

ГЕНЕТИКА

УДК 577.3

© 1992 г. Е. В. КОРОТКОВ

MB1-СЕМЕЙСТВО ПОВТОРОВ В КЛОНАХ ИЗ ГЕНОМОВ НЕКОТОРЫХ МЛЕКОПИТАЮЩИХ

Обнаружены члены MB1-семейства повторов в геномах различных млекопитающих из банка данных EMBL-22. Проведена классификация членов MB1-семейства повторов по степени их подобия друг другу. Обнаружено разбиение MB1-повторов из генома человека на два подсемейства. В геномах кролика и быка MB1-повторы зеркальны по пуринам и пиримидинам семействам повторов кролика и быка типа SINE, известных ранее. Обсуждаются возможные механизмы образования MB1-семейства повторов, общего для многих млекопитающих.

Ранее в клонах из генома человека было обнаружено новое свойство зеркальной симметрии последовательностей оснований нуклеиновых кислот и новое семейство MB1-повторов (Коротков, 1991). Это MB1-семейство повторов насчитывает несколько сотен тысяч копий на геном человека и значительно отличается от повторов семейства SINE. У членов MB1-семейства повторов отсутствуют последовательности поли А на 3'-конце, и члены этого семейства не фланкированы короткими прямыми повторами. Члены MB1-семейства повторов зеркальны по пуринам и пиримидинам относительно членов семейства B1-повторов мыши или относительно первого мономера членов семейства Alu-повторов из генома человека. Эта выявленная зеркальная симметрия не может быть обнаружена экспериментально и не может быть создана ни одним из известных генетических процессов. Предполагается, что такой вид зеркальной симметрии между последовательностями оснований связан с синтезом двунитевой ДНК с параллельными нитями (Коротков, 1991).

Выявление нового семейства повторов произошло благодаря использованию в качестве меры связи между последовательностями взаимной информации. Это позволяет выявлять не только гомологичные или же комплементарные последовательности, но и последовательности, взаимосвязанные более сложно, в частности только по местоположению пуринов и пиримидинов (Коротков, Мамаева, 1989). В данной работе этот подход был использован для анализа степени распространенности MB1-повторов в клонах из геномов млекопитающих, позвоночных, беспозвоночных, а также прокариот из банка данных EMBL-22 (Hamp, Cameron, 1986). Проведена также классификация обнаруженных ранее членов MB1-семейства повторов из генома человека. Показано, что MB1-семейство повторов в геноме человека представлено в виде двух подсемейств.

Ранее (Коротков, 1990, 1991) члены MB1-семейства повторов кроме генома человека были также выявлены и в геномах мыши и крысы. В настоящей работе удалось показать присутствие членов MB1-семейства повторов в геномах некоторых млекопитающих, в том числе в геноме сумчатых млекопитающих. Полученные результаты показывают, что MB1-семейство повторов является общим для многих видов млекопитающих.

Для компьютерного анализа последовательностей оснований ДНК использовали клоны млекопитающих, позвоночных, беспозвоночных и прокариот, накопленных в банке данных EMBL-22. В качестве последовательностей для сравнения брали 16 первых MB1-последовательностей, выявленных ранее (Коротков, 1991). Это последовательности, которые содержатся в некодирующих районах клонов HSA1ATP (Long et al., 1984), HSAPC3A (Protter et al., 1984), HSBNGF (Ullrich et al., 1983), HSFIXG (Yoshitake, 1985), HSAAPDP (Hanauer, Mandel, 1984), HSHP201 (Maeda et al., 1984), HSIFNG (Gray, Goedde, 1982), HSIL05 (Holbrook et al., 1984), HSMG01 (Meller et al., 1984), HSMG02 (Meller et al., 1984), HSOPS (Nathaus, Hogness, 1984). В этих клонах содержатся разнообразные гены человека, такие, как α_1 -антитрипсин, кортикотропин β -липотропин, ген Аро СIII, и другие гены. В случае выявления всевозможных членов MB1-семейства повторов из генома человека этот набор MB1-повторов был достаточен для выявления почти всех членов MB1-семейства (Коротков, 1991).

Поскольку все ранее выявленные 49 MB1-последовательностей подобны друг другу в большинстве случаев по местоположению пуринов и пиримидинов, то поиск MB1-повторов проводился для ускорения компьютерного анализа только с использованием матрицы $M1$ (2.2). Четыре элемента этой матрицы показывают совпадение между двумя сравниваемыми последовательностями вида $PuPu$, $PuPy$, $PyPu$ и $PyPy$, где Pu — пурин, а Py — пиримидин.

Как и ранее, рассчитывали величину (Коротков, 1991)

$$2I = 2 \left(\sum_{i=1}^2 \sum_{j=1}^2 m_{ij} \ln m_{ij} - \sum_{i=1}^2 x_i \ln x_i - \sum_{j=1}^2 y_j \ln y_j + L \ln L \right), \quad (1)$$

где x_i — число Pu и Py на первой последовательности, $i = 1, 2$, а y_j — число Pu и Py во второй последовательности, m_{ij} — элементы матрицы $M1$, L — длина сравниваемых последовательностей.

Поскольку в реальном генетическом тексте наблюдается попарная скоррелированность оснований (Компьютерный анализ ..., 1990), то распределение величины $2I$ может отличаться от распределения $\chi^2(1)$. Для проверки соответствия распределения величины $2I$ распределению $\chi^2(1)$ функция распределения величины $2I$ определялась экспериментально. Для этого сравнивали различные некодирующие последовательности из клонов человека со всеми последовательностями клонов человека из банка данных EMBL-22. Общий объем такого сравнения превышал 10^7 . Такой анализ показал, что $2I$ имеет распределение $\chi^2(1)$, если исключить из рассмотрения последовательности с попарной скоррелированностью оснований. Доля таких последовательностей составляет более 60% от всех последовательностей, накопленных в банке данных EMBL-22. Для того чтобы исключить влияние попарной скоррелированности оснований в генетическом тексте на значение величины $2I$, определяли также матрицу $M'(4,4)$. Эта матрица представляет собой частоты совпадения пар $PuPy$, $PyPu$, $PuPu$ и $PyPy$ между сравниваемыми последовательностями. Матрицу $M'(4,4)$ подсчитывали таким образом, что каждое следующее совпадение пар $PuPy$, $PyPu$, $PuPu$ или $PyPy$ между двумя сравниваемыми последовательностями получалось из предыдущего сдвигом вправо на одно основание. Элементами матрицы M' являются частоты совпадений типа A_1B_1/A_2B_2 , где A_1 и B_1 — пурин или пиримидин на первой последовательности, а A_2 и B_2 — пурин или пиримидин на второй последовательности.

Определяли величину

$$2I' = 2 \left(\sum_{i=1}^4 \sum_{j=1}^4 m'_{ij} \ln m'_{ij} - \sum_{i=1}^4 x'_i \ln x'_i - \sum_{j=1}^4 y'_j \ln y'_j + L \ln L \right), \quad (2)$$

где m'_{ij} — элементы матрицы $M'(4,4)$, x'_i — число $PuPu$, $PyPy$, $PuPy$ и $PyPu$ на

$M1$		M'				
	P_u	P_y	$P_u P_u$	$P_y P_u$	$P_u P_y$	$P_y P_y$
P_u	50	0	0	0	0	0
P_y	0	50	0	50	0	0

Рис. 1. Вид матриц $M1$ и M' для двух внутренне скоррелированных последовательностей (см. текст статьи)

первой последовательности, y_j — число тех же пар на второй последовательности. На величины x_i' наложены следующие ограничения:

$$\begin{aligned} x_1' (P_u P_u) + x_2' (P_u P_y) &= x_1 (P_u), \\ x_3' (P_y P_u) + x_4' (P_y P_y) &= x_2 (P_y), \\ x_1' (P_u P_u) + x_3' (P_y P_u) &= x_1 (P_u), \\ x_3' (P_y P_u) + x_4' (P_y P_y) &= x_2 (P_y). \end{aligned}$$

На величины y_j' наложены такие же ограничения, и поэтому величина $2I'$ будет распределена как χ^2 с одной степенью свободы (Кульбак, 1967). Для случайных последовательностей $2I = I'$.

Безусловной взаимной информации (или взаимной энтропии) соответствует условная взаимная информация (или взаимная энтропия), которую можно оценить как $F' = (I' - I)$. Взаимная информация является условной взаимной информацией следующего элемента между двумя сравниваемыми последовательностями, если известна взаимная информация предыдущего элемента, умноженная на длину сравниваемых последовательностей. Расчет F' позволяет определить такую меру подобия между сравниваемыми последовательностями, на которую будет полностью исключено влияние попарной скоррелированности пуринов и пиримидинов в сравниваемых последовательностях. Например, рассмотрим две периодические последовательности длиной в 100 н. п.:

5'-ATCGATCGATCGATCGATCG...-3' — первая последовательность,
5'-GCATGCATGCATGCATGCAT...-3' — вторая последовательность.

Матрицы $M1$ и M' принимают вид, показанный на рис. 1. Величина $I = 100 \cdot \ln 2 \approx 69,3$, а величина $I' = 100 \cdot \ln 2 \approx 69,3$. В таком случае $F' = I' - I = 0$. Это показывает, что подобие между двумя последовательностями связано только с видом их внутренней структуры, т. е. со скоррелированностью оснований между собой.

Для величины F' была также получена экспериментальная функция распределения посредством сравнения разнообразных последовательностей с последовательностями из банка данных EMBL-22. Оказалось, что эта функция распределения для величины F' очень хорошо согласуется с распределением χ^2 с одной степенью свободы. Таким образом, вычисление F' позволяет достаточно точно определить вероятность того, что взаимосвязь между сравниваемыми последовательностями обусловлена чисто случайными факторами.

В качестве взаимосвязанных выбирали такие последовательности, которые имели вероятность случайной взаимосвязи α , определенной с учетом числа независимых испытаний, менее 5%. Расчет величины α проводили так же, как

и ранее (Коротков, Мамаева, 1989; Коротков, 1991), только вместо величины I для такого подсчета использовали величину F' . Для поиска MB1-повторов по клонам позвоночных, млекопитающих, беспозвоночных, прокариот банка данных EMBL-22 использовали величину $2F' \geq 42$, что дает вероятность α менее 5%.

Проводили также попарное сравнение 49 MB1-повторов, полученных ранее (Коротков, 1991). Такое попарное сравнение MB1-повторов позволило заполнить матрицу A (49,49) — матрицу подобия между всеми MB1-повторами. В матрице A записаны величины $2F'$ при сравнении любой пары MB1-повторов. Матрица A является симметричной относительно главной диагонали. Главную диагональ, показывающую подобие любой MB1-последовательности самой себе, не заполняли.

На основе матрицы A был построен граф максимальных взаимосвязей 49 MB1-последовательностей между собой. Для построения этого графа по матрице A для каждой MB1-последовательности определяли другую MB1-последовательность, имеющую с ней наибольшую взаимосвязь. Эту связь на графе изображали в виде стрелки. Ее направление показывает, что для той MB1-последовательности, из которой исходит стрелка, другая MB1-последовательность, в которую упирается стрелка, является наиболее подобной ей из всех 49 последовательностей. Учитывались также минорные связи, но только те, которые меньше основной связи по величине χ^2 на четыре-пять единиц. Такое уменьшение силы связи соответствует увеличению вероятности того, что связь между MB1-последовательностями обусловлена чисто случайными факторами не более чем в 10 раз. Случаи, когда i -элемент имеет наибольшую связь с j -элементом и j -элемент имеет наибольшую взаимосвязь с i -элементом, показаны на графе двумя стрелками. Для такой пары MB1-последовательностей минорные связи показывали в любом случае. Сила связи в величинах $2F'$ показана на графе рядом со связью.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

16 MB1-повторов из генома человека, обнаруженные в предыдущей работе (Коротков, 1991), были сравнены с клонами из геномов млекопитающих, позвоночных, беспозвоночных и прокариот из банка данных EMBL-22. Члены MB1-семейства повторов были выявлены только в клонах в классе млекопитающих — в инфраклассе высших и низших зверей. В клонах из геномов прокариот, позвоночных и беспозвоночных MB1-повторы не были обнаружены. Класс млекопитающих включает в свой состав 17 отрядов. Члены MB1-семейства повторов были ранее выявлены в отряде приматов (человек) и грызунов (мышь и крыса) (Коротков, 1991). В настоящей работе они также выявлены в отряде приматов (зеленая мартышка), зайцеобразных (кролик), хищных (собака), непарнокопытных (лошадь), парнокопытных (бык, овца), а также в инфраклассе низших зверей — в отряде сумчатых (опоссум). Список клонов, где обнаружены MB1-последовательности, показан в табл. 1, а сами последовательности вместе с координатами в клоне — на схеме 1.

Все MB1-последовательности обнаружены в некодирующих районах (5'- или 3'-фланкирующие районы и интроны разнообразных генов). Величина $2F'$ для всех случаев подобия MB1-последовательностей генома человека вновь выявленным членам MB1-семейства из геномов млекопитающих превышает 42. Это давало вероятность возникновения данной взаимосвязи из-за случайных факторов с учетом объема проанализированных клонов из геномов млекопитающих менее 5%.

Таким образом, из 17 отрядов высших зверей MB1-повторы выявлены в шести отрядах. Можно с достаточно большой долей вероятности считать, что члены MB1-семейства повторов можно выявить в геномах всех млекопитающих. Только отсутствие достаточно большого количества клонов с известными последовательностями из геномов различных млекопитающих в банке данных EMBL-22, по-видимому, не позволило выявить MB1-семейство повторов в остальных 11 отрядах млекопитающих.

Особенно интересно обнаружение MB1-последовательности в клоне DVNEBEV

Клоны млекопитающих, где выявлены МВ1-повторы

Номер клона	Шифр клона	$\chi^2(1)$	Вид соответствия по- следовательностей	Ссылка
1	HSA1ATP*	49	—	Long et al., 1984
2	ВТРОМС1		ATCG	Watanabe, 1982
3	CAREP3	48	ATCG	Saffer, Lerman, 1983
4	OAMTIB	43	ATCG	Peterson et al., 1988
5	OAMTH	77	ATCGACTG	
6	HSAPC3A*	57	TAGCCAGT	То же
7	BTACHR		AGCT	Protter et al., 1984
8	OCILIR	58	TCGA	Tanabe et al., 1984
9	HSMG02*	54	—	Furutani et al., 1985
10	OCIG05		TGACCAG	Meller et al., 1984
11	DVHBBB	46	ACTGACT	
12	HSILO5*	47	ACTG	Emorine, Max, 1983
13	BTNA2		ACTG	Koop, Goodman, 1988
14	BTTHBNA2	45	GATCCT	
15	CFSRP54	50	CTAGAG	Holbrook et al., 1984
16	ECIFN10	44	AGCCTT	
			AGTCCT	Lochrine et al., 1985
			—	
			ATCGATCG	Irwin et al., 1985
			ATCGGCTA	
			ACTGGATT	Romisch et al., 1989
			TGACACCG	
			ATCG	Himmler et al., 1986
			TAGC	
			ACTGTAGA	
			TAACGCTG	

* Шифры клонов человека, МВ1-последовательности которых были использованы для сравнения.

Схема 1

1	TGTGTATTTA	ATTTCACAAT	AGOTCTATGA	CAAAAGTCCAG	CTTTCTCATC	TCCAGGAAAC	TGAGGTTCAAG	AGAGGTTAAG
2	242-catotattggg	gtogtsoaga	gtoggacacg	aotgaagcaa	ottagoagaa	gaagcagcag	agaaactcaa	agaaatttgaa
3	630-cttggtactc	ttcacccagg	accatttga	ggtaggcoot	caagttacag	atggcgaaac	TCCGGAOCAG	AGAGGTTAAAG
4	3066-gacatacag	ggttacggaa	atcacgtgaa	ooteotacac	AGGGGAAAT	GGACTCTTCG	GTTCGAGTC	TCTCCATTGG
5	216-taaatttaga	ggtaatttgg	acacotcatt	oocgatcagt	agtaagcgta	AAATCTGATA	CCTCCGGGTC	TCGCCAATTC
6	TTAGTCCOOC	GAACCGGCTT	TGGGGTAGGT	GTTATTTTCT	CACTTTTCAG	ATGAGAAAAAT	TGAGGCTCAG	AGCGATTAGG
7	2262-toagatoccc	gggtotaagt	ttatgaacat	tataatcgaa	GTAATAATGC	TACCCCTTTG	ACTCCGTATC	TGTCAAAATC
8	1372-ctcttggtaa	ttccgtcatt	gttaCATOCA	AAATAAGAAAG	GTAGAATTTTTG	TATTATTG	ACTCCAAAGT	CCTTGAATAC
9	ACTATCTCAC	TTAATCCTTC	AACTAGGCAC	AACTTATCCC	CATCCCTTAT	ATGAGGAAGC	TGAGGCACAG	AGAGGTGAAG
10	961-ctaataaag	agataaatg	aaaaaccacg	ccggtcotta	oogagtaaat	tgaatttgta	gtgtttgaa	gggtccacta
11	1171-catagattta	agttagaag	gaactccAGT	AACTGCTTTT	TATTTTACAG	TTGAGGAGCC	TGAGGCCCTA	TCATGATAAG
12	CAGATTATAT	TCAGAAATGC	TGAACACTTC	ACTACCGAGG	ATAAACGCTC	TAATTTTTTG	ACTCCGAGTC	TCTTCAATTC
13	1586-agtaatttca	TTGGAACACT	TCATCATCTC	TTTAATGGAA	GTAATAATATC	TATTCCTGTA	GTTCGGAGTC	TGTTTACTTT
14				1-gaTC	CCTTTTCACAG	GGAAACCGAA	TCCCGGTCAG	AGAGGTTAAG
15	157-gaaagactoc	coggcooett	tggggcagc	ggtagcggcg	ctTTTTTCGG	GTGAGGAAAG	TAAGGCTCAG	AGAGATTAAAG
16	980-ggaactagaaa	tgtoctgata	tagaATGAA	TAAGAGCGCG	TGTTTTTAA	ATCAAAACAGT	TAAATCAAA	ATAACGAGA

1	TAACCTCTCC	AAAGTCACAC	AGCTAAATAGC	AACTTGACGT	GGAGCAATCT	GGCCTCAGAG	CCTTAAATTT
2	TAACCTCTCC	AAAGTCACAG	AGCTAAATAGT	ttctccctcc	aaacccaaot	gocccotttag	gottctocag-391
3	ATACTGTATC	AGAACGACAG	GCTTGGToga	tggcagaago	aggattcaaa	gtgaggatac	ccagacattt-789
4	ACTGGACAAA	TTCCatgagt	caagaatcc	cogttgacaa	atocaggtct	tgggttogg	gacgtaaag-2917
5	ATTGAAACGG	TTCCATGTCT	TCGATCACTc	acgtctctgt	acacctaagt	ttagtcaacc	gggttgcteta-67
6	TGACCTGCCC	CAGATCACAC	AACTAAATCAA	TCTTCAATG	ACTTTTCCAAA	TGAGAGGCTG	CCTCCCTCTG
7	ATCGAACAGG	TTTCAactc	gtggtttcga	ccttgaaott	caaattttcc	acacgaaaaa	taggagaat-2113
8	GATAAACTAA	TTCTtagatt	gtogaattg	tgctcttcog	tttataogct	ttactataag	acacatogaa-1222

9 TGAATGGCCC AAGGTCACAC AGCTGGGAAG ACAGGGAGCT AAACCTGAAAC TCTAGTCTGG CTGCCCCAG
 10 ggtcccoctgg TTCTGGAGTG TAAACCTGTC TTTCCTGAA TCCAAACTAA AGGTCAAACC AGGGAccccg-812
 11 TAAATCATCT AAGGCCACAG GGCAATGAAT GACCAATATG CAATTTGAAT CCATACCCCT TGACTCCaaa-1320
 12 ATTAACCGA GATCCAGTGA GTCCGGGAATA ACGGTCATTG CCGGTCTCTT CTTAAACTTT TCTCTTACAC
 13 ATaatgttaa aaacgtatgt atcattoaac gtotoggtta ttacctagga cccagaagac taaaggacca-1437
 14 TGACTTGaaa togttaacgt caaaggttaag ggggaagago ttcccatoot gtotggotoc aotgtggtoc-114
 15 TAACTTGCCG aaagtaaac agccagtatg taacagagct gggatttgaa tcaaggtctg toagactcgc-306
 16 CGAGTTACCT CTtttaataa ettataaate atgaacgggt agtaatgaaa otogaacocg atatggttot-831

Последовательности MB1-повторов в геномах некоторых млекопитающих. Цифры слева и справа от последовательности показывают ее координаты в соответствующем клоне. Нумерация последовательностей совпадает с нумерацией соответствующих им клонов в табл. 1

Схема 2

Зеркальное подобие консенсус-последовательности SINE-повтора быка (1), полученной в настоящей работе (табл. 2), и консенсус-последовательности MB1-повторов (2), полученной ранее (Коротков, 1991). Зеркально совпадающие основания выделены заглавными буквами. Зеркальное подобие SINE-повтора быка в клоне BTIRBP (Borst et al., 1989) (3) и MB1-повтора в клоне HSMGO2 (4) (Meller et al., 1984). Зеркально совпадающие основания также выделены заглавными буквами

5'-ACCccATGAA CTgCAGCACA CCAAgCtTCC CCgTCC-3' (1)
 5'-TAAcTGCCC AAgGTCACAC AGCTaGtAAG GGgGAG-3' (2)
 6686-AGACTgCAGC aTgCCAAGCC TCCCGTCTCTT CA-6717 (3)
 88-CCCAAAGTCA gAgAGCTgGG AAGACAGGGA GC-119 (4)

Схема 3

1 3'-CTTGCAATG TTACCTTTTT TGATTAATG TGAGGGAGG GCACCTCTCT ATCACTTTGT COTCTGAGT TCTCCACTAT
 2 2936 3'-taaaatctct atctcaatgt ctctctctct cTCTGTTTC AGAAGATAGA CGACGAAGTG AGGGGTCTAc cggcgttctc
 3 2047 3'-ctctctctct cagctctctct ctctctctCT CTCTCTCTCT AGAAGGTAGA CGACCAAGTG AGGGGTCTTAc CTGCATTATC
 4 1872 3'-atctcaatgt ccaactttct cctctctCTCT CTCTCTCTCT AGAAGGTAGG TAACCAAGTG AGGGGTCTAC CGGTCTTACC
 5 1165 3'-ctctctctct ctctctctct ctctctctCT CTCTCTCTCT TTTCGGTGGG TGACTTAAGTG AGAGGTTTAC GGACCGTCTC
 6 451 3'-caataagctt ctctctctct CTCTCTCTCT CTCTCTTTTG AGAAGATAGA CGACCAAGTA AGGGGTCTAc CGGGCTTACC

1 CCTGGACCAG TTCCAGGGTG TGGACTTTTG ACCCAATCT TGGGTACCCA GGACTGACAC AGGTACAGTG- 5'
 2 gaoccgagoc cagtoeggtt togtctcccg gtotggaaa aaggctccag aggtactct caggtcccg- 5' 2786
 3 GGTCCCGGCC cgtccogctt togtctcccg aagtagaacc aacccgggga aggagggcto caggaagga- 5' 1898
 4 GGTCTTAAGT CGTCAGGCT togtctcccg gtotaggaa aagaoccaaga ggtatacaaa atccocgggt- 5' 1723
 5 GACCCGAGCT CCGTCCGATT ttggtatca ggtcttgagt tagacctga gggcacaga aggtccctg- 5' 1016
 6 GGCTCGAGC gactaggtt togtctcccg gaggaggoc agaagggtgt accccagtc cggggttact- 5' 302

Зеркальное подобие MB1-последовательности из клона HSACTH с повтором типа SINE из генома кролика. Зеркально совпадающие основания показаны заглавными буквами. Координаты слева и справа от последовательностей показывают расположение последовательностей в соответствующих клонах: 1 — MB1-повтор клона HSACTH (6073—5824) (Takahachi et al., 1983). Повторы типа SINE из клонов: 2 — OCSPARO, $\chi^2(1) = 30$ (Boggaram et al., 1988); 3 — OCCYP45E, $\chi^2(1) = 32$ (Kagawa et al., 1987); 4 — OCMHRLAB, $\chi^2(1) = 33$ (Rebier et al., 1987); 5 — OCPLA2, $\chi^2(1) = 37$ (Narayanau et al., 1988); 6 — OCPG65, $\chi^2(1) = 32$ (Misrachi et al., 1988)

(Koop, Goodman, 1988) из генома опоссума, так как считается, что сумчатые млекопитающие являются более древними, чем плацентарные млекопитающие. С учетом того, что сумчатые млекопитающие образовались в меловом периоде мезозоя, время жизни MB1-семейства повторов можно оценить в 100 млн. лет.

В настоящее время среднечастотные семейства повторов типа SINE обнаружены у многих млекопитающих. К их числу относят Alu повторы из генома человека и некоторых обезьян (Quentin, 1988; Daniels, Deininger, 1983), B1-повторы из геномов мыши и крысы (Quentin, 1989), повторы из геномов кролика и быка (Molecular Evolutionary..., 1985). Все эти семейства повторов имеют длину в несколько сотен пар оснований, они окаймлены короткими повторами и содержат на 3'-конце район поли A. Однако между этими семействами повторов из

Шифры клонов, содержащих повторы быка и координаты повторов в клоне

Номер клона	Шифр клона	Координаты повтора	Ссылка
1	BTIRBP	6649—6748	Borst et al., 1989
2	BTACHRA1	2618—2717	Noda et al., 1983
3	BTALUL1	14—113	Duncan, 1987
4	BTALUL2	14—113	To же
5	BTCASK35	630—531	Alexander, 1988
	To же	3894—3795	To же
	»	2447—2546	»
	»	2590—2689	»
6	BTGLO1	400—499	Schimenti, Duncan, 1984
7	BTNBP	4552—4651	Brunner et al., 1986
8	BTNBPP3	1137—1236	To же
9	BTGLO2	1573—1474	Schimenti, Duncan, 1984
10	BTGLO1	556—655	To же
11	BTCASK35	4910—5009	Alexander, 1988
	To же	683—584	To же

различных отрядов млекопитающих не отмечена сколько-нибудь значительная гомология. В этом смысле MB1-семейство повторов является первым случаем семейства, которое можно найти у многих, если не у всех, млекопитающих. Также надо заметить, что MB1-семейство повторов значительно отличается от семейства типа SINE. У членов MB1-семейства повторов, как это отмечалось ранее (Коротков, 1991), отсутствует последовательность поли А на 3'-конце, и они не фланкированы короткими повторами. Степень дивергенции между членами MB1-семейства повторов внутри генома человека значительно превышает таковую у повторов типа SINE. Также степень эволюционной дивергенции нуклеотидов между членами MB1-семейства повторов внутри одного вида, как это видно из результатов этой работы и проделанной ранее (Коротков, 1991), сравнима со степенью дивергенции членов MB1-семейства между различными видами. Полученные результаты позволяют предположить, что образование видов, которые предшествовали млекопитающим, сопровождалось возникновением MB1-семейства повторов в начале мелового периода или же раньше. В таком случае возникновение отдельных видов млекопитающих может быть связано с расселением различных семейств повторов типа SINE в процессе видообразования.

В настоящее время не ясна роль зеркальной симметрии между MB1-семейством повторов и семейством SINE человека и B1-семействами мыши и крысы. Однако поскольку такая зеркальная симметрия наблюдается только по местоположению пуринов и пиримидинов, то члены семейств повторов типа SINE в различных видах также могут быть зеркальны MB1-семейству повторов без сохранения значительной гомологии между этими семействами. Такой тип зеркальной симметрии был выявлен для повторов типа SINE быка (Molecular Evolutionary..., 1985) и MB1-повторов человека. На схеме 2 показано зеркальное совпадение между консенсус-последовательностью MB1-повторов, выделенной ранее (Коротков, 1991), и консенсус-последовательностью SINE-семейства повторов быка, полученной в настоящей работе по 15 повторам, которые содержатся в клонах, показанных в табл. 2. Район зеркальной симметрии имеет длину в 36 оснований с $\chi^2(1) = 16,5$ и приходится на наиболее консервативный участок MB1-повторов. Общая длина наиболее консервативной части всех MB1-повторов составляет, как это было показано ранее (Коротков, 1991), около 80 нуклеотидных пар. Повтор быка был выделен в клонах из генома быка как имеющий длину в 115 оснований. Вероятность такого зеркального совпадения вследствие случайных факторов α между консенсус-последовательностью SINE-повторов быка и консенсус-последовательностью MB1-повторов человека составляет менее $2 \cdot 10^{-2}$.

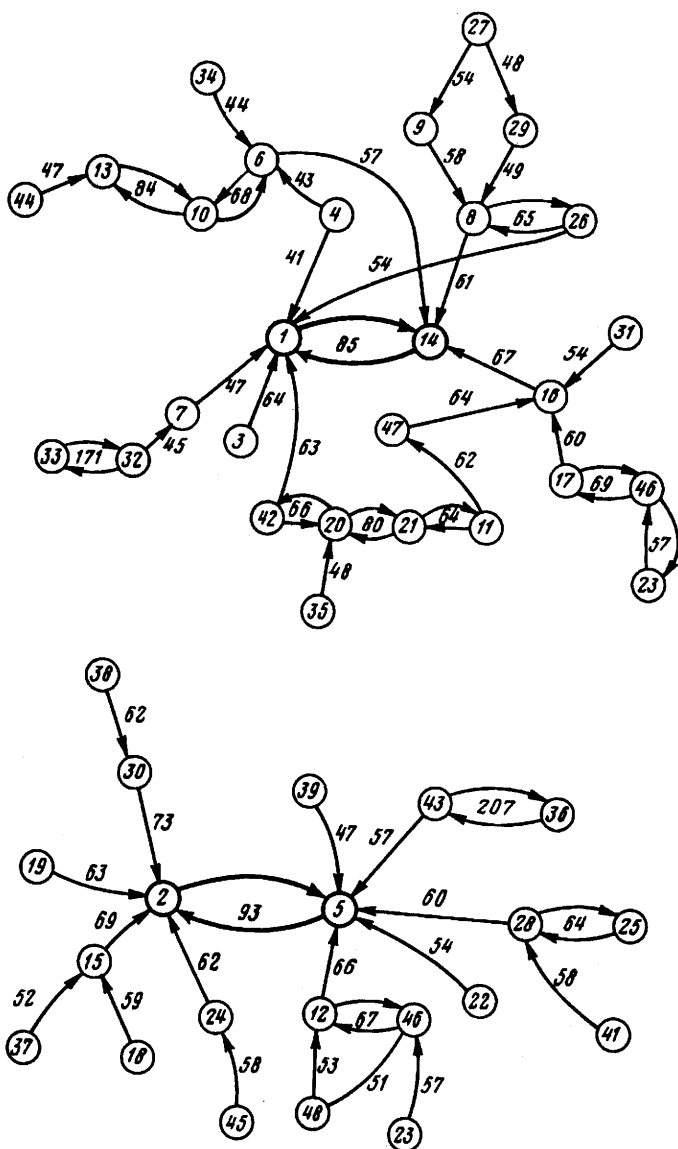


Рис. 2. Граф наибольшего подобия MB1-последовательностей из генома человека. Список использованных MB1-последовательностей дан в табл. 3. Алгоритм построения графа описан в тексте статьи

В геноме кролика также удалось выявить зеркальную симметрию между MB1-повтором человека и повтором кролика (схема 3). Выявлено зеркальное подобие MB1-повтора из клона HSACTH (Takahashi et al., 1983), содержащего ген кортикотропин-липотропина человека с повтором типа SINE кролика, содержащимся в пяти различных клонах кролика. Такое зеркальное соответствие наблюдается с величинами $\chi^2(1)$ в диапазоне от 30 до 37. Однако в данном случае выделить зеркальное соответствие консенсус-последовательности MB1-повторов человека и консенсус-последовательности повторов кролика на уровне $\alpha \leq 0,05$ не удалось. По-видимому, это может быть связано с тем, что MB1-повторы могут сами достаточно значительно дивергировать между различными видами млекопитающих. Поэтому консенсус-последовательности MB1-повторов из генома

Шифры клонов и координаты в клонах MB1-последовательностей,
используемых при построении графа наибольшего подобия
MB1-последовательностей

Номер клона	Шифр клона	Координаты MB1-последовательности	Авторы
1	HSA1ATP	3236—3385	Long et al., 1984
2	HSA1ATP	610—759	To же
3	HSA1ATP	1207—1356	*
4	HSACTH	6073—5924	Takahashi et al., 1983
5	HSAPC3A	2491—2640	Protter et al., 1984
6	HSAPC3A	4087—4236	To же
7	HSBNGF	2873—2624	Ullrich et al., 1983
8	HSFIXG	11482—11631	Yoshitake et al., 1985
9	HSGAPDP	1513—1364	Hanauer, Mandel, 1984
10	HSHP201	1072—1221	Maeda et al., 1984
11	HSIFNG	4185—4334	Gray, Goedde, 1982
12	HSILO5	19—168	Holbrook et al., 1984
13	HSMGO1	3317—3168	Meller et al., 1984
14	HSMGO2	1064—1263	To же
15	HSMGO2	1763—1912	*
16	HSOPS	4902—4753	Nathaus, Hogness, 1984
17	HS1ATP	2405—2554	Long et al., 1984
18	HS1ATP	2714—2863	To же
19	HS1ATP	7080—6931	*
20	HSACHR8	1744—1893	Noda et al., 1983
21	HSACHR8	1793—1942	To же
22	HSARS1	134—383	Montiel et al., 1984
23	HSENKPH2	2567—2716	Comb et al., 1983
24	HSFIXG	36574—36425	Yoshitake et al., 1985
25	HSIFNG	5857—5708	Gray, Goedde, 1982
26	HSIL1R	1396—1247	Furutani et al., 1985
27	HSMGO2	1045—901	Meller et al., 1984
28	HSOPS	1345—1489	Nathaus, Hogness, 1984
29	HSPROL1	266—117	Truong et al., 1984
30	HSAPC3A	475—624	Protter et al., 1984
31	HS1ATP	6233—6382	Long et al., 1984
32	HSENKPH2	3965—4114	Comb et al., 1983
33	HSBB4R1	151—2	Mager, Heuthorn, 1984
34	HSIGK16	1151—1002	Pech et al., 1984
35	HSOPS	5838—5694	Meller et al., 1984
36	HSHP201	4360—4509	Maeda et al., 1984
37	HSILLO5	2291—2147	Holbrook et al., 1984
38	HSILL1R	1125—1274	Furutani et al., 1985
39	HSILO5	644—793	Holbrook et al., 1985
40	HSREN04	948—799	Hardman et al., 1984
41	HSREP10	2667—2519	Lee et al., 1984
42	HSBNGF	3561—3710	Ullrich et al., 1983
43	HSGL06	1240—1091	Sawada et al., 1983
44	HSIGK16	245—96	Pech et al., 1984
45	HS1ATP	3666—3517	Longee et al., 1984
46	HSREP10	3684—3833	Lee et al., 1984
47	HS1HR2	1762—1912	Friemer Degen et al., 1983
48	HSENKB1	344—493	Horikawa et al., 1982
49	HSDHFR02	398—249	Yang et al., 1984

человека. кролика и быка могут достаточно сильно между собой различаться, и только некоторые члены MB1-семейства повторов из генома человека будут зеркальными повторам типа кролика.

У остальных видов млекопитающих выделить повторы, зеркальные MB1-повторам человека, в банке данных EMBL-22, не удалось. Это может быть связано с тем, что объем известных клонов для многих млекопитающих недостаточен для проведения такого анализа. Кроме того, необходимо определить консенсус-

последовательность MB1-повторов, характерную для данного вида млекопитающих. Все это можно будет сделать по мере накопления последовательностей из геномов различных млекопитающих в банке данных нуклеотидных последовательностей.

Граф наибольшего подобия MB1-последовательностей из генома человека позволил провести классификацию MB1-повторов (рис. 2, табл. 3). Как видно из рис. 2, члены MB1-семейства повторов четко разбились на две группы. Связи между членами этих групп всегда носят минорный характер по сравнению со связями MB1-последовательностей внутри группы. Центром первой группы является MB1-последовательность около α_1 -антитрипсина человека и гаптоглобина HP2. Центром второй группы является другой MB1-повтор около α_1 -антитрипсина человека и MB1-повтор в интроне III Aро CII гена. Можно предполагать существование двух различных центров расселения MB1-повторов. Причем источниками такого расселения могут быть не обязательно центры выделенных групп, а какие-либо другие MB1-последовательности, с которыми центры выделенных групп могут быть связаны. Следует отметить, что подобное же разбиение на подмножества наблюдается у семейства Alu повторов из генома человека (Quentin, 1988) и у B1-повторов из генома мыши (Quentin, 1989).

В целом полученные в настоящей работе результаты показывают, что в геномах многих млекопитающих существует общее для них семейство повторов. Такое семейство повторов обнаружено впервые. Также можно предполагать, что образование нового вида млекопитающих связано с возникновением и расселением по геному нового семейства повторов типа SINE, зеркального по местоположению пуринов и пиримидинов семейству MB1-повторов.

СПИСОК ЛИТЕРАТУРЫ

- Компьютерный анализ генетических текстов/Под ред. Франк-Каменецкого М. Д. М.: Наука, 1990. 263 с.
- Коротков Е. В. Семейство зеркальных B1-подобных повторов в геноме человека//Докл. АН СССР. 1990. Т. 311. С. 238.
- Коротков Е. В. Новое семейство широко распространенных MB1-повторов в геноме человека//Мол. биол. 1991. Т. 25. С. 250.
- Коротков Е. В., Мамаева Е. Ф. ID-подобные последовательности в клонах ДНК некоторых видов//Мол. биол. 1989. Т. 23. С. 1113.
- Кульбак С. Теория информации и статистика. М.: Наука, 1967. 408 с.
- Alexander L. J., Stewart A. F., Mackinlay A. G., Kapenskaya T. V., Tkach T. M., Gorodetsky S. I. Isolation and characterization of the bovine k-casein gene//Eur. J. Biochem. 1988. V. 178. P. 395.
- Boggaram V., Qing K., Mendelson C. R. The major apoprotein of rabbit pulmonary surfactant//J. Biol. Chem. 1988. V. 263. P. 2939.
- Borst D. E., Redmond T. M., Elser J. E., Conda M. A., Wiggert B. W., Chader G. J., Nickerson J. M. Interphotoreceptor retinoid-binding protein//J. Biol. Chem. 1989. V. 264. P. 1115.
- Brunner A. M., Schimenti J. C., Duncan C. H. Dual evolutionary models in the bovine globine locus//Biochemistry. 1986. V. 25. P. 5028.
- Comb M., Rosen H., Seftburg P., Adelman J., Herbert E. Primary structure of the human proenkephalin gene//DNA. 1983. V. 2. P. 213.
- Daniels G. R., Deininger P. L. A second major class of Alu family repeated DNA sequences in a primate genome//Nucl. Acids Res. 1983. V. 21. P. 7995.
- Duncan C. H. Novel Alu-type repeat in artidactyls//Nucl. Acids Res. 1987. V. 15. P. 1340.
- Emorine L., Max E. E. Structural analysis of a rabbit immunoglobulin J-C locus reveals multiple deletions//Nucl. Acids Res. 1983. V. 11. P. 8877.
- Frierer Degen S. J., Macgillivray R. T. A., Davie F. W. Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin//Biochemistry. 1983. V. 22. P. 2087.
- Furutani Y., Notake M., Yamayoshi M., Yamagishi Y.-I., Nomura N., Ohne M., Furuta R., Fukui T., Yamada M., Nakamura S. Cloning and characterization of the cDNA for human and rabbit interleukin-1 precursor//Nucl. Acids Res. 1985. V. 16. P. 5868.
- Gray P. W., Goedde D. V. Structure of the human immune interferon gene//Nature. 1982. V. 298. P. 859.
- Ham G. H., Cameron G. H. The EMBL date library//Nucl. Acids Res. 1986. V. 14. P. 5.
- Hanauer A., Mandel J. The glyceraldehyde 3-phosphate dehydrogenase gene family: structure of a human cDNA and X-chromosome linked pseudogene: amazing complexity of the gene family in mouse//EMBO J. 1984. V. 3. P. 2627.

- Hardman J. A., Hort Y. J., Catanzaro D. F., Tellam J. T., Baxter J. D., Morris B. J., Shine J. Primary structure of the human renin gene//DNA. 1984. V. 3. P. 457.
- Himmler A., Hauptmann R., Adolf G. R., Swetly P. Molecular cloning and expression in *E. coli* of equine type 1 interferons//DNA. 1986. V. 5. P. 345.
- Holbrook N. J., Lieber M., Crabtree G. R. DNA sequence of the 5'-flanking region of the human interleukin 2 gene: homologies with adult T-cell leukemia virus//Nucl. Acids Res. 1984. V. 12. P. 5005.
- Horicawa S., Takai T., Tomosato M., Takahashi H., Noda M., Kakidani H., Kubo T., Hirose T., Thagama S., Hagahashida H., Miyata T., Numa S. Isolation and structural organization of the human preproenkephalin gene//Nature. 1982. V. 297. P. 451.
- Irwin D. M., Ahern K. G., Pearson G. D., McGillivray R. T. Characterization of the bovine prothrombin gene//Biochemistry. 1985. V. 24. P. 6854.
- Kagawa N., Mihara K., Sato R. Structural analysis of cloned cDNA for polycyclic hydrocarbon-inducible forms of rabbit liver microsomal cytochrome P-450//J. Biochem. 1987. V. 101. P. 1471.
- Koop B. F., Goodman M. Evolutionary and developmental aspects of two hemoglobin β -chain genes (ϵ^m and β^m) of opossum//Proc. Natl Acad. Sci. USA. 1988. V. 85. P. 3893.
- Lee M. G., Loomis C., Cowan N. J. Sequence of an expression human β -tubulin gene containing ten Alu-family members//Nucl. Acids Res. 1984. V. 12. P. 5823.
- Lochrie M. A., Hurley J. B., Simon M. I. Sequence of the alpha subunit of photoreceptor G protein: Homologies between transducin, ras and elongation factors//Science. 1985. V. 228. P. 96.
- Long G. L., Chandra T., Woo S. L. C., Davie E. M., Kurachi K. Complete sequence of the cDNA for human α_1 -antitrypsin and the gene for the S-variant//Biochemistry. 1984. V. 23. P. 4828.
- Maeda N., Yang F., Barnett D. R., Bowman B. H., Smithies O. Duplication within the haptoglobin HP2 gene//Nature. 1984. V. 309. P. 131.
- Mager D. L., Heuthorn P. S. Identification of a retrovirus-like repetitive element in human DNA//Proc. Natl Acad. Sci. USA. 1984. V. 81. P. 7510.
- Meller P., Jeffreys A. J., Wilson V., Blanchetot H. Organization of the human myoglobin gene//EMBO J. 1984. V. 3. P. 439.
- Misrachi M., Loosfelt H., Atger M., Meriel C., Zerach V., Dessen P., Milgrom E. Organization of the entire rabbit progesterone receptor mRNA and of the promoter and 5'-flanking region of the gene//Nucl. Acids Res. 1988. V. 16. P. 5459.
- Molecular Evolutionary Genetics/Ed. McIntyre R. J. N. Y.; L.: Plenum Press, 1985.
- Montiel L. F., Norburg C. J., Tuite M. F., Dobson M. J., Mills J. J., Kingsman A. J., Kingsman S. M. Characterization of human chromosomal DNA sequences which replicate autonomously in *saccharomyces cervisie*//Nucl. Acids Res. 1984. V. 12. P. 1049.
- Narayanan V., Barbosa E., Rees R., Tennekoon G. Characterization of a cloned cDNA encoding rabbit myelin P2 protein//J. Biol. Chem. 1988. V. 263. P. 8332.
- Nathaus J., Higness D. S. Isolation and nucleotide sequence of the gene encoding human rhodopsin//Proc. Natl Acad. Sci. USA. 1984. V. 81. P. 4851.
- Noda M., Furutani Y., Takahashi H., Toysato M., Tanabe T., Shimizu S., Kikuyotani S., Kayano T., Hirose T., Inayama S., Numa S. Cloning and sequence analysis of the calf cDNA and human genomic DNA encoding α -subunit precursor of muscle anticholine receptor//Nature. 1983. V. 305. P. 818.
- Pech M., Jacnidhen H. R., Rohenz H. D., Neumaier P. S., Klobeck H. G., Zachau H. G. Organization and evolution of a gene cluster for human immunoglobulin variable regions of the kappa type//J. Mol. Biol. 1984. V. 176. P. 189.
- Peterson M. G., Hannan F., Merser J. F. B. The sheep metallothionein gene family//Eur J. Biochemistry. 1988. V. 174. P. 417.
- Protter A. A., Levy-Vilson B., Miller J., Bencen G., White T., Seilhamer J. J. Isolation and sequence analysis of the human apolipoprotein CIII gene and the intergenic region between the Apo AI and Apo CIII genes//DNA. 1984. V. 3. P. 449.
- Quentin Y. The Alu family developed through successive waves of fixation closely connected with primate lineage history//J. Mol. Evol. 1988. V. 27. P. 194.
- Quentin Y. Successive waves of fixation of B1-variants in rodent lineage history//J. Mol. Evol. 1989. V. 28. P. 299.
- Rebier M.-C., Marche P. N., Kindt T. J. A rabbit class I major histocompatibility complex gene with a T-cell-specific expression pattern//J. Immunol. 1987. V. 139. P. 2066.
- Romisch K., Webb J., Herz J., Prehn S., Frank R., Vingron M., Dobberstein B. Homology of 54K-protein of signal-recognition particle, docking protein and two *E. coli* proteins with putative GTP-binding domains//Nature. 1989. V. 340. P. 478.
- Saffer J. D., Lerman M. I. Unusual class of Alu sequences containing a potential Z-DNA segment//Mol. Cell. Biol. 1983. V. 3. P. 960.
- Sawada I., Beal M. P., Chen C., Champ B., Wilson A. C., Schmid C. Intergenic DNA sequences flanking the pseudo- α -globin genes of human and chimpanzee//Nucl. Acids Res. 1983. V. 11. P. 8067.
- Shimienti J. C., Duncan C. H. Ruminant globin gene structures suggest an evolutionary role for Alu-type repeats//Nucl. Acids Res. 1984. V. 12. P. 1641.
- Takahachi H., Hakamata Y., Watanabe Y., Kikino R., Miyata T., Numa S. Complete nucleotide sequence of the human corticotropin — β -lipotropin precursor gene//Nucl. Acids Res. 1983. V. 11. P. 6847.
- Tanabe T., Noda M., Furutani Y., Takai I., Takahashi H., Tanake K., Hirose T., Nuyama S. I., Numa S. Primary structure of β -subunit precursor of calf muscle acetylcholine receptor deduced from cDNA sequence//Eur. J. Biochem. 1984. V. 144. P. 11.

- Truong A. T., Tuez C., Belayew A., Renard A., Pictet R., Bell G. I., Martial J. A. Isolation and characterization of the human prolactin gene//EMBO. J. 1984. V. 3. P. 429.
- Ullrich A., Gray A., Berman C., Dull T. J. Human β -nerve growth factor gene sequence highly homologous to that of mouse//Nature. 1983. V. 303. P. 821.
- Watanabe Y., Tsukada T., Nataka M., Nakanishi S., Numa S. Structural analysis of repetitive DNA sequences in the bivariate corticotropin- β -lipotropin precursor gene region//Nucl. Acids Res. 1982. V. 10. P. 1459.
- Yang G. K., Masters J. N., Attardi G. Human dihydrofolate reductase gene organization//J. Mol. Biol. 1984. V. 176. P. 169.
- Yoshitake S., Schach B. G., Foster D. C., Davie E., Kurachi K. Nucleotide sequences of the gene for human factor IX (Antihemophilic factor B)//Biochemistry. 1985. V. 24. P. 3736.

Институт химической физики
им. Н. Н. Семенова РАН,
Москва

Поступила в редакцию
5.I.1991

E. V. KOROTKOV

MB1 FAMILY REPEATS IN GENOMES OF MANY MAMMALS

N. N. Semenov Institute of Chemical Physics,
Russian Academy of Sciences, Moscow

Members of MB1 family repeats are revealed in genomes of many mammals (cow, rabbit, opossum, horse, ...) The MB1 repeats from cow and rabbit genomes are mirror-reflected about the SINE families repeats from cow and rabbit genomes. The life time of MB1 repeats are no less than 100 million years. Classification of MB1 repeats from human genome using the information similarity was performed. This classification has revealed two subfamily MB1 repeats in human genome. Possible processes of creation of MB1 family repeats common for many mammals are discussed.