

МОДЕЛЬ СОВМЕСТНОГО ВЛИЯНИЯ ЕСТЕСТВЕННОГО ОТБОРА И МУТАЦИОННОГО ДАВЛЕНИЯ НА АМИНОКИСЛОТНЫЙ СОСТАВ БЕЛКОВ И ИСПОЛЬЗОВАНИЕ КОДОНОВ В МРНК У БАКТЕРИЙ

Сообщение 1.

Белорусский государственный медицинский университет

Оригинальная модель совместного влияния естественного отбора и мутационного давления на аминокислотный состав белков была создана в результате тщательного анализа зависимостей между частотами использования каждого кодона от общей GC-насыщенности бактериальных геномов.

Ключевые слова: мутационное давление, бактериальный геном, использование кодонов, GC-содержание, аминокислотные замены, нуклеотидные замены.

V. V. Khrustalev, E. V. Barkovsky

THE ORIGINAL MODEL OF COMBINED INFLUENCE OF NATURAL SELECTION AND MUTATIONAL PRESSURE ON THE AMINO ACID AND CODON USAGE IN BACTERIA. PART 1

After accurate analysis of dependences between the usage of every codon and total GC-content in bacterial genomes the original model of combined influence of natural selection and mutational pressure on the amino acid and codon usage in bacteria has been formulated.

Key words: mutational pressure, bacterial genome, codon usage, gc-content, amino acid substitutions, nucleotide substitutions.

На последовательность и частоты использования аминокислот в белке могут влиять два основных фактора: мутационное давление и естественный отбор. Последний является экзогенным фактором, который действует на всех уровнях реализации генетической информации. Согласно теории нейтральности М. Кимуры [5], естественный отбор на молекулярном уровне в популяции проявляется главным образом отрицательным отбором.

Сущность отрицательного отбора заключается в том, что носители отрицательных (вредных) мутаций, как правило, погибают, не оставляя потомства. Тем самым в популяции не фиксируются несовместимые с жизнью мутации. Как известно [3,6], наиболее жестко лимитируется аминокислотный состав функционально важных участков белка. Аминокислотные замены в остальных частях белка относительно нейтральны, процесс их фиксации имеет вероятностный характер. Нейтральные мутации могут быть фиксированы в популяции при определенном стечении множества обстоятельств. Процесс их фиксации называется случайным генетическим дрейфом [5].

Согласно теории Н. Суеоки [9,10] мутационное давление возникает из-за повышенной частоты замен аденина и тимина на гуанин и цитозин (GC давление), или наоборот (AT давление) в кодирующих участках ДНК (мРНК) соответствующих белков.

Следует уточнить, что мутационное давление имеет определенное направление по вполне естественным причинам. Как известно, аденин и гуанин по своей структуре являются пуриновыми основаниями, а цитозин и тимин (урацил в мРНК) являются пиримидиновыми основаниями. Вероятность фиксации в по-

пуляции транзиции, то есть, замены пуринового основания на пуриновое или пиримидинового на пиримидиновое, примерно в два – четыре раза выше, чем вероятность фиксации трансверсии (замены пуринового основания на пиримидиновое, или наоборот) [1,4]. По нашему мнению, именно повышенная частота фиксации транзиций является молекулярной основой мутационного давления. По нашим данным, в генах с GC-давлением повышенной частотой фиксации обладают как синонимичные, так и несинонимичные транзиции А на G и Т на С [1]. В генах с AT-давлением повышена частота фиксации транзиций С на Т и G на А [4].

Общие частоты использования гуанина и цитозина в ДНК равны и изменяются сочетано, так как эти нуклеотиды комплементарны друг другу: при замене аденина на гуанин в одной из цепочек ДНК, во время первой же репликации на дочерней цепочке появляется комплементарный гуанину цитозин.

Из всего вышесказанного можно сделать вывод, что для возникновения мутационного давления в направлении AT на GC достаточно, чтобы частота возникновения замены исключительно аденина на гуанин или исключительно тимина на цитозин превышала частоты возникновения остальных нуклеотидных замен.

В пользу существования направленного мутационного давления говорит тот факт, что далеко не все нуклеотидные замены фиксируются в популяции, причём не столько из-за их элиминации отбором, сколько из-за непредсказуемости, присущей случайному генетическому дрейфу. Даже при наличии таких ограничений, у отдельных микроорганизмов общая GC-насыщенность в ходе эволюции снизилась до 24% (*Mycoplasma mycoides*)

и повысилась до 72% (*Streptomyces coelicolor*) [2].

Материал и методы

В качестве объектов исследования были выбраны 85 совокупностей кодирующих участков (СКУ) бактериальных мРНК из электронной базы данных Codon Usage Database [8]. Данные СКУ мРНК принадлежат следующим видам бактерий: *Mycoplasma mycoides*; *Fusobacterium nucleatum*; *Chlamydomonas pneumoniae*; *Borrelia burgdorferi*; *Clostridium perfringens*; *Campylobacter jejuni*; *Rickettsia conorii*; *Staphylococcus aureus*; *Wolbachia endosymbiont*; *Bacillus cereus*; *Streptococcus agalactiae*; *Parachlamydia sp.*; *Oceanobacillus iheyensis*; *Lactococcus lactis*; *Helicobacter hepaticus*; *Leptospira interrogans*; *Thermoanaerobacter tengcongensis*; *Listeria innocua*; *Enterococcus faecalis*; *Treponema denticola*; *Haemophilus ducreyi*; *Legionella pneumophila*; *Bartonella henselae*; *Pasteurella multocida*; *Pyrococcus furiosus*; *Acinetobacter sp.*; *Nostoc sp.*; *Coxiella burnetii*; *Bacteroides thetaiotaomicron*; *Bacillus halodurans*; *Bacillus subtilis*; *Lactobacillus plantarum*; *Photobacterium profundum*; *Mannheimia succiniciproducens*; *Aquifex aeolicus*; *Photorhabdus luminescens*; *Vibrio parahaemolyticus*; *Thermotoga maritima*; *Shewanella oneidensis*; *Bacillus licheniformis*; *Vibrio vulnificus*; *Desulfotalea psychrophila*; *Vibrio cholerae*; *Wolinella succinogenes*; *Synechocystis sp.*; *Yersinia pestis*; *Porphyromonas gingivalis*; *Escherichia coli*; *Bdellovibrio bacteriovorus*; *Nitrosomonas europaea*; *Shigella flexneri*; *Prochlorococcus marinus*; *Erwinia carotovora*; *Neisseria meningitidis*; *Xylella fastidiosa*; *Klebsiella pneumoniae*; *Thermosynechococcus elongatus*; *Corynebacterium glutamicum*; *Rhodospirillum rubrum*; *Brucella suis*; *Pseudomonas syringae*; *Agrobacterium tumefaciens*; *Propionibacterium acnes*; *Bifidobacterium longum*; *Gloeobacter violaceus*; *Mezorhizobium loti*; *Sinorhizobium meliloti*; *Desulfovibrio vulgaris*; *Methylococcus capsulatus*; *Bradyrhizobium japonicum*; *Xanthomonas axonopodis*; *Rhodospseudomonas palustris*; *Chromobacterium violaceum*; *Halobacterium sp.*; *Deinococcus radiodurans*; *Ralstonia solanacearum*; *Caulobacter crescentus*; *Burkholderia pseudomallei*; *Leifsonia xyli*; *Bordetella bronchiseptica*; *Symbiobacterium thermophilum*; *Thermus thermophilus*; *Mycobacterium avium*; *Nocardia farcinica*; *Streptomyces coelicolor*.

Общее число нуклеотидных последовательностей в 85 СКУ составило 273 590, общее число кодонов в них составило 85 366 443.

По общей частоте содержания гуанина и цитозина (G+C) 85 СКУ мРНК распределились следующим образом: 38 из них имеют показатель G+C > 50%, а у 47 G+C-содержание < 50%, диапазон вариации G+C – от 0,24 до 0,72. Каждый вид бактерий представлен не менее чем одной тысячей кодирующих участков.

Расчёты коэффициентов корреляции (R), достоверности аппроксимации (R²), наклона тренда (N), частот использования каждой аминокислоты и каждого кодона, а также разницы между средними арифметическими двух выборочных совокупностей производились с помощью оригинальной программы VVK 3.2.

Результаты и обсуждение

Наша модель не отрицает роли положительного дарвиновского отбора в процессе адаптивной эволюции на молекулярном уровне, однако, она предполагает, что адаптивную природу имеет лишь незначительная часть мутаций [5].

Фиксация неблагоприятной мутации – событие очень редкое. Тем не менее, нельзя полностью отрицать возможности её фиксации при определенном стечении обстоятельств, главным из которых является малая численность популяции. Живые организмы обладают огромным запасом приспособляемости к изменяющимся условиям как внешней, так и внутренней среды.

Главной идеей нашей гипотезы является следующее утверждение: большинство мутаций, которые фиксируются в популяции, являются фенотипически «молчащими», то есть, не влияющими на выживание и воспроизведение особи.

В генетическом коде есть двукратно, четырехкратно, шестикратно, трехкратно вырожденные и невырожденные серии кодонов.

В двукратно, трехкратно и четырехкратно вырожденных сериях синонимичные триплеты отличаются друг от друга исключительно по нуклеотидам в третьем положении. Все вырожден-

ные серии содержат гетерогенные по GC-насыщенности триплеты (табл. 1).

Как видно из таблицы 2, зависимость частот использования аминокислот Cys, Gln, Glu, His, Asp, Thr, Met, Ser, Leu от G+C не является линейной [7], однако, «стабильными» [9] эти аминокислоты с нашей точки зрения назвать нельзя. Средний уровень GC-насыщенности триплетов, кодирующих данные аминокислоты, действительно приводит к стабилизации частот их использования при 0,36 > G+C < 0,56. Тем не менее, при G+C < 0,36 частоты использования Thr, His и Gln статистически достоверно снижаются. При G+C > 0,56 статистически достоверно снижаются частоты использования Gln и Glu, а при G+C > 0,66 статистически достоверно снижаются частоты использования Met и Cys. Однако снижение частот использования данных аминокислот при сильном GC-или AT-давлении не может быть описано линейной зависимостью, скорее, можно предположить стохастический характер влияния мутационного давления на их использование. Частота использования Asp, наоборот, достоверно увеличивается в группах микроорганизмов с G+C < 0,36 и G+C > 0,56. Несмотря на средний уровень GC-насыщенности триплетов серии Val, частота использования данной аминокислоты достоверно возрастает по мере увеличения G+C.

В сериях триплетов, кодирующих аминокислоты Cys, Gln, Glu, Thr, Leu, His, Asp, Ser и Val, при GC-давлении частоты использования GC-богатых триплетов возрастают пропорционально снижению частот использования GC-бедных. При AT-давлении частоты использования GC-бедных триплетов возрастают пропорционально снижению частот использования GC-богатых. Такая закономерность наблюдается в отсутствие сильного GC-или AT-давления. Частоты использования GC-бедных и GC-богатых триплетов образуют характерный перекрест на уровне G+C близком к 0,5. В результате, при G+C в пределах от 0,4 до 0,6 частоты использования вышеперечисленных аминокислот относительно стабильны [7].

Частоты использования аминокислот Tyr, Phe, Ile, Asn, Lys находятся в обратной линейной зависимости от общей GC-насыщенности [7]. Этот факт объясняется тем, что вышеперечисленные аминокислоты кодируются GC-бедными триплетными.

В сериях триплетов, кодирующих аминокислоты Tyr, Phe, Ile, Asn и Lys, также существует перекрест между частотами использования относительно GC-бедных и относительно GC-богатых триплетов, но при GC-давлении частота использования последних возрастает в гораздо меньшей степени, чем падает частота использования первых. При AT-давлении частоты использования GC-бедных триплетов резко возрастают.

Прямую линейную зависимость от G+C частот использования аминокислот Gly, Arg, Ala и Pro [7] также можно объяснить исключительно характером кодирующих их триплетов (абсолютно GC-насыщенные и GC-насыщенные по первому и второму положениям).

В сериях триплетов, кодирующих Gly, Arg, Ala и Pro, при AT-давлении частоты использования GC-бедных триплетов возрастают в гораздо меньшей степени, чем убывают частоты использования GC-богатых. При GC-давлении резко возрастают частоты использования GC-богатых триплетов.

Тем не менее, зависимость частот использования триплетов от GC-содержания в СКУ мРНК близка к линейной только для абсолютно GC-бедных и абсолютно GC-богатых кодонов. Для большинства остальных триплетов наиболее точно характеризует такую зависимость полиномиальная кубическая функция.

Для поиска глубинных причин, вызывающих изменение частот использования аминокислотных остатков в бактериальных белках, необходимо проанализировать зависимость частоты использования каждого кодона от G+C в СКУ мРНК.

В нашей модели для описания характера зависимости частоты использования кодона от уровня GC-насыщенности мы использовали следующие показатели: наклон тренда частоты использования кодона относительно G+C (N) и величину достоверности аппроксимации зависимости частоты его использования от G+C к прямой линии тренда (R²).

Поскольку при усилении мутационного давления не только увеличивается показатель R², но и N зачастую меняет свой знак, мы исследовали характер вышеописанной зависимости

Таблица 1
Генетический код мРНК, насыщенность кодонов гуанином и цитозином

АК	GC-нас-ть	АК	GC-нас-ть	АК	GC-нас-ть	АК	GC-нас-ть
кодон	%	кодон	%	кодон	%	кодон	%
Phe		Ser		Tyr		Cys	
UUU	0	UCU	33,3	UAU	0	UGU	33,3
UUC	33,3	UCC	66,6	UAC	33,3	UGC	66,6
Leu		UCA	33,3	Term		Term	
UUA	0	UCG	66,6	UAA	0	UGA	33,3
UUG	33,3			UAG	33,3	Trp	
						UGG	66,6
Leu		Pro		His		Arg	
CUU	33,3	CCU	66,6	CAU	33,3	CGU	66,6
CUC	66,6	CCC	100	CAC	66,6	CGC	100
CUA	33,3	CCA	66,6	Gln		CGA	66,6
CUG	66,6	CCG	100	CAA	33,3	CGG	100
				CAG	66,6		
Ile		Thr		Asn		Ser	
AUU	0	ACU	33,3	AAU	0	AGU	33,3
AUC	33,3	ACC	66,6	AAC	33,3	AGC	66,6
AUA	0	ACA	33,3	Lys		Arg	
Met		ACG	66,6	AAA	0	AGA	33,3
AUG	33,3			AAG	33,3	AGG	66,6
Val		Ala		Asp		Gly	
GUU	33,3	GCU	66,6	GAU	33,3	GGU	66,6
GUC	66,6	GCC	100	GAC	66,6	GGC	100
GUA	33,3	GCA	66,6	Glu		GGA	66,6
GUG	66,6	GCG	100	GAA	33,3	GGG	100
				GAG	66,6		

Таблица 2

Коэффициенты корреляции (R) частот использования аминокислот бактериальных белков с общей GC-насыщенностью кодирующих их участков ДНК и достоверность аппроксимации (R²) полученной линейной зависимости

АК	R	R ²	АК	R	R ²
Tyr	-0,87	0,7520	Asn	-0,93	0,8652
Phe	-0,84	0,7110	Lys	-0,91	0,8314
Ile	-0,95	0,9005			
Pro	0,92	0,8543	Gly	0,94	0,8924
Ala	0,95	0,9068	Arg	0,96	0,9187
Trp	0,74	0,5477			
Cys	-0,08	0,0070	His	0,38	0,1444
Gln	-0,01	0,0104	Asp	0,34	0,1129
Glu	-0,48	0,2347	Ser	-0,59	0,3504
Thr	0,18	0,0316	Ser (дуплет)	-0,39	0,1566
Leu	0,20	0,0385	Ser (квартет)	-0,39	0,1590
Met	-0,26	0,0655	Val	0,74	0,5443

на определенных участках.

Для относительно GC-богатых кодонов N и R² были вычислены на участках G+C<0,6 и G+C>0,6. Для относительно GC-бедных кодонов N и R² были вычислены на участках G+C<0,4 и G+C>0,4. Также нами были рассчитаны средние частоты использования каждого кодона при G+C>0,6; в пределах G+C от 0,45 до 0,55; при G+C<0,4.

По мере увеличения G+C от 0,5 до 0,6 частота использования GC-богатого кодона, как правило, возрастает, причем пропорционально падению частоты использования GC-бедного кодона. По мере уменьшения G+C от 0,5 до 0,4, наоборот, возрастает частота использования GC-бедного кодона при падении частоты использования GC-богатого. Этот феномен легко может быть объяснен фиксацией преимущественно синонимичных замен в условиях малого мутационного давления.

В рамках нашей модели, при G+C>0,6 и при G+C<0,4 мутационное давление возрастает настолько, что в популяции зна-

чительно увеличивается частота фиксации несинонимичных нуклеотидных замен. В данном случае речь идет преимущественно о симметричном (распространяющемся равномерно как на одну, так и на другую цепочку ДНК) мутационном давлении. На существование симметричного мутационного давления указывает тенденция к равномерному снижению частот использования кодонов с G и C в третьем положении в четырехкратно вырожденных сериях при АТ-давлении и равномерному увеличению частот их использования при GC-давлении.

Анализ средних частот использования кодонов при сильном мутационном давлении в совокупности с показателями N и R² позволяют определить наличие селекционных ограничений на несинонимичные нуклеотидные замены, вызываемые мутационным давлением. Говорить следует лишь о той или иной степени селекционных ограничений на аминокислотную замену.

Характер зависимости частоты использования кодона от G+C при сильном мутационном давлении позволяет делать выводы о возможности фиксации несинонимичных нуклеотидных замен, которые могут происходить в этом кодоне под воздействием соответствующего мутационного давления. Каждый кодон под воздействием сильного мутационного давления может превратиться путем транзиции либо в один, либо в два несинонимичных кодона.

● На наличие жесткого отрицательного отбора относительно фиксации несинонимичной нуклеотидной замены указывает значение R² > 0,25 при сохранении или возрастании N при сильном мутационном давлении (G+C<0,4 или G+C>0,6) по сравнению с N в его отсутствии. В таблице 4 обозначается трижды перечеркнутой стрелочкой.

● На наличие умеренного отрицательного отбора относительно фиксации несинонимичной нуклеотидной замены указывает значение R² < 0,25 при сохранении или незначительном изменении N при сильном мутационном давлении. В таблице 4 обозначается дважды перечеркнутой стрелочкой.

● На наличие малого влияния отрицательного отбора на закрепление в популяции несинонимичной нуклеотидной замены указывает R² < 0,25 при N меньше (вплоть до изменения знака) при сильном мутационном давлении. Обозначается стрелочкой, перечеркнутой один раз.

● В случае значительного снижения N с переменной его знака под воздействием сильного мутационного давления при R² > 0,25 можно говорить о почти полном отсутствии отрицательного отбора относительно закрепления соответствующей нуклеотидной замены в популяции. Обозначается неперечеркнутой стрелочкой.

В большинстве случаев при сильном мутационном давлении нуклеотидные замены постепенно фиксируются в популяции, оставляя «след» в виде понижения частот использования определенных кодонов. В каком-то отдельном случае аминокислотная замена в определенном участке белка может быть радикальной (с выраженным неблагоприятным эффектом) или летальной. Микроорганизм с такой мутацией погибает, и данная мутация элиминируется из популяции отрицательным отбором [5]. Та же самая аминокислотная замена может зафиксироваться в популяции случайным генетическим дрейфом в том случае, если произойдет в другой, не выполняющей критической функции, части белка [3,5,6]. С течением времени фиксируются именно нейтральные (или почти нейтральные) замены.

Совместное влияние мутационного давления и естественного отбора на частоты использования кодонов в СКУ мРНК бактериальных белков (Часть 1).

Аминокислота Phe (рис. 1а) кодируется двумя триплетами: UUU и UUC. Триплет UUU является абсолютно и относительно GC-бедным, и поэтому зависимость частоты его использования от G+C близка к обратной линейной. Тем не менее, при G+C<0,4 несколько снижен N и уменьшен R² относительно таковых при G+C>0,4 (табл. 3-4). Частота использования триплета UUC демонстрирует прямую линейную зависимость от G+C при G+C<0,6, однако, при G+C>0,6 резко уменьшается N и увеличивается разброс частот использования этого кодона (малое влияние отрицательного отбора). Триплет UUC под действием GC-давления может превращаться толи в CUC (Leu4), толи в UCC (Ser4). Судя по вышеприведенным данным, фиксация подобных

Таблица 3

Влияние мутационного давления на частоты использования кодонов (Часть 1)

AK	Показатели для частот использования относительно GC-богатых кодонов				Средние значения частот использования кодонов в группах бактерий с G+C				Показатели для частот использования относительно GC-бедных кодонов					
	N		R ²		> 0,6		0,45 - 0,55		< 0,4		N		R ²	
	G+C < 0,6	G+C > 0,6							G+C < 0,4	G+C > 0,4				
Phe	-	-	-	-	22,97	39,60	-93,71	0,31	-104,26	0,89	-	-	-	-
UUU	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UUC	55,66	0,53	4,49	0,002	28,40	17,35	-	-	-	-	-	-	-	-
Leu2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UUA	-	-	-	-	13,62	44,38	-195,08	0,42	-109,94	0,67	-	-	-	-
UUG	30,86	0,16	-81,31	0,359	9,2	20,78	-	-	-	-	-	-	-	-
Leu4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CUU	-	-	-	-	15,77	21,23	-20,86	0,01	-54,59	0,5	-	-	-	-
CUC	64,32	0,39	93,13	0,05	30,19	16,44	-	-	-	-	-	-	-	-
CUA	-	-	-	-	7,67	9,78	6,97	0,01	-34,30	0,54	-	-	-	-
CUG	170,07	0,72	220,34	0,19	55,73	31,57	-	-	-	-	-	-	-	-
Ser4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UCU	-	-	-	-	10,06	17,14	-24,83	0,06	-50,26	0,74	-	-	-	-
UCC	40,45	0,52	-24,49	0,04	13,11	10,67	-	-	-	-	-	-	-	-
UCA	-	-	-	-	8,60	15,36	-58,43	0,31	-39,59	0,81	-	-	-	-
UCG	36,85	0,6	-9,11	0,003	17,16	8,26	-	-	-	-	-	-	-	-
Pro	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCU	-	-	-	-	10,07	12,76	17,43	0,06	-39,46	0,76	-	-	-	-
CCC	42,29	0,41	114,14	0,202	18,87	10,24	-	-	-	-	-	-	-	-
CCA	-	-	-	-	9,88	12,38	2,42	0,001	-37,01	0,53	-	-	-	-
CCG	67,47	0,64	68,81	0,111	29,16	13,02	-	-	-	-	-	-	-	-

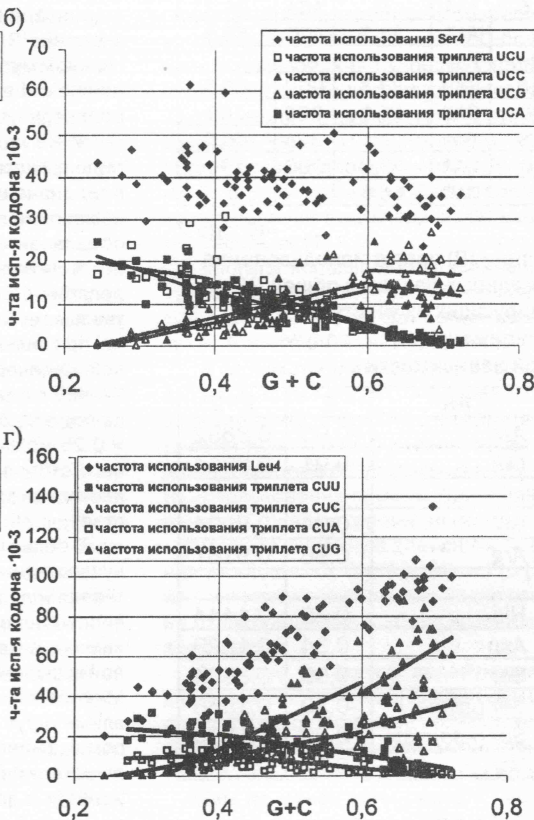
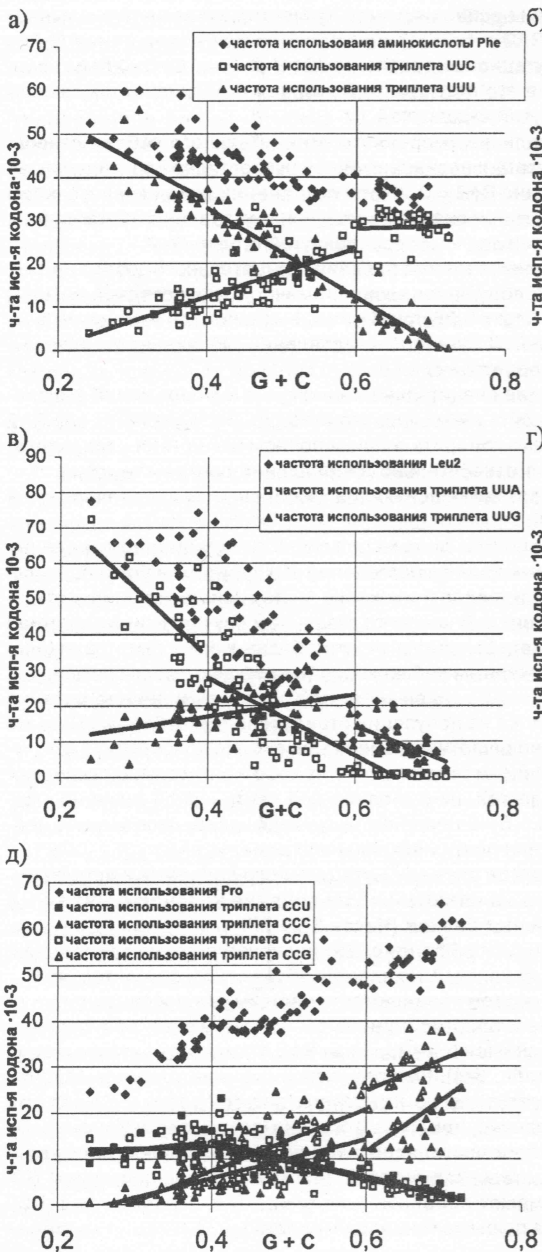


Рис. 1. Зависимость частот использования аминокислот и кодирующих их триплетов от GC-насыщенности SKU мРНК (Часть 1)

этого можно сделать вывод о том, что замены UCA на UUA (Leu2) почти не фиксируются, возможно, из-за радикальности аминокислотной замены Ser4 на Leu2.

При G+C > 0,6 значения N для кодонов UCC и UCG отрицательны, то есть, до G+C = 0,6 частоты использования UCC и UCG демонстрируют прямую линейную зависимость от G+C, а после – отрицательную и отнюдь не линейную (значения R² весьма низкие). Оба этих триплета под воздействием GC-давления могут превращаться в триплеты из серии Pro (CCC и CCG). Судя по вышеприведенным данным, такие замены хорошо фиксируются случайным генетическим дрейфом в SKU мРНК бактерий. Исходя только из строения серина и пролина, по-

Таблица 4

Степень влияния отрицательного отбора на фиксацию несинонимичных нуклеотидных замещений, возникающих под воздействием сильного мутационного давления (Часть 1)

GC-давление		AT-давление	
Phe	→ Ser4, Leu4	Ser4	→ Phe
Ser4	→ Pro	Ser4	→ Leu2
Leu4	→ Pro	Leu4	→ Phe
Leu2	→ Ser4	Leu4	→ Leu2
Leu2	→ Leu4	Pro	→ Leu4, Ser4

замен подвергается малому влиянию отрицательного отбора (табл. 3).

Аминокислота Ser кодируется одной четырехкратно вырожденной серией (Ser4) (рис. 16) и одной двукратно вырожденной серией (Ser2). Они должны рассматриваться как две отдельные серии, так как кодоны «дуплета» не могут превращаться в кодоны «квартета» в результате одной аминокислотной замены. Серия Ser4 включает триплеты UCU, UCC, UCA, UCG.

Триплет UCU может при сильном AT-давлении превращаться в UUU (Phe), а триплет UCA – в UUA (Leu). При G+C > 0,4 частоты использования обоих триплетов демонстрируют обратную линейную зависимость от G+C. Для триплета UCU при G+C < 0,4 показатель N в два раза меньше (по модулю), чем при G+C > 0,4; R² также мал при сильном AT-давлении. Этот факт дает основания для предположения, что на фиксацию замены Ser4 на Phe отрицательный отбор оказывает малое влияние, несмотря на то, что фенилаланин содержит ароматический радикал, а серин – OH-группу.

Для кодона UCA при G+C < 0,4 значение N значительно больше по модулю, чем при G+C > 0,4; а величина R² = 0,31 свидетельствует о небольшом отклонении от линейной зависимости. Из

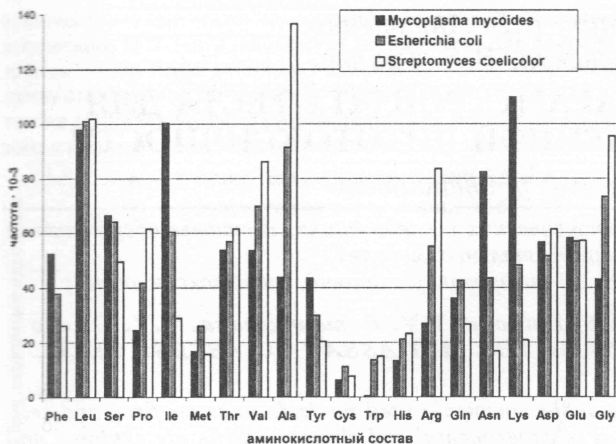


Рис. 2. Зависимость частот использования аминокислот от GC-насыщенности бактериальных геномов

добное предположение казалось бы абсурдным: пролин обладает специфическим имино-радикалом, который существенно изменяет вторичную структуру белка.

Аминокислота Leu (рис. 1в-1г) кодируется шестью кодонами, которые образуют одну шестикратно вырожденную серию: UUA, UUG, CUU, CUC, CUA, CUG. Кодон UUA, являясь абсолютно и относительно GC-бедным, демонстрирует обратную прямую зависимость от G+C. Кодон UUG при $G+C < 0,6$ демонстрирует прямую зависимость от G+C ($R^2 = 0,161$). При $G+C > 0,6$ значение N для UUG отрицательно и велико по модулю, а частота использования UUA стремится к нулю. Отсутствие всяких ограничений на синонимичные замены UUG на CUG подтверждает и $R^2 = 0,39$. То есть, у всех микроорганизмов с сильным GC-давлением такие замены свободно фиксируются, а некоторый разброс значений обусловлен, скорее всего, вероятностным характером закрепления синонимичной замены в популяции.

Степень влияния отрицательного отбора на замену UUG на UCG определить невозможно, так как замена UUG на CUG априори абсолютно нейтральна.

Кодон CUA свободно переходит в UUA при AT-давлении (показатель N при $G+C < 0,4$ для CUA положителен). В то же время, показатель N для кодона CUU при $G+C < 0,4$ отрицателен. Естественно, ограничения для замены Leu на Phe несколько больше, чем для синонимичной замены, однако разброс в частотах использования как CUA, так и CUU при $G+C < 0,4$ достаточно велик.

При $G+C > 0,6$ показатели N как для кодона CUC, так и для CUG больше, чем при $G+C < 0,6$, однако показатели R^2 невелики, особенно для CUC. В данном случае речь может идти об умеренном влиянии отрицательного отбора на замены Leu на Pro: у некоторых микроорганизмов с сильным GC-давлением пониженная частота использования CUC и CUG всё таки зафиксировалась. Рост частот использования CUC и CUG при GC-давлении также может быть связан с превращением в них кодонов дуплета лейцина (UUA и UUG).

В точке равновесия квартета лейцина обнаруживается явная асимметрия: частота использования CUG больше, чем частота использования CUC. Такая асимметрия не только сохраняется, но и увеличивается при $G+C > 0,6$. Возможно, это вызвано отсутствием ограничений на превращение UUG в CUG. Интересно отметить, что подобная асимметрия в использовании кодонов присуща и кодонам из серии Pro (рис. 1д). Именно этот факт может трактоваться как еще одно, косвенное, доказательство возможности аминокислотных замен Leu на Pro.

Несинонимичные замены в кодонах CCU и CCA из четырехкратно вырожденной серии пролина при сильном AT-давлении могут относительно свободно фиксироваться. Эти кодоны могут превращаться в менее GC-насыщенные кодоны из серии Leu4 и Ser4. Об этом свидетельствует закрепление низких частот использования CCU и CCA при $G+C < 0,4$. Кодоны CCC и CCG демонстрируют зависимость от G+C, близкую к прямой линейной.

Знание закономерностей влияния мутационного давления

и отрицательного отбора на состав бактериальных белков может быть использовано на практике. Перспективным, с нашей точки зрения, является моделирование путём биотехнологии аминокислотного состава отдельных белков, в том числе ферментов. В качестве примера изменений аминокислотного состава белков под воздействием AT-и GC-давления приведены (см. рис. 2) средние частоты использования аминокислот у *Mycoplasma mycoides* ($G+C = 0,24$) и *Streptomyces coelicolor* ($G+C = 0,72$) в сравнении с *Escherichia coli* ($G+C = 0,5$).

Выводы

На базе анализа 273 590 нуклеотидных последовательностей из 85 SKU бактериальных мРНК разработана модель сочетанного влияния естественного отбора и мутационного давления на аминокислотный состав бактериальных белков. Её основные положения:

1. Аминокислотные замены в белках происходят в результате преимущественно одношаговых несинонимичных нуклеотидных замен в триплетных кодирующих их участках ДНК. Нуклеотидные замены вызываются мутационным давлением определенного направления, наибольший вклад в создание которого вносит повышенная частота возникновения соответствующих транзиций.

2. Аминокислотные замены в белках фиксируются в популяции случайным генетическим дрейфом в том случае, если они являются нейтральными (или почти нейтральными) по отношению к естественному отбору. Отрицательный отбор препятствует фиксации аминокислотных замен в том случае, если замены являются радикальными (вредными).

3. Для каждой аминокислоты количество и характер возможных замен, происходящих под воздействием мутационного давления определенного направления, детерминированы главным образом GC-насыщенностью кодирующих её триплетов.

4. При наличии сильного мутационного давления вероятность возникновения нуклеотидных замен определенного направления значительно возрастает, что приводит к повышению вероятности возникновения и, как следствие, фиксации или элиминации соответствующих аминокислотных замен.

5. О степени консервативности или радикальности аминокислотных замен, происходящих в результате точечных мутаций в кодонах под воздействием мутационного давления, можно судить исключительно по изменению частот использования триплетов при сильном мутационном давлении. Снижение частоты использования данного триплета по мере увеличения мутационного давления говорит о том, что произошедшие в таком триплете нуклеотидные замены были фиксированы в популяции. В случае роста частоты использования данного триплета по мере усиления мутационного давления, следует считать, что все возможные несинонимичные замены, которые могли произойти в таком триплете, фиксированы не были.

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., Хрусталев В.В. Сравнительная характеристика мРНК аденилатциклаз актиномицетов // Белорусский мед. журнал. – 2004. – №3. – стр.27-30.
3. Барковский Е.В., Хрусталев В.В. Сравнительная характеристика матричных РНК и аминокислотных последовательностей аденилатциклаз цианобактерий // Белорусский мед. журнал. – 2005. – №1. – стр.23-27.
4. Бутвиловский А.В., Хрусталев 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. Хрусталев В.В. эволюционные взаимоотношения и стратегии кодирования аденилатциклаз актиномицетов и цианобактерий // Студенты БГМУ – медицинской науке и здравоохранению Беларуси. – 2005. – стр.53-54
7. Хрусталев В.В. Влияние мутационного давления на состав бактериальных белков / Актуальные проблемы современной медицины 2005. – 2005. – стр.359 – 362.
8. Nakamura, Y., Gojbori, T. and Ikemura, T. Codon usage tabulated from the international DNA sequence databases: status for the year 2000. // Nucl. Acids Res. – 2000. – 28, 292.
9. Sueoka N. Directional mutation pressure and neutral molecular evolution // Proc. Natl. Acad. Sci. USA. – 1988. – Vol.85. – P.2653-2657.
10. 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.