = 12$ ) 4 варіанти замін відносяться до транзицій і 8 до трансверсій (табл. 2). Враховуючи назви типів нуклеотидів, наведені в розділі «Матеріали і методи», ми виділили 2 типи транзицій ( $w \rightarrow s$  та  $s \rightarrow w$ ) з чотирма варіантами ( $a \rightarrow g$ ,  $u \rightarrow c$ ,  $g \rightarrow a$ ,  $c \rightarrow u$ ), а також 2 типи трансверсій ( $y \rightarrow r$  та  $r \rightarrow y$ ) з чотирма підтипами ( $u \rightarrow g$ ,  $c \rightarrow g$ ,  $g \rightarrow u$  та  $g \rightarrow c$ ) і восьми варіантами ( $u \rightarrow a$ ,  $u \rightarrow g$ ,  $c \rightarrow a$ ,  $c \rightarrow g$ ,  $a \rightarrow u$ ,  $g \rightarrow u$ ,  $a \rightarrow c$ ,  $g \rightarrow c$ ). Порівняння виділених типів, підтипів і варіантів показує, що сумарна кількість транзицій (85,4 %) у 5,8 разів перевищує кількість трансверсій (14,6 %). Відсотки двох типів сумарних транзицій в усіх генах (44,5 і 40,9), а також двох типів (7,2 і 7,4) та двох підтипів трансверсій (4,4 і 4,2) майже однакові між собою, проте сумарна кількість трансверсій підтипу  $c \rightarrow g$  (4,4 %) в 1,6 разів перевищує кількість трансверсій підтипу  $u \rightarrow g$  (2,8 %). Отже, заміни цитозина на пуринові нуклеотиди значно частіші, ніж тиміна.

Таблиця 2

**Сумарна кількість транзицій і трансверсій у генах п'яти штамів Х-вірусу картоплі**

Гени	Типи замін нуклеотидів											
	Транзиції				Трансверсії							
					y → r				r → y			
	w → s		s → w		u → r		c → r		r → u		r → c	
a → g	u → c	g → a	c → u	u → a	u → g	c → a	c → g	a → u	g → u	a → c	g → c	
Re	375	463	378	384	18	23	51	6	16	40	54	14
Cp	29	68	51	55	6	0	14	1	3	0	7	3
Tgb1	23	52	23	47	7	14	18	4	10	8	13	0
Tgb2	18	21	6	18	1	0	4	2	0	0	7	2
Tgb3	9	16	9	14	0	0	6	0	0	0	0	0
Сума замін	454	620	467	518	32	37	93	13	29	48	81	19
% замін	18,8	25,7	19,4	21,5	1,3	1,5	3,9	0,5	1,2	2,0	3,4	0,8
	44,5		40,9		2,8		4,4		3,2		4,2	
					7,2				7,4			
	85,4				14,6							

За зменшенням сумарної частоти нуклеотидні заміни розподіляються у такий спосіб:  $u \rightarrow c$  (620 замін),  $c \rightarrow u$  (518),  $g \rightarrow a$  (467),  $a \rightarrow g$  (454),  $c \rightarrow a$  (3,9),  $a \rightarrow c$  (3,4),  $g \rightarrow u$  (2,0),  $u \rightarrow g$  (1,5),  $u \rightarrow a$  (1,3),  $a \rightarrow u$  (1,2),  $g \rightarrow c$  (0,8),  $c \rightarrow g$  (0,5). Такий розподіл свідчить про близьку кількість прямих і зворотних замін у всіх варіантах пар нуклеотидів. Максимальна величина відношення частот прямих і зворотних транзицій складає 1,2 ( $u \rightarrow c$  (620 замін),  $c \rightarrow u$  (518)), а трансверсій – 1,6 ( $g \rightarrow c$  (0,8),  $c \rightarrow g$  (0,5)). Найбільша величина відношення частот у варіантів транзицій становить 1,4 ( $u \rightarrow c$  (620),  $a \rightarrow g$  (454)), а серед трансверсій – 7,8 ( $c \rightarrow a$  (3,9),  $c \rightarrow g$  (0,5)). Отже, трансверсії мають набагато меншу, але значно більш варіабельну частоту, ніж транзиції.

*Визначення кількості нуклеотидних замін у семи кодонних позиціях п'яти генів ХВК.* Порівняння сумарної кількості нуклеотидних замін у різних генах п'яти штамів вірусу показує, що переважно їх більшість (від 71,2 до 100 %) складають однонуклеотидні еквівалентні заміни в третій кодонній позиції, які не призводять до замін амінокислот (табл. 3). Еквівалентні заміни в першій кодонній позиції генів Re становлять лише 5,1 % (78/1517), генів Cp – 3,0% (6/200), генів Tgb1 – 2,6 % (4/156), генів Tgb3 – 8,9 % (4/45), а в генах Tgb2 такі заміни відсутні. Двохнуклеотидні еквівалентні заміни в першій і третій кодонних позиціях зустрічаються лише в довгих генах Re (1,6 %, 24/1517) та Cp (2,0 %, 4/200). Одно-, дво- чи трьохнуклеотидні еквівалентні заміни у позиціях 2, 1 і 2, 2 і 3 та 1, 2 і 3 відсутні в усіх генах вірусу.

## Заміни нуклеотидів у різних позиціях кодонів у генах п'яти штамів Х-вірусу картоплі

Позиції замін у кодонах	Кількість еквівалентних замін у генах*:					Кількість нееквівалентних замін у генах*:				
	Re	Ср	T1	T2	T3	Re	Ср	T1	T2	T3
1	78	6	4	0	4	87	20	20	0	0
2	0	0	0	0	0	84	7	14	0	7
3	1415	190	152	79	41	37	0	4	0	0
1 і 2	0	0	0	0	0	24	0	6	0	0
1 і 3	24	4	0	0	0	32	8	8	0	0
2 і 3	0	0	0	0	0	26	2	8	0	2
1, 2 і 3	0	0	0	0	0	15	0	3	0	0
Сума замін	1517	200	156	79	45	305	37	63	0	9
Замін на 1000 нуклеотидів	34,7	28,1	22,9	22,7	21,1	7,0	5,2	9,3	0,0	4,2
% замін	83,3	84,4	71,2	100	83,3	16,7	15,6	28,8	0,0	16,7
	85,4					14,6				

\* - T1, T2, T3 – скорочені назви генів Tgb1, Tgb2 та Tgb3

Сумарний відсоток нееквівалентних замін нуклеотидів (14, 6) у 5,8 разів менший, ніж еквівалентних (85,4). Кількість нееквівалентних замін у різних генах варіює значно більше, ніж еквівалентних (від 0 до 9,3 на 1000 нуклеотидів) і не має прямої пропорційної залежності від довжини гена, яка чітко проявляється щодо еквівалентних замін.

*Розподіл нуклеотидних замін на ділянках генів.* Порівняння замін нуклеотидів на різних ділянках генів (від повної довжини до 1/30 частини) показало, що на великих ділянках однакового розміру кількість замін практично однакова. Так, у гені Re довжиною 4371 нуклеотид середня кількість замін складає  $182,2 \pm 29,4$ ; на половині гена (2035 нуклеотидів) –  $91,7 \pm 15,8$ ; на чвертині (1093 нуклеотидів) –  $46,4 \pm 10,4$ . З подальшим зменшенням довжини ділянок варіювання кількості нуклеотидних замін на них швидко збільшується (рис. 1). Так, на 146-нуклеотидних ділянках гена Re (1/30 частина гена) кількість замін варіює від  $15,1 \pm 11,4$  (ділянка 29) до  $93,5 \pm 29,6$  (ділянка 12), а на 21-нуклеотидних ділянках гена Tgb3 (1/10 гена) чітко проявляються зони з високою частотою мутацій (6, 7, 8), з низькою – (1, 4), а також стабільні ділянки (2, 3, 5, 9, 10). Для з'ясування природи варіювання частоти нуклеотидних замін у різних зонах генів (наявність так званих високо мутабельних гарячих точок чи випадкова локалізація мутацій на певних ділянках) ми дослідили розподіл позицій замін у гені Tgb3 (213 нуклеотидів), а також на 213-нуклеотидних 5'-кінцевих ділянках чотирьох інших генів ХВК.

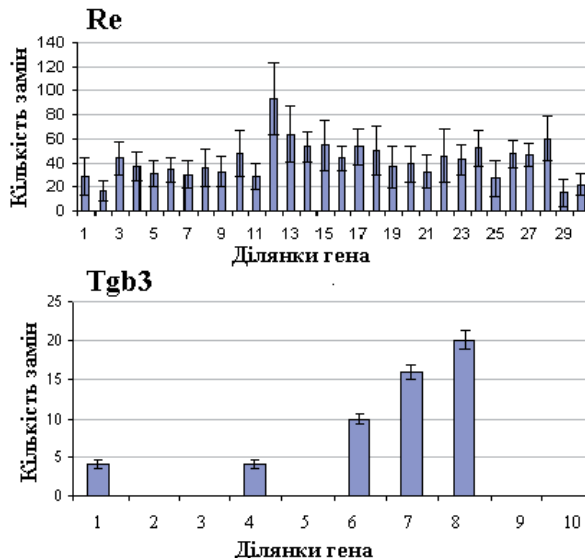
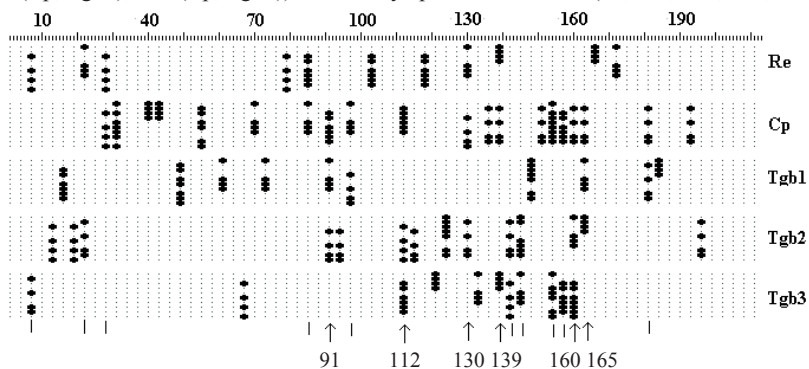


Рис. 1. Розподіл нуклеотидних замін на ділянках двох генів п'яти штамів Х-вірусу картоплі. Re – 30 ділянок гена реплікази довжиною по 145 нуклеотидів. Tgb3 – 10 ділянок гена транспортного білка довжиною по 23 нуклеотиди

Встановлено, що на досліджених ділянках заміни нуклеотидів локалізовані в 66 позиціях: 11 позицій у генах Re, 20 – у Cp, 10 – у Tgb1, 14 – у Tgb2 та 11 у гені Tgb3 (рис. 2). Більшість (28 із 66) позицій замін різні у різних типах генів, однак 10 з них збігаються у двох типах (позиції: 7 (гени Re, Tgb3), 22 (Re, Tgb2); позиції: 28, 85 (Re, Cp), 97 (Cp, Tgb1); 142, 145 (Tgb2, Tgb3); 154, 157 (Cp, Tgb3); 180 (Cp, Tgb1)) і шість – у трьох типах генів (91, 112, 130, 139, 160, 163).



**Рис. 2.** Позиції замін нуклеотидів в генах 5-ти штамів Х-вірусу картоплі. 10...190 – маркери позицій 213-нуклеотидних ділянок генів. Re – ген реплікази; Cp – ген білка оболонки; Tgb1, Tgb2, Tgb3 – гени транспортних білків. Вертикальні ряди по 10 точок – збіжність (світлі точки) і заміни нуклеотидів (темні кружки) в однакових позиціях генів у 10-ти пар штамів: ks/os, ks/s1, ks/rb, ks/y1, os/s1, os/rb, os/y1, s1/rb, s1/y1, rb/y1 (зверху вниз). ↑(92-164) – збіжні позиції замін нуклеотидів у трьох генах. | – збіжність позицій замін у двох генах

Моделювання випадкового збігання нуклеотидних замін у одній, двох і трьох кодонних позиціях, проведене шляхом генерування по 100 повторів 11, 20, 14 і 11 випадкових чисел, показало, що більшість позицій випадкових замін (25,7; 41,6 та 48,2) локалізовані лише в одному гені (у різних генах ці позиції не збігаються). Найбільша кількість незбіжних позицій (48,2) виявляється за індукування замін у трьох кодонних позиціях, найменша – у одній позиції (25,7). Кількість модельованих позицій випадкових нуклеотидних замін, збіжних у двох генах, варіює від 7,4 до 12,4; збіжних у трьох генах – від 1,0 до 5,2. Збіжність позицій випадкових замін у чотирьох генах зустрічається дуже рідко, а в усіх п'яти генах – його не виявлено в численних модельних дослідженнях. Порівняння наведених (фактичних) величин збіжності позицій нуклеотидних замін з отриманими методом комп'ютерного моделювання показує, що модельована збіжність найкраще відповідає фактичній за умов випадкових замін нуклеотидів в одній із кодонних позицій (першій, другій чи третій).

Таблиця 4

**Розподіл дійсних і модельованих позицій нуклеотидних замін на 213-нуклеотидних ділянках п'яти генів Х-вірусу картоплі**

Збіжність позицій замін*	Частота замін			
	дійсна	модельована за випадкових замін у кодонних позиціях:		
		в одній	у двох	у трьох
1	28	25,7 ± 2,4	41,6 ± 6,9	48,2 ± 3,7
2	10	12,4 ± 3,8	9,8 ± 3,0	7,4 ± 1,5
3	6	5,2 ± 2,0	1,7 ± 1,1	1,0 ± 1,2
4	0	0	0	0,1 ± 0,3
5	0	0	0	0
Сума замін	66	66,1	66,3	66,4

\* - Кількість генів, що мають позиції нуклеотидних замін, відсутні у інших генах.

**Обговорення результатів.** Для порівняльного аналізу спонтанних нуклеотидних замін у генах вірусів рослин нами відібрано 5 близько споріднених штамів ХВК з однаковою довжиною геномів і генів, що виключає можливість кількох типів мутацій у дослідних штамів (делецій, інсерцій, дублікацій, транслокацій тощо) і дозволяє аналізувати нуклеотидні заміни у чистому вигляді. Такий аналіз вбачається доцільним [8].

Всі порівняльні аналізи нуклеотидних замін за їх типами (еквівалентні, нееквівалентні, транзиції, трансверсії, одно-, дво- та трьохнуклеотидні) і місцем локалізації (пари штаблів, геноми, гени, генні та кодонні позиції) були проведені за єдиним показником – сумарною кількістю замін у генах десяти пар штаблів, що спрощує інтерпретацію результатів досліджень.

Оскільки аналізи нуклеотидних замін були проведені за власними вузько спеціалізованими комп'ютерними програмами, написаними для кожної конкретної задачі досліджень, коректність роботи програм постійно контролювалась нами за такими основними тестами: 1) програма виявляє всі заміни нуклеотидів, штучно зроблені в одній з двох копій довільного нуклеотидного сиквенсу і точно визначає їх генні та кодонні позиції; 2) типи і позиції нуклеотидних замін, знайдені програмно, відповідають таким у файлах сиквенсів; 3) результати аналізу, отримані за різними програмами, повністю узгоджуються між собою. Як приклад останнього тесту наводимо сумарну кількість нуклеотидних замін у генах Tgb3, яка дорівнює 54 за такими програмами: визначення замін у парах штаблів (табл. 1); аналіз транзицій і трансверсій (табл. 2:  $9 + 16 + 9 + 14 + 6 = 54$ ); визначення замін у кодонних позиціях (табл. 3:  $45 + 9 = 54$ ); визначення позицій замін (рис. 2: 54 темні кружки). Аналогічна однозначність даних проявляється також щодо всіх інших генів вірусу, що свідчить про коректність роботи всіх використаних комп'ютерних програм.

На додаток до численних відомих факторів, від яких залежить частота нуклеотидних замін [3, 7, 11, 16, 19], нами виявлена залежність її від довжини гена (табл. 1). Поряд із прямо пропорційною залежністю сумарної кількості замін від довжини гена, що цілком природно, подібна залежність чітко проявляється також щодо питомої величини: середньої кількості замін на 1000 нуклеотидів. Причиною цього явища може бути більша ймовірність спонтанних мутацій у великих генах, ніж у малих, через формування нуклеотидних контекстів, CpG-сайтів чи інших компонентів, що підвищують частоту нуклеотидних замін [19]. З іншого боку, оскільки мутації, як правило, шкідливі для вірусу [1], то більш ймовірною (менш шкідливою) буде заміна половини нуклеотидів у 10-нуклеотидному сайті, ніж у двохнуклеотидному.

Поряд із наведеною пропорційною залежністю проявляються також помітні відхилення від неї як у генах (частота замін у Tgb3 більша, ніж у Tgb2), так і в парах штаблів: пара s1 / x3 має більше замін у гені Tgb1, ніж у Ср. Ці дані узгоджуються з високою варіабельністю частоти точкових мутацій у вірусів рослин [3, 10, 14, 15].

Загальновідомо, що частота мутацій у некодуючих ділянках геномів значно більша за таку в функціональних генах, однак у 3'- і 5'-некодуючих ділянках геномів п'яти штаблів вірусу нами не виявлено жодної нуклеотидної заміни. Враховуючи малі розміри ділянок (72–84 нуклеотиди), це явище можна пояснити уже обговореною нами малою ймовірністю та великою шкідливістю мутацій у малих генах (нуклеотидних ділянках). Оскільки 3'- і 5'-некодуючі ділянки містять регуляторні елементи, важливі для репродукції вірусу, заміни нуклеотидів на цих ділянках не менш шкідливі, ніж у функціональних генах і тому не фіксуються.

Результати наших досліджень (табл. 2) узгоджуються з загальноизвестним положенням про превалювання частоти транзицій над трансверсіями, однак принципово відрізняються від співвідношення цих типів замін у псевдогенах коника [7]. Знайдене нами значне перевищення частоти транзицій у 5,8 разів (84,4/14,6) свідчить про наявність ефективних механізмів обмеження трансверсій у генах ХВК. Оскільки з 12 можливих комбінацій замін нуклеотидів 4 (33,3 %) відносяться до транзицій, а 8 (66,7 %) – до трансверсій, то за відсутності обмежень частота трансверсій повинна бути вдвічі більша, а не в 5,8 разів менша, ніж транзицій. Розподілення транзицій і трансверсій на типи, підтипи та варіанти вперше проведене нами дозволило виявити додаткові показники обмеження нуклеотидних замін. Так, частоти обох типів транзицій ( $w \rightarrow s / s \rightarrow w = 1,1$ ) і трансверсій ( $y \rightarrow r / r \rightarrow y = 1,02$ ) майже не відрізняються, однак частота підтипу  $c \rightarrow g$  у 1,6 разів більша, ніж  $u \rightarrow g$  (4,4/2,8). Питання про розповсюдженість цього явища та його біологічну роль залишається відкритим.

Встановлений нами розподіл нуклеотидних замін за кодонними позиціями (табл. 3) узгоджується з даними про їх розподіл у 15-ти зразках ортологічних генів людини та миші [16]. Найбільша кількість замін у третій кодонній позиції (94,0 % від еквівалентних замін і 9,9 % від нееквівалентних) пояснюється виродженістю генетичного коду. Оскільки кодони однієї амінокислоти (синонімічні кодони) відрізняються між собою лише за третім нуклеотидом, заміни у третій кодонній позиції не призводять до замін амінокислот і тому називаються нейтральними чи мовчазними мутаціями або трансляційно незалежними, синонімічними чи еквівалентними



нуклеотидними замінами. Через повне збереження структури і функцій білків за таких мутацій вони не видаляються добром.

Наявність незначної кількості еквівалентних замін у першій кодонній позиції (4,6 %) та подвійних (двохнуклеотидних) замін у першій і третій позиціях (1,4 %) пояснюється тим, що по 2 кодони двох амінокислот: аргініна (AGA/CGA, AGG/CGG) та лейцина (UUA/CUA, UUG/CUG) містять сайти еквівалентних замін не в третій, а в першій кодонній позиції. Цікаво, що дві з цих замін (C $\leftrightarrow$ T) є транзиціями, а дві інші (C $\leftrightarrow$ A) – найбільш частими трансверсіями (табл. 2). Оскільки 4 кодона з 61 складають 6,6 %, сумарний відсоток еквівалентних замін у першій та першій і третій кодонних позиціях (4,6+1,4) дуже близький до очікуваного (6,6), особливо з урахуванням поправки на меншу частоту трансверсій, ніж транзицій. Відсутність одно-, дво- та трьохнуклеотидних еквівалентних заміні у кодонних позиціях 2, 1 і 2, 2 і 3 та 1, 2 і 3 зумовлена структурою генетичного коду: всі синонімічні кодони однієї амінокислоти мають однакові два перших нуклеотиди, через що еквівалентні заміни нуклеотидів у першій і другій кодонній позиції неможливі, за винятком раніше наведених кодонів аргініна та лейцина. Цікаво, що 9 двохкодонних амінокислот у третій позиції мають або пуринові, або піримідинові нуклеотиди, що сприяє еквівалентним замінам і обмежує нееквівалентні через більшу ймовірність транзицій, ніж трансверсій. Сумарна кількість всіх транзицій і трансверсій в третій кодонній позиції досліджених штамів ХВК складає 79,6 % і перевищує кількість нуклеотидних замін у першій позиції у 8,7 разів, у другій – у 17,3; у першій і другій – у 66,3; у першій і третій – у 24,9; у другій і третій – у 49,8 і в усіх трьох позиціях – у 113,7 разів.

Розподіл нуклеотидних замін на ділянках генів досліджених штамів ХВК цілком узгоджується з висновком, зробленим за результатами аналізу мутацій у геномних сиквенсах приматів [19], однак принципово відрізняється від розподілу замін нуклеотидів у генах трьох видів тетрахеми [11]. В геномах і на великих ділянках генів ХВК частоти нуклеотидних замін кількісно подібні, що наводить на думку про однакову ймовірність мутацій на всіх ділянках генома. Розподіли замін на коротких ділянках, представлені на рисунках 1 та 2, не суперечать цьому припущенню, оскільки через малу кількість і випадкову локалізацію мутацій, у генах повинні зустрічатись короткі ділянки без замін, а також з малою і великою їх кількістю.

Найбільш інтригуючим результатом наших досліджень є шість випадків збіжності позицій нуклеотидних замін на однакових ділянках трьох різних генів (рис. 2), що можна інтерпретувати як наявність у всіх (зовсім різних за властивостями) генах вірусу високо мутабельних точок у позиціях 91, 112, 130, 139, 160 та 163 від початку гена. Перевірка такої можливості шляхом моделювання збіжності позицій замін на основі отриманих експериментальних даних (фактичної кількості замін на ділянках кожного гена) і припущення про однакову ймовірність заміни кожного нуклеотида ділянки показала, що результати моделювання найкраще відповідають експериментальним даним за умов програмної генерації замін лише в одній (будь-якій) із трьох кодонних позицій. В свою чергу дані моделювання узгоджуються з тим, що основну кількість всіх нуклеотидних замін складають еквівалентні заміни саме в одній (у третій) кодонній позиції. Переважна більшість нуклеотидних замін відбувається в одній позиції.

Оскільки наша модель не враховує обмеження типів замін у певних позиціях, з точки зору проведеного моделювання всі кодонні позиції є рівнозначними і мають однакові середні значення замін. При цьому модельовані позиції замін, які збігаються на ділянках двох чи трьох генів, не відповідають (і не повинні відповідати) дійсним позиціям за умов проведеного моделювання. Важливо лише те, що наша модель переконливо показує можливість випадкового збігу позицій нуклеотидних замін у двох, трьох і дуже рідко в чотирьох генах при наявності випадкових мутацій лише в одній кодонній позиції. Через те, що переважно більшість точкових мутацій становлять нуклеотидні заміни в третій кодонній позиції, збіжність позицій замін на ділянках двох-трьох генів може бути випадковою. В той же час на основі проведених досліджень ми не можемо стверджувати, що в геномах штамів ХВК відсутні ділянки зі статистично значимими відмінностями частоти мутацій. Для з'ясування цього питання потрібні подальші дослідження.

За результатами проведених досліджень можна вважати, що основним типом спонтанних точкових мутацій в геномах штамів ХВК є транзиції, спричинені помилками включення пуринових (G замість A, або A замість G), а також піримідинових нуклеотидів (U замість C, або C замість U) в процесі синтезу вірусних РНК.

Сумарна кількість всіх нуклеотидних замін у генах десяти пар штамів, а також їх кількість на 1000 нуклеотидів мають прямо пропорційну залежність від довжини гена. Кількість замін

на 1000 нуклеотидів у різних генах варіює від 41,7 до 25,4, однак кінцеві некодуючі ділянки геномів 3'UTR та 5'UTR ідентичні в усіх штамів вірусу.

За зменшенням сумарної частоти нуклеотидні заміни розподіляються у такий спосіб: u→c (620 замін), c→u (518), g→a (467), a→g (454), c→a (3,9), a→c (3,4), g→u (2,0), u→g (1,5), u→a (1,3), a→u (1,2), g→c (0,8), c→g (0,5). Такий розподіл свідчить про близьку кількість прямих і зворотних замін у всіх варіантах пар нуклеотидів, а також про набагато меншу, але значно більш варіабельну частоту трансверсій, ніж транзицій.

Переважає більшість нуклеотидних замін (79,6 % від загальної кількості) локалізується в третій кодонній позиції. В першій позиції кодонів їх частота менша в 8,7 разів, а в другій – в 17,3 рази.

В геномах ХВК і на великих ділянках однакових генів однакового розміру частоти нуклеотидних замін практично однакові, але в цих же генах зустрічаються короткі ділянки без замін нуклеотидів, а також з малою і великою їх кількістю. Моделювання розподілу нуклеотидних замін на 213-нуклеотидних ділянках свідчить про можливість випадкового збігу позицій нуклеотидних замін у двох, трьох і дуже рідко в чотирьох генах за умови випадкових мутацій лише в одній кодонній позиції.

*А. Н. Кириченко, І. С. Щербатенко*

*Институт микробиологии и вирусологии НАН Украины им. Д.К.Заболотного, Киев*

## **СПОНТАННЫЕ ЗАМЕНЫ НУКЛЕОТИДОВ В ГЕНАХ X-ВИРУСА КАРТОФЕЛЯ**

### **Резюме**

Выяснение факторов, определяющих типы и темпы замен нуклеотидов в геномах имеет важное значение для совершенствования распространенных моделей молекулярной эволюции, исследования эволюционного процесса, распознавание кодирующих и не кодирующих РНК, идентификации биологически важных мотивов таких как репликационные сигналы, регуляторные участки и тому подобное. Однако, многие вопросы относительно факторов, влияющих на замены нуклеотидов в геномах вирусов растений все еще остаются не выясненными.

В нашей работе проведен сравнительный компьютерный анализ замен нуклеотидов в геномных сиквенсах десяти пар близко родственных штаммов X-вируса картофеля (ХВК) с одинаковой длиной геномов и генов.

Выявлено значительное варьирование частоты замен в зависимости от пар нуклеотидов и штаммов, протяженности и локализации участков геномов и генов, а также типа замен и их кодонных позиций. Наибольшие отношения частот замен найдено в разных парах нуклеотидов (u→c/ c→g, 1240), кодонных позициях (вторая/третья, 17,3) и типах (транзиции / трансверсии, 5,8); наименьшие - в прямых и обратных замен нуклеотидов одной пары (1,03-1,6). Частота трансверсий намного меньше, но значительно более вариабельна, чем транзиций и не имеет четкой зависимости от длины гена (нуклеотидного участка).

**К л ю ч е с л о в а:** X-вирус картофеля, геномные сиквенсы, замены нуклеотидов, транзиции, трансверсии, компьютерный анализ

*A.N. Kyrychenko, I.S. Shcherbatenko.*

*Zabolotny Institute of Microbiology and Virology, National Academy of Sciences of Ukraine, Kyiv*

## **SPONTANEOUS NUCLEOTIDE SUBSTITUTION IN POTATO VIRUS X GENES**

### **S u m m a r y**

Elucidation of the factors underlying nucleotide substitution types and rates in genomes is of significance for improving the prevalent models for molecular evolution, and is essential in studying the evolution process, distinguishing the coding and noncoding RNAs, identifying biologically significant motifs such as replication signals, control regions and so on. However, many questions concerning the factors that affect the nucleotide substitutions in plant virus genomes still remain unclear.

In this paper a comparative computer analysis of nucleotide substitution in genomic sequences of 10 pairs of closely related potato virus X (PVX) strains having equal genome and gene length has been carried out.

A significant frequency variation of nucleotide substitutions, depending on nucleotide and strain pairs, length and localization of genome and gene regions as well as substitution types and their codon positions, was



found. The highest relations of substitution frequency are found in different nucleotide pairs (u→c/c→g, 1240), codon positions (third/second, 17.3) and substitution types (transitions/transversions, 5.8), the least ones – in forward and back substitutions in the same nucleotide pairs (1.03-1.6). The transversion frequencies are generally significantly lower but more variable than the transition one, and there is no direct relationship between gene (region) length and transversion frequency.

The paper is presented in Ukrainian.

**Key words:** potato virus X, genomic sequences, genomic substitutions, conservative sites, transitions, transversions, computer analysis.

**The authors address:** *Shcherbatenko I.S.*, Zabolotny Institute of Microbiology and Virology, National Academy of Sciences of Ukraine; 154 Acad. Zabolotny St., Kyiv, MSP, D 03680, Ukraine.

1. Carrasco P., de la Iglesia F., Elena S.F. The distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco etch virus // *J. Virol.* – 2007. – **81**, N 23. – P. 12979–12984.
2. Chowda-Reddy R.V., Sun H., Chen H., Poysa V., Ling H., Gijzen M., Wang A. Mutations in the P3 protein of soybean mosaic virus G2 isolates determine virulence on Rsv4-genotype soybean // *Mol. Plant Microbe Interact.* – 2011. – **24**, N1. – P. 37–43.
3. Grigoras I., Timchenko T., Grande-Perez A., Katul L., Vetten H.J., Gronenborn B. High variability and rapid evolution of a nanovirus // *J. Virol.* – 2010. – **84**, N 18. – P. 9105–9117.
4. Hasiow-Jaroszevska B., Borodynyo N., Jackowiak P., Figlerowicz M., Pospieszny H. Single mutation converts mild pathotype of the Pepino mosaic virus into necrotic one // *Virus Res.* – 2011. – **159**, N 1. – P. 57–61.
5. Hirata H., Lu X., Yamaji Y., Kagiwada S., Ugaki M., Namba S. A single silent substitution in the genome of Apple stem grooving virus causes symptom attenuation // *J. Gen. Virol.* – 2003. – **84**. – P. 2579–2583.
6. Jayaswal V., Jermini L.S., Poladian L., Robinson J. Two stationary nonhomogeneous Markov models of nucleotide sequence evolution // *Syst. Biol.* – 2011. – **60**, 1. – P. 74–86.
7. Keller I., Bensasson D., Nichols R.A. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes // *PLoS Genet.* – 2007. – **3**, N 2. – e22.
8. Lebre S., Michel C.J. A stochastic evolution model for residue Insertion-Deletion Independent from Substitution // *Comput. Biol. Chem.* – 2010. – **34**, N 5-6. – P. 259–267.
9. Lopez C., Aramburu J., Galipienso L., Soler S., Nuez F., Rubio L. Evolutionary analysis of tomato Sw-5 resistance-breaking isolates of Tomato spotted wilt virus // *J. Gen. Virol.* – 2011. – **92**. – P. 210–215.
10. Malpica J.M., Fraile A., Moreno I., Obies C.I., Drake J.W., Garcia-Arenal F. The rate and character of spontaneous mutation in an RNA virus // *Genetics.* – 2002. – **162**, 4. – P.1505–1511.
11. Moradian M.M., Beglaryan D., Skozylas J.M., Kerikorian V. Complete mitochondrial genome sequence of three tetrahymena species reveals mutation hot spots and accelerated nonsynonymous substitutions in Ymf genes // *PLoS ONE.* – 2007. – **2**, 7. – e650.
12. Moreno A., Hebrard E., Uzest M., Blanc S., Fereres A. A single amino acid position in the helper component of cauliflower mosaic virus can change the spectrum of transmitting vector species // *J. Virol.* – 2005. – **79**, 21. – P.13587–13593.
13. Nakken S., Rodland E.A., Hovig E. Impact of DNA physical properties on local sequence bias of human mutation // *Hum. Mutat.* – 2010. – **31**, N 12. – P. 1316–1325.
14. Pagan I., Firth C., Holmes E.C. Phylogenetic analysis reveals rapid evolutionary dynamics in the plant RNA virus genus tobamovirus // *Mol. Evol.* – 2010. – **71**, N 4. – P. 298–307.
15. Roossinck M.J., Schneider W.L. Mutant clouds and occupation of sequence space in plant RNA viruses // *Curr. Top. Microbiol. Immunol.* – 2006. – **299**. – P.337–348.
16. Wang G.Z., Chen L.L., Zhang H.Y. Phase-dependent nucleotide substitution in protein-coding sequences // *Biochem. Biophys. Res. Commun.* – 2007. – **355**, N 3. – P. 599–602.
17. Wang H. Confidence intervals for the substitution number in the nucleotide substitution models // *Mol. Phylogenet. Evol.* – 2011. – **60**, N 3. – P. 472–479.
18. Washietl S., Findeiss S., Muller S.A., Kalkhof S., von Bergen M., Hofacker I.L., Stadler P.F., Goldman N. RNACode: robust discrimination of coding and noncoding regions in comparative sequence data // *RNA.* – 2011. – **17**, N 4. – P. 578–594.
19. Zhang W., Bouffard G.G., Wallace S.S., Bond J.P. NISC comparative sequencing program. Estimation of DNA sequence context-dependent mutation rates using primate genomic sequences // *J. Mol. Evol.* – 2007. – **65**, N 3. – P. 207–214.

Отримано 14.09.2011