

Оригинальная компьютерная программа для изучения стратегии кодирования белков в матричных рнк – VVK 3.2

Белорусский государственный медицинский университет

В течение второй половины XX века специалистами в области молекулярной эволюции были разработаны многочисленные методы вычисления эволюционных дистанций [3] и филогенетических взаимоотношений [7] между молекулами генетических биополимеров (ДНК, РНК и белками). Все они основаны на определении индексов сходства между выровненными последовательностями нуклеотидов (аминокислотных остатков) в гомологичных генах (белках). Тем не менее, получаемые с их помощью результаты не являются исчерпывающими.

Общепринятым является тот факт, что наибольшей ценностью обладает первичная информация. В генетике же первичной информацией является [2] стратегия кодирования – картина использования кодонов в кодирующих участках ДНК (мРНК).

Согласно представлениям Н. Суеоки, мутационное давление может приводить не только к перераспределению частот использования синонимичных кодонов, но и вызывать направленные аминокислотные замены [8]. В случае если последние являются нейтральными, они могут фиксироваться в популяции случайным генетическим дрейфом [5]. Руководствуясь этими представлениями, можно сказать, что движущая сила эволюции, дающая материал для естественного отбора и случайного генетического дрейфа, эндогенна. Эта движущая сила (мутационное давление) является результатом повышенной частоты возникновения и фиксации в популяции нуклеотидных замен (транзиций) аденина на гуанин и тимина на цитозин (GC-давление) или гуанина на аденин и цитозина на тимин (AT-давление). В течение второй половины XX века специалистами в области молекулярной эволюции были разработаны многочисленные методы вычисления эволюционных дистанций [3] и филогенетических взаимоотношений [7] между молекулами генетических биополимеров (ДНК, РНК и белками). Все они основаны на определении индексов сходства между выровненными последовательностями нуклеотидов (аминокислотных остатков) в гомологичных генах (белках). Тем не менее, получаемые с их помощью результаты не являются исчерпывающими.

Общепринятым является тот факт, что наибольшей ценностью обладает первичная информация. В генетике же первичной информацией является [2] стратегия кодирования – картина использования кодонов в кодирующих участках ДНК (мРНК).

Согласно представлениям Н. Суеоки, мутационное давление может приводить не только к перераспределению частот использования синонимичных кодонов, но и вызывать направленные аминокислотные замены [8]. В случае если последние являются нейтральными, они могут фиксироваться в популяции случайным генетическим дрейфом [5]. Руководствуясь этими представлениями, можно сказать, что движущая сила эволюции, дающая материал для естественного отбора и случайного генетического дрейфа, эндогенна. Эта движущая сила (мутационное давление) является результатом повышенной частоты возникновения и фиксации в популяции нуклеотидных замен (транзиций) аденина на гуанин и тимина на цитозин (GC-давление) или гуанина на аденин и цитозина на тимин (AT-давление). В недавних

работах [8] был доказан вклад замен (трансверсий) аденина на цитозин в создание GC-давления, однако, частота закрепления подобных замещений меньше, чем вышеупомянутых транзиций [1,4]. Силу и направление мутационного давления для определенного организма или участка его генома можно определить, зная стратегию кодирования соответствующего белка в ДНК (мРНК).

Компьютерная программа VVK 3.2. была создана для изучения стратегии кодирования как отдельных ферментов и их филогенетических рядов, так и для широкомасштабных исследований с использованием нескольких десятков совокупностей кодирующих участков (СКУ) мРНК бактерий.

В специальной литературе описаны показатели, которые не вычисляются программой VVK 3.2. По нашему мнению, все эти показатели являются вторичными по отношению к тем, которые рассчитываются данной программой. К тому же, часть из них не учитывает гипотезы Н.Суеоки о направленном мутационном давлении.

Для демонстрации возможностей программы VVK 3.2. были использованы:

- 85 СКУ бактерий с общим числом нуклеотидных последовательностей 273 590.
- 59 нуклеотидных последовательностей бактериальных мРНК, кодирующих аденилатциклазы актиномицетов, цианобактерий, спирохет, протеобактерий, планктомицетов и фирмикутов.

- Нуклеотидные последовательности, соответствующие кофермент-связывающему домену алкогольдегидрогеназы III класса 10 хордовых животных [7].

Программа VVK 3.2. работает с известной нуклеотидной последовательностью кодирующих участков мРНК. Нуклеотидная последовательность вводится латинскими символами в рабочую ячейку программы без пробелов, знаков препинания и знаков конца абзаца. Для обозначения нуклеотидов используются общепринятые однобуквенные обозначения: аденин – А, гуанин – G, урацил – U, цитозин – C. После введения нуклеотидной последовательности программа рассчитывает все описанные ниже показатели: G+C; 1GC; 2GC; 3GC; ПТК; C 1,2; C 4,6; доля ПТК среди C 4,6; а также частоты использования каждого кодона и каждой аминокислоты.

Стратегию кодирования определенного белка обычно сравнивают с таковой для гомологичных ему белков, образующих один филогенетический ряд. Для этого необходимо владеть нуклеотидными последовательностями, соответствующими гомологичным белкам нескольких видов живых организмов. При работе с филогенетическими рядами программа VVK 3.2. создаёт базы данных с частотами использования кодонов для каждой последовательности. На основании этой базы данных можно строить три типа графиков, примеры которых приведены ниже. Для филогенетических рядов также рассчитываются коэффициенты корреляции всех показателей с величиной G+C, их наклон относительно G+C и достоверность аппроксимации данных зависимостей к линии тренда.

Стратегия кодирования одинакова для всех белков данного организма, если мутационное давление равномерно распространяется на весь его геном. При изучении стратегии кодирования у таких организмов желательно использовать СКУ мРНК. Частота использования каждого кодона для СКУ мРНК является среднестатистической для всех кодирующих участков данного генома. Такой подход к изучению стратегии кодирования гарантирует высокую степень достоверности полученных данных. После вычисления средних значений частот использования кодонов для каждого изучаемого вида создается база данных, аналогичная таковой для филогенетических рядов.

Наиболее ценным показателем для определения стратегии кодирования белка в ДНК (мРНК) является показатель G+C. Данный показатель равен доле гуанина и цитозина в нуклеотидной последовательности. Также важны показатели 1GC, 2GC и 3GC. Каждый из этих показателей равен доле гуанина и цитозина в нуклеотидной последовательности только в первых, только во вторых или только в третьих положениях кодонов, соответственно.

Ценность этих показателей продемонстрирована нами при изучении 85 СКУ бактерий, общее число нуклеотидных последовательностей в которых составило 273 590.

На рисунке 1 изображён график, отражающий корреляцию показателей 1GC, 2GC и 3GC с общим содержанием G+C. Наличие прямой зависимости между каждым из этих трёх показателей и общей GC-насыщенностью указывает на то, что мутационное давление оказывает огромное влияние на стратегию кодирования. Показатель 3GC демонстрирует самый высокий наклон (1,84) относительно G+C. Это является следствием того, что отрицательный отбор далеко не всегда может оказывать влияние на содержание гуанина и цитозина в третьем положении: большинство нуклеотидных замен по третьему положению не приводит к изменению состава кодируемого белка [5,8]. Показатель 1GC имеет гораздо меньший наклон (0,71), чем показатель 3GC. Этот факт можно объяснить тем, что лишь некоторая часть замен по первому положению – синонимична. Несмотря на то, что все замены по второму положению несинонимичны, показатель 2GC демонстрирует определенную линейную зависимость от общего содержания G+C (наклон = 0,45). Объяснить это можно, приняв во внимание тот факт, что не все замены аминокислотных остатков приводят к потере или ухудшению функции белка. Какую-то часть аминокислотных замен можно считать нейтральной. В том случае, если мутация нейтральна, отрицательный отбор на неё не влияет, и она закрепляется, запечатлевая в себе основное направление нуклеотидных замен. Из этого следует вывод, что наиболее «чувствителен» к мутационному давлению показатель 3GC.

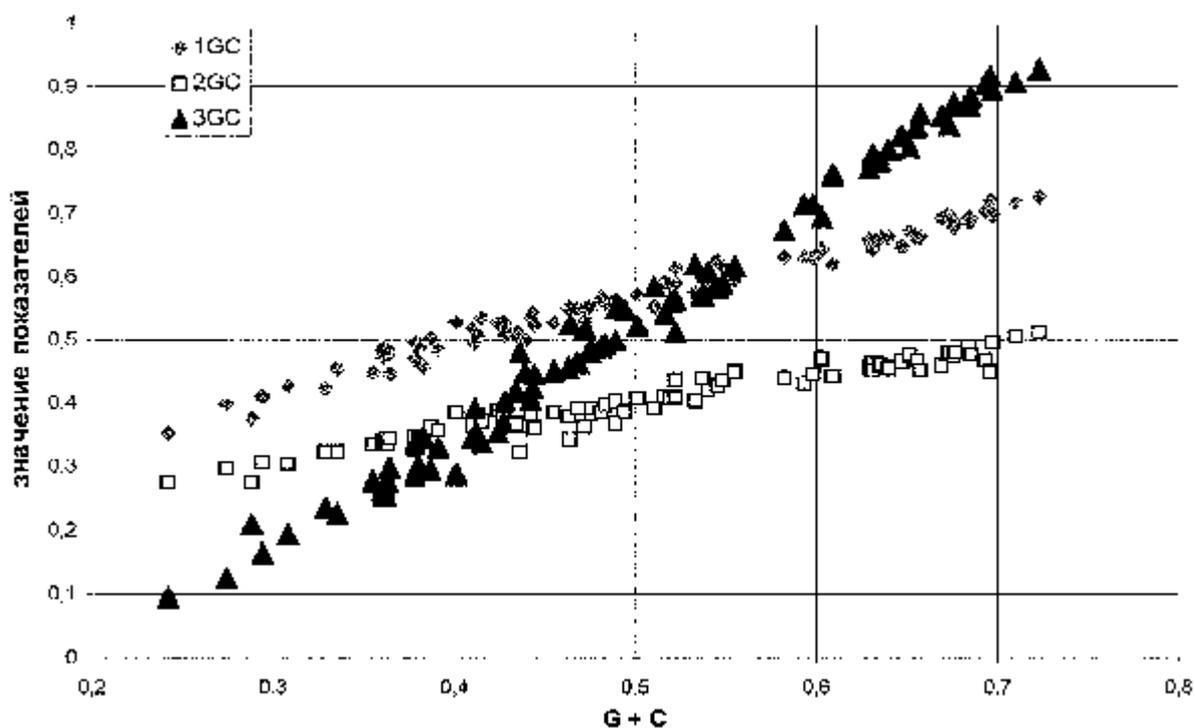


Рис.1. Зависимость показателей 1GC, 2GC и 3GC от G+C в СКУ мРНК бактерий
Отсутствие наклона или отрицательный наклон 2GC относительно G+C

указывает на отсутствие влияния мутационного давления на аминокислотный состав соответствующего белка или его участка. Для нуклеотидной последовательности кофермент-связывающего домена АДГ III класса [7] наклон $2GC = -0,21$ (см. рис.2). Отрицательная линейная зависимость 2GC от общей GC-насыщенности свидетельствует о существовании жесткого отрицательного отбора относительно фиксации в популяции аминокислотных замен, происходящих в данном домене под воздействием мутационного давления.

В процессе изучения влияния естественного отбора и мутационного давления на использование претерминальных кодонов [2] мы пользовались следующими оригинальными показателями: ПТК – частота использования претерминальных кодонов в нуклеотидной последовательности; С 1,2 – частота использования претерминальных кодонов из 2-кратно вырожденных серий, кодирующих аминокислотные остатки Tyr, Gln, Lys, Glu, Cys, и невырожденного кодона UGG, кодирующего Trp; С 4,6 – частота использования кодонов 4-и 6-кратно вырожденных серий (среди которых имеются ПТК), кодирующих аминокислотные остатки: Leu, Arg, Gly и Ser; Доля ПТК в С 4,6 – доля претерминальных триплетов в 4-и 6-кратно вырожденных сериях, кодирующих аминокислотные остатки: Leu, Arg, Gly и Ser.

На примере бактериальных аденилатциклаз нами были получены следующие выводы (см. рис.3). Частота использования ПТК в кодирующих участках мРНК бактериальных аденилатциклаз находится в строгой обратной зависимости от G+C. Коэффициент корреляции G+C и ПТК $= -0,9$. Обнаружена высокая степень корреляции ($r=0,9$) между показателем ПТК и С 1,2 и обратная зависимость между С 1,2 и G+C ($r=-0,74$). Показатель С 4,6 в мРНК бактериальных аденилатциклаз не проявляет сильной корреляции с общим содержанием ПТК ($r=-0,53$), но коэффициент корреляции С 4,6 с G+C равен 0,65. Тем не менее, значительное влияние на общее использование ПТК оказывает перераспределение частот использования кодонов, кодирующих Leu, Arg, Gly и Ser (см. рис.2). Доля ПТК в С 4,6 коррелирует с показателем ПТК ($r=0,83$) и демонстрирует обратную зависимость от G+C ($r=-0,91$).

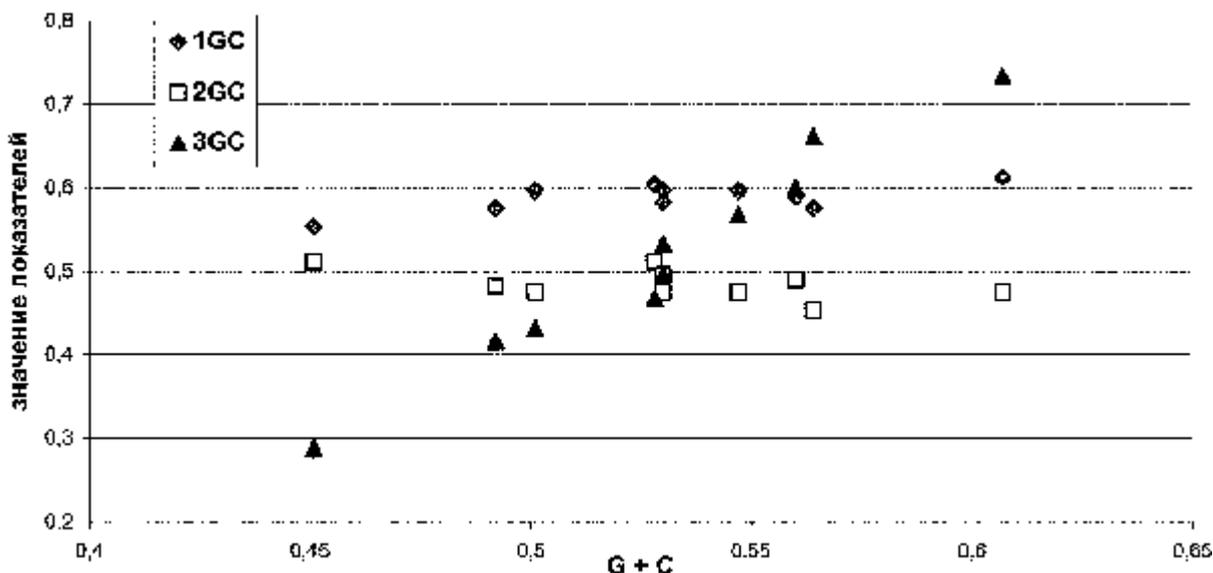


Рис.2. Зависимость показателей 1GC, 2GC и 3GC от G+C в нуклеотидных последовательностях кофермент-связывающего домена АДГ III класса 10 хордовых животных

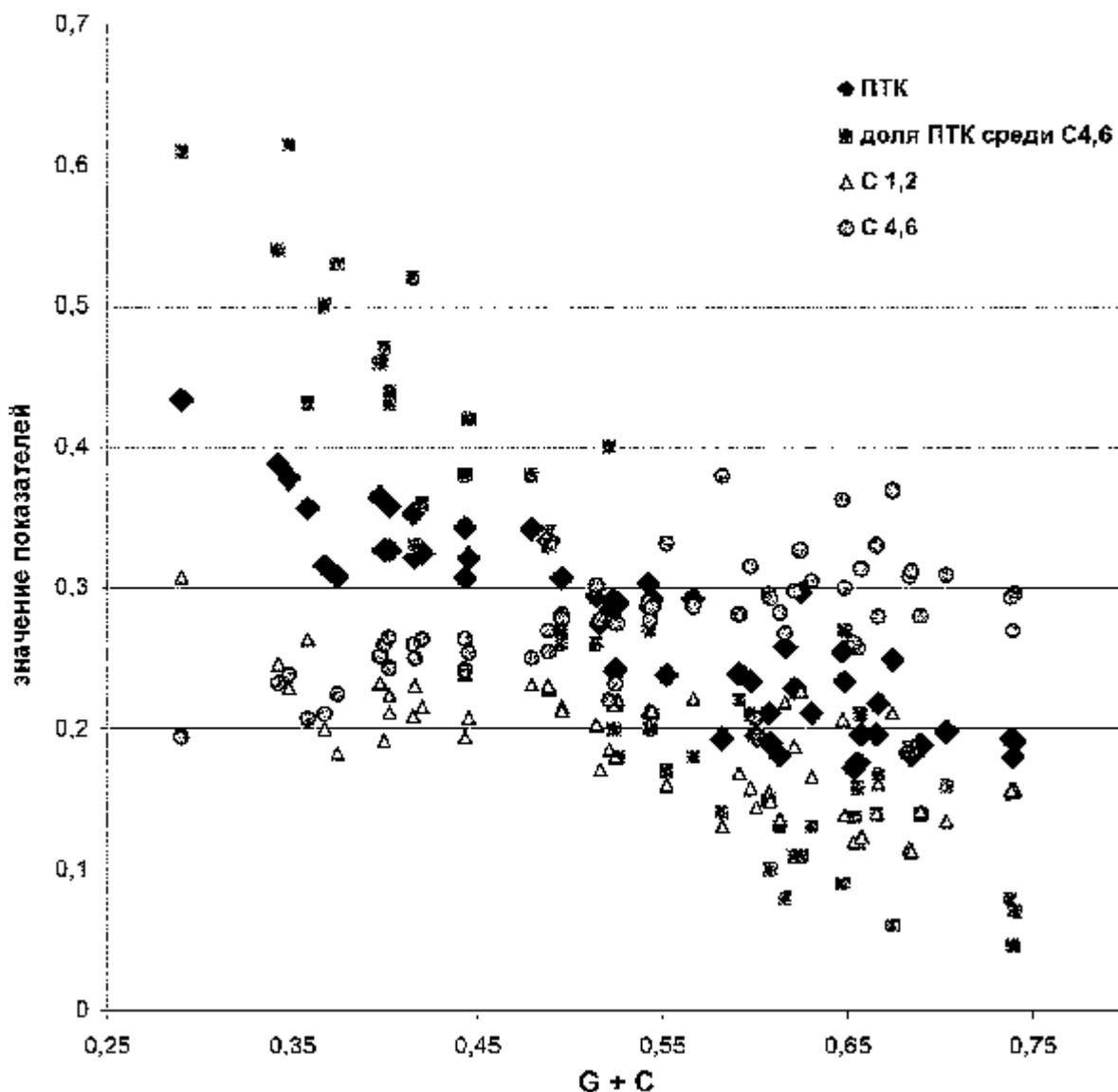


Рис.3. Зависимость показателей ПТК, С 1,2, С 4,6, Доля ПТК среди С 4,6 от G+C в кодирующих участках бактериальных аденилатциклаз

Как видно из рисунка 3, частота использования ПТК в кодирующих участках мРНК бактериальных аденилатциклаз находится в строгой обратной зависимости от G+C. Этот факт свидетельствует о большей степени влияния мутационного давления на использование ПТК, нежели чем отрицательного естественного отбора [7]. Наибольший вклад в этот процесс вносит перераспределение частот использования кодонов из четырех-и шестикратно вырожденных серий.

Для изучения влияния мутационного давления и естественного отбора на частоту использования отдельной аминокислоты нами используются следующие показатели: наклон частоты использования кодона относительно G+C (N) и величина достоверности аппроксимации зависимости частоты его использования от G+C к прямой линии тренда (R^2). Для относительно GC-богатого кодона N и R^2 вычисляются на участках G+C0,6. Для относительно GC-бедного кодона N и R^2 вычисляются на участках G+C0,4.

Целесообразность использования данных показателей продемонстрирована на примере использования кодонов из серии цистеина в SKU бактериальных мРНК (см. рис.4).

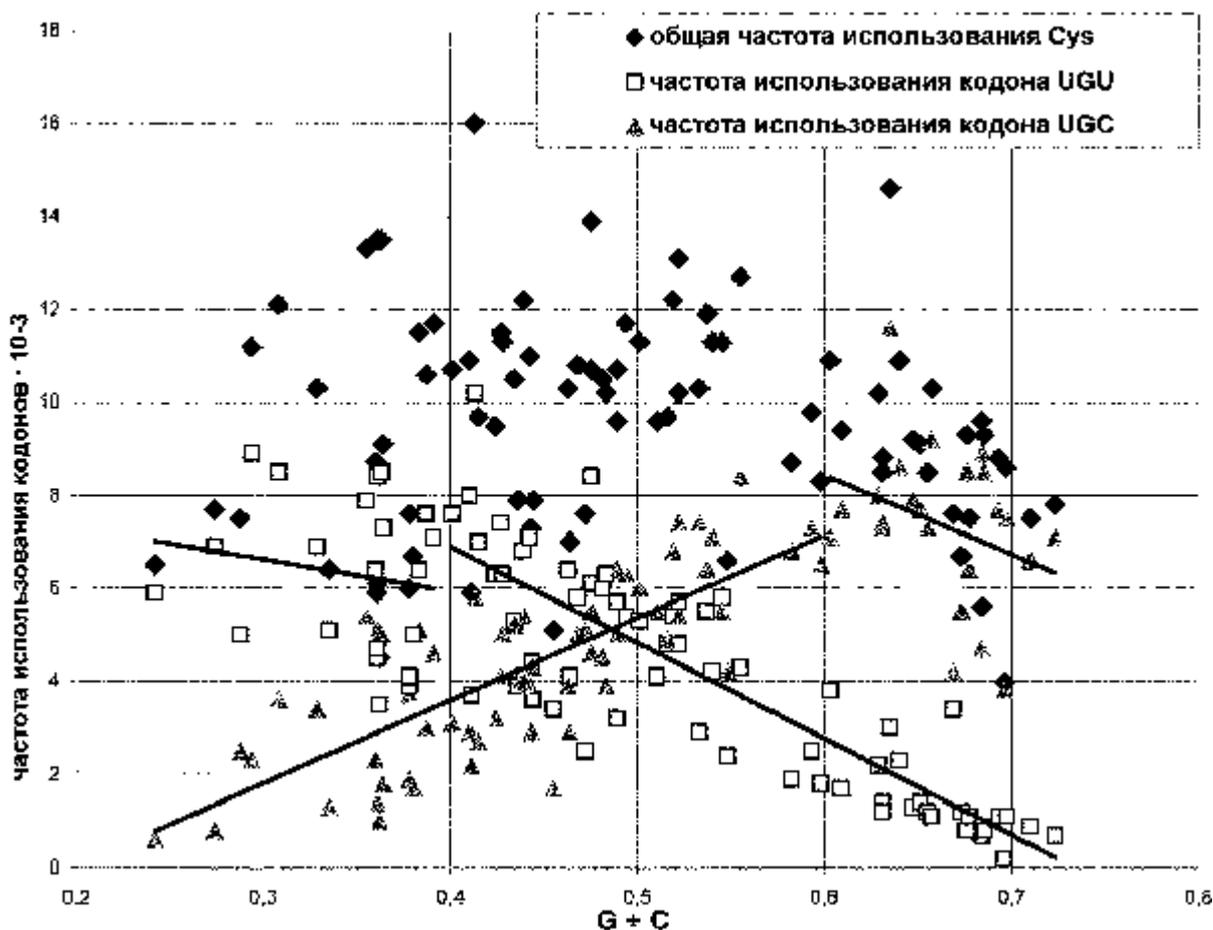


Рис.4. Зависимость частот использования цистеина и кодирующих его триплетов от G+C в SKU мРНК бактерий

Аминокислота цистеин кодируется двумя кодонами: UGU и UGC. При $G+C > 0,4$ для кодона UGU показатель $N = -20,6$, а $R^2 = 0,723$. Это говорит о существовании обратной линейной зависимости между частотой использования UGU и показателем G+C для микроорганизмов с $G+C > 0,4$. При $G+C < 0,6$ для кодона UGC показатель $N = 17,8$, а $R^2 = 0,587$. Это говорит о существовании прямой линейной зависимости между частотой использования UGC и показателем G+C для микроорганизмов с $G+C < 0,6$. Это объясняется тем, что при G+C от 0,4 до 0,6 кодоны из одной и той же серии под воздействием мутационного давления могут свободно превращаться друг в друга. Структурной основой такого превращения в данном случае являются транзиции U на C и C на U по третьему положению. Такие нейтральные замены легко могут быть фиксированы в популяции случайным генетическим дрейфом.

При сильном GC-давлении ($G+C > 0,6$) возрастает вероятность возникновения мутации UGC в CGC. В результате этой мутации на месте цистеина в молекуле белка появляется аргинин. Если подобная замена нейтральна, она может быть зафиксирована в популяции. Действительно, при $G+C > 0,6$ для кодона UGC показатель $N = -17,1$ при $R^2 = 0,103$. Значительный разброс частот использования UGC среди различных видов бактерий при сильном GC-давлении и общая тенденция к снижению использования этого кодона говорят о том, что процесс замещения цистеина

аргинином действительно имеет место, причём носит стохастический характер. С другой стороны, при АТ-давлении кодон UGU может превращаться в UAU, вызывая замену цистеина на тирозин. Определенную степень нейтральности таких замен у бактерий подтверждает большой разброс частоты использования кодона UGU при $G+C < 0,4$ ($R^2 = 0,028$). Наклон частоты использования UGU относительно $G+C$ значительно меньше при $G+C > 0,4$.

На основании такого относительно простого анализа можно сделать следующий вывод: аминокислотные замены цистеина на аргинин (происходящие под воздействием сильного GC-давления) и цистеина на тирозин (происходящие под воздействием сильного АТ-давления) подвергаются малому влиянию отрицательного отбора, то есть, данные аминокислотные замены относительно нейтральны.

Таким образом, показатели, вычисляемые с помощью программы VVK 3.2., предоставляют исчерпывающую информацию о стратегии кодирования белка в мРНК; программа VVK 3.2. может использоваться для анализа динамики изменения стратегии кодирования белков как в нуклеотидных последовательностях гомологичных белков в широком филогенетическом ряду, так и в SKU мРНК обширной выборки микроорганизмов.

Литература

1. Барковский Е.В. Мембраносвязанные аденилатциклазы: монография / Е.В. Барковский, О.В. Ачинович; под общ. ред. Е.В. Барковского. – Мн.: БГМУ, 2005. 134 с.
2. Барковский Е.В., Хрусталеv В.В. Сравнительная характеристика матричных РНК и аминокислотных последовательностей аденилатциклазы цианобактерий//Белорусский мед. журнал. – 2005. – №1. – стр.23-27.
3. Барковский Е.В., Козюлевич С.Р., Бутвиловский А.В., Хрусталеv В.В., Ачинович О.В. Характеристика методов определения эволюционных расстояний между нуклеотидными последовательностями генетических макромолекул//Здравоохранение. – 2005.-№5. – стр.37-43.
4. Бутвиловский А.В., Хрусталеv В.В. Транзиции и трансверсии в нуклеотидных последовательностях мРНК алкогольдегидрогеназ классов I, II, III и IV мыши и человека//Молодежная наука и современность. – Курск, 2005. – часть I. – стр.53-54.
5. Кимура М. Молекулярная эволюция: теория нейтральности / М. Кимура. М., 1985. 398 с.
6. Хрусталеv В.В., Барковский Е.В., Бутвиловский А.В., Козюлевич С.Р., Ачинович О.В. Основные методы анализа эволюционных отношений между последовательностями генетических макромолекул//Здравоохранение. – 2005.-№8. – стр.11-13.
7. Хрусталеv В.В., Бутвиловский А.В. Влияние мутационного давления на использование претерминальных кодонов в мРНК алкогольдегидрогеназ III класса хордовых животных//Молодежная наука и современность. – Курск, 2005. – часть I. – стр.81-82
8. Sueoka N. Wide intra-genomic G+C heterogeneity in human and chicken is mainly due to strand-symmetric directional mutation pressures: dGTP-oxidation and symmetric cytosine-deamination hypotheses//Gene. – 2002. – Vol.300. – P.141-154