

*В.Э. Бутвиловский, Е.В. Барковский, А.В. Бутвиловский, Ю.И. Линник*  
О сходстве стратегий кодирования митохондриальных белков  
человека и трихинеллы. Часть 2. Картина использования  
синонимичных кодонов. Содержание аминокислот групп GARP и  
FYMINK

*Белорусский государственный медицинский университет*

Изучены 12 ферментов дыхательной цепи и кодирующие их мРНК человека, трихинеллы и цианорабдитис. Установлено, что аминокислотный состав изученных белков и картина использования синонимичных кодонов в мРНК человека более сходны с таковыми трихинеллы. Ключевые слова: стратегия кодирования, картина использования синонимичных кодонов, дистанция МакАйнерни.

Стратегия кодирования белка определяется картиной использования кодонов в соответствующих ему мРНК и ДНК. Анализ картин использования кодонов в последовательностях нуклеиновых кислот стал возможным в 70-80-х годах прошлого столетия, когда в международных базах данных появилось достаточное количество секвенированных последовательностей РНК и ДНК [4]. С тех пор был установлен ряд важных закономерностей стратегий кодирования белков [1, 3, 5].

Продолжающееся в настоящее время секвенирование последовательностей геномов различных организмов позволяет изучать картины использования кодонов в большем числе нуклеотидных последовательностей РНК и ДНК, что обеспечивает биологическую значимость получаемых выводов. Практическое применение результатов исследований такого плана связано с проведением обратной трансляции аминокислотных последовательностей белков в соответствующие им последовательности нуклеиновых кислот, а также определением белок-кодирующих и белок-некодирующих участков геномов [8].

Цель данной части исследования: сравнить картины использования синонимичных кодонов в мРНК, а также аминокислотный состав кодируемых ими митохондриальных белков человека и трихинеллы.

Материал и методы

Проанализированы взятые с сервера NCBI (National Center for Biotechnology information, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) последовательности ряда митохондриальных белков (субъединицы 1 – 4, 4L, 5, 6 НАДН-дегидрогеназы, цитохром b, субъединицы 1 – 3 цитохром – с – оксидазы, субъединица 6 АТФ-синтазы), а также кодирующих их мРНК человека (*Homo sapiens*, [9]) и трихинеллы (*Trichinella spiralis*, [7]). В качестве контроля использованы аналогичные белки и соответствующие им мРНК свободноживущего круглого червя цианорабдитис (*Caenorhabditis elegans*, [11]). Показатели относительного использования синонимичных кодонов (RSCU, relative synonymous codon's usage) и аминокислотный состав изучены

при помощи программы MEGA 3 [6, 10]. Сходство стратегий кодирования изучаемых белков в мРНК определено по дистанции МакАйнерни:

$$D_{jk} = \sum_{i=1}^n \frac{abs(RSCU_{ji} - RSCU_{ki})}{n}$$

где n – число синонимичных вырожденных кодонов для определенной таблицы генетического кода,  $j_i$  – кодон i последовательности j,  $k_i$  – кодон i последовательности k [8].

Полученные результаты обработаны статистически [2] с помощью пакета программ Microsoft Excel, достоверность различий определена по критерию Стьюдента.

#### Результаты и обсуждение

Первоначально для каждой из 20 аминокислот вычислены суммарные RSCU соответствующих ей ГЦЗ-кодонов. Средние значения показателя относительного использования синонимичных кодонов в мРНК, кодирующих ряд митохондриальных белков человека, трихинеллы и цианорабдитис, представлены в таблице 1.

Таблица 1

Средние значения RSCU для мРНК, кодирующих ряд митохондриальных белков человека, трихинеллы и цианорабдитис

Аминокислота (соответствующие ГЦЗ-кодоны)	Среднее RSCU для мРНК человека	Среднее RSCU для мРНК трихинеллы	Среднее RSCU для мРНК цианорабдитис	p для RSCU по мРНК человека и мРНК трихинеллы	p для RSCU по мРНК человека и мРНК цианорабдитис
Фен (УУЦ)	1,08±0,146	0,84±0,198	0,10±0,033	>0,05	<0,001
Лей (ЦУЦ, ЦУГ, УУГ)	2,13±0,161	2,00±0,441	0,93±0,099	>0,05	<0,001
Иле (АУЦ)	1,07±0,128	0,79±0,179	0,18±0,037	>0,05	<0,001
Мет (АУГ)	0,45±0,119	0,65±0,235	0,53±0,082	>0,05	>0,05
Вал (ГУЦ, ГУГ)	1,53±0,180	1,26±0,282	0,52±0,075	>0,05	<0,001
Сер (УЦЦ, УЦГ, АГЦ +АГГ для круглых червей)	3,29±0,265	2,40±0,129	1,04±0,180	<0,01	<0,001
Про (ЦЦЦ, ЦЦГ)	2,08±0,263	0,75±0,173	0,64±0,239	<0,001	<0,001
Тре (АЦЦ, АЦГ)	1,77±0,196	1,09±0,143	0,58±0,343	<0,05	<0,01
Ала (ГЦЦ, ГЦГ)	2,20±0,201	1,07±0,355	0,62±0,221	<0,05	<0,001
Тир (УАЦ)	1,29±0,140	0,88±0,212	0,29±0,053	>0,05	<0,001
Гис (ЦАЦ)	1,38±0,207	0,92±0,253	0,31±0,177	>0,05	<0,001
Гли (ЦАГ)	0,13±0,049	0,37±0,233	0,33±0,117	>0,05	>0,05
Асн (ААЦ)	1,55±0,120	0,90±0,208	0,11±0,036	<0,05	<0,001
Лиз (ААГ)	0,24±0,084	0,67±0,253	0,18±0,062	>0,05	>0,05
Асп (ГАЦ)	1,37±0,226	0,65±0,224	0,45±0,212	<0,05	<0,05
Глу (ГАГ)	0,45±0,154	0,70±0,267	0,52±0,111	>0,05	>0,05
Цис (УГЦ)	1,19±0,276	0,50±0,229	0,08±0,087	>0,05	<0,001
Три (УГГ)	0,21±0,087	0,57±0,221	0,32±0,177	>0,05	>0,05
Арг (ЦГЦ, ЦГГ)	2,17±0,327	0,90±0,369	0,05±0,050	<0,05	<0,001
Гли (ГГЦ, ГГГ)	2,41±0,162	1,09±0,154	0,56±0,148	<0,001	<0,001

Установлено, что по 15-ти аминокислотам (75,0±9,93%) средние значения RSCU для мРНК человека достоверно ( $p<0,001$ ) ближе к таковым трихинеллы по сравнению с контролем. Это свидетельствует о наличии коэволюции компонентов изучаемой системы “паразит-хозяин” на молекулярно-генетическом уровне. В пользу такой точки зрения говорит и то, что по 12-ти аминокислотам (60,0±11,24%) средние значения RSCU для мРНК человека и трихинеллы не являются статистически различимыми. Для

сравнения, статистически незначимыми являются отличия средних значений RSCU для мРНК человека и цианорабдитис лишь по 5-ти аминокислотам ( $25,0 \pm 9,93\%$ ).

Для учета особенностей кодирования отдельных белков в мРНК нами вычислены дистанции МакАйнерни для мРНК, кодирующих каждый из 12 изученных ферментов дыхательной цепи (табл. 2).

Таблица 2

Дистанции МакАйнерни для мРНК, кодирующих митохондриальные белки человека и трихинеллы (опыт), а также человека и цианорабдитис (контроль)

Фермент / организм	Опыт	Контроль	p
НАДН-дегидрогеназа-1	$0,010 \pm 0,0015$	$0,018 \pm 0,0020$	$<0,01$
НАДН-дегидрогеназа-2	$0,022 \pm 0,0016$	$0,016 \pm 0,0016$	$<0,05$
НАДН-дегидрогеназа-3	$0,013 \pm 0,0022$	$0,017 \pm 0,0021$	$>0,05$
НАДН-дегидрогеназа-4	$0,024 \pm 0,0017$	$0,017 \pm 0,0018$	$<0,05$
НАДН-дегидрогеназа-4L	$0,023 \pm 0,0023$	$0,020 \pm 0,0024$	$>0,05$
НАДН-дегидрогеназа-5	$0,023 \pm 0,0013$	$0,020 \pm 0,0021$	$>0,05$
НАДН-дегидрогеназа-6	$0,021 \pm 0,0021$	$0,015 \pm 0,0019$	$<0,05$
Цитохром b	$0,009 \pm 0,0013$	$0,018 \pm 0,0017$	$<0,001$
Цитохром-с-оксидаза-1	$0,008 \pm 0,0009$	$0,016 \pm 0,0016$	$<0,001$
Цитохром-с-оксидаза-2	$0,010 \pm 0,0014$	$0,017 \pm 0,0016$	$<0,01$
Цитохром-с-оксидаза-3	$0,010 \pm 0,0014$	$0,017 \pm 0,0016$	$<0,01$
АТФ-синтаза 6	$0,008 \pm 0,0013$	$0,018 \pm 0,0019$	$<0,001$

Из приведенных в таблице 2 данных видно, что достоверные отличия дистанций МакАйнерни характерны для 9 изученных мРНК, кодирующих ( $75,0 \pm 13,06\%$ ) белки человека и трихинеллы (по сравнению с таковыми для мРНК человека и цианорабдитис). При этом полученные дистанции для шести из них (для мРНК, соответствующих субъединице 1 НАДН-дегидрогеназы, цитохрому b, субъединицам 1-3 цитохром-с-оксидазы, субъединице 6 АТФ-синтазы) меньше таковых контроля, а для трех из них (соответствующих субъединицам 2, 4, 6 НАДН-дегидрогеназы) выше таковых контроля.

Однако при коэволюции человека и трихинеллы на молекулярно-генетическом уровне следовало бы ожидать достоверно меньших дистанций МакАйнерни опыта по сравнению с контролем для всех изученных мРНК. Это наводит на мысль о существовании некоторых особенностей такой методики изучения картин использования кодонов.

Для их определения нами проведен тщательный сравнительный анализ показателей RSCU для кодонов изучаемых мРНК человека, трихинеллы и цианорабдитис. Установлено, что особенностью анализа МакАйнерни является то, что биологическая значимость результатов напрямую связана с числом кодонов нуклеиновой кислоты, по которым вычисляется дистанция.

Рассмотрим эту особенность на примере кодонов УГУ и УГЦ (соответствуют цистеину) мРНК, кодирующих субъединицу 6 НАДН-дегидрогеназы человека, трихинеллы и цианорабдитис (табл. 3).

Таблица 3

Показатели RSCU и количество кодонов УГУ и УГЦ в мРНК, кодирующих ряд митохондриальных белков человека, трихинеллы и цианорабдитис

Организм/ показатель	мРНК, кодирующие субъединицу 6 НАДН-дегидрогеназы		12 изученных мРНК	
	Кодон УГУ	Кодон УГЦ	Кодон УГУ	Кодон УГЦ
Человек	n=1 RSCU=2,0	n=0 RSCU=0	n=4 RSCU=0,4	n=17 RSCU=1,6
Трихинелла	n=0 RSCU=0	n=2 RSCU=2,0	n=44 RSCU=1,7	n=7 RSCU=0,3
Цианорабдитис	n=3 RSCU=2,0	n=0 RSCU=0	n=39 RSCU=1,9	n=2 RSCU=0,1

Анализируя значения показателя относительного использования синонимичных кодонов в мРНК, кодирующих субъединицу 6 НАДН-дегидрогеназы, можно прийти к выводу, что картины использования кодонов УГУ и УГЦ в данной мРНК человека сходны с таковыми в мРНК цианорабдитис и диаметрально противоположны таковым мРНК трихинеллы. На этом примере четко видно, что данные значения RSCU получены по малому числу синонимичных кодонов (1-3) и, следовательно, обладают низкой репрезентативностью. Такие значения RSCU будут вносить существенный вклад в вычисляемую на их основании дистанцию МакАйнерни, анализ которой может привести к ошибочным выводам.

Очевидным путем устранения данного недостатка методики МакАйнерни является совместный анализ и вычисление дистанций максимально возможного (в пределах поставленной цели исследования) количества последовательностей нуклеиновых кислот. При учете установленной особенности дистанция МакАйнерни может быть рекомендована в качестве эффективного способа определения сходства картин использования синонимичных кодонов в последовательностях РНК и ДНК.

В таблице 2 видно, что значения показателей RSCU по 12 изученным мРНК человека ближе к таковым трихинеллы по сравнению с контролем. Вычисленная нами общая дистанция МакАйнерни для 12 изучаемых мРНК человека и трихинеллы ( $0,010 \pm 0,0009$ ) в 1,5 раза меньше таковой для мРНК человека и цианорабдитис ( $0,015 \pm 0,0017$ ; различия достоверны,  $p < 0,05$ ). Этот факт свидетельствует о статистически значимом сходстве картин использования синонимичных кодонов в мРНК, кодирующих исследуемые митохондриальные белки человека и трихинеллы. Наличие такого сходства, с нашей точки зрения, является признаком сходства аминокислотного состава кодируемых белков.

Таблица 4

Доля аминокислот GARP в митохондриальных белках человека, трихинеллы и цианорабдитис

Фермент / показатель	Доля GARP у человека, %	Доля GARP у трихинеллы, %	Доля GARP у цианорабдитис, %	p по доле GARP у человека и трихинеллы	p по доле GARP у человека и цианорабдитис
НАДН-дегидрогеназа-1	21,4±2,30	14,7±2,05	13,8±2,02	<0,05	<0,05
НАДН-дегидрогеназа-2	17,3±2,03	7,8±1,56	5,0±1,30	<0,001	<0,001
НАДН-дегидрогеназа-3	18,3±3,61	9,5±2,72	8,1±2,59	>0,05	<0,05
НАДН-дегидрогеназа-4	16,3±1,72	14,6±1,74	10,8±1,53	>0,05	<0,05
НАДН-дегидрогеназа-4L	16,3±1,72	11,1±3,49	6,5±2,81	>0,05	<0,05
НАДН-дегидрогеназа-5	18,3±1,57	14,1±1,53	9,1±1,25	>0,05	<0,001
НАДН-дегидрогеназа-6	25,3±3,30	8,3±2,21	7,6±2,21	<0,001	<0,001
Цитохром <i>b</i>	20,9±2,09	15,9±1,90	14,9±1,85	>0,05	<0,05
Цитохром-с-оксидаза-1	24,0±1,89	20,4±1,78	18,7±1,70	>0,05	<0,05
Цитохром-с-оксидаза-2	20,3±2,67	13,8±2,30	13,4±2,24	>0,05	>0,05
Цитохром-с-оксидаза-3	20,0±2,48	12,8±2,08	14,9±2,23	<0,05	>0,05
АТФ-синтаза <i>b</i>	20,8±2,70	8,7±1,70	10,1±2,14	<0,001	<0,01

В таблице 4 показана доля аминокислот группы GARP, кодируемых ГЦ-богатыми кодонами – глицина (G; ГГУ, ГГЦ, ГГА, ГГГ), аланина – (A; ГЦУ, ГЦЦ, ГЦА, ГЦГ), аргинина (R; ЦГУ, ЦГЦ, ЦГА, ЦГГ) и пролина (P; ЦЦУ, ЦЦЦ, ЦЦА, ЦЦГ).

Установлено, что содержание аминокислот группы GARP неразличимо статистически в 7-ми изученных белках ( $58,3 \pm 14,87\%$ ) человека и трихинеллы и лишь в 2-х белках человека и цианорабдитис ( $16,7 \pm 11,25\%$ ;  $p < 0,05$ ). Исходя из предположения о коэволюции человека и трихинеллы (как компонентов системы «паразит-хозяин», формирующейся при трихинеллезе) на молекулярно-генетическом уровне следовало бы ожидать, что данный параметр у человека и трихинеллы будет неразличимым для всех изученных белков. Отличия полученных данных от теоретически ожидаемых связаны, вероятно, с функциональными особенностями изучаемых ферментов. Проанализируем зависимость содержания аминокислот группы GARP в изученных белках от насыщенности гуанином и цитозином кодирующих их мРНК (рис. 1).

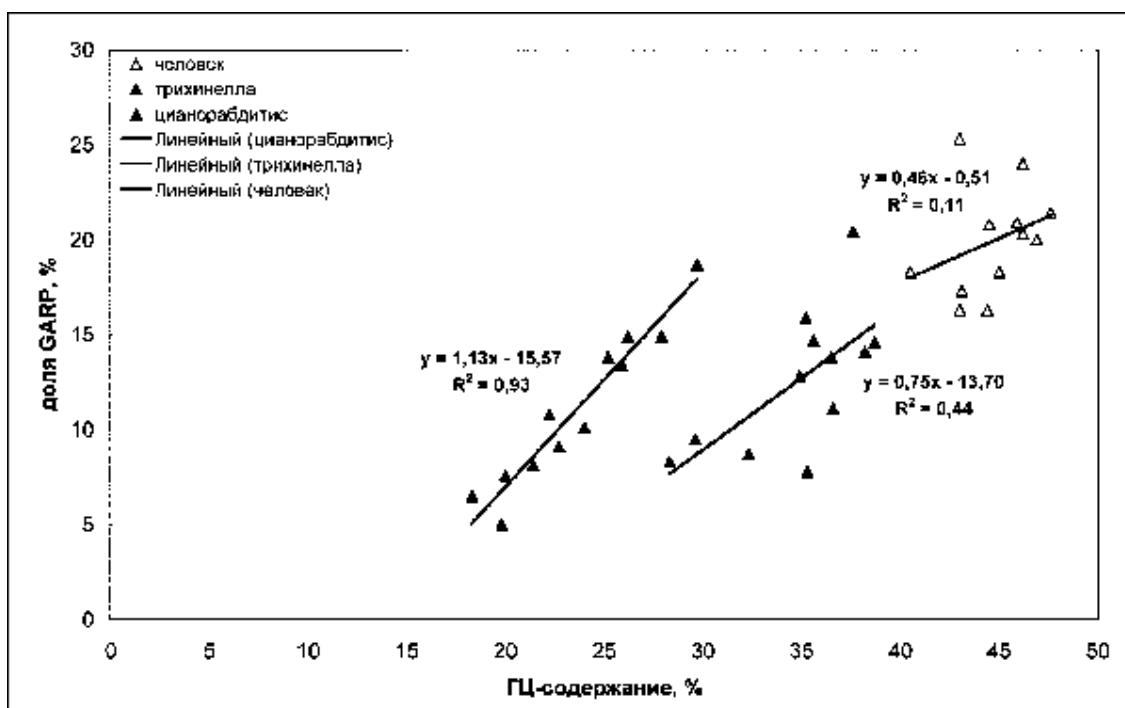


Рис. 1. Зависимость содержания аминокислот группы GARP в ряде митохондриальных белков человека, трихинеллы и цианорабдитис от насыщенности гуанином и цитозином кодирующих их мРНК

На данном рисунке видно, что при увеличении ГЦ-насыщенности мРНК всех организмов происходит линейный рост содержания аминокислот группы GARP в соответствующих им белках. Максимальный наклон линии линейного тренда характерен для цианорабдитис (1,13), меньший – для трихинеллы (0,75), а минимальный – для человека (0,46). Установлено, что между содержанием GARP и ГЦ-насыщенностью цианорабдитис и трихинеллы существуют достоверные сильные прямые корреляционные связи ( $r=0,96\pm 0,089$ ,  $p<0,001$  и  $r=0,66\pm 0,238$ ,  $p<0,05$ , соответственно). Это иллюстрируют и высокие значения коэффициентов достоверности аппроксимации линейных трендов на рисунке ( $R^2=0,93$  и  $R^2=0,44$ , соответственно). Корреляционная связь между содержанием аминокислот группы GARP и ГЦ-насыщенностью кодирующих их мРНК человека не является достоверной ( $r=0,33\pm 0,299$ ,  $p>0,05$ ).

Перейдем к анализу закономерностей содержания кодируемых ГЦ-бедными кодонами аминокислот FУMINK (табл. 5) – фенилаланина (F; УУУ, УУЦ), тирозина – (Y; УАУ, УАЦ), метионина (M; АУА, АУГ), изолейцина (I; АУУ, АУЦ), аспарагина (N; ААУ, ААЦ) и лизина (K; ААА, ААГ)

Таблица 5

Доля аминокислот FУMINK в митохондриальных белках человека, трихинеллы и цианорабдитис

Фермент / показатель	Доля FYMINK у человека, %	Доля FYMINK у трихинеллы, %	Доля FYMINK у цианорабдитис, %	p по доле FYMINK у человека и трихинеллы	p по доле FYMINK у человека и цианорабдитис
НАДН-дегидрогеназа-1	28,0±2,52	34,1±2,74	39,5±2,87	>0,05	<0,01
НАДН-дегидрогеназа-2	32,6±2,52	30,2±2,67	46,1±2,97	>0,05	<0,001
НАДН-дегидрогеназа-3	30,5±4,29	44,0±4,61	45,9±4,73	<0,05	<0,05
НАДН-дегидрогеназа-4	29,0±2,12	25,1±2,14	42,3±2,44	>0,05	<0,001
НАДН-дегидрогеназа-4L	30,6±4,66	29,6±5,07	53,3±5,69	>0,05	<0,01
НАДН-дегидрогеназа-5	31,0±1,88	29,5±2,00	43,3±2,16	>0,05	<0,001
НАДН-дегидрогеназа-6	27,0±3,37	43,6±3,97	44,5±4,14	<0,01	<0,01
Цитохром b	31,5±2,39	36,7±2,50	37,3±2,51	>0,05	>0,05
Цитохром-с-оксидаза-1	31,2±2,05	33,1±2,08	30,9±2,02	>0,05	>0,05
Цитохром-с-оксидаза-2	27,3±2,96	30,2±3,06	32,9±3,09	>0,05	>0,05
Цитохром-с-оксидаза-3	26,2±2,73	32,3±2,92	36,1±3,01	>0,05	<0,05
АТФ-синтаза b	31,0±3,08	42,4±2,97	36,7±3,42	<0,05	>0,05

Содержание аминокислот FYMINK неразличимо статистически в 9-ти изученных белках ( $75,0 \pm 13,06\%$ ) человека и трихинеллы и лишь в 4-х белках человека и цианорабдитис ( $33,3 \pm 14,21\%$ ,  $p < 0,05$ ). При регрессионном анализе зависимости содержания FYMINK в изученных белках от ГЦ-насыщенности кодирующих их мРНК обнаружено, что при увеличении содержания гуанина и цитозина в мРНК всех организмов происходит линейное падение содержания аминокислот FYMINK в соответствующих им белках (рис. 2).

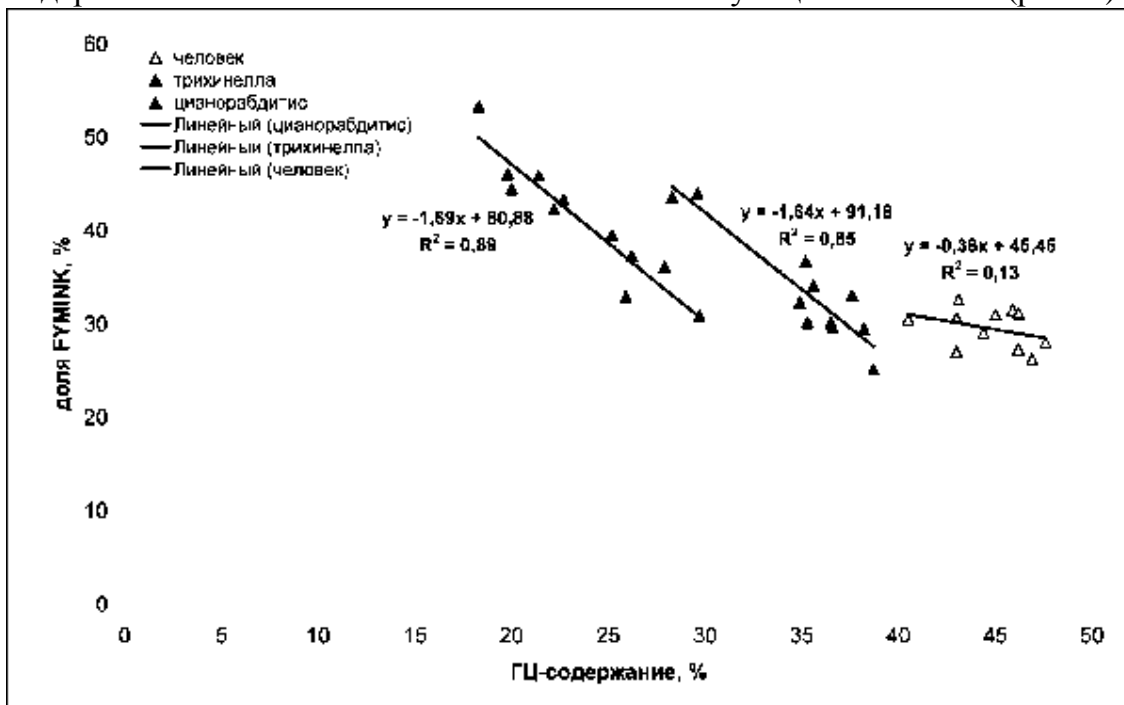


Рис. 2. Зависимость содержания аминокислот FYMINK в ряде митохондриальных белков человека, трихинеллы и цианорабдитис от насыщенности гуанином и цитозином кодирующих их мРНК

Наиболее выраженный наклон линии линейного тренда характерен для цианорабдитис ( $-1,69$ ), менее – для трихинеллы ( $-1,64$ ), а наименее выраженный – для человека ( $-0,36$ ). Установлено, что между содержанием FYMINK и ГЦ-насыщенностью мРНК цианорабдитис и трихинеллы существуют достоверные сильные обратные корреляционные связи ( $r = -0,94 \pm 0,114$ ,  $p < 0,001$  и  $r = -0,92 \pm 0,131$ ,  $p < 0,001$ , соответственно). Связь между данными показателями человека не является достоверной.

Наблюдаемые различия характера связи между содержанием аминокислот (групп GARP и FYMINK) и ГЦ-насыщенностью, могут быть объяснены с разной удаленностью ГЦ-насыщенности мРНК от точки равновесия с мутационным давлением (50,0%). Так, по мере удаления от данной точки наблюдается более выраженный наклон линий линейных трендов и модулей коэффициентов корреляции между содержанием изучаемых групп аминокислот и ГЦ-насыщенностью кодирующих их мРНК. По наклону линейных трендов и характеру корреляционной связи между содержанием данных групп аминокислот и ГЦ-насыщенностью мРНК, ближе к человеку находится трихинелла, а не цианорабдитис.

Полученные данные по содержанию аминокислот FYMINK и GARP свидетельствуют о том, что аминокислотный состав изучаемых белков также является объектом коэволюции человека и трихинеллы как компонентов системы «паразит-хозяин», формирующейся при трихинеллезе на молекулярно-генетическом уровне.

#### Выводы

1. Картина использования синонимичных кодонов в мРНК, кодирующих исследуемые митохондриальные белки человека достоверно более сходна с таковой трихинеллы по сравнению с контролем.
2. С ростом ГЦ-насыщенности изученных мРНК наблюдается линейный рост содержания аминокислот группы GARP и линейное падение содержания аминокислот группы FYMINK в кодируемых ими белках.
3. Содержание аминокислот групп GARP и FYMINK в изученных митохондриальных белках человека достоверно ближе к таковым трихинеллы по сравнению с цианорабдитис.

#### Литература

1. Бутвиловский, А.В. Динамика изменений нуклеотидных последовательностей матричных РНК и аминокислотных последовательностей алкогольдегидрогеназ в процессе эволюции: автореф. ... дис. канд. мед. наук: 03.00.04. – Минск, 2007. – 22 с.
2. Медик, В.А. Статистика в медицине и биологии: руководство в 2-х томах / В.А. Медик, М.С. Токмачев, Б.Б. Фишман // Под ред. Ю.М. Комарова. Т.1. Теоретическая статистика. – М.: Медицина, 2000. – С. 283 – 289.
3. Черноус, Е.А., Бутвиловский, А.В., Линник, Ю.И., Давыдов, А.В. Стратегия кодирования NADH-дегидрогеназ круглых червей // Матеріали XI



Міжнародного медичного конгресу студентів і молодих вчених – Тернопіль: Укрмедкнига, 2007. – С. 290.

4. Codon catalog usage is a genome strategy modulated for gene expressivity / R.Grantham [et. al] // *Nucleic Acids Res.*-1981.-Vol. 9. – P. 43-74.

5. Codon usage: mutational bias, translational selection or both? / P.M. Sharp [et. al] // *Biochem. Soc. Trans.* – 1993. – Vol. 21. – P. 835-841.

6. Kumar, S. 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.

7. Lavrov, D.V. *Trichinella spiralis* mtDNA. A nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans / D.V. Lavrov, W.M. Brown // *Genetics.* – 2001. – Vol. 157 (2). – P. 621-637.

8. McInerney, J.O. GCUA: General codon usage analysis / J.O. McInerney // *Bioinformatics.* – 1998. – Vol. 14(4). – P.372-373.

9. Mitochondrial genome variation and the origin of modern humans / Ingman M. [et. al] // *Nature.* – 2000. – Vol. 408. – P. 708-713.

10. Nei, M. *Molecular Evolution and Phylogenetics* / M. Nei, S. Kumar // Oxford University Press, New York, 2000.

11. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum* / R. Okimoto [et. al] // *Genetics.* – 1992. – Vol. 130 (3). – P. 471-498.