

Рассмотрены показатели, наиболее часто анализируемые при изучении аминокислотного состава белков и нуклеотидного состава нуклеиновых кислот, представлены примеры их использования для решения прикладных задач. Ключевые слова: сайт, виды сайтов, транзиции, трансверсии, ГЦ-насыщенность, аминокислоты групп GARP и FYMINK.

Определение нуклеотидного состава нуклеиновых кислот и аминокислотного состава белков является начальным этапом исследований в области молекулярной эволюции и проводится преимущественно с помощью специальных компьютерных программ (например, серии MEGA [14]). На полученных в ходе этих исследований данных основан анализ характера аминокислотных замен и стратегий кодирования белков в мРНК и ДНК [1], определение степени нейтральности замен нуклеотидов [16], вычисление скорректированных эволюционных дистанций [13, 17], построение дендрограмм [7] и другие исследования в области молекулярной эволюции и эволюционной биохимии.

При анализе нуклеотидного состава ДНК и РНК наиболее часто анализируются следующие показатели: частота транзиций, частота трансверсий, относительные вероятности трансверсий и транзиций, соотношение трансверсий и транзиций (наблюдаемое и расчетное), общая ГЦ-насыщенность, ГЦ-насыщенность отдельных положений кодона, зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности, количество синонимичных и несинонимичных сайтов, количество вырожденных и невырожденных сайтов.

Транзиция – это мутация, обусловленная заменой одного пуринового основания на другое ($A \leftrightarrow G$) или одного пиримидинового на другое ($U, T \leftrightarrow C$) [10]. Транзиции – простые замены, поскольку изменений ориентации пурин-пиримидин в мутантном сайте двухцепочечной молекулы ДНК не происходит. Частота транзиций (P) вычисляется по формуле:

$$P = \alpha / L,$$

где α – число наблюдаемых транзиций; L – общее число нуклеотидных сайтов, по которым сравниваются последовательности.

Трансверсия – это мутация, обусловленная заменой пуринового основания на пиримидиновое и наоборот ($A, G \leftrightarrow U, T, C$) [10]. Трансверсии – сложные или перекрестные замены, поскольку происходит изменение ориентации пурин-пиримидин в мутантном сайте двухцепочечной молекулы ДНК. Частота трансверсий (Q) определяется по формуле:

$$Q = \beta / L,$$

где β – число наблюдаемых трансверсий; L – общее число нуклеотидных сайтов, по которым сравниваются последовательности.

Средние относительные вероятности трансверсий (w) и транзиций ($1-w$) рассчитываются по формуле:

где p – соотношение трансверсии / транзиции согласно таблице генетического кода, q – соотношение трансверсии / транзиции в сравниваемых нуклеотидных последовательностях [6].

Многочисленными исследованиями было показано, что транзиции более вероятны, чем трансверсии, поскольку приводят к более консервативным аминокислотным заменам [1, 3]. Анализ мутаций в кодирующих участках мРНК позволяет определять частоты отдельных

видов транзиций и трансверсий для синонимичных и несинонимичных замен нуклеотидов. Например, при анализе 138 одношаговых мутаций в мРНК, кодирующих алкогольдегидрогеназы класса 4 мыши и человека, установлено, что для несинонимичных замен транзиции $A \leftrightarrow G$ значительно преобладают над транзициями $C \leftrightarrow U$, а для синонимичных замен их частоты приблизительно равны; частоты трансверсий для несинонимичных замен могут быть расположены в ряд $A \leftrightarrow C > G \leftrightarrow C > G \leftrightarrow U > A \leftrightarrow U$, в то время как для синонимичных замен их частоты также приблизительно равны [1].

Соотношение наблюдаемых транзиций и трансверсий (k) определяется по формуле:

$$k = \alpha / \beta.$$

Расчетное соотношение транзиций и трансверсий (R) отличается от соотношения наблюдаемых транзиций и трансверсий k [15]. Определение этого показателя основано на том, что по каждому нуклеотиду может произойти одна транзигция и две различных трансверсии. Тогда пропорция транзиций к общему количеству замен выглядит как:

$$\alpha / (\alpha + 2\beta) = R / (1 + R).$$

При равных частотах транзиций и трансверсий их расчетное соотношение равно 0,5. Известно, что для ядерных генов этот показатель колеблется в пределах от 0,5 до 2 [15]. В отличие от наблюдаемого соотношения транзиций и трансверсий расчетное соотношение может использоваться не только для анализа нуклеотидного состава РНК и ДНК, но и для вычисления синонимичных и несинонимичных дистанций модифицированным методом Нея-Годжбори [7, 13].

Исследования средних расчетных соотношений транзиций и трансверсий в отдельных положениях кодонов мРНК, кодирующих алкогольдегидрогеназы человека, показали их значительные отличия от общего по всем положениям (3,193). Так, наибольшее расчетное соотношение транзиций и трансверсий характерно для третьего положения кодонов (5,836), меньшее – для второго (2,032) и наименьшее – для первого (1,708) [4].

Примером практического использования расчетных соотношений транзиций и трансверсий является разработка структурно-функциональной классификации алкогольдегидрогеназ. В ней в качестве одного из аргументов объединения в один класс алкогольдегидрогеназ 1А, 1В и 1С использованы наибольшие (по сравнению с другими изоферментами) значения расчетных соотношений транзиций и трансверсий в кодирующих их мРНК (1,664, 1,543, 1,407, соответственно) [4].

Общая ГЦ-насыщенность – это общее содержание гуанина и цитозина. Повышение ГЦ-насыщенности мРНК или ДНК в процессе эволюции свидетельствует об увеличении термодинамической стабильности ДНК за счет образования большего количества водородных связей (трех между гуанином и цитозином вместо двух между аденином и тиминном). Так, рост ГЦ-насыщенности в процессе эволюции установлен в мРНК, кодирующих мембраносвязанные аденилатциклазы позвоночных и беспозвоночных животных [3]. Имеются данные и об уменьшении ГЦ-насыщенности мРНК в ходе эволюции (например, мРНК, кодирующих алкогольдегидрогеназы класса 3 хордовых), что объясняется необходимостью увеличения содержания аминокислот группы FУMINK в кодируемом белке [5].

Определение ГЦ-насыщенности информативно и в отдельных положениях кодона – первом, втором и третьем (ГЦ1, ГЦ2 и ГЦ3-насыщенность, соответственно). Однако наибольшее значение имеет определение ГЦ3-насыщенности как маркера мутационного давления [15].

Зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности была описана С. Кумаром и М. Неем [15]. Аналогичная зависимость была

получена нами на примере мРНК, кодирующих алкогольдегидрогеназу класса 3 человека (рис. 1) [1].

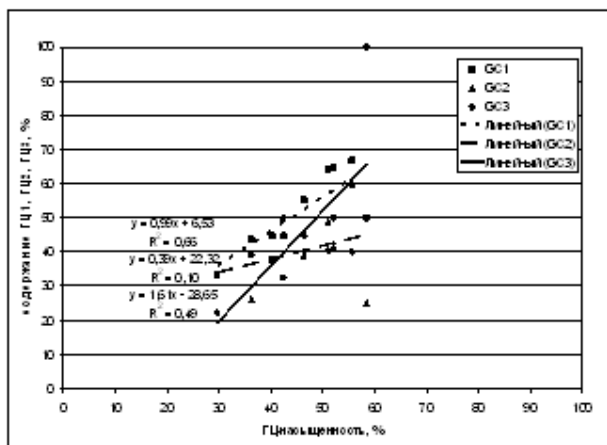


Рисунок 1. Классическая зависимость содержания гуанина и цитозина в отдельных положениях кодона от общей ГЦ-насыщенности на примере экзонов гена, кодирующего алкогольдегидрогеназу класса 3 человека

При анализе этой зависимости было установлено, что с увеличением общей ГЦ-насыщенности наблюдается различной степени линейный рост содержания гуанина и цитозина во всех положениях нуклеотида в кодоне. Наибольший рост характерен для значений ГЦ3 (наклон тренда – 1,61), что свидетельствует о выраженном влиянии на замены в данном положении кодона мутационного давления и слабом влиянии отрицательного отбора. Это объясняется тем, что 72% замен по третьему положению не приводят к изменению кодируемой аминокислоты. Меньший наклон тренда по значениям ГЦ1, равный 0,99, связан с меньшей долей синонимичных замен (5%). Несмотря на то, что все замены по второму положению несинонимичны, наклон тренда по значениям ГЦ2 является положительным (0,39). Объяснить этот факт можно тем, что часть замен аминокислот является нейтральной, то есть не приводит к изменению функции белка. В этом случае отрицательный отбор не оказывает влияния на данную мутацию, и она закрепляется, отражая основное направление нуклеотидных замен. Таким образом, в экзонах гена, кодирующего алкогольдегидрогеназу класса 3 человека, “разрешены” преимущественно замены в третьем положении кодона, в меньшей степени – в первом, а в наименьшей степени – во втором, что связано с вероятностью синонимичных и несинонимичных замен.

Имеются данные о том, что выраженное предпочтение синонимичных замен свидетельствует о структурно-функциональных ограничениях, налагаемых на молекулу, а получение такого графика может выступать в качестве одного из критериев ее эволюционной сформированности. Получение графиков с иным характером зависимости ГЦ-насыщенности отдельных положений кодонов от общей ГЦ-насыщенности мРНК и ДНК, напротив, наводят на мысль, что соответствующие гены находятся в стадии активных эволюционных изменений, то есть кодируют эволюционно “несовершенные” белки [1].

Сайт – это участок небольшого размера молекулы нуклеиновой кислоты или белка, обычно равный одному нуклеотиду или одному аминокислотному остатку, соответственно. Существуют 2 системы классификации сайтов нуклеиновых кислот в связи с их положением в кодоне – система вырожденности и синонимичности/несинонимичности (табл. 1) [2].

Таблица 1. Системы вырожденности и синонимичности/несинонимичности сайтов

Система классификации сайтов	
по вырожденности	по синонимичности/несинонимичности
невырожденный	абсолютно несинонимичный
двухкратновырожденный	несинонимичный на $\frac{2}{3}$ синонимичный на $\frac{1}{3}$
трехкратновырожденный	несинонимичный на $\frac{1}{3}$ синонимичный на $\frac{2}{3}$
четырекратновырожденный	абсолютно синонимичный

Нолькратно вырожденный сайт (невырожденный сайт) – это нуклеотидный сайт, в котором любая возможная замена является несинонимичной. Пример: в первом положении кодона ЦУУ (Лей) возможны 3 несинонимичные замены: ЦУУ → УУУ (Фен), ЦУУ → АУУ (Иле) и ЦУУ → ГУУ (Вал). Следовательно, первый сайт кодона ЦУУ является невырожденным.

Двухкратно вырожденный сайт – это нуклеотидный сайт, в котором одна из трех возможных замен является синонимичной. Пример: в третьем положении кодона УУУ (Фен) возможны 2 несинонимичные и 1 синонимичная замена: УУУ → УУЦ (Фен), УУУ → УУА (Лей) и УУУ → УУГ (Лей). Следовательно, третий сайт кодона УУУ является двухкратно вырожденным.

Трехкратно вырожденный сайт – это нуклеотидный сайт, в котором две замены из трех возможных синонимичны. Пример: в третьем положении кодона АУУ (Иле) возможны 1 несинонимичная и 2 синонимичные замены: АУУ → АУЦ (Иле), АУУ → АУА (Иле) и АУУ → АУГ (Мет). Следовательно, третий сайт кодона АУУ является трехкратно вырожденным. Многие исследователи не выделяют трехкратновырожденные сайты в отдельную группу, подсчитывая их количество вместе с двухкратновырожденными.

Четырекратно вырожденный сайт – это нуклеотидный сайт, в котором возможны лишь синонимичные замены. Пример: в третьем положении кодона ЦУУ (Лей) возможны 3 синонимичные замены: ЦУУ → ЦУЦ (Лей), ЦУУ → ЦУА (Лей) и ЦУУ → ЦУГ (Лей). Следовательно, третий сайт кодона ЦУУ является четырекратно вырожденным.

В соответствии со стандартной таблицей генетического кода доля невырожденных сайтов по всем положениям кодона равна 63,39%, двухкратновырожденных – 19,13% и четырекратновырожденных сайтов – 7,49%. Близкие данные о распределении вырожденных сайтов получены и для последовательностей мРНК, кодирующих алкогольдегидрогеназы человека [2]. При анализе содержания "0"-кратно вырожденных сайтов в отдельных положениях кодона (в соответствии с таблицей генетического кода) установлено, что 86,89% сайтов в первом положении, 100,0% сайтов во втором положении и 3,28% сайтов в третьем положении являются невырожденными. Двухкратновырожденные сайты встречаются в первом и третьем, а четырекратновырожденные сайты – лишь в третьем положении кодона (их теоретическое содержание для этого положения составляет 52,46%) [2].

Несинонимичный сайт – это сайт, в котором возможна несинонимичная замена. Синонимичный сайт – это сайт, в котором возможна синонимичная замена. Если все замены в данном сайте несинонимичны, сайт называется абсолютно несинонимичным. Если все замены в данном сайте синонимичны, сайт называется абсолютно синонимичным. Если одна из замен в данном сайте несинонимична, сайт называется несинонимичным на $\frac{1}{3}$

(соответственно, синонимичным на 2/3). Если две замены в данном сайте несинонимичны, сайт называется несинонимичным на 2/3 (соответственно, синонимичным на 1/3) [10].

В соответствии с таблицей генетического кода теоретически ожидаемое содержание синонимичных и несинонимичных сайтов составляет 25,52% и 74,48%, соответственно. Имеются исследования [2], посвященные сравнению доли вырожденных и невырожденных, синонимичных и несинонимичных сайтов в последовательностях мРНК, кодирующих ряд белков, с теоретически ожидаемыми.

При анализе аминокислотного состава белков чаще анализируются:

- в связи с зарядом аминокислот – содержание положительнозаряженных, отрицательнозаряженных и нейтральных аминокислот;
- в связи с химической структурой – содержание ароматических, серосодержащих, гидроксильных аминокислот и др.;
- в связи с ГЦ-насыщенностью – содержание аминокислот GARP (кодируемых ГЦ-богатыми кодонами) и FYMINK (кодируемых ГЦ-бедными кодонами, табл. 2).

Таблица 2. Аминокислоты групп GARP и FYMINK и соответствующие им кодоны мРНК

Аминокислоты группы GARP		Аминокислоты группы FYMINK			
G	глицин	ГГУ			
		ГГЦ			
		ГГА			
		ГГГ	F	фенилаланин	УУУ
A	аланин	ГЦУ	F	фенилаланин	УУЦ
		ГЦЦ	Y	тирозин	УАУ
		ГЦА	Y	тирозин	УАЦ
		ГЦГ	M	метионин	АУГ
R	аргинин	ЦГУ	I	изолейцин	АУУ
		ЦГЦ			АУЦ
		ЦГА			АУУ
		ЦГГ	N	аспарагин	ААУ
		АГА			ААЦ
		АГГ			ААА
P	пролин	ЦЦУ	K	лизин	ААГ
		ЦЦЦ			
		ЦЦА			
		ЦЦГ			

При анализе мРНК, кодирующих алкогольдегидрогеназы класса 3 хордовых (рис. 2 [5]) установлено, что с ростом их ГЦ-насыщенности наблюдается линейный рост содержания аминокислот группы GARP ($r = 0,72 \pm 0,310$, $p < 0,05$) и линейное падение содержания аминокислот группы FYMINK ($r = -0,88 \pm 0,211$, $p < 0,001$). Аналогичная закономерность была получена и для 85 совокупностей кодирующих участков бактериальных мРНК, 74 совокупностей кодирующих участков двухцепочечных РНК-вирусов и мРНК, кодирующих ферменты дыхательной цепи человека, трихинеллы и цианорабдитис [8, 9, 11].

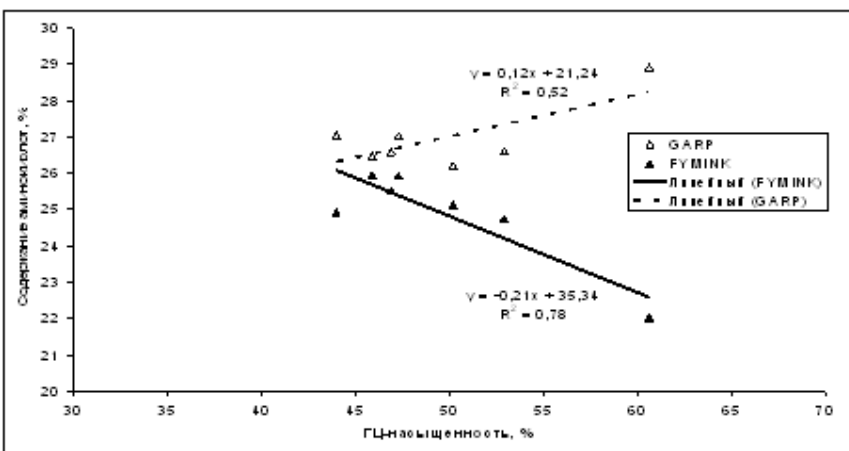


Рисунок 2. Зависимость содержания GARP и FYMINK в алкогольдегидрогеназах класса 3 хордовых от GC-насыщенности кодирующих их мРНК

Данное наблюдение объясняется тем, что все четырехкратно вырожденные серии кодонов, кодирующих глицин, аланин, аргинин и пролин, являются абсолютно GC-насыщенными или GC-насыщенными по первому и второму положениям (за исключением кодонов АГА и АГГ, соответствующих аргинину). Следовательно, с увеличением GC-насыщенности мРНК данные кодоны будут использоваться чаще, что и обеспечит рост содержания глицина, аланина, аргинина и пролина [1].

Вместе с тем аминокислоты группы FYMINK кодируются либо абсолютно GC-бедными либо относительно GC-богатыми кодонами. И поэтому с увеличением GC-насыщенности данные кодоны будут использоваться реже, что повлечет уменьшение содержания фенилаланина, тирозина, метионина, изолейцина, аспарагина и лизина [1].

Еще одной группой показателей, определяемых при изучении аминокислотного состава белков и нуклеотидного состава нуклеиновых кислот на основании предварительно проведенных выравниваний (например, с помощью алгоритма ClustalW [12]), являются количество константных, переменных, синглетонных и парсимоничных сайтов.

Константные сайты – это нуклеотидные или аминокислотные сайты выровненных последовательностей, в которых не наблюдаются замены.

Переменные сайты – это нуклеотидные или аминокислотные сайты выровненных последовательностей, в которых наблюдаются замены.

Синглетонные сайты – это переменные сайты не менее 5 выровненных последовательностей, в которых преобладает частота одного нуклеотида или аминокислотного остатка [10]. Пример: сайты № 192, 207 на рисунке 3.

Парсимоничные сайты – это переменные сайты не менее 5 выровненных последовательностей, в которых как минимум два нуклеотида или аминокислотных остатка встречаются хотя бы два раза [10]. Пример: сайты №191, 192 на рисунке 3.

190	200	210	220	230	240	250
H. s.	EKEQQQLIDDHFLFDKPVSP	LLASGMARDWPDARGIWHNDNKSFLVWVNEEDHLRVISM				
C. f.	EQEQQQLIDDHFLFDKPVSP	LLASGMARDWPDARGIWHNDNKTFLLVWVNEEDHLRVISM				
B. t.	EQEQQQLIDDHFLFDKPVSP	LLASGMARDWPDARGIWHNDNKSFLVWVNEEDHLRVISM				
O. c.	EQEQQQLIDDHFLFDKPVSP	LLASGMARDWPDARGIWHNDNKSFLVWVNEEDHLRVISM				
M. m.	EQEQQQLIDDHFLFDKPVSP	LLASGMARDWPDARGIWHNDNKSFLVWVNEEDHLRVISM				
R. n.	EQEQQQLIDDHFLFDKPVSP	LLASGMARDWPDARGIWHNDNKSFLVWVNEEDHLRVISM				
G. g.	EQEQQQLIDDHFLFDKPVSP	LLASGMARDWPDARGIWHNDNKTFLLVWVNEEDHLRVISM				
Z. d.	DAEQQLIDDHFLFDKPI	SPLLASGMARDWPDARGIWHNDNKTFLLVWVNEEDHLRVISM				
I. p.	DAEQQLIADHFLFDKPVSP	LLAAGMARDWPDARGIWHNDEKTFLLVWVNEEDHLRVISM				
B. f.	DAEQQLIADHFLFDKPVSP	LLTCAGMARDWPDARGIWHNNEKSFLIWIINEEDHLRVISM				
C. i.	EEHQQLINDHFLFDKPVSP	LLASGMARDWPDARGIWHNDKKNFLVWVNEEDHLRVISM				
:	.*:***	*****:****	..:*****	*****:*	.*:*	*****

Рисунок 3. Выровненные участки каталитических доменов М-изоферментов креатинкиназ позвоночных и соответствующие участки креатинкиназ ланцетника и оболочника.

Анализ синглетонных и парсимоничных сайтов является перспективным направлением исследований в области молекулярной эволюции и эволюционной биохимии. Определение парсимоничных сайтов является основой метода максимальной парсимонии (экономии [7]), являющегося наиболее часто используемым дискретным методом построения дендрограмм (рис. 4).

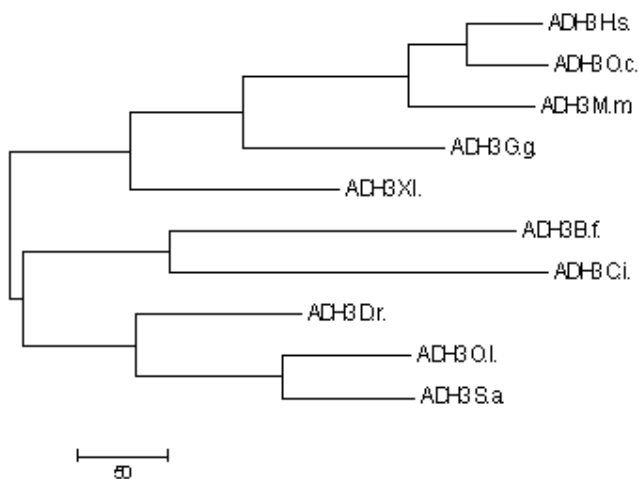


Рисунок 4. Дендрограмма для мРНК, кодирующих алкогольдегидрогеназы класса 3 хордовых, построенная методом максимальной парсимонии.

Интенсивное пополнение международных банков данных секвенированными последовательностями нуклеиновых кислот и белков обуславливает сохранение актуальности изучения их состава. Вместе с тем появляются и новые оригинальные подходы [5] к изучению состава белков и нуклеиновых кислот, открывающие новые перспективы для исследований в этой области молекулярной эволюции как науки.

Литература

1. Алкогольдегидрогеназы хордовых животных: монография / А. В. Бутвиловский, Е. В. Барковский, В. Э. Бутвиловский; под общ. ред. Е.В. Барковского. Минск: БГМУ, 2007. 144 с.
2. Барковский, Е. В. Вырожденность, синонимичность и несинонимичность нуклеотидных сайтов в последовательностях мРНК алкоголь-дегидрогеназ человека 1–7 типов / Е. В. Барковский, А. В. Бутвиловский // Материалы международного симпозиума “Молекулярные механизмы регуляции функции клетки”. Тюмень: Издательство “ВекторБук”, 2005. С. 273–275.
3. Барковский, Е. В. Мембраносвязанные аденилатциклазы: монография / Е. В. Барковский, О. В. Ачинович; под общ. ред. Е. В. Барковского. Минск: БГМУ, 2005. 134 с.
4. Бутвиловский, А. В. Эволюционные взаимоотношения и структурно-функциональная классификация алкогольдегидрогеназ человека / А. В. Бутвиловский, Е. В. Барковский, В. Э. Бутвиловский // Вестник ВГМУ. Витебск, 2006. № 2. С. 22–27.
5. Изменения в процессе эволюции мутационного давления в последовательностях мРНК, кодирующих алкогольдегидрогеназы класса 3 хордовых животных / А. В. Бутвиловский [и др.] // Медицинский журнал. Минск, 2007. № 1. С. 22–25.
6. Кимура, М. Молекулярная эволюция: теория нейтральности / М. Кимура. М., 1985. 398 с.
7. Методы молекулярной эволюции и филогенетики: учеб.-метод. пособие / Е. В. Барковский [и др.]. Минск: БГМУ, 2005. 63 с.
8. О сходстве стратегий кодирования митохондриальных белков человека и трихинеллы. Часть 1. ГЦ-насыщенность, доля ГЦЗ-кодонов и частота использования претерминальных кодонов / В. Э. Бутвиловский, Ю. И. Линник, А. В. Бутвиловский, Е. В. Барковский // Медицинский журнал. Минск, 2007. № 3. С. 39–41.
9. О сходстве стратегий кодирования митохондриальных белков человека и трихинеллы. Часть 2. Картина использования синонимичных кодонов. Содержание аминокислотных групп GARP и FYMINK / В. Э. Бутвиловский, Е. В. Барковский, А. В. Бутвиловский, Ю. И. Линник // Медицинский журнал. Минск, 2007. № 4. С. 39–42.
10. Справочник терминов молекулярной эволюции и филогенетики. Учебно-методическое пособие / В. Э. Бутвиловский [и др.] // Минск: БГМУ, 2006. 40 с.
11. Хрусталева, В.В. Частота использования претерминальных кодонов в матричных РНК бактериальных аденилатциклаз / В. В. Хрусталева, Е. В. Барковский // Здравоохранение. 2006. № 2. С. 17–20.
12. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions – specific gap penalties and weight matrix choice / J.D. Tompson, D.G. Higgins, T.J. Gibson // Nucl. Acids Res. 1994. Vol. 22. P. 4673–4680.
13. Ina, Y. New methods for estimating the numbers of synonymous and nonsynonymous substitutions / Y. Ina // Mol. Biol. Evol. 1995. Vol. 40. P. 190–226.
14. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment / S. Kumar, K. Tamura, M. Nei // Brief. Bioinform. 2004. Vol. 5. P. 150–163.
15. Nei, M. Molecular Evolution and Phylogenetics / M. Nei, S. Kumar. Oxford University Press, New York, 2000. 333 p.
16. Sueoka, N. Directional mutation pressure and neutral molecular evolution / N. Sueoka // Proc. Natl. Acad. Sci. USA. 1988. Vol. 85. P. 2653–2657.

17. Tamura, K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C – content biases / K. Tamura // Mol. Biol. Evol. 1992. Vol. 9. P. 678–687.