

ХАРАКТЕРИЗАЦІЯ ГЕНІВ ТА ПРОТЕЇНІВ ВІРУСУ ІМУНОДЕФІЦИТУ ВЕЛИКОЇ РОГАТОЇ ХУДОБИ

Балак О. К.

Харківський національний медичний університет, Харків, Україна

Лиманська О. Ю.

Національний науковий центр «Інститут експериментальної і клінічної ветеринарної медицини», Харків, Україна, e-mail: olgaliman@ukr.net

Метою роботи було визначення дії природного відбору на вірус імунодефіциту великої рогатої худоби (ВІ ВРХ) через ідентифікацію поліморфізмів шляхом порівняння низки генів, відкритих рамок зчитування та протеїнів і аналіз впливу несинонімічних замін в протеїні Vif ізолятів ВІ ВРХ на конформаційні параметри Vif. Тиск відбору на гени оцінено за допомогою тесту Таджими, а також визначення співвідношення коефіцієнтів несинонімічних замін (K_a) до синонімічних (K_s) K_a/K_s . Коефіцієнти K_a та K_s , їхнє співвідношення (K_a/K_s) розраховано на основі вирівнювань амінокислотних та нуклеотидних послідовностей гена vif ізолятів ВІ ВРХ. Відношення частот несинонімічних (dN) до синонімічних (dS) замін на нуклеотидний сайт dN/dS для оцінки впливу відбору на ген vif розраховували з використанням методу SLAC на сервері Datamonkey. Кількість водневих зв'язків, α -спіралей, β -шарів, β -поворотів для третинних структур протеїнів визначено за допомогою сервера I-TASSER. Ідентифіковано гени, відкриті рамки зчитування (ORF) та регуляторну область U3 геному ВІ ВРХ, які перебувають під дією відбору. Гени gag, pol, s, vif, відкриті рамки зчитування ORF W та ORF Y перебувають під тиском негативного (очищуючого) відбору. На ген env та регуляторну область U3 діє позитивний відбір. Аналіз ентропії Шеннона (піків, які є специфічними до позицій амінокислотних залишків), що інтерпретують як поліморфізми, дозволив виявити 16, 8 та 4 несинонімічних замін для протеїнів Pol, Gag, Vif ВІ ВРХ відповідно. Через аналіз впливу чотирьох несинонімічних замін в протеїні Vif на конформаційні параметри Vif двох ізолятів ВІ ВРХ визначено істотні зміни числа водневих зв'язків, α -спіралей, β -шарів, β -поворотів. Показано різнонаправлений вплив відбору на гени ВІ ВРХ

Ключові слова: позитивний відбір, молекулярна еволюція, амінокислотна заміна, одноступінчастий поліморфізм

Вірус імунодефіциту великої рогатої худоби (ВІ ВРХ, bovine immunodeficiency virus, BIV) — лентівірус сімейства Retroviridae — вперше був ізольований у 1969 році, і з цього часу серопозитивних тварин виявляли у багатьох країнах світу. Вірус є причиною прогресуючих та персистентних інфекцій ВРХ, втрати ваги та зниження надоїв молока, виникнення вторинних інфекцій. Аналіз гематологічного профілю BIV-інфікованих тварин свідчить про лімфаденопатію, лімфоцитоз, дисфункцію нейтрофілів та моноцитів. Антигенно та генетично ВІ ВРХ є близьким до вірусів імунодефіциту людини та мавпи та, отже, може бути моделлю при вивченні інших ретровірусів [1].

Одним з основних факторів еволюції вірусів є природний відбір, який підтримує молекулярну адаптацію, тобто виникнення, накопичення та поширення корисних для реплікації та трансмісії мікроорганізмів мутацій (позитивний відбір), або направлений на елімінацію несприятливих генетичних змін (негативний, або очищуючий, відбір) [2].

Найчастішою молекулярно-генетичною подією в ході молекулярної еволюції є нуклеотидні заміни: синонімічні заміни (нейтральні мутації), які змінюють первинну структуру гена, але не підпадають під дію відбору на рівні білків, та несинонімічні заміни, результатом яких є зміни на всіх рівнях організації білка. Зміна первинної структури білка може впливати на конформаційні параметри молекули (наприклад, кількість α -спіралей та β -шарів), упакування просторової структури, функціональні та фізико-хімічні властивості білка.

Для дослідження молекулярної еволюції генів та відхилення від моделі нейтральної еволюції [3] розроблено низку статистичних методів [4], призначених як для порівняння коротких

послідовностей (наприклад, тест Таджими [5]), так і для аналізу великих масивів даних у масштабі геномів (наприклад, метод оцінювання відносного часу дивергенції — RelTime (relative time) метод [6]).

Для вивчення механізму еволюції білків важливим є питання дії відбору на ті чи інші заміни амінокислот, які можуть бути консервативними або радикальними залежно від їх впливу на фізико-хімічні властивості амінокислоти (зокрема, заряд, полярність). Радикальна заміна веде до значної зміни структури та функції протеїну. Цей термін є вельми відносним, оскільки існує велика кількість визначень радикальних замін, які суперечать одне одному.

Для визначення тиску відбору застосовують співвідношення швидкостей радикальних та консервативних амінокислотних замін [7, 8], яке є прямо пропорційним співвідношенню частот несинонімічних (dN_a) та синонімічних (dN_s) нуклеотидних замін, що до теперішнього часу широко використовують для визначення типу відбору при вивченні еволюції генів, кодуєних білки [9].

Стимування багатьох інфекційних захворювань ускладнено виключною здатністю патогенів адаптуватися шляхом еволюційних змін [10]. Так, розвиток резистентності до противірусних препаратів вірусу імунodefіциту людини є прикладом паралельної еволюції [11]. Великий розмір популяцій та висока швидкість мутування багатьох патогенів, в тому числі, ретровірусів, дозволяє їм ефективно уникати тиску імунної системи [12].

GenBank містить інформацію щодо 3 ізолятів ВІ ВРХ з повним геномом (нуклеотидні послідовності двох з яких (NC_001413 та M32690) є ідентичними), відомості щодо яких надано до цієї бази даних з 1990 по 2006 рік. Відсутність нових повногеномних сиквенсів ізолятів ВІ ВРХ спонукали автора порівняти існуючі біомолекули ізолятів цього вірусу та спробувати відповісти на запитання, чому за останні понад 15 років в GenBank не депоновано жоден ізолят ВІ ВРХ з повним геномом.

В даній роботі визначено дію природного відбору на ВІ ВРХ через ідентифікацію мутацій шляхом порівняння послідовностей низки генів, відкритих рамок зчитування та протеїнів та проведено порівняльний аналіз впливу несинонімічних замін в протеїні Vif ізолятів ВІ ВРХ на конформаційні параметри цієї молекули.

Матеріали і методи. Нуклеотидні послідовності ізолятів ВІ ВРХ з повним геномом (таксономічний ідентифікатор (txid) 11657) отримано з бази даних GenBank (www.ncbi.nlm.nih.gov/genbank).

Вирівнювання послідовностей ВІ ВРХ проведено за допомогою програми Molecular Evolutionary Genetics Analysis (MEGA) (версія 6.06) [13]. Програму BioEdit (версія 7.2.5) [14] використано для маніпулювання з нуклеотидними та амінокислотними послідовностями, мутаційного аналізу протеїнів, побудови графіків ентропії Шеннона. Тиск відбору на гени оцінено через проведення тесту Таджими на нейтральність, який вбудовано в програму MEGA 6.

Коефіцієнти несинонімічних замін (K_a) та синонімічних замін (K_s), їхнє співвідношення (K_a/K_s) розраховано на основі вирівнювань амінокислотних та нуклеотидних послідовностей гена *vif* ізолятів ВІ ВРХ. Для цього спочатку обчислено загальне число несинонімічних відмінностей для пари послідовностей, що порівнюють, - реальних N_a (реальн.) та теоретично можливих N_a (теор.) несинонімічних замін на даному фрагменті. K_a дорівнює співвідношенню N_a (реальн.)/ N_a (теор.). Таким же чином визначено K_s .

Крім того, додатково для порівняння гомологічних послідовностей гена ізолятів ВІ ВРХ розраховували відношення частот несинонімічних (dN) до синонімічних (dS) замін на нуклеотидний сайт dN/dS . $dN/dS > 1$ вказує на позитивний відбір, $dN/dS \approx 1$ — немає відбору, тобто нейтральна еволюція, $dN/dS < 1$ — очищуючий відбір. Оцінку впливу відбору на ген *vif/vif* проводили з використанням методу SLAC (single likelihood ancestor counting) на сервері Datamonkey (www.datamonkey.org) [15].

Нуклеотидні послідовності генів та відкритих рамок зчитування ВІ ВРХ (ORF, послідовностей нуклеотидів, які не мають стоп-кодону після стартового кодону (сайту ініціації трансляції) та потенційно можуть кодувати протеїни) отримано з повновимірних геномів 4 ізолятів ВІ ВРХ.

Кількість водневих зв'язків, α -спіралей, β -шарів, β -поворотів для третинних структур протеїнів визначено за допомогою сервера I-TASSER (Iterative Threading Assembly Refinement, <http://zhanglab.ccmb.med.umich.edu/I-TASSER>) за стандартних параметрів [16].

Результати роботи. Проведеним пошуком по базі даних GenBank визначено наступна кількість повногеномних сиквенсів для лентівірусів за період з 01.01.2006 р. по 28.09.2023 р.: вірус артриту-енцефаліту кіз — 65, вірус інфекційної анемії коней — 32, вірус вісна-маєді — 4, вірус хвороби Джембрана — 0, ВІ ВРХ — 0.

За час 01.01.1990 р. по 31.12.2006 р. до GenBank депоновано 2 варіанти одного і того ж ізоляту ВІ ВРХ (НХВ3) з повним геномом, нуклеотидні послідовності яких збігаються, за різними номерами (NC_001413 та M32690). Програми для біоінформатичного аналізу, наприклад, SLAC на сервері Datamonkey, враховують тільки один з зазначених вище варіантів ВІ ВРХ.

За допомогою аналізу ентропії Шеннона ідентифіковано 4, 8 та 16 амінокислотних варіабельних сайтів для протеїнів Vif, Gag та Pol ВІ ВРХ відповідно (рис. 1).

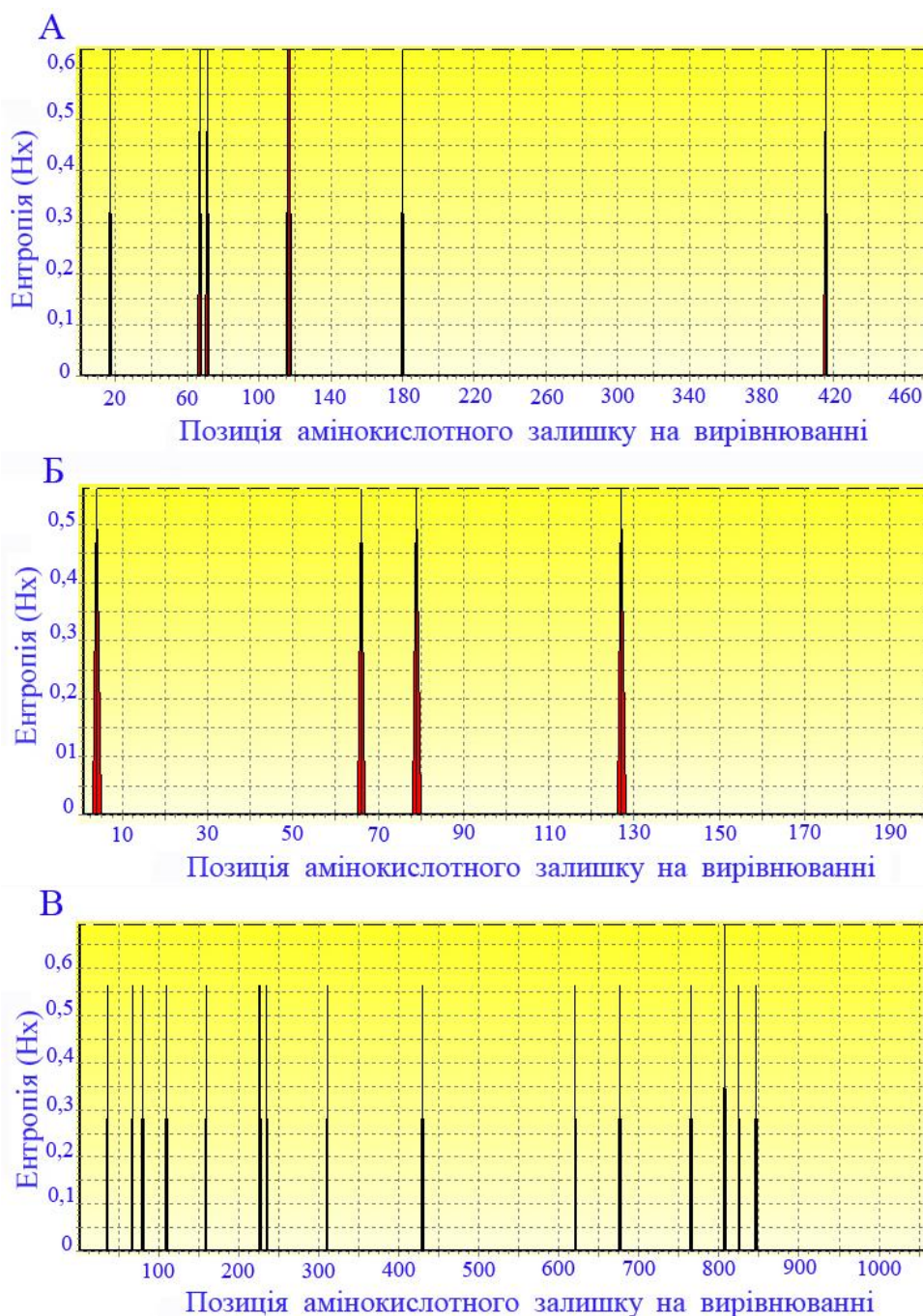


Рис. 1. Графік ентропії Шеннона для протеїнів, які кодують гени ВІ ВРХ *gag* (А), *vif* (Б) та *pol* (В). Міру мінливості кожного сайту амінокислотної послідовності отримано на підставі множинного вирівнювання нуклеотидних послідовностей та наступної трансляції для 4 ізолятів ВІ ВРХ.

Варіабельність розраховано як ентропію для позиції кожного амінокислотного залишка (а.з.). Відносна кількість варіабельних сайтів для протеїнів Pol, Gag, Env, Vif ВІ ВРХ та протеїнів, що транскрибуються з відкритих рамок зчитування, становила: Pol — 1.5 % (16/1056), Gag — 1.6 % (8/476), Vif — 2.0 % (4/198), Env — 2.6 % (24/904), ORF Y — 2.5 % (2/80), ORF W — 3.6 % (2/55).

Аналіз ентропії Шеннона, а саме, піків, які є специфічними до позицій а.з., що інтерпретують як поліморфізми, показав, що в протеомі ВІ ВРХ найменшою варіабельністю характеризується протеїн Pol, а найбільшою — протеїн ORF W.

Найбільш вивченим ретровірусом є вірус імунодефіциту людини типу 1 (ВІЛ-1), для якого встановлено, що протеїн Vif (virus infectivity factor) протидіє ферментам родини АРОВЕС. Ці ензими призводять до мутацій геному ВІЛ через дезамінування нуклеотидів. Протеїн Vif блокує включення АРОВЕС3F та АРОВЕС3G при синтезі віріону. Внаслідок цього процесу утворюються гіпермутовані та нежиттєздатні вірусні геноми [17].

Одним з методів виявлення дії відбору на рівні послідовностей є визначення співвідношення коефіцієнту несинонімічних замін (K_a) до такого синонімічних замін (K_s).

Значення $K_a/K_s < 1$ свідчить про дію переважно очищувачого відбору, при якому значимі мутації є шкідливими. $K_a/K_s > 1$ вказує на дію позитивного відбору.

На основі вирівнювань амінокислотних та нуклеотидних послідовностей гена *vif* ізолятів ВІ ВРХ та їхніх графіків ентропії Шеннона визначено кількість несинонімічних (ведуть до заміни амінокислоти) та синонімічних (не ведуть до заміни амінокислоти) замін та розраховано співвідношення K_a/K_s . Для протеїну Vif ізоляту L04972 порівняно з ізолятом NC_001413 виявлено дві несинонімічні заміни а.з. (N_a (реальн.) дорівнює 2), одна з яких (Т (Р), позиція 4 табл. 1) відповідає заміні неполярного а.з. (Т) на інший також неполярний а.з. (Р), а друга (Т (І), позиція 66 табл. 1) — заміні неполярного а.з. (Т) на гідрофобний а.з. (І).

Полярність а.з. та гідрофобні взаємодії відіграють важливу роль при формуванні третинної структури протеїну. Неполярні а.з. формують ядро протеїнової глобули, що уникає контактів з молекулами води. Зазначені дві заміни двох неполярних а.з. на неполярний та гідрофобний а.з. ведуть до суттєвого зменшення числа водневих зв'язків молекули Vif ізоляту L04972 порівняно з таким ізоляту NC_001413 (табл. 2).

Число теоретично можливих несинонімічних замін N_a (теор.) становило 12, K_a — 0,17. Число синонімічних замін N_s (реальн.) становило 3, N_s (теор.) — 7, K_s — 0,43. Відношення K_a/K_s становило 0,39, що свідчить про дію очищувачого відбору (рис. 2) на протеїн Vif ізоляту L04972.

Відбір не діє на синонімічні заміни, і патоген (в даному випадку ВІ ВРХ) з синонімічною заміною неможливо відрізнити від патогена з відсутньою синонімічною заміною без аналізу його геному. Вони відбуваються тільки через випадкові причини, і такий процес називається генетичним дрейфом.

Якщо несинонімічні заміни відбуваються рідше порівняно з синонімічними, як у випадку Vif ізоляту L04972, імовірність закріпитися для несинонімічної заміни нижче, ніж у синонімічної. Іншими словами, серед несинонімічних мутацій існують шкідливі, і негативний відбір елімінує такі мутації.

Для протеїну Vif ізоляту L04974 порівняно з ізолятом NC_001413 виявлено також дві несинонімічні заміни а.з. (Е (К), позиція 79 та Т (А) позиція 127 табл. 1), які відповідають заміні неполярних а.з. (Е та Т) на інші також неполярні а.з. (К та А). Зазначені дві заміни двох неполярних а.з. на два неполярні а.з. ведуть до істотних змін конформаційних параметрів молекули Vif ізоляту L04974 (табл. 2), а саме, до збільшення числа водневих зв'язків, α -спіралей, β -поворотів.

Число теоретично можливих несинонімічних замін N_a (теор.) становило 10, K_a — 0,20. Синонімічні заміни не виявлено, що вказує на відсутність генетичного дрейфа. Отже, відношення $K_a/K_s > 1$ (як і відношення $dN/dS > 1$), що свідчить про дію позитивного відбору (рис. 2) на протеїн Vif ізоляту L04974.

Якщо несинонімічні заміни відбуваються частіше, ніж синонімічні, як у випадку Vif ізоляту L04974, то це означає, що серед несинонімічних мутацій зустрічаються корисні з підвищеною імовірністю закріплення, і, отже, діє позитивний відбір. У найпростішому випадку позитивний відбір збільшує частоту ізолятів (алелей), які мають перевагу в частині пристосованості порівняно з іншою частиною популяції. Позитивний відбір збільшує генетичну різноманітність популяції, а негативний відбір видаляє шкідливі ізоляти (алелі).

Розділ 2. Ветеринарна вірусологія та мікробіологія

Таблиця 1 — Позиції амінокислотних залишків (а.з.) протеїнів Gag, Vif, Pol BI ВРХ з несинонімічними замінами. Амінокислотні послідовності протеїнів ізоляту NC_001413 використано як референтні

Протеїн	Позиція, а.з.	Номер ізоляту BI ВРХ у GenBank	Заміна амінокислоти
Gag	17	L04974, L04972	P (L) - Pro (Leu)
	67	L04974, L04972	N (D) - Asn (Asp)
	71	L04974, L04972	R (K) - Arg (Lys)
	116	L04974, L04972	A (T) - Ala (Thr)
	117	L04974, L04972	D (E) - Asp (Glu)
	180	L04974, L04972	V (I) - Val (Ile)
	416	L04974	R (K) - Arg (Lys)
	454	L04972	P (S) - Pro (Ser)
Vif	4	L04972	T (P) - Thr (Pro)
	66	L04972	T (I) — Thr (Ile)
	79	L04974	E (K) - Glu (Lys)
	127	L04974	T (A) - Thr (Ala)
Pol	35	L04972	T (I) - Thr (Ile)
	66	L04974	V (M) - Val (Met)
	79	L04974	T (I) - Thr (Ile)
	109	L04974	G (E) - Gly (Glu)
	158	L04972	V (I) - Val (Ile)
	225	L04974	G (E) - Gly (Glu)
	226	L04972	R (K) - Arg (Lys)
	234	L04974	P (K) Pro (Lys)
	310	L04972	V (I) - Val (Ile)
	429	L04974	Y (H) — Tyr (His)
	620	L04974	E (K) - Glu (Lys)
	676	L04972	V (I) - Val (Ile)
	765	L04974	H (Q) — His (Gln)
	807	L04974, L04972	K (R) - Lys (Arg)
	825	L04972	T (A) - Thr (Ala)
846	L04974	W (-) - Trp (-)	

Таблиця 2 — Параметри тривимірних структур протеїну Vif для ізолятів NC_001413, L04972 та L04974 BI ВРХ.

Кількість	Ізолят BI ВРХ		
	NC_001413	L04974	L04972
водневих зв'язків	115	124	101
α-спіралей	7	10	8
β-шарів	6	4	4
β-поворотів	23	28	27

Третинну структуру протеїну Vif (розташування елементів вторинної структури в просторі один по відношенню до іншого) ізолятів BI ВРХ охарактеризовано через кількість водневих зв'язків, α-спіралей (регулярних спіралей, що утворюються через міжпептидні водневі зв'язки в межах одного поліпептидного ланцюга), β-шарів (мають слабо зігнуту конфігурацію поліпептидного ланцюга та формуються через міжпептидні водневі зв'язки в межах окремих ділянок одного поліпептидного ланцюга або суміжних поліпептидних ланцюгів), β-поворотів (структур, що представляють собою повернуту на 180° петлю) (табл. 2).

До факторів, що впливають на утворення α-спіралей, віднесено, зокрема, (i) схильність амінокислот утворювати α-спіраль, (ii) взаємодія між R-групами на відстані 3-х або 4-х а.з., (iii) близьке розташування а.з. з R-групами великого розміру (Asn, Trp, Cys), (iv) наявність залишків Gly, Pro.

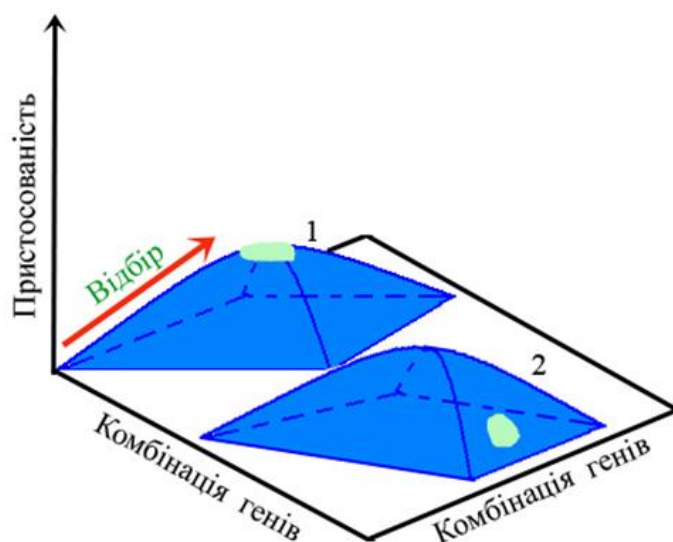


Рис. 2. Графік функції пристосованості від комбінації генів (адаптивний ландшафт). 1 — негативний (очищуючий) відбір, при якому найпристосованіший ізолят ВІ ВРХ є частим. 2 — позитивний відбір, при якому найпристосованіший ізолят ВІ ВРХ є рідкісним. Відбір завжди штовхає популяцію до вершини.

Для пошуку слідів позитивного відбору, що діяв в минулому [18], або відбувається за теперішнього часу [19, 20], застосовано низку різних методів, проте всі вони не є досконалими та мають певні недоліки.

Один з них, тест Таджими, для визначення форми природного відбору використовує середні значення відмінностей між послідовностями та кількість варіабельних сайтів (табл. 3). Негативні значення коефіцієнту D тесту Таджими свідчать про дію очищуючого відбору та зниження варіабельності, а $D > 0$ — про дію позитивного (балансуючого) відбору [21, 22].

Таблиця 3 — Параметри мутаційного аналізу (тесту Таджими) для генів, відкритих рамок зчитування ВІ ВРХ. Виділено значення коефіцієнту тесту Таджими для дії позитивного (балансуючого) відбору ($D > 0$).

Ген/протеїн ВІ ВРХ	m	S	p_s	Θ	π	D
	4	92	0.013145	0.007170	0.007096	-0.107352
<i>env</i>	4	42	0.016006	0.008731	0.008892	<u>0.192369</u>
<i>pol</i>	4	31	0.009776	0.005332	0.005151	-0.352193
<i>gag</i>	4	5	0.014205	0.007748	0.007576	-0.212492
ген <i>s</i> *	4	8	0.025890	0.014122	0.012945	-0.824072
<i>vif</i>	4	7	0.011725	0.006396	0.005863	-0.817336
ORF W	4	3	0.018182	0.009917	0.009091	-0.754451
ORF Y	4	3	0.012346	0.006734	0.006173	-0.754451
U3 область	4	4	0.008114	0.004426	0.004733	<u>0.650102</u>

Примітки: *ген *s* — ген, що кодує гіпотетичний протеїн S, m — число послідовностей для аналізу, S — число поліморфних сайтів в послідовностях, $p_s = S/n$ (n — загальне число сайтів), $\Theta = p_s/a_1$ — мутаційний параметр, π — нуклеотидна різноманітність, D — коефіцієнт тесту Таджими (оцінка відповідності характеру нуклеотидних замінів гіпотезі нейтральності).

Ретровірусні геноми фланковано довгими кінцевими повторами (LTR), які розділено на три області: U3, R та U5. Для регуляторної області U3 ізолятів L04972 та L04974 виявлено по 3 одонуклеотидних поліморфізми (ОНП) та дві делеції порівняно з ізолятом NC_001413. Для U3 ізоляту L04972 виявлено дві транзиції і одну трансверсію, а для L04974 — дві трансверсії і одну транзицію. ОНП, що виникають в області U3, можуть вести до значних змін її функцій. Зокрема, для області U3 провірусної ДНК ВІЛ-1, яка містить промотор, енхансер для вірусної транскрипції, продемонстровано, що її G-багата послідовність приймає структуру G-квадруплексу. Варіабельність нуклеотидної послідовності серед різних видів вірусів для області U3 перевищує таку для генів, що кодують протеїни. Проте варіабельні послідовності області U3 зберігають здатність утворювати G-квадруплес. Це вказує на ту обставину, що промоторна

область U3, важливим елементом якої є G-квадруплес, перебуває під сильним еволюційним тиском [23], що підтверджується отриманими даними щодо U3 ВІ ВРХ.

Висновки. Ідентифіковано гени, відкриті рамки зчитування (ORF) та регуляторну область U3 геному ВІ ВРХ, які перебувають під дією природного відбору. Гени *gag*, *pol*, *s*, *vif*, відкриті рамки зчитування ORF W та ORF Y перебувають під тиском негативного (очищуючого) відбору, на що вказують негативне значення коефіцієнту D тесту Таджими на нейтральність, співвідношення $K_a/K_s < 1$ та $dN/dS < 1$. На ген *env* та регуляторну область U3, для яких отримано $D > 0$, $K_a/K_s > 1$, діє позитивний відбір.

Проведено порівняльний аналіз впливу несинонімічних замінів в протеїні Vif ізолятів ВІ ВРХ на конформаційні параметри молекули Vif. Виявлені чотири поліморфізми в геномах двох ізолятів ВІ ВРХ ведуть до істотних змін конформаційних параметрів молекул Vif, а саме, до зміни числа водневих зв'язків, α -спіралей, β -поворотів.

Відсутність в базі даних GenBank нових повногеномних сиквенсів ВІ ВРХ за період з 2007 р. до теперішнього часу, в той час, як за цей же період до GenBank депоновано, наприклад, 241 новий ізолят ретровірусу вірусу лейкозу ВРХ, 65 нових ізолятів лентівірусу вірусу артрити-енцефаліту кіз, ставить під сумнів природний характер виникнення ВІ ВРХ.

Перспективи використання отриманих результатів. Під позитивним відбором звичайно перебуває лише невелика частина гена, в той час як інша його частина перебуває під негативним відбором.

Отримані відомості щодо дії позитивного відбору на гени ВІ ВРХ є основою для вивчення еволюційної стабільності ВІ ВРХ за допомогою нових програм, а також для уточнення, на які саме сегменти гена *env* та регуляторної області U3 діє позитивний відбір.

Отримані результати щодо дії негативного відбору на гени ВІ ВРХ, що знайдено на підставі тесту Таджими, є основою для застосування критерію Макдональда–Крейтмана, який дозволяє визначити позитивний відбір в тих випадках, коли його неможливо виявити за допомогою визначення відношення частоти несинонімічних замінів dN до частоти синонімічних замінів dS (dN/dS).

Фінансування. Роботу виконано за фінансової підтримки гранту 34.02.01.01Ф Національної академії аграрних наук України.

Список літератури

1. Keshavarz H., Mohammadi A., Morovati S. Evidence of bovine immunodeficiency virus: a molecular survey in water buffalo populations of Iran. *Veterinary Medicine and Science*. 2022. Vol. 8, No 5. P. 2167–2172. DOI: <https://doi.org/10.1002/vms3.872>.
2. Gutierrez B., Escalera-Zamudio M., Pybus O. G. Parallel molecular evolution and adaptation in viruses. *Current Opinion in Virology*. 2019. Vol. 34. P. 90–96. DOI: <https://doi.org/10.1016/j.coviro.2018.12.006>.
3. Kimura M. The neutral theory of molecular evolution: a review of recent evidence. *Japanese Journal of Genetics*. 1991. Vol. 66, No 4. P. 367–386. DOI: <https://doi.org/10.1266/jjg.66.367>.
4. Kumar S., Hedges B. Advances in time estimation methods for molecular data. *Molecular Biology and Evolution*. 2016. Vol. 33, No 4. P. 863–869. DOI: <https://doi.org/10.1093/molbev/msw026>.
5. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989. Vol. 123, No 3. P. 585–595. DOI: <https://doi.org/10.1093/genetics/123.3.585>.
6. Tamura K. et al. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences*. 2012. Vol. 109, No 47. P. 19333–19338. DOI: <https://doi.org/10.1073/pnas.1213199109>.
7. Hanada K., Shiu S.-H., Li W.-H. The nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. *Molecular Biology and Evolution*. 2007. Vol. 24, No 10. P. 2235–2241. DOI: <https://doi.org/10.1093/molbev/msm152>.
8. Wernegreen J. J. Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide. *PLoS One*. 2011. Vol. 6, No 12. P. e28905. DOI: <https://doi.org/10.1371/journal.pone.0028905>.
9. Sohal V. K. Comparative study: nonsynonymous and synonymous substitution of SARS-CoV-2, SARS-CoV, and MERS-CoV genome. *Genomics and Informatics*. 2021. Vol. 19, No 2. P. e15. DOI: <https://doi.org/10.5808/gi.20058>.
10. Palumbi S. R. Humans as a world's greatest evolutionary force. *Science*. 2001. Vol. 293, No 5536. P. 1786–1790. DOI: <https://doi.org/10.1126/science.293.5536.1786>.
11. Wensing A. M. et al. 2017 update of the drug resistance mutations in HIV-1. *Topics in Antiviral Medicine*. 2016. Vol. 24, No 4. P. 132–133. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/pmc5677049/>.
12. Maldarelli F. et al. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *Journal of Virology*. 2013. Vol. 87, No 18. P. 10313–10323. DOI: <https://doi.org/10.1128/jvi.01225-12>.
13. Tamura K. et al. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*. 2013. Vol. 30, No 12. P. 2725–2729. DOI: <https://doi.org/10.1093/molbev/mst197>.

14. Hall T. A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999. Vol. 41. P. 95–98.
15. Kosakovsky Pond S. L., Frost S. D. W. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*. 2005. Vol. 22, No 5. P. 1208–1222. DOI: <https://doi.org/10.1093/molbev/msi105>.
16. Yang J., Zhang Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*. 2015. Vol. 43, No W1. P. W174–W181. DOI: <https://doi.org/10.1093/nar/gkv342>.
17. Suspene R. et al. Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. *Nucleic Acids Research*. 2006. Vol. 34, No 17. P. 4677–4684. DOI: <https://doi.org/10.1093/nar/gkl555>.
18. Eyre-Walker A. The genomic rate of adaptive evolution. *Trends in Ecology and Evolution (Amsterdam)*. 2006. Vol. 21, No 10. P. 569–575. DOI: <https://doi.org/10.1016/j.tree.2006.06.015>.
19. Grossman S. R. et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*. 2010. Vol. 327, No 5967. P. 883–886. DOI: <https://doi.org/10.1126/science.1183863>.
20. Ben-Jemaa S. et al. Whole genome sequencing reveals signals of adaptive admixture in Creole cattle. *Scientific Reports*. 2023. Vol. 13, No 1. P. 12155. DOI: <https://doi.org/10.1038/s41598-023-38774-7>.
21. Jadhav A. et al. Genomic diversity and evolution of quasispecies in Newcastle disease virus infection. *Viruses*. 2020. Vol. 12, No 11. P. 1305–1323. DOI: <https://doi.org/10.3390/v12111305>.
22. Furuse Y. Identifying potentially beneficial genetic mutations associated with monophyletic selective sweep and a proof-of-concept study with viral genetic data. *mSystems*. 2021. Vol. 6, No 1. P. e01151-20. DOI: <https://doi.org/10.1128/msystems.01151-20>.
23. Piekna-Przybylska D. et al. U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry*. 2014. Vol. 53. P. 2581–2593. DOI: <https://doi.org/10.1021/bi4016692>.

CHARACTERIZATION OF BOVINE IMMUNODEFICIENCY VIRUS GENES AND PROTEINS

Balak O. K.

Kharkiv National Medical University, Kharkiv, Ukraine

Lymanska O. Yu.

National Scientific Center “Institute of Experimental and Clinical Veterinary Medicine”, Kharkiv, Ukraine

The goal of the study was determining the effect of natural selection on the bovine immunodeficiency virus (BIV) through the identification of polymorphisms by comparing a number of genes, open reading frames and proteins and analyzing the effect of nonsynonymous substitutions in the Vif protein of BIV isolates on the Vif conformational parameters. The selection pressure on the genes was estimated by Tajima test, as well as the determination of the ratio of nonsynonymous substitutions (K_a) to synonymous (K_s) K_a/K_s . K_a and K_s coefficients, their ratio (K_a/K_s) were calculated on the basis of alignments of amino acid and nucleotide sequences of the vif gene of BIV isolates. The nonsynonymous (dN) to synonymous (dS) substitution rate ratio per nucleotide site dN/dS was calculated to estimate the effect of selection on the vif gene by SLAC method on Datamonkey server. The number of hydrogen bonds, α -helices, β -sheets, β -turns for the protein tertiary structures of proteins was determined by I-TASSER server. Genes, open reading frames (ORFs) and the U3 regulatory region of BIV genome which are under the influence of selection have been identified. Gag, pol, s, vif genes, ORF W and ORF Y are under the negative (purifying) selection pressure. Env gene and the U3 regulatory region are under positive selection. Shannon entropy analysis (peaks that are specific to the positions of amino acid residues), interpreted as polymorphisms, revealed 16, 8, and 4 nonsynonymous substitutions for Pol, Gag, and Vif BIV proteins, respectively. Significant changes in the number of hydrogen bonds, α -helices, β -layers, β -turns were determined by the analysis of the effect of four nonsynonymous substitutions in the Vif protein on the conformational parameters of Vif of two BIV isolates. The multidirectional effect of selection on BIV genes is shown

Keywords: positive selection, molecular evolution, amino acid substitution, SNP