

**ВИБІРКОВІСТЬ КОДОНІВ І ЗАМІНИ НУКЛЕОТИДІВ У ГЕНАХ
ВІРУСУ КАРЛИКОВОСТІ СОЇ**

Проведено комп'ютерний аналіз вибірковості кодонів і спонтанних замін нуклеотидів у шести шта- мів вірусу карликовості сої.

Встановлено, що частота використання синонімічних кодонів у вірусних генах широко варіює за- лежно від гена, перекривання генів, кодона, послідовності перших двох кодонних нуклеотидів, однонук- леотидних контекстів, локалізованих перед кодоном і за ним, а також від вмісту GC (G+C) у генах і в третій позиції кодонів.

Під впливом перекривання генів загальна кількість нуклеотидних замін зменшується в 2,5 разів, кіль- кість еквівалентних замін у третій позиції кодонів – у 2,8–4,3 рази, вміст дикодонів у генах – у 1,4–1,6 разів. Поряд із цим спостерігається збільшення кількості нееквівалентних замін нуклеотидів у другій кодонній позиції, а також висока вибірковість кодонів ($F_{or} = 0,94-1,0$) і кореляція між вмістом нуклео- тидів у генах і третій позиціях кодонів ($r = 0,73-0,74$).

Одержані результати свідчать про наявність селекції дикодонів у вірусних генах.

Ключові слова: вірус карликовості сої, перекривання генів, вибірковість синонімічних кодонів, за- міни нуклеотидів.

Генетичний код містить 64 триплетні кодони, з яких 61 кодують 20 стандартних амінокис- лот, а 3 останніх виконують роль сигналів зупинки трансляції (термінальні кодони). Оскільки кількість кодонів втричі більша за кількість амінокислот, генетичний код є виродженим: три амінокислоти (Аргінін, Лейцин, Серин) мають по 6 кодонів; п'ять амінокислот (Аланін, Глі- цин, Пролін, Треонін, Валін) – по 4 кодони; одна амінокислота (Ізолейцин) – 3 кодони; дев'ять амінокислот (Аспарагін, Глютамін, Аспарагінова та Глютамінова кислота, Гістидин, Лізин, Тирозин, Фенілаланін, Цистеїн) – по 2 кодони і дві амінокислоти (Метіонін, Триптофан) ма- ють по 1 кодону.

Частота використання синонімічних кодонів, що кодують одну амінокислоту, значно варіює [2, 4, 6, 18]. Дуже часто для кожної амінокислоти існує переважний кодон, який ви- користовується найчастіше [16]. Це явище вибіркового використання кодонів, що отримало назву codon bias (нахил, схил, відхилення), залежить від нуклеотидного складу, довжини і рівня експресії гена; набору транспортних РНК, структурних властивостей кодованої аміно- кислоти, вторинної структури та стабільності мРНК, нуклеотидного вмісту в третій позиції кодонів тощо [1, 10, 12, 13, 14, 15, 17]. Дані стосовно кореляції вибіркової вибірковості кодонів із назва- ними факторами у різних представників еукаріотів, прокаріотів і вірусів дуже суперечливі, а вибірковість кодонів у генах, що перекриваються, та її залежність від нуклеотидних замін майже не досліджена.

Зважаючи на це, наша робота присвячена вивченню вибіркової вибірковості використання синоніміч- них кодонів і замін нуклеотидів у генах шести штамів вірусу карликовості сої (ВКС), які ма- ють 190-нуклеотидні ділянки перекривання генів капсидного і транспортного білків.

Матеріали і методи. *Модельні об'єкти.* В роботі використані геномні сиквенси шес- ти близько споріднених штамів вірусу карликовості сої (ВКС, Soybean dwarf virus, родина Luteoviridae, рід Luteovirus), отримані з генетичного банку NCBI за такими номерами досту- пу: штам YS ізолят M95-1 (ys1) – AB038147; YP M94-1 (yp1) – AB038148; DS HS97-8 (ds8) – AB038149; DP M96-1 (dp1) – AB038150; DS HS99-5 (ds5) – AB076038 та штам Tas-1 (tas) – L24049. (В дужках наведені скорочені назви дослідних штамів, подані в таблицях).

Дослідження проведені на 190-нуклеотидних ділянках чотирьох генів: ген транспортного білка (Tr), який має 190 нуклеотидів і повністю перекривається з геном капсидного білка (Ср) зі зсувом на 1 нуклеотид; 5'-кінцева ділянка гена Ср, що перекривається з геном Tr; ділянка гена Ср, локалізована за зоною перекривання та 5'-кінцева ділянка гена реплікази (Re).

Методи досліджень. Вибірковість синонімічних кодонів (frequency of optimal codons, For) визначали за відносною частотою їх використання [8], і виражали як відношення кіль- кості переважних кодонів до синонімічних.

Очікувану частоту кодонів визначали двома способами – середню очікувану та нуклеотиднозалежну очікувану. Середню частоту кодонів вираховували як середнє значення їх вмісту (кількість у гені поділену на 61), а середню нуклеотиднозалежну – за формулою: $EnF = Len/3 \times Fn1 \times Fn2 \times Fn3$, де EnF – очікувана нуклеотиднозалежна частота кодона, Len – довжина сиквенса (кількість нуклеотидів), $Fn1$, $Fn2$ та $Fn3$ – частота першого, другого та третього нуклеотида кодона [11].

Кількість замінів нуклеотидів визначали в 15-ти парах комбінацій шести штамів вірусу, з'являючи перший штам з п'ятьма іншими, другий з чотирма, третій з трьома, четвертий з двома і п'ятий з шостим. У трьох генах (Re , Tr , Cr) кожної пари підраховували кількість еквівалентних і нееквівалентних одонуклеотидних замінів в 1-ій, 2-ій та 3-ій позиціях кодонів, двохнуклеотидних замінів – в 1 і 2, 1 і 3 та 2 і 3 позиціях, а також кількість трьохнуклеотидних замінів.

Вміст нуклеотидів, кодонів та дикодонів в генах і кодонних позиціях визначали за власними короткими вузькоспеціалізованими комп'ютерними програмами (утилітами), написаними мовою QBasic для вирішення кожної конкретної задачі досліджень.

Статистичну обробку даних та кореляційний аналіз проводили за вбудованими статистичними функціями прикладного програмного пакету Microsoft Excel 2002.

Результати досліджень. *Вибірковість синонімічних кодонів у генах вірусу карликовості сої.* Порівняння переважних кодонів на 190-нуклеотидних ділянках вірусних генів показує, що вибірковість кодонів за перекривання генів значно вища, ніж без нього (рис. 1). Так, у гені Cr всі 18 амінокислот з синонімічними кодонами мають лише 1 переважний кодон (18 кружків на графіку, $For = 1$). В гені Tr , локалізованому в гені Cr зі зміщенням на 1 нуклеотид, лише глютамін (Q) має однакову частоту обох кодонів (caa , cag). For гена становить 0,94. На відміну від цього, 3 амінокислоти гена Re (L, T, C) і 4 амінокислоти гена $Cr1$ (K, H, G, E) мають по 2 рівночастотних кодона, а пролін (P) у гені $Cr1$ кодує три таких кодона (ccc , cct , cca). Отже, For гена $Cr1$ становить 0,83, а гена Re – 0,72.

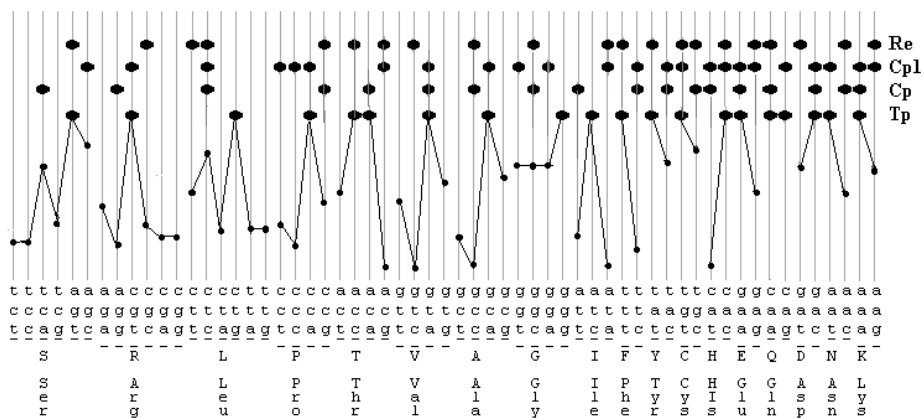


Рис. 1. Переважні кодони в чотирьох генах вірусу карликовості сої. $Re...Tr$ – 190-нуклеотидні ділянки генів реплікази (Re), капсидного білка ($Cr1$) та генів капсидного (Cr) і транспортного (Tr) білків, що перекриваються. $tct...aag$ – кодони стандартних амінокислот. S, Ser...K, Lys – амінокислоти

Типи переважних кодонів різних амінокислот значно варіюють у різних генах незалежно від наявності чи відсутності перекривання. Зокрема, agt є переважним кодоном серина в генах Tr і Re , кодон cgt – переважає синонімічні кодони в генах Tr і $Cr1$, а ctg – в генах Cr , $Cr1$ та Re . Найбільша варіабельність переважних кодонів виявлена у треоніна, у якого acc переважає три інші синонімічні кодони в генах Tr і Re , aca – в генах Tr і Cr , а acg – у генах $Cr1$ і Re .

Порівняння вибірковості кодонів за критерієм відхилення частоти їх використання від очікуваної показало, що алгебраїчна сума таких відхилень варіює від -6 до -21,3, а сума їх абсолютних значень – від 86,7 до 115,7. Сумарні відхилення від середнього значення, очікуваного за відсутності вибірковості кодонів, були дещо більшими (110–115,7) порівняно з очі-

куваними за процентним вмістом нуклеотидів (86,7 – 93,8), однак суттєвої різниці відхилень частоти кодонів від очікуваної на ділянках перекривання генів і поза ними не виявлено.

Цікаво, що відхилення частоти використання кодонів від очікуваної за процентним вмістом нуклеотидів майже однакове у генах Cp і Tr (88,8 і 93,5), а також у Cp1 і Re (86,7 і 93,8) незважаючи на те, що нуклеотидний склад генів Cp і Tr є ідентичним за рахунок їх перекривання, а генів Cp1 і Re – зовсім різний. Отримані дані свідчать про можливість вибіркової селекції кодонів за постійного кількісного співвідношення нуклеотидів.

Вибірковість стоп-кодонів зводиться до того, що кодуєча рамка містить лише один термінальний кодон, а дві інші рамки – різну кількість різних комбінацій трьох кодонів (tga, tag та taa), причому, одна з них більш насичена стоп-кодонами, ніж інша [11]. Наші дослідження показали, що кодуєча рамка гена Tr всіх шести штамів вірусу має найбільш сильний термінальний кодон taa. Наступна рамка зі зсувом +1 містить 74 стоп-кодони (23 taa, 15 tag і 36 tga), а остання рамка зі зсувом +2 зовсім не має термінальних кодонів, оскільки вона збігається з кодуєчою рамкою гена Cp. У свою чергу рамка Cp зі зсувом +1 має термінальний кодон taa (збігається з кодуєчою рамкою гена Tr), а рамка зі зсувом +2 збігається з рамкою гена Tr зі зсувом +1. За відсутності перекривання генів рамки зі зсувом +1 містять 48 (ген Cp1) і 43 стоп-кодони (ген Re), а рамки зі зсувом +2, відповідно – 90 і 120 термінальних кодонів. Стоп-кодон tga у цих генах зустрічається вдвічі частіше за tag і taa.

Таким чином, на ділянках перекривання генів всі позарамкові термінальні кодони зосереджені в одній некодуєчій рамці і їх кількість близька до середньої величини стор-кодонів у двох некодуєчих рамках генів, що не перекриваються. Ділянки перекривання генів мають більшу вибіркковість кодонів, ніж ділянки аналогічних генів поза зонами перекривання. За напрямком зменшення вибіркковості синонімічних кодонів гени вірусу карликовості сої утворюють наступну послідовність: Cp, Tr, Re і Cp1 і мають значення For 1; 0,94; 0,83 та 0,72, відповідно.

Заміни нуклеотидів у штамів вірусу карликовості сої. Визначали сумарну кількість еквівалентних і нееквівалентних нуклеотидних замін у семи кодонних позиціях (1, 2, 3, 1 і 2, 1 і 3, 2 і 3 та 1, 2 і 3) трьох генів шести штамів ВКС, а також середню кількість замін у семи кодонних позиціях 15-ти комбінацій пар шести дослідних штамів вірусу. Встановлено, що середня кількість замін у пар штамів варіює від 0,1 до 8,8 і суттєво відрізняється в різних генах (рис. 2). В генах Re і Cp вона значно більша за ту, яка спостерігається на ділянках перекривання генів.

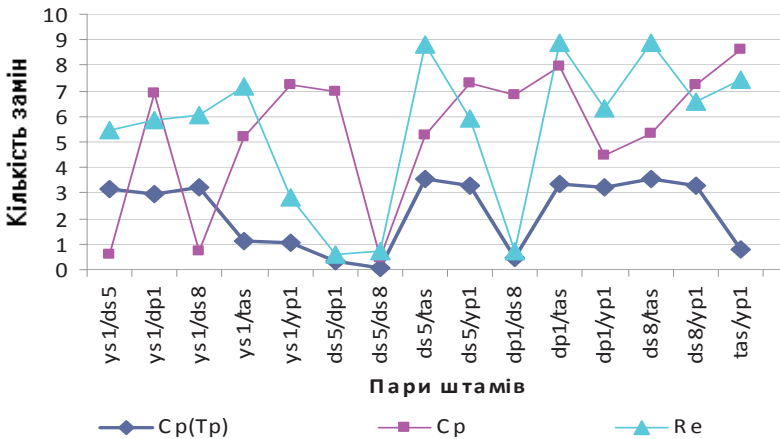


Рис. 2. Заміни нуклеотидів у генах шести штамів вірусу карликовості сої. Cp(Tr) – 190-нуклеотидна ділянка перекривання генів капсидного (Cp) і транспортного (Tr) білків. Cp, Re – 190-нуклеотидні ділянки генів капсидного білка та реплікази

Найбільшу сумарну кількість замін нуклеотидів складають еквівалентні одонуклеотидні заміни у третій кодонній позиції генів Cp і Re, незначну кількість – одно-, дво- та трьох нуклеотидні нееквівалентні заміни в позиціях 1, 2, 1 і 2, 1 і 3, 2 і 3 та 1, 2 і 3, а також еквівалентні заміни в позиціях 1 та 1 і 3. Еквівалентні заміни в «2-вмістних» позиціях (2, 1 і 2, 2 і 3 та 1, 2 і 3) зовсім відсутні, що і очікується за структурою генетичного коду: еквівалентні

(синонімічні) кодони у більшості випадків відрізняються між собою за третім нуклеотидом і лише аргінін, лейцин та серин мають по 2 кодони, відмінні за першим нуклеотидом.

В зоні перекривання генів Tr і Cr спостерігається різке зменшення еквівалентних заміни у кодонній позиції 3 і збільшення нееквівалентних у позиції 2. Причиною цього явища, очевидно, є збіг позиції 3 у гені Cr з позицією 2 у гені Tr через зсув рамки зчитування на 1 нуклеотид. Оскільки еквівалентні заміни в позиції 3 гена Cr, які становлять переважну більшість спонтанних нуклеотидних заміни, не є еквівалентними в позиції 2 гена Tr, загальна кількість нуклеотидних заміни на ділянках перекривання генів значно зменшується.

Досить значна кількість двох- та трьохнуклеотидних заміни у генах Cr і Re, очевидно, є наслідком кількох послідовних одонуклеотидних заміни, оскільки одночасна заміна двох і, особливо, трьох нуклеотидів в одному кодоні мало ймовірна. Відсутність двох- і трьохнуклеотидних заміни у гені Cr (на ділянці перекривання генів) спричинене різким зменшенням одонуклеотидних заміни на цій ділянці.

Залежність вибіркової кодони від контекстів. В таблицях генетичного коду синонімічні кодони, як правило, мають або всі 4 нуклеотиди в третій кодонній позиції або по 2 пуринові (a і g) чи піримідинові нуклеотиди (t і c) в третій і зрідка в першій позиціях. Звідси випливає, що вибірковість кодони фактично зводиться до помилкового включення в полінуклеотиди a замість g та t замість c і навпаки, а дані про залежність вибіркової кодони від контексту [1, 5] свідчать про значний вплив сусідніх нуклеотидів на реакцію полімеризації. Зважаючи на це, нами досліджено вплив на вибірковість кодони кількох контекстів: два перші нуклеотиди кодони, перший нуклеотид перед кодоном і за ним, а також комбінації дикодони.

Дослідження впливу двох перших нуклеотидів генетичних кодони на частоту їх використання в чотирьох генах шести штамів ВКС показало високу залежність частоти від послідовності нуклеотидів. Так, ga-початкові кодони (gap, де n будь-який нуклеотид) мають найбільший вміст у досліджених геномах (524), а кодони agn зустрічаються лише 284 рази. Така закономірність характерна для всіх кодони, що мають не однакові перші 2 нуклеотиди. Зокрема, частоти використання кодони tcn і ctn становлять, відповідно, 405 і 117, atn і tan – 352 і 159, can і acn – 327 і 250.

Нуклеотиди, розташовані безпосередньо перед і за кодонами (передкодонні та закодонні одонуклеотидні контексти), суттєво впливають на вибірковість кодони, що проявляється як збільшення частоти їх використання в 2–5 разів порівняно з очікуваною. Ефективність контекстів широко варіює залежно від їх позиції, гена і кодона. Так, передкодонний контекст C кодона act (Cact) збільшує частоту використання act в гені Re в 5,3 рази, а в гені Cr – в 3,3 рази. Контекст Cacc підвищує частоту кодона в гені Cr у 4,2 рази, а accA – в 3,8 разів. Впливу перекривання генів на активність контекстів не виявлено.

Частота використання (сумарна кількість повторів) дикодони у зоні перекривання генів Tr і Cr (233 і 268) суттєво відрізняється від такої в генах Re і Cr1 (327 і 452), має найбільшу величину в кодуєчій рамці (зсув 0) і дещо варіює в некодуєчих рамках (табл. 1). Із загальної кількості варіантів дикодони (61 x 61 = 3721) у генах зустрічаються лише від 6,3 до 12,1 відсотків можливих варіантів. Більшість з них (94,1 % у генах Tr і Cr, а також 99,2 % у генах Re і Cr1) повторюються від 1 до 6 разів, решта – від 7 до 11 (Re, Cr1) або від 7 до 25 разів (Tr, Cr). Найбільшу кількість повторів мають дикодони генів, що перекриваються: 25 – дикодон aagacg (ген Tr, зсув рамки +1; Cr, зсув +2), 21 – aagatc (ген Tr, зсув рамки +2), 18 – cagaca (ген Tr, зсув +2; Cr зсув 0). Наявність численних повторів дикодони підтверджена шляхом ідентифікації їх позицій у генах. Зокрема, дикодон aagatc містять всі 6 штамів у позиціях 295, 532, 550, а також 3 штами у позиції 502 (3 x 1 + 6 x 3 = 21).

Через збіг різних рамок за перекривання генів деякі позиції зчитування представлені в таблиці двічі (3051 і 3052). Ці позиції, а також позиції 3050 і 3053 являють собою ідентичні ділянки зони перекривання.

Вплив GC-вмісту генів на вибірковість кодони. Через перекривання генів Tr і Cr вміст GC (G + C) в їх кодонних позиціях має три значення – 31,0; 34,1 та 34,9 % замість шести (2 гена по 3 позиції). Подібне незначне варіювання вмісту GC (від 26,0 до 37,8 %) спостерігається також у шести кодонних позиціях генів Re та Cr1. Незважаючи на близький вміст GC в усіх позиціях кодони, проявляється чітка кореляція між вибірковістю кодони і відсотком GC-вмісту в третій кодонній позиції (GC3). Так, ген Cr, у якого кожна із 18 амінокислот кодується лише одним

із можливих синонімічних кодонів ($F_{or} = 1$), містить 34,9 % GC у третій кодонній позиції, а ген Tr, локалізований в гені Cr, має 17 амінокислот з одним синонімічним кодоном ($F_{or} = 0,94$) і 31,0 % GC3. Ген Re має більшу вибірковість кодонів порівняно з Cr1 (17 амінокислот з одним синонімічним кодоном проти 15-ти), а також більший вміст GC3 (35,2 % проти 28,5 %). Незважаючи на чітку кореляцію, залежність між F_{or} і GC3 у генах Cr, Tr, Cr1 і Re не є лінійною: 1; 0,94; 0,83; 0,72 та 34,9; 31,0; 35,2; 28,5 %, відповідно.

Таблиця 1

Повторність дикодонів у трьох відкритих рамках зчитування генів шести штамів вірусу карликовості сої

Повторність, крат	Зсув рамок зчитування у генах											
	Tr			Cr			Cr1			Re		
	0	+1	+2	0	+1	+2	0	+1	+2	0	+1	+2
1	39	31	35	35	39	31	195	178	156	130	112	114
2	27	21	24	24	27	21	71	73	66	72	65	49
3	84	80	89	88	84	80	62	53	44	114	98	86
4	8	6	11	11	8	6	62	60	52	31	20	18
5	15	9	11	11	15	9	26	24	25	20	26	22
6	88	71	81	81	88	71	31	26	23	36	37	35
7	2	0	2	2	2	0	3	2	5	4	2	0
8	1	0	2	2	1	0	0	0	0	1	1	1
9	5	8	3	4	5	8	1	2	2	4	4	2
10	1	1	3	3	1	1	0	1	1	0	0	0
11	0	1	1	1	0	1	1	0	0	0	1	0
12	3	2	3	3	3	2	0	0	0	0	0	0
14	1	0	1	0	1	0	0	0	0	0	0	0
15	1	2	0	1	1	2	0	0	0	0	0	0
18	0	0	1	1	0	0	0	0	0	0	0	0
21	0	0	1	1	0	0	0	0	0	0	0	0
25	0	1	0	0	0	1	0	0	0	0	0	0
Сума	275	233	268	268	275	233	452	419	374	412	366	327
Позиції*	3051	3052	3053	3050	3051	3052	3638	3639	3640	144	145	146

* Початкові позиції зчитування кодонів у генах штама Ys-1

Шляхом порівняння вмісту нуклеотидів у трьох кодонних позиціях чотирьох генів вірусу карликовості сої виявлено кореляції нуклеотидного вмісту в третій кодонній позиції з його вмістом у окремих генах, сумарним вмістом у чотирьох генах та вмістом у N1 кодонному контексті (табл. 2). Кореляція між вмістом нуклеотидів у третій кодонній позиції гена та в цьому гені більш висока і більш стабільна за перекривання генів ($r = 0,73$ і $0,74$), ніж без нього ($r = 0,43$ і $-0,53$), оскільки в останньому випадку кореляція не лише менша за рівнем, але й різна за знаком у генах Re і Cr1. Сумарний вміст нуклеотидів у чотирьох генах найбільш чітко корелює з їх сумарним вмістом у третій кодонній позиції генів ($r = 0,83$) і значно менше – з сумарним вмістом у кодонній позиції +3 ($r = 0,37$), яка відповідає позиції нуклеотида, наступного за кодоном, і має назву контекст N1 (від англійського next). Сумарний вміст нуклеотидів у кодонних позиціях 3 та +3 має слабку негативну кореляцію ($r = -0,18$), що свідчить про обмеження в генах таких пар кодонів, у яких третій нуклеотид першого кодона збігається з першим нуклеотидом другого.

Обговорення результатів. Проведені нами порівняння переважних кодонів на 190-нуклеотидних ділянках генів вірусу карликовості сої показали більш високу вибірковість кодонів за перекривання генів, ніж без нього (рис. 1), а також більшу вибірковість у гені Cr ($F_{or} = 1$), ніж Tr ($F_{or} = 0,94$). Подібне явище – домінування за вибірковістю кодонів одної з кодуючих рамок, що перекриваються, виявлено в геномах еукаріотів [9]. Широке варіювання видів переважних кодонів різних амінокислот, виявлене нами у різних генах ВКС незалежно від наявності чи відсутності перекривання рамок, показано також у 12 видів дрозофіли за відсутності перекривання генів [18].

Вміст нуклеотидів у генах і третій кодонних позиціях чотирьох генів вірусу карликовості сої

Гени	Кодонні позиції	Нуклеотиди				Коефіцієнт кореляції, r
		a	t	g	c	
Re	3	245	311	271	313	-0,53 ¹⁾
	1, 2, 3	897	882	839	802	
Cp1	3	322	346	213	259	0,43 ²⁾
	1, 2, 3	1075	777	754	814	
Cp	3	325	237	193	385	0,74 ³⁾
	1, 2, 3	1038	725	754	903	
Tp	3	393	233	284	230	0,73 ⁴⁾
	1, 2, 3	1038	725	754	903	
4 gena	3	1285	1127	961	1187	0,83 ⁵⁾
	1, 2, 3	4048	3109	3101	3422	0,37 ⁶⁾
	+3	1299	961	1315	961	-0,18 ⁷⁾

¹⁾ - ⁴⁾ Вміст нуклеотидів у третій кодонній позиції гена та в гені (у трьох позиціях). ⁵⁾ Сумарний вміст нуклеотидів у третій кодонних позиціях чотирьох генів та в чотирьох генах. ⁶⁾ Сумарний вміст нуклеотидів у кодонних позиціях +3 чотирьох генів та в чотирьох генах. ⁷⁾ Сумарний вміст нуклеотидів чотирьох генів у кодонних позиціях 3 та +3

Нами встановлено високу залежність частоти використання кодонів у геномах вірусу карликовості сої від послідовності першим двох нуклеотидів цих кодонів. Так, сумарна кількість у геномах шести штамів вірусу ga-початкових кодонів (аспарагінової – gat, gas і глутамінової кислоти – gaa, gag) становить 524, а ag-початкових (аргініна – aga, agg і серина – agc, agt) – 284; tc-початкових кодонів (серина – tca, tcc, tcg, tct) – 405, а ct-початкових (лейцина – cta, ctc, ctg, ctt) – 117; at-початкових кодонів (ізолейцина – ata, atc, att і метіоніна – atg) – 352, а та-початкових (тірозина – tat, tac і термінальних кодонів tag і taa) – 159. За винятком останнього варіанта, де низький вміст та-початкових кодонів порівняно з at-початковими зумовлений відсутністю термінальних кодонів tag, tga і taa у кодуєчих рамках, причина виявленої залежності залишається незрозумілою. Оскільки така залежність спостерігається в усіх шести можливих комбінаціях пар динуклеотидів (at-ta, ag-ga, ac-ca, gt-tg, tg-gt, gc-cg), з'ясування її причин представляє інтерес для вивчення механізмів вибіркості кодонів як основи точності та ефективності експресії генів.

Наші дані про значно меншу кількість нуклеотидних заміни на ділянках перекривання генів Cp1 і Tp (468) порівняно з такою в генах Re (1149) і Cp1 (1136) узгоджуються з даними про більш високу консервативність генів полеровірусів, що перекриваються, ніж за відсутності перекривання [7].

За даними літератури обидві некодуєчі рамки зчитування генів містять велику кількість термінальних кодонів, що забезпечує швидке припинення трансляції за мутацій зсуву рамки [11]. Кількість стоп кодонів має асиметричний розподіл між некодуєчими рамками і корелює з відсотком GC у генах. Відповідно до цього рамка Cp1 зі зсувом +1 містить 48 стоп-кодонів (16, 23, 9), зі зсувом +2 – 90 (20, 13, 57), а рамки Re – 43 і 120 кодонів, відповідно. Сумарна кількість стоп-кодонів у рамках Re (163) і Cp1 (138) корелює з вмістом GC у цих генах – 1641 і 1568, відповідно (табл. 2). За перекривання генів існує лише одна некодуєча рамка через збіг рамки Cp1 зі зсувом +2 з рамкою Tp зі зсувом +1.

Результати наших досліджень підтверджують дані про значний вплив контекстів на вибіркості кодонів, отримані на моделі генів еукаріотів [5]. Однуклеотидні контексти дуже широко варіюють за рівнем збільшення частоти використання кодонів порівняно з очікуваною залежно від гена, кодона і контекста, проте вплив позиції контекста відносно кодона (до чи після), а також вплив перекривання генів на активність контекстів нами детально не досліджено.

Нами підтверджено наявність високої вибіркості дикодонів у генах вірусу карликовості сої, встановленої іншими дослідниками на моделях бактерій, архей, еукаріотів [16] та

поліовірусу [3], а також встановлено високу залежність видів (варіантів) переважних дикодонів від перекривання генів (табл. 1). У зоні перекривання сумарний вміст дикодонів (233-280) значно менший, ніж у генах, що не перекриваються (327-452), однак зони перекривання містять більше переважних дикодонів. Деякі з них повторюються до 18-25 разів у генах шести штамів вірусу (зустрічаються 3-4 рази в кожному гені), чого не спостерігається поза зонами перекривання генів. Переважні пари кодонів не збігаються з очікуваними за кількістю можливих комбінацій таких пар. Так, дикодон треоніна асгасг зустрічається 25 разів, а кількість комбінацій дикодонів треоніна, який має 4 кодона, становить 16 (4 x 4). В той же час дикодон аргініна сгасга зустрічається лише 13 разів за наявності 36 комбінацій дикодонів у 6-кодонній амінокислоті (6 x 6). Наведені розрахунки відповідають максимальній кількості дикодонів (36), названій у роботі [3], однак вони вірні лише для дуплетів однакових кодонів. Загальна кількість дикодонів 6-кодонних амінокислот становить 324 (3 амінокислоти x 6 кодонів x 3 амінокислоти x 6 кодонів), а 4-кодонних амінокислот – 400 (5 x 4 x 5 x 4). За такого розрахунку частота асгасг (25) і сгасга (13) корелюють з числом комбінацій (400 і 324, відповідно), однак частота аагаатс (21), що кодує пару амінокислот лізин-ізолейцин не відповідає числу комбінацій дикодонів 2- і 3-кодонних амінокислот, що становить 108 (9 x 2 x 1 x 3 x 2).

Отже, висока частота деяких дикодонів у зонах перекривання генів є результатом селективного відбору, а не наслідком найбільшої ймовірності знаходження у геномах представників найбільш численних комбінацій дикодонів.

А.Н. Кириченко, О.І. Гордейчик, І.С. Щербатенко

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

ИЗБИРАТЕЛЬНОСТЬ КОДОНОВ И ЗАМЕНЫ НУКЛЕОТИДОВ В ГЕНАХ ВИРУСА КАРЛИКОВОСТИ СОИ

Резюме

Проведено комп'ютерний аналіз избирательности кодонов и спонтанных замен нуклеотидов в генах шести штаммов вируса карликовости сои. Установлено, что частота использования синонимических кодонов в вирусных генах широко варьирует в зависимости от гена, перекрывания генов, кодона, последовательности двух первых кодонных нуклеотидов, однонуклеотидных контекстов, локализованных перед кодоном и после него, а также от содержания GC (G+C) генах и в третьей позиции кодонов.

Под влиянием перекрывания генов общее количество нуклеотидных замен уменьшается в 2,5 раз, количество эквивалентных замен в третьей позиции кодонов – в 2,8–4,3 раза, содержание дикодонов в генах – в 1,4–1,6 раз. Наряду с этим наблюдается увеличение количества неэквивалентных замен нуклеотидов во второй кодонной позиции, а также высокая избирательность кодонов ($F_{op} = 0,94-1,0$) и корреляция между содержанием нуклеотидов в генах и в третьих позициях кодонов ($r = 0,73-0,74$). Полученные результаты свидетельствуют о наличии селекции дикодонов в вирусных генах.

Ключевые слова: вирус карликовости сои, перекрывание генов, избирательность синонимических кодонов, замены нуклеотидов.

A.N. Kyrychenko, O.I. Gordeychik, I.S. Shcherbatenko

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

CODON BIAS AND NUCLEOTIDE SUBSTITUTIONS IN SOYBEAN DWARF VIRUS

S u m m a r y

Computational analysis of codon usage bias and spontaneous nucleotide substitutions in six strains of soybean dwarf virus was performed.

It was shown that synonymous codon usage in the virus genes varies widely depending on the gene, gene overlapping, codon, codon's two first nucleotides, mononucleotide context located upstream and downstream of the codon, GC-content in virus-encoded genes and in the third codon position.

Overlapping of genes causes a 2.5-fold decrease of the total number of nucleotide substitutions, 2.8-4.3-fold decrease of the number of synonymous substitutions in the third codon position, 1.4-1.6-fold decrease of the dicodon content in genes. At the same time there is a significant increase of nonsynonymous substitutions in the second codon position as well as a high codon bias ($F_{op} = 0.94-1.0$) and correlation between the nucleotide

content in genes and in the third codon positions ($r=0.73-0.74$). The results obtained evidence for selection of dicodons in the viral genes.

The paper is presented in Ukrainian.

К е y w o r d s: soybean dwarf virus, gene overlapping, synonymous codon usage, nucleotide substitutions.

T h e a u t h o r ' s a d d r e s s: *Shcherbatenko I.S.*, Zabolotny Institute of Microbiology and Virology, National Academy of Sciences of Ukraine; 154 Acad. Zabolotny St., Kyiv, 03143, Ukraine.

1. *Carlini D.B.* Context-dependent codon bias and messenger RNA longevity in the yeast transcriptome // *Mol. Biol. Evol.* – 2005. – **22**, N 6. – P. 1403–1411.
2. *Chen R., Yan H., Zhao K.N., Martinac B., Liu G.B.* Comprehensive analysis of prokaryotic mechanosensation genes: their characteristics in codon usage // *DNA Seq.* – 2007. – **18**, N 4. – P. 269–278.
3. *Coleman J.R., Papamichail D., Skiena S., Futcher B., Wimmer E., Mueller S.* Virus attenuation by genome-scale changes in codon pair bias // *Science.* – 2008. – **320**, N 5884. – P. 1784–1787.
4. *D'Andrea L., Pinto R.M., Bosch A., Musto H., Cristina J. A.* Detailed comparative analysis on the overall codon usage patterns in Hepatitis A virus // *Virus Res.* – 2011. – **157**, N 1. – P. 19–24.
5. *Fedorov A., Saxonov S., Gilbert W.* Regularities of context-dependent codon bias in eukaryotic genes // *Nucleic Acids Res.* – 2002. – **30**, N 5. – P. 1192–1197.
6. *Fu M.* Codon usage bias in herpesvirus // *Arch. Virol.* – 2010. – **155**, N 3. – P. 391–396.
7. *Guyader S., Ducray D.G.* Sequence analysis of Potato leaf roll virus isolates reveals genetic stability, major evolutionary events and differential selection pressure between overlapping reading frame products // *J. Gen. Virol.* – 2002. – **83**. – P. 1799–1807.
8. *Ikemura T.* Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system // *J. Mol. Biol.* – 1981. – **151**. – P. 389–409.
9. *Lee S., Weon S., Lee S., Kang C.* Relative codon adaptation index, a sensitive measure of codon usage bias // *Evol Bioinform Online.* – 2010. – **6**. – P. 47–55.
10. *Lee Y., Zhou T., Tartaglia G.G., Vendruscolo M., Wilke C.O.* Translationally optimal codons associate with aggregation-prone sites in proteins // *Proteomics.* – 2010. – **10**, N 23. – P. 4163–4171.
11. *Rima B.K., McFerran N.V.* Dinucleotide and stop codon frequencies in single-stranded RNA viruses // *J. Gen. Virol.* – 1977. – **78**. – P. 2859–2870.
12. *Roymondal U., Das S., Sahoo S.* Predicting gene expression level from relative codon usage bias: an application to *Escherichia coli* genome // *DNA Res.* – 2009. – **16**, N 1. – P. 13–30.
13. *Shabalina S.A., Ogurtsov A.Y., Spiridonov N.A.* A periodic pattern of mRNA secondary structure created by the genetic code // *Nucleic Acids Res.* – 2006. – **34**, N 8. – P. 2428–2437.
14. *Sharp P.M., Li W.H.* The codon Adaptation Index - a measure of directional synonymous codon usage bias, and its potential applications // *Nucleic Acids Res.* – 1987. – **15**, N 3. – P. 1281–1295.
15. *Tatarinova T.V., Alexandrov N.N., Bouck J.B., Feldmann K.A.* GC3 biology in corn, rice, sorghum and other grasses // *BMC Genomics.* – 2010. – **11**. – P. 308–326.
16. *Tats A., Tenson T., Remm M.* Preferred and avoided codon pairs in three domains of life // *BMC Genomics.* – 2008. – **9**. – P. 463–478.
17. *Tuller T., Waldman Y.Y., Kupiec M., Ruppin E.* Translation efficiency is determined by both codon bias and folding energy // *Proc. Natl. Acad. Sci. U S A.* – 2010. – **107**, N 8. – P. 3645–3650.
18. *Vicario S., Moriyama E.N., Powell J.R.* Codon usage in twelve species of *Drosophila* // *BMC Evol Biol.* – 2007. – **7**. – P. 226–243.

Отримано 17.05.2011