

УДК 577.21

В. В. ЩЕРБИК, Л. П. БУЧАЦЬКИЙ

Київський національний університет імені Тараса Шевченка
вул. Володимирська, 64, Київ, Україна, 01033

СТАТИСТИЧНІ ЗАКОНОМІРНОСТІ ВИКОРИСТАННЯ СИНОНІМІЧНИХ КОДОНІВ У ГЕНОМІ РІЗНИХ ОРГАНІЗМІВ

На основі інформації, узятої з бази даних Codon Usage Database, встановлено, що для досліджуваних шести груп організмів статистика використання кодонів визначається двома випадковими процесами: процесом Орнштейна – Уленбека і Вінерівським процесом.

Ключові слова: вибірковість кодонів, сімейства синонімічних кодонів, GC_3 - показник, інформаційна ентропія, ансамблі організмів

В універсальному генетичному коді кожному триплету нуклеотидів (кодону) відповідає певна амінокислота, але не навпаки – генетичний код вироджений. Це означає, що амінокислоти генетичного коду кодуються кількома кодонами, які є синонімічні. Відповідно до генетичного коду амінокислоти M, W кодуються одним кодоном, амінокислоти C, D, E, F, H, K, N, Q, Y – двома кодонами, амінокислота I і термінатор – трьома кодонами, амінокислоти A, G, P, T, V – чотирма кодонами і амінокислоти L, R, S – шістьма кодонами. Таке розбиття кодонів представляє сімейства синонімічних кодонів, які позначаються як SF (1) ... SF (6) [1].

Нерівне використання синонімічних кодонів серед SF-сімейств отримало назву “вибірковість кодонів” (codon usage bias). Вибірковість кодонів варіює як серед різних геномів, так і усередині одного геному [2]. Згідно з геномною гіпотезою [3], невідповідна перевага кодонів визначається деякою загальною стратегією. Ця стратегія строго індивідуальна для кожного організму. При випадковому розподілі десяти тисяч кодонів їх частоти змінюються на 50%. Для коротких генів варіація частот кодонів ще більша. Нерівномірне використання кодонів – це невід’ємна властивість генів. Тому виділення на цьому фоні зв'язку вибірковості кодонів з яким-небудь біологічним процесом є непростим завданням.

Молекулярні дослідження дозволяють припустити, що вибірковість кодонів виявляється в результаті балансу між мутацією і селекцією трансляції генів; це явище широко розповсюджене серед видів і може внести свій вклад в еволюцію геному [4].

Вибірковість кодонів асоціюється з трьома головними біохімічними процесами, пов'язаними з трансляцією мРНК: ефективність, точність і регуляція [5, 6]. Оптимальна ефективність трансляції досягається взаємною еволюцією вибірковості кодонів і кількістю відповідних тРНК [7 – 10]. Вибірковість кодонів корелює з рівнем експресії генів [11–15] і впливає на точність трансляції [16, 17]. Вибір кодонів визначає вторинну структуру мРНК, яка пов'язана з її стабільністю та ініціацією трансляції [18–20]. Виявлений зв'язок розподілу кодонів із стійкістю до точкових мутацій [21].

Слід зазначити, що кореляція вибірковості кодонів з окремими біохімічними процесами не є постійною для усіх геномів, але змінюється уздовж філогенетичних дерев [22].

Одним з найбільш важливих параметрів, який пояснює відмінність у використанні кодонів між організмами є GC - вміст генів [23–27]. Гени, в яких GC - вміст більший, мають більше кодонів, що закінчуються на G або C. GC -вміст генів є параметром, який пов'язаний з внутрішніми і зовнішніми силами, що формують будову геному. Оскільки GC -багаті гени мають більшу довжину, то і дія на них еволюційних сил більша, ніж на GC - бідні гени. GC -вміст може бути корисним для порівняння геномів при визначенні областей горизонтального перенесення генів.

Метою цієї роботи є визначення можливих випадкових процесів, які призводять до вибірковості кодонів.

Матеріал і методи досліджень

Початкові дані щодо використання кодонів узяті з бази даних Codon Usage Database [28]. Усі організми були розподілені по шести групах (табл. 1). Органели хлоропласти і мітохондрії визначені як несамостійні організми.

Нижній рівень кількісного вмісту кодонів у геномі для вибраного організму визначався за кількістю нульових значень у таблиці з 64-х кодонів генетичного коду. Як правило, число нулів не перевищувало трьох.

Такий підхід не є абсолютно точним за визначенням розміру вибірки числа організмів у кожній групі, але збільшення нижнього порогу значущості може тільки зменшити кількість вибраних організмів; тим більше для більшої частини організмів вміст кодонів явно більше порогу. Для хлоропластів поріг менший, оскільки вміст кодонів у них розподілений більш рівномірно за генетичним кодом.

Таблиця 1

Групи вибраних організмів

Групи організмів	Мінімум кодонів	Кількість організмів
Хребетні	1000	706
Безхребетні	1000	921
Рослини	1000	1524
Хлоропласти	500	5680
Бактерії	1000	2682
Мітохондрії	1000	2326

Результати досліджень та їх обговорення

GC- вміст кодонів в геномі різних груп організмів

На рис. 1 наводиться статистика відсоткового вмісту кодонів, що закінчуються на G або C ($GC_3\%$) у різних груп організмів.

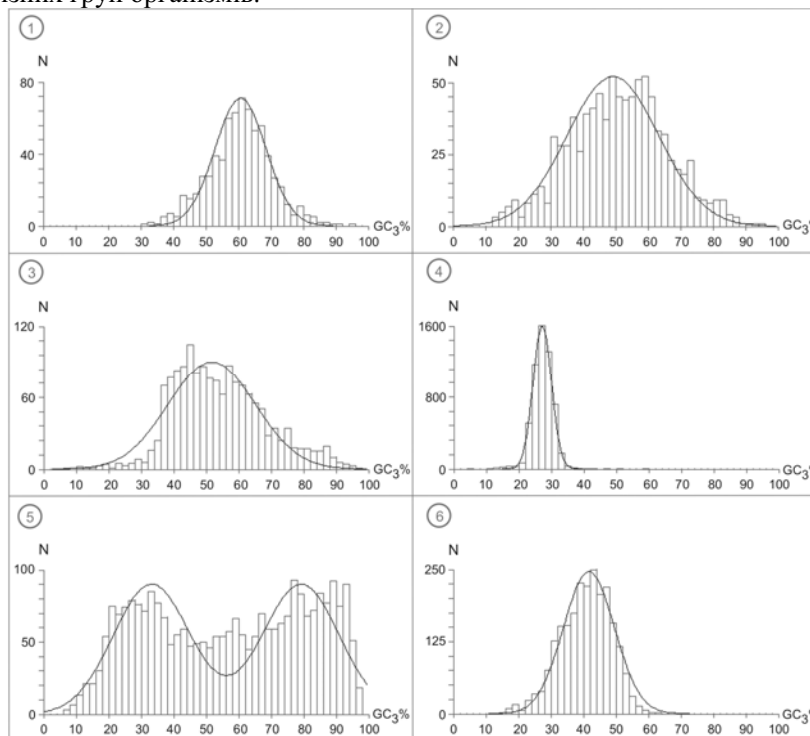


Рис. 1. Розподіл організмів (N) по $GC_3\%$ -вмісту кодонів (Гаусові криві). Крок квантування 2 %. 1 – хребетні, 2 – безхребетні, 3 – рослини, 4 – хлоропласти, 5 – бактерії, 6 – мітохондрії хребетних.

Розподіл по GC_3 -вмісту кодонів визначається процесом Орнштейна – Уленбека, який є стаціонарним процесом [29]. При відхиленні числа організмів від стаціонарного значення 50 % дисперсія розподілу зменшується. Для бактерій чітко видно бімодальний розподіл – тут стаціонарним значенням 50 % є абсолютна різниця між $GC_3\%$ та $AT_3\%$ у геномах двох груп бактерій.

Графіки розподілів груп організмів за відсотковим вмістом кодонів, що закінчуються на G або C, займають 32 кодони. Можна довільним чином зафіксувати 32 кодони і побудувати для них статистику розподілу організмів за відсотковим вмістом цих кодонів. Більш раціональним є вибір 32-х кодонів, які кодують амінокислоти, приєднані до тРНК аміноацил-тРНК – синтетазами класу II. Амінокислоти F, S, P, T, A, H, N, K, D, G визначають статистику організмів за відсотковим вмістом ($C_{II}\%$) кодонів, які обробляються аміноацил-тРНК – синтетазами класу II [30].

На рис. 2 наводиться статистика по $C_{II}\%$ -вмісту кодонів для різних груп організмів.

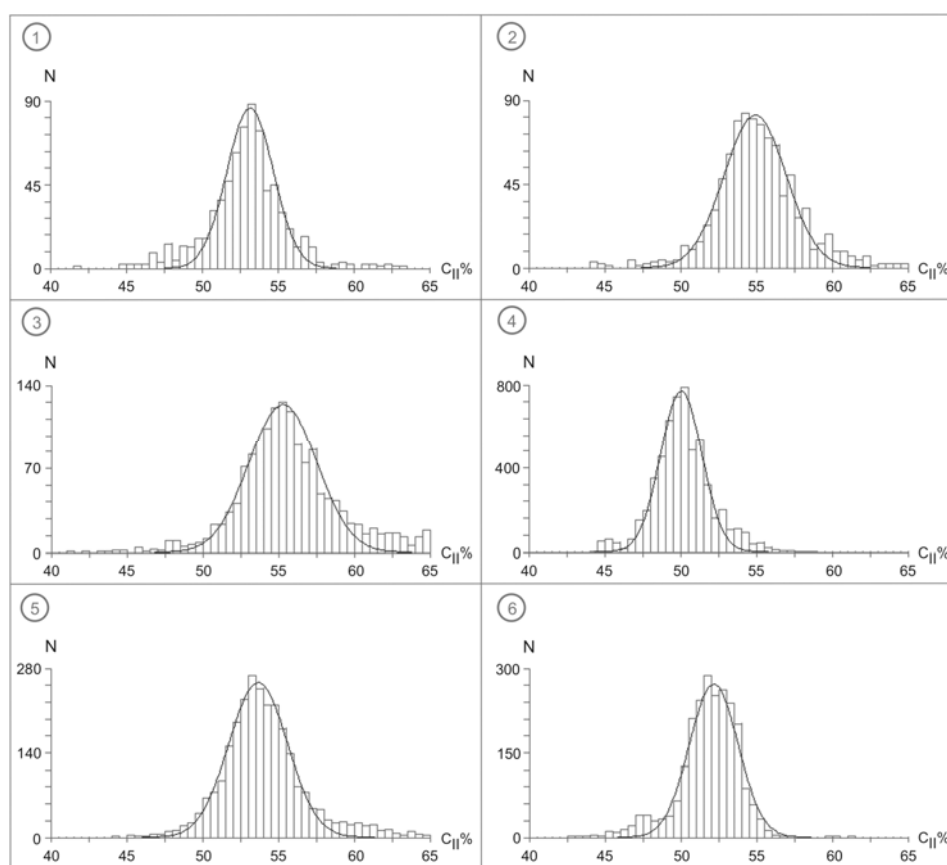


Рис. 2. Розподіл організмів (N) за $C_{II}\%$ -вмістом кодонів (Гаусові криві). Крок квантування 0,5 %. 1 – хребетні, 2 – безхребетні, 3 – рослини, 4 – хлоропласти, 5 – бактерії, 6 – мітохондрії хребетних.

Розподіли для класу II дуже вузькі; вони визначаються Вінерівським процесом, який є нестационарним [29]. При відхиленні від 50 % дисперсія розподілу зростає. Є тенденція до збільшення відсоткового вмісту кодонів, які обробляються аміноацил-тРНК – синтетазами класу II.

Інформаційна ентропія розподілу кодонів

Ентропія розглядається як одна з найважливіших і універсальних характеристик частотних розподілів статистичних величин різної природи. Ентропійний метод дослідження вибірковості кодонів є одним з найпотужніших. Відомо, що ентропія – це міра хаотичності, невпорядкованості фізичної системи. Але не лише; ентропія – це ще і міра структурної організованості фізичного об'єкту: чим більше ентропія його структур, тим більше можливостей розвитку в об'єкта.

Ентропія Шенона [34] є кількісною мірою інформації, що міститься в повідомленні. Розподіл кодонів у геномі кожного організму можна вважати повідомленням, складеним із триплетів нуклеотидів як неділимих слів. У підході Шенона розглядається тільки частота слів у повідомленні, але не контекст вмісту. Щоб врахувати генетичний код, необхідно прийняти, що синонімічні кодони мРНК, що кодують амінокислоту, є окремими словами, і ентропія генома визначається як сума з 21-ої ентропії незалежних повідомлень, складених довільним чином з синонімічних кодонів.

Ентропія генома з урахуванням генетичного коду обчислюється за формулою [35]:

$$S = \sum_i \sum_c -p_c(A_i) \text{Log}_2(p_c(A_i))$$

де

$i = 1 \dots 21$ – число амінокислот плюс термінатор, $c = 1 \dots n_a$;

n_a – число синонімічних кодонів амінокислоти A_i , $1 \leq n_a \leq 6$;

$p_c(A_i) = n_c(A_i) / \sum_c n_c(A_i)$ – частота синонімічних кодонів c ;

$n_c(A_i)$ – кількість синонімічних кодонів c амінокислоти A_i .

Якщо усі частоти кодонів для A_i рівні, то ентропія $S(A_i) = \text{Log}_2(z(A_i))$, де $z(A_i)$ – число виродженості для A_i . Максимальна ентропія геному, визначена за вибірковості кодонів, дорівнює $S_{\max}(\text{Uni}) = 22 + 5\text{Log}_2 3 \approx 29.924813$ для універсального генетичного коду і $S_{\max}(\text{Mit}) = 28 + 2\text{Log}_2 3 \approx 31.169925$ для генетичного коду мітохондрій хребетних.

Логарифмічна функція від кількості кодонів сильно змінює форму розподілу організмів за величиною ентропії. Замість гаусових кривих спостерігаються криві, які добре описуються асиметричним розподілом Лапласа (Рис. 3).

Щільність асиметричного розподілу Лапласа [36] задається формулою

$$p(S) = [\lambda_L \lambda_R / (\lambda_L + \lambda_R)] \text{Exp}\{\lambda_L(S < S_a) \cup -\lambda_R(S \geq S_a)\}$$

де λ_L і λ_R – лівий і правий коефіцієнти загасання, S_a – ентропія положення максимуму. Щільність асиметричного розподілу Лапласа на рис. 3-1,2,3,5 має форму, при якій $\lambda_L < \lambda_R$, а на рис. 3-4: $\lambda_L > \lambda_R$. Складнішим є графік щільності на рис. 3-6. Він є сумою трьох щільностей: двох дзеркальних асиметричних розподілів Лапласа і одного центрального симетричного розподілу Лапласа.

Якщо прийняти, що зміна ентропії геному в групі організмів визначається Вінерівським процесом, то при усередненні за інтервалами часу спостереження Δt з щільністю ймовірності $\sigma \times \text{Exp}\{-\sigma \Delta t\}$, де σ – виробництво ентропії в одиницю часу, щільність функції розподілу організмів за ентропією визначатиметься асиметричним розподілом Лапласа.

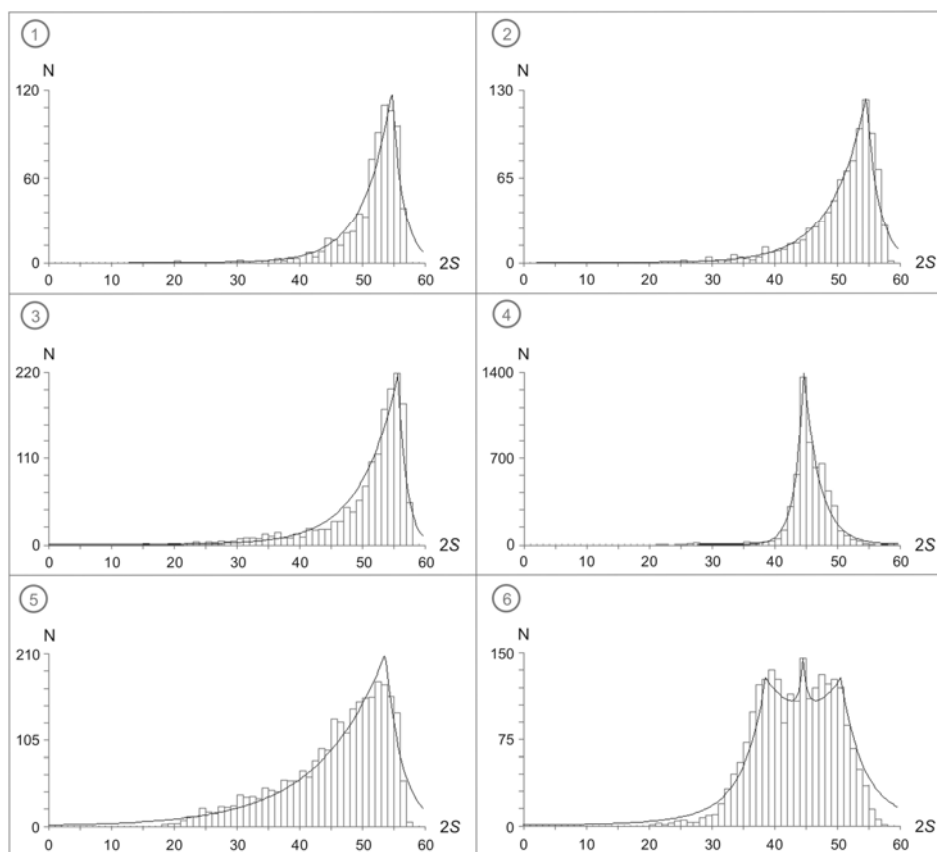


Рис. 3. Розподіл організмів (N) за ентропією Шенона S. 1 – хребетні, 2 – безхребетні, 3 – рослини, 4 – хлоропласти, 5 – бактерії, 6 – мітохондрії хребетних.

Ентропія сімейств синонімічних кодонів

При побудові гістограм розподілу кодонів на рис.1, 2, 3 для кожного організму визначалося лише одне значення параметра, над яким виконувалася процедура квантування. Якщо врахувати, що ентропія є адитивною величиною, то можна побудувати гістограми розподілу організмів за ентропією кодонів окремо для кожного сімейства амінокислот. Сімейство амінокислот $SF(n_a)$ складається з амінокислот, які мають рівне число синонімічних кодонів n_a . Значення ентропій за кожною амінокислотою сімейства $SF(n_a)$ об'єднуються, але не сумуються, в загальну множину відліків ентропії S для цієї групи організмів :

$$S = \bigcup_{A_i} \sum_c -p_c(A_i) \text{Log}_2(p_c(A_i))$$

де

$c = 2 \dots n_a, 2 \leq n_a \leq 6, A_i \subset SF(n_a)$;

n_a – число синонімічних кодонів амінокислоти A_i ,

$p_c(A_i) = n_c(A_i) / \sum_c n_c(A_i)$ – частота синонімічних кодонів c,

$n_c(A_i)$ – кількість синонімічних кодонів c для амінокислоти A_i .

На рис. 4, 5 приведені розподіли організмів по ентропії Шенона S сімейств синонімічних кодонів SF (2)...SF (6).

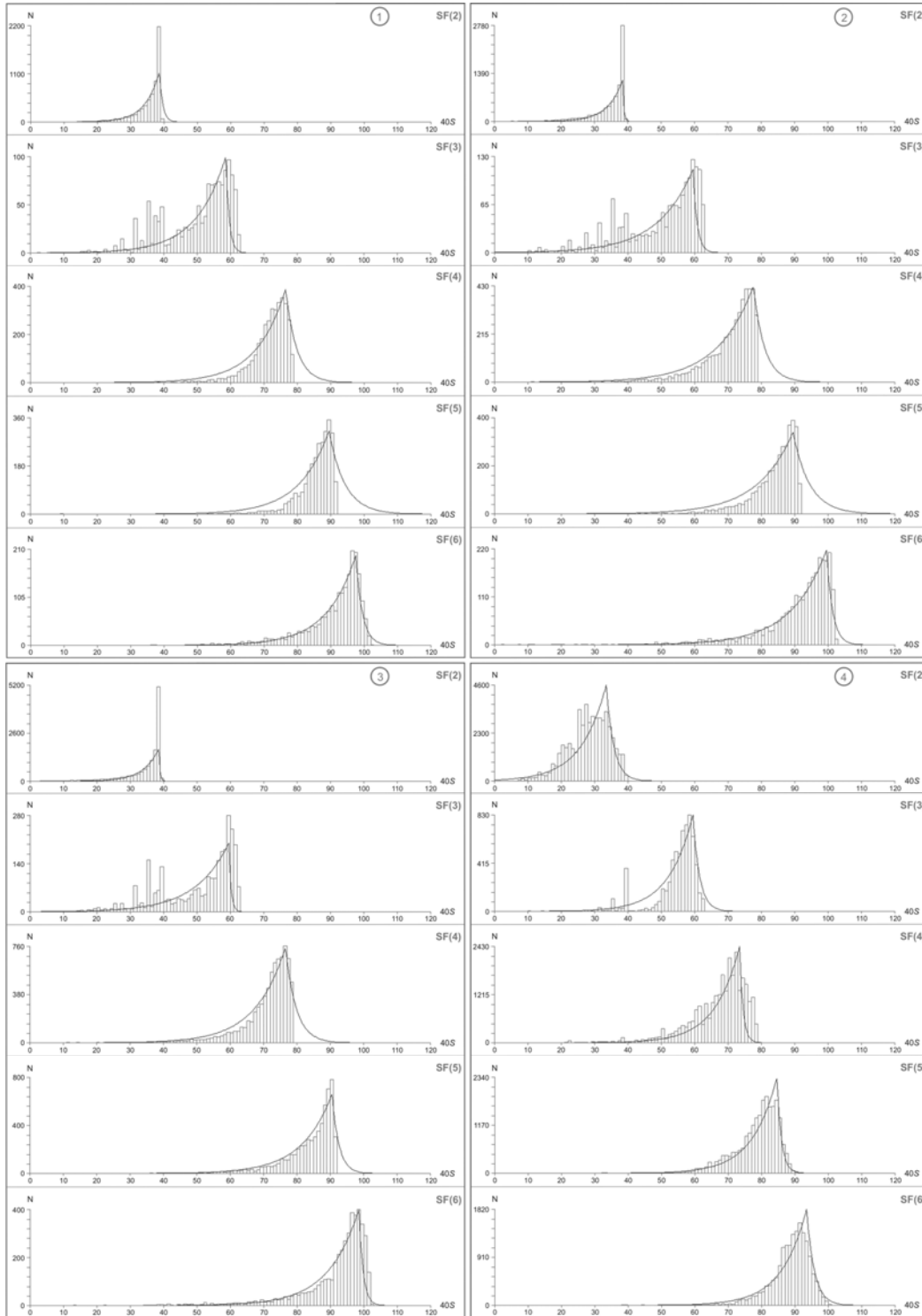


Рис. 4. Розподіли організмів (N) за ентропією Шенона S сімейств синонімічних кодонів SF (2)...SF (6). 1 – хребетні, 2 – безхребетні, 3 – рослини, 4 – хлоропласти.

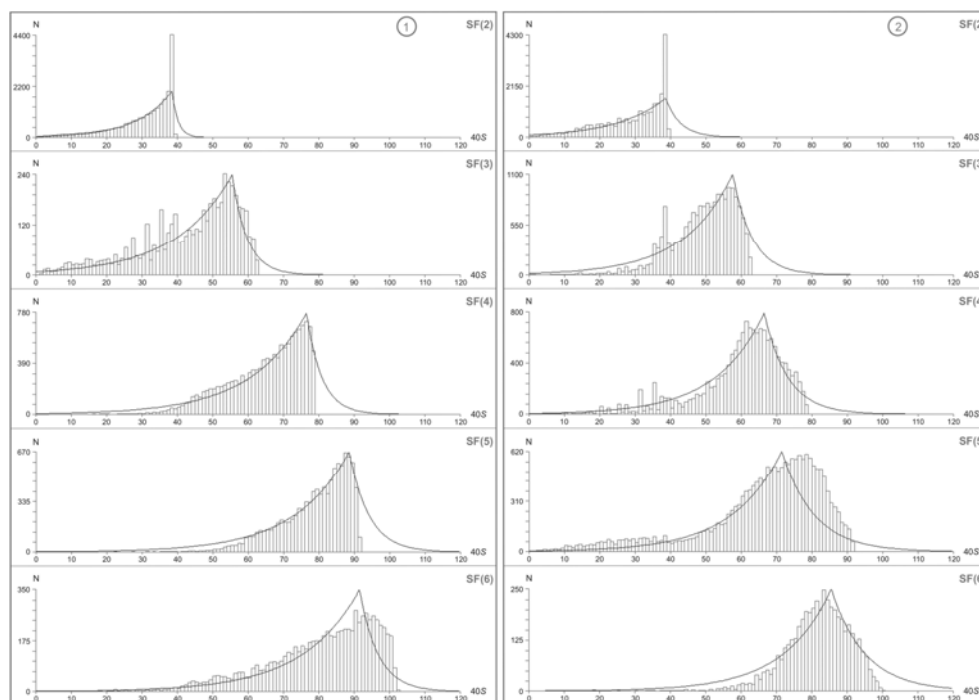


Рис. 5. Розподіли організмів (N) за ентропією Шенона S сімейств синонімічних кодонів SF (2)...SF (6). 1 – бактерії, 2 – мітохондрії хребетних.

Головним результатом такого розбиття кодонів по сімействах є наявність інваріантної ентропії, зворотної коефіцієнту λ_L асиметричного розподілу Лапласа по сімействах SF(3)...SF(6). Для сімейств SF(2) розподіл організмів має сильний викид біля $S = 1$ і може мати інше значення λ_L . У хлоропластів і мітохондрій хребетних λ_L має постійне значення для усіх сімейств SF(2)...SF(6).

Якщо сімейства SF(1) і SF(2), що мають в сумі 20 кодонів, замінити віртуальним сімейством SF(5), тобто чотири амінокислотами по 5 кодонів, то при великій кількості можливого розбиття ми отримаємо для SF(5) той же коефіцієнт λ_L (табл. 2). Цей результат для віртуальних сімейств доводить, що інваріант λ_L не залежить від структури генетичного коду.

Кодони віртуальних амінокислот $A5_1...A5_4$ сімейства SF(5):

- $A5_1 = \{UUU, UUC, AUG, UAU, UAC\}$
- $A5_2 = \{CAU, CAC, CAA, CAG, UGG\}$
- $A5_3 = \{AAU, AAC, AAA, AAG, GAU\}$
- $A5_4 = \{GAC, GAA, GAG, UGU, UGC\}$.

Таблиця 2

Коефіцієнти λ_L сімейств кодонів (40S)

Групи організмів	Сімейства кодонів				
	SF(2)	SF(3)	SF(4)	SF(5)	SF(6)
Хребетні	0.230	0.113	0.116	0.112	0.115
Безхребетні	0.220	0.094	0.093	0.094	0.095
Рослини	0.210	0.100	0.110	0.110	0.110
Хлоропласти	0.133	0.139	0.135	0.136	0.134
Бактерії	0.100	0.062	0.063	0.065	0.062
Мітохондрії	0.073	0.072	0.073	0.071	0.074

Для мітохондрій хребетних віртуальні амінокислоти $A3_1...A3_8$ сімейства SF(3) визначені шляхом перегрупування кодонів сімейства SF(2) :

A3₁ = {UGU, UGC, GAU}
 A3₂ = {GAC, GAA, GAG}
 A3₃ = {UUU, UUC, CAU}
 A3₄ = {CAC, AUU, AUC}
 A3₅ = {AAA, AAG, AUA}
 A3₆ = {AUG, AAU, AAC}
 A3₇ = {CAA, CAG, UGA}
 A3₈ = {UGG, UAU, UAC}

Для мітохондрій хребетних віртуальні амінокислоти A5₁...A5₈ сімейства SF(5) визначені шляхом перегрупування кодонів сімейств SF(4) і SF(6):

A5₁ = {GCU, GCC, GCA, GCG, GGU}
 A5₂ = {GGC, GGA, GGG, CCU, CCC}
 A5₃ = {CCA, CCG, CGU, CGC, CGA}
 A5₄ = {CGG, ACU, ACC, ACA, ACG}
 A5₅ = {GUU, GUC, GUA, GUG, UAA}
 A5₆ = {UAG, AGA, AGG, UUA, UUG}
 A5₇ = {CUU, CUC, CUA, CUG, UCU}
 A5₈ = {UCC, UCA, UCG, AGU, AGC}

Для сімейства SF(*n_a*) з числом організмів N_{OR} кількість відліків ентропії дорівнює N_S = N_{OR} × *n_a*. Середнє квантоване значення ентропії S_{mv}(SF) для сімейства SF(*n_a*):

$$S_{mv}(SF) = 1/N_S \sum_{S=1... \{S_0\}} S \times N(S)$$

де S₀ = Log₂(*n_a*), {S₀} – цілочисельне значення S₀ з масштабним множником.

Теоретичне середнє значення ентропії S_{th}(SF) визначається за формулою:

$$S_{th}(SF) = \int_{-\infty}^{+\infty} S \times p(S) dS = S_a - 1/\lambda_L + 1/\lambda_R$$

Теоретичні функції розподілу Лапласа побудовані так, щоб значення S_{th}(SF) співпадало з вибіркоким квантованим середнім S_{mv}(SF) при рівні значущості 1% за критерієм Стьюдента. Найкраще значення λ_L визначається окремо методом найменших квадратів при різних S_a.

Значення λ_R розглядається як очікуване при збільшенні об'єму вибірки груп організмів. Інваріант λ_L можна отримати при великому, але кінцевому числі розбиття кодонів по різних сімействах SF(*n_a*). Це означає, що сімейства SF(*n_a*) є кластерними множинами, представниками яких є амінокислоти генетичного коду.

Висновки

Розподіл організмів за GC₃%-вмістом кодонів визначається стаціонарним процесом Орнштейна – Уленбека. Статистика 32-х кодонів, які кодуєть амінокислоти, що приєднуються до тРНК, аміноацил-тРНК – синтетазами класу II визначається нестаціонарним Вінерівським процесом. Розподіл організмів за ентропією Шенона також визначається Вінерівським процесом. Розподіл організмів за ентропією сімейств синонімічних кодонів описується асиметричним розподілом Лапласа, у якого лівий коефіцієнт загасання є інваріантом для окремої групи організмів.

1. *Wright F.* The effective number of codons' used in a gene. / F. Wright. // *Gene* — 1990, Vol. 87. — P. 23—29.
2. *Sharp Paul M.* Forces that influence the evolution of codon bias. / Paul M. Sharp, Laura R. Emery and Kai Zeng. // *Phil. Trans. R. Soc. B* — 2010, Vol. 365, — P. 1203—1212.
3. *Grantham R.C.* Codon catalog usage and the genome hypothesis. / Grantham, R.C., Gautier, C., Gouy, M., Mercier, R., and Pavé A. // *Nucleic Acids Res.* — 1980, Vol. 8, — P. 49—79.
4. *Behura Susanta K.* Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. / Susanta K. Behura and David W. Severson. // *Biol. Rev.* — 2013, Vol. 88, — P. 49—61.
5. *Plotkin J. B.* Synonymous but not the same: the causes and consequences of codon bias. / J. B. Plotkin and G. Kudla. // *Nat. Rev. Genet.*, — 2011, Vol. 12, — P. 32—42.
6. *Gingold H.* Determinants of translation efficiency and accuracy. / H. Gingold and Y. Pilpel. // *Mol. Syst. Biol.*, — 2011, Vol. 7, — P. 481.

7. *Bulmer M.* Coevolution of codon usage and transfer rna abundance. / M. Bulmer. // Nature — 1987, Vol. 325, — P. 728—730.
8. *Dong H.* Co-variation of trna abundance and codon usage in escherichia coli at different growth rates. / H. Dong, L. Nilsson, and C. G. Kurland. // J.Mol. Biol. — 1996, Vol. 260, N.5, — P. 649—663.
9. *Cannarozzi G.* A role for codon order in translation dynamics. / G. Cannarozzi, N. N. Schraudolph, M. Faty, P. von Rohr, et al. // Cell — 2010, Vol. 141, N.2, — P. 355—367.
10. *Tuller T.* Translation efficiency is determined by both codon bias and folding energy. / T. Tuller, Y. Y. Waldman, M. Kupiec, and E. Rupp. // Proc. Natl. Acad. Sci. U. S. A. — 2010, Vol. 107, N. 8 — P. 3645—3650.
11. *Grantham R.* Codon catalog usage is a genome strategy modulated for gene expressivity. / R. Grantham, C. Gautier, M. Gouy, M. Jacobzone, et al. // Nucleic Acids Res., — 1981, Vol. 9, N. 1 — P. 213.
12. *Gouy M.* Codon usage in bacteria: correlation with gene expressivity. / M. Gouy and C. Gautier. // Nucleic Acids Res. — 1982, Vol. 10, N.2 2 — P. 7055—7074.
13. *Kudla G.* Coding sequence determinants of gene expression in escherichia coli. / G. Kudla, A. W. Murray, D. Tollervey, and J. B. Plotkin. // Science, — 2009, Vol. 324 — P. 255—258.
14. *Gustafsson C.* Codon bias and heterologous protein expression. / C. Gustafsson, S. Govindarajan, and J. Minshull. // Trends Biotechnol. — 2004, Vol. 22, N. 7 — P. 346—353.
15. *Boel Gregory.* Codon influence on protein expression in E. coli correlates with mRNA levels. / Gregory Boel, Reka Letso, Helen Neely, et al. // Nature — 2016, Vol. 529 — P. 358—363.
16. *Drummond D. A.* Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. / D. A. Drummond and C. O. Wilke. // Cell — 2008, Vol. 134, N.2 — P. 341—352.
17. *Eyre-Walker A.* Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy? / A Eyre-Walker. // Mol. Biol. Evol. — 1996, Vol. 13 — P. 864—872.
18. *Tuller T.* An evolutionarily conserved mechanism for controlling the efficiency of protein translation. / T. Tuller, A. Carmi, K. Vestsigian, S. Navon, et al. // Cell — 2010, Vol. 141, N. 2 — P. 344—354.
19. *Ma J.* Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. / Ma J, Campbell A, Karlin S. // J. Bacteriol. — 2002, Vol. 184 — P. 5733—5745.
20. *Novoa1 Eva Maria.* Speeding with control: codon usage, tRNAs, and ribosomes. / Eva Maria Novoa1 and Lluís Ribas de Pouplana. // Trends Genet. — 2012, Vol. 28, N. 11 — P. 574—581.
21. *Mahdi Rami N.* Codon usage bias as a function of generation time and life expectancy. / Rami N. Mahdi and Eric C. Rouchka. // Bioinformatics — 2012, Vol. 8, N. 3 — P. 158—162.
22. *Nekrutenko A.* Assessment of compositional heterogeneity within and between eukaryotic genomes. / Nekrutenko A., Li W. H. // Genome Res. — 2000, Vol. 10, N. 12 — P. 1986—1995.
23. *Sueoka N.* DNA G+C content of the third codon position and codon usage biases of human genes. / Sueoka N, Kawanishi Y. // Gene — 2000, Vol. 261 — P. 53—62.
24. *Epstein RJ.* A functional significance for codon third bases. / Epstein RJ, Lin K, Tan TW. // Gene — 2000, Vol. 245, N. 2 — P. 291—298.
25. *Wan X.-F.* Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. / Wan, X.-F., Xu, D., Kleinhofs, A., Zhou J. // BMC Evol. Biol. — 2004, Vol. 4, N. 19, — P. 1—11.
26. *Hershberg R.* General Rules for Optimal Codon Choice. / Hershberg R, Petrov D.A. // PLoS Genet. — 2009, Vol. 5, N. 7: e1000556. — P. 1—10.
27. *Hui-Qi Zhou.* Analysis of the Relationship between Genomic GC Content and Patterns of Base Usage, Codon Usage and Amino Acid Usage in Prokaryotes: Similar GC Content Adopts Similar Compositional Frequencies Regardless of the Phylogenetic Lineages. / Hui-Qi Zhou, Lu-Wen Ning, Hui-Xiong Zhang, Feng-Biao Guo. // PLoS ONE — 2014, Vol. 9, N.9: e107319.
28. Codon Usage Database. Режим доступа: <http://www.kazusa.or.jp/codon/>.
29. *Репке Г.* Неравновесная статистическая механика. / Г. Репке. — Москва: Мир, 1990 — 320с.
30. *Cavarelli J.* Recognition of tRNAs by aminoacyl-tRNA synthetases. / Cavarelli J. // FASEB — 1993, Vol. 7, — P. 79—86.
31. *Frappat L.* Universality and Shannon entropy of codon usage. / Frappat L, Minichini C, Sciarrino A., Sorba P. // Phys. Rev. E — 2003, Vol. 68, 061910.
32. *Suzuki H.* The ‘weighted sum of relative entropy’: a new index for synonymous codon usage bias. / Suzuki, H., Saito, R. and Tomita, M. // Gene — 2004, Vol. 335, — P. 19—23.
33. *Wan X.-F.* Codon O: a new informatics method measuring synonymous codon usage bias. / Wan, X.-F., J. Zhou, and D. Xu. // International Journal of General Systems — 2006, Vol. 35, — P. 109—125.
34. *Shannon C.E.* A mathematical theory of communication. / Shannon C.E. // The Bell System Technical Journal — 1948, Vol. 27, — P. 379—423.

35. Zeeberg B. Shannon information theoretic computation of synonymous codon usage biases in coding regions of human and mouse genomes. / Zeeberg B. // *Genome Res.* —2002, Vol. 12, — P. 944—955.
36. Kozubowski T. J. Log-Laplace distributions. / Kozubowski T. J., Podgorski K. // *Int. Math. J.* — 2003, Vol. 3, N. 4, — P. 467—495.

V. V. Stcherbic, L. P. Buchatsky

Taras Shevchenko National University of Kyiv, Ukraine

STATISTICAL REGULARITIES OF SYNONYMOUS CODON USAGE IN THE GENOME
OF DIFFERENT ORGANISMS

On the basis of information taken from the Codon Usage Database, it is established that for six of the investigated groups of organisms, codon usage statistics is determined by two random processes: the Ornstein-Uhlenbeck process and the Wiener process.

Key words: codon usage bias, synonymous codon families, GC3- index, informational entropy, ensembles of organisms

Рекомендує до друку

Н. М. Дробик

Надійшла 29.11.2018