

## РОЛЬ КОНТЕКСТА В ВОЗНИКНОВЕНИИ МУТАЦИЙ В ГЕНАХ МИТОХОНДРИАЛЬНОЙ ДНК ЧЕЛОВЕКА

© 2005 г. Б. А. Малярчук

*Институт биологических проблем Севера Дальневосточного отделения Российской академии наук,  
Магадан 685000; факс: (41322) 3-44-63; e-mail: malyar@ibpn.kolyma.ru*

Поступила в редакцию 04.02.2004 г.

Основываясь на данных о распределении мутаций в генах митохондриальной ДНК (мтДНК), проведен контекстный анализ участков, включающих позиции мутирования, которые характеризуются появлением более чем двух параллельных мутаций. Показано, что механизм дислокационного мутагенеза, приводящий к появлению неспаренных оснований на участках смещения праймерной или матричной цепи мтДНК в процессе репликации, объясняет происхождение 21% нестабильных позиций в генах мтДНК. Контекстный анализ показал, что наибольшее влияние на возникновение мутаций в G-позициях генов мтДНК имеют пиримидиновые основания в позициях +1 и +2 (консенсусы gYRNS, gYY и gR, где g – позиция мутирования). Наибольшее влияние на мутагенез в T-позициях оказывают основания в положениях –1 и +1 (консенсусы RtY и tA, где t – позиция мутирования). Полученные данные свидетельствуют в целом о преобладании контекст-зависимых механизмов возникновения мутаций в митохондриальном геноме человека.

Исследования мутационных спектров главной некодирующей области митохондриальной ДНК (мтДНК) человека показали, что появление мутаций в гипервариабельных последовательностях (ГВС1 и 2) этого участка мтДНК в значительной мере обусловлено контекстом (т.е. особенностями первичной структуры ДНК) вблизи вариабельных позиций [1]. Процессы возникновения мутаций характеризуются неравномерностью, поскольку в ряде исследований показано, что индуцированные и спонтанные мутации происходят с повышенной частотой лишь в определенных позициях ДНК, называемых “горячими точками” [2–5]. Выраженное влияние соседних оснований, отстоящих от вариабельных позиций на расстоянии  $\pm 2$  нуклеотида, на возникновение мутаций было показано в исследованиях мутационных спектров генов человека в выборках с наследственной патологией [2, 3]. Хотя нельзя исключить влияние на мутагенез и более удаленных оснований, но именно ближайшие основания, по всей видимости, оказывают наибольший эффект на мутационную скорость [6]. Выраженное влияние контекста ДНК на скорость мутаций, индуцированных мутагенами, зарегистрировано в исследованиях мутационных спектров генов *E. coli* [6, 7] и млекопитающих [8].

Влияние контекста ДНК на мутагенез (как спонтанный, так и индуцированный) может проявиться на каждом этапе процессов репликации и репарации ДНК [4, 9]. Предполагается, например, что появление неспаренных оснований в процессе репликации может быть обусловлено особенностями первичной структуры дуплексов “матрица-

праймер” в активном участке ДНК-полимеразного комплекса [10]. Соседние основания могут стабилизировать неспаренные основания, способствуя тем самым увеличению мутационной скорости [3]. Большое значение имеет также дислокационный мутагенез, вызывающий появление неспаренных оснований (и, как следствие, мутаций) в результате смещения праймерной или матричной цепи ДНК на монотонных нуклеотидных трактах в процессе репликации [11]. Анализ мутационных спектров главной некодирующей области мтДНК человека показал, что модель дислокационного мутагенеза способна объяснить появление более 25% “горячих точек” в ГВС1 и 2. В ГВС1 мтДНК выявлены также консенсусные последовательности “горячих точек” CC и KTNC-NK (“горячие точки” подчеркнуты) [1]. Таким образом, результаты исследований мутационных спектров главной некодирующей области мтДНК позволили заключить, что в этом участке митохондриального генома преобладают контекст-зависимые механизмы генерирования мутаций.

Целью настоящей работы является исследование мутационных спектров генов мтДНК, кодирующих две рРНК, 22 тРНК и 13 субъединиц белков дыхательной цепи, с помощью контекстного анализа ближайшего нуклеотидного окружения наиболее вариабельных позиций митохондриальных генов.

### МАТЕРИАЛЫ И МЕТОДЫ

В анализе использованы мутационные спектры генов мтДНК человека, полученные на осно-

ваний анализа данных [12–14] об изменчивости 794 нуклеотидных последовательностей длиной 15445 пн (между позициями 577–16023 согласно нумерации нуклеотидов в кембриджской последовательности мтДНК [15]). Мутационные спектры реконструированы с помощью метода медианных сетей, основанного на алгоритме “максимальной экономии” [16]. Под мутационным спектром понимается распределение частот появления независимых мутаций каждого типа (транзиций, трансверсий, делеций и инсерций) в каждой нуклеотидной позиции анализируемого участка ДНК [4].

Для описания характера распределения позиций мтДНК в зависимости от их вариабельности использовались статистические методы (“non-parametric statistics”, “distribution fitting”), реализованные в пакете программ STATISTICA 5.0 (StatSoft, Inc., Tulsa, USA). Мутационные спектры анализировались в соответствии с методическими подходами, описанными в работах [1, 6]. Для изучения роли соседних оснований в возникновении нуклеотидных замен анализировались участки ДНК на расстоянии  $\pm 5$  оснований от вариабельной позиции (участки мутирования). Эти участки выравнивались вручную относительно анализируемых вариабельных позиций (позиций мутирования) определенного типа (A, G, C, T). Консенсусные последовательности участков мутирования строились для вариабельных оснований G и T, т.е. замены для оснований C и A преобразовывались в комплементарную форму. Для обозначения консенсусных нуклеотидов в участках мутирования использовалась общепринятая номенклатура однобуквенных обозначений нуклеотидов: N = A, T, G, C; W = A или T; S = G или C; R = A или G; Y = T или C; M = A или C; K = G или T; B = T или G или C; V = A или G или C; H = A или T или C; D = A или T или G [17].

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ изменчивости нуклеотидных последовательностей генов белков, рРНК и тРНК (суммарной длины 15366 пн) показал, что вариабельными в кодирующей области мтДНК являются 1454 нуклеотидные позиции (9.4% от длины исследованного участка мтДНК). Применение метода филогенетического анализа позволяет различать параллельные мутации, произошедшие независимо в одних и тех же нуклеотидных позициях, но на разных этапах эволюции мтДНК и в различных ветвях ее филогенетического дерева. Поэтому с учетом параллельных мутаций мутационные спектры генов мтДНК характеризуются 1849 независимыми мутационными событиями, т.е. на каждую позицию приходится в среднем 1.3 мутации, хотя значения этого показателя изменяются в широком интервале – до 14 парал-

лельных мутаций на позицию. Анализ показал, что характер распределения позиций мтДНК в зависимости от их вариабельности является геометрическим ( $p < 0.01$ ) – более 90% позиций кодирующих участков мтДНК являются мономорфными, в 7% позиций наблюдаются единичные мутации, и лишь в оставшихся менее чем 3% позиций наблюдается гомоплазия (от 2 до 14 параллельных мутаций на позицию).

Наибольший интерес для выявления влияния факторов контекста ДНК на мутагенез представляют нуклеотидные позиции, характеризующиеся повышенной частотой мутаций [1, 4]. Поэтому в дальнейшем нами анализируются участки мтДНК, включающие позиции мутирования, которые характеризуются появлением более двух параллельных мутаций на позицию. В табл. 1 приводятся результаты контекстного анализа, проведенного на расстоянии  $\pm 5$  оснований от вариабельных G-нуклеотидов на L-цепи мтДНК. Участки, в которых вариабельными являются нуклеотиды C, записаны в комплементарном виде по последовательности H-цепи мтДНК. Полученные данные показывают, что все без исключения участки мутирования могут быть подразделены в несколько типов коротких консенсусных последовательностей: gYRNS, gYY и gR (g – позиция мутирования). Следует отметить, что в составе последовательностей, описываемых консенсусом gR, подавляющее большинство (9 из 12 последовательностей) имеют вид gRY. Замены в участках gY возникают почти в три раза чаще, чем в gR ( $p = 0.003$ ), а в участках gC – достоверно чаще, чем в gT или gR ( $p = 0.03$ ). Замены в участках gNY также возникают достоверно чаще, чем в gNR ( $p = 0.003$ ), а в участке gNC (но не gNT) – достоверно чаще, чем в gNR ( $p = 0.02$ ). В 5'-положениях по отношению к позиции мутирования нами не выявлено никаких контекстных закономерностей, поэтому наибольшее влияние на возникновение мутаций в G-позициях генов мтДНК имеют пиримидиновые основания в положениях +1 и +2. Модель дислокации цепей ДНК в процессе репликации способна объяснить происхождение нестабильных позиций лишь в пяти из 42 проанализированных участков (12%).

В табл. 2 приводятся результаты анализа участков мутирования T-нуклеотидов. Примечательно, что в данном случае модель дислокационного мутагенеза способна объяснить происхождение мутаций в 32.4% позиций мутирования. Контекстный анализ положения +1 показал, что участки мутирования подразделяются в две основные группы: tY и tA (t – позиция мутирования). Анализ положения +2 показал, что консенсус tNY распространен достоверно чаще, чем tNR ( $p = 0.0004$ ). В положении –1 достоверно чаще наблюдаются отатки аденина и гуанина (частота At значимо выше, чем остальных вариантов последовательнос-

**Таблица 1.** Контекстный анализ участков мтДНК, характеризующихся переменными G-основаниями

Позиция	Число параллельных мутаций	Последовательность мтДНК											
		C	C	A	C	A	GG	C	A	C	C	A	
14364	3	C	C	A	C	A	GG	C	A	C	C	A	
15110	4	T	G	C	T	T	GG	C	A	A	C	T	
12346H	5	A	G	T	G	T	GG	C	A	T	G	G	
13708	7	G	C	C	T	G	GG	C	A	G	C	C	
3705	4	G	C	A	C	T	GG	C	G	A	G	C	
15043	4	A	T	C	G	G	GG	C	G	A	G	G	
9449H	3	A	T	C	C	C	GG	T	A	T	C	G	
10373	3	T	A	T	G	A	GG	T	G	A	C	T	
3666	3	T	C	A	G	G	GG	T	G	A	G	C	
							GG	Y	R		S		
11176	3	A	A	C	C	A	GG	C	C	A	G	A	
8251	3	A	T	A	G	G	GG	C	C	C	G	T	
1888	4	G	G	A	G	A	GG	C	C	A	A	A	
15884	4	A	A	T	G	G	GG	C	C	T	G	T	
9554	4	C	A	C	T	G	GG	C	C	C	C	C	
1719	6	C	C	T	T	A	GG	C	C	A	A	A	
3316	6	C	C	A	T	G	GG	C	C	A	A	C	
5460	8	T	C	A	T	C	GG	C	C	C	T	T	
5046	3	T	A	G	C	A	GG	T	T	C	T	A	
11914	7	A	C	C	A	C	GG	T	T	C	T	C	
709	14	T	C	C	C	C	GG	T	T	C	C	A	
12630	3	A	C	A	T	G	GG	T	C	C	A	T	
3918	4	G	G	G	G	A	GG	T	C	C	G	A	
3438	3	T	A	C	G	G	GG	C	T	A	C	T	
6023	4	G	C	C	G	A	GG	C	T	G	G	G	
14569	4	A	C	A	C	C	GG	C	T	A	A	C	
							GG	Y	Y				
9266	3	A	C	A	G	G	GG	G	C	C	C	T	
2706	3	G	G	C	A	T	GG	A	C	A	C	A	
12372	3	A	C	C	C	T	GG	A	C	T	T	C	
15257	3	C	A	G	T	A	GG	A	C	A	G	T	
5147	5	A	C	C	A	C	GG	A	C	C	C	T	
8697	3	C	A	A	A	T	GG	A	T	A	A	C	
15930	3	C	C	G	G	A	GG	A	T	G	A	A	
3010	6	A	T	C	C	C	GG	A	T	G	G	T	
10685	4	G	C	A	G	C	GG	G	T	G	G	G	
3915	3	G	A	A	G	G	GG	G	A	G	T	C	
12007	5	C	A	A	T	G	GG	G	G	C	T	C	
1438	4	A	A	A	C	T	GG	A	G	A	G	T	
							GG	R					
7337	3	G	C	T	T	C	GG	A	A	G	C	G	d
1598	4	T	G	G	A	C	GG	A	A	C	C	A	d
6455H	4	C	A	G	A	C	GG	A	A	G	A	G	d
930	3	T	A	G	A	A	GG	C	C	G	G	C	d
6260	5	G	T	G	G	A	GG	G	C	C	G	G	d

Примечание. Приводятся последовательности L-цепи мтДНК в направлении 5'-3'. Добавлением буквы "H" к номерам позиций отмечены те случаи, когда анализируются комплементарные H-цепи мтДНК. Выделены консенсусы. Буквой "d" отмечены участки мтДНК, нестабильность которых объясняется моделью дислокационного мутагенеза.

Таблица 2. Контекстный анализ участков мтДНК, характеризующихся переменными Т-основаниями

Позиция	Число параллельных мутаций	Последовательность мтДНК											
14766	6	C	A	A	A	A	t	T	A	A	C	C	
13105H	3	G	A	A	G	A	t	T	C	C	T	G	
14233H	3	T	A	T	G	A	t	T	A	T	G	G	
15236H	3	T	C	A	G	A	t	T	C	A	T	T	
15758H	3	T	C	C	G	A	t	T	C	A	G	G	
14769H	3	G	G	G	G	G	t	T	A	G	T	T	
15924H	7	T	C	C	G	G	t	T	T	A	C	A	
13020	3	T	T	A	G	G	t	C	T	C	C	A	
10915	3	A	G	C	T	G	t	T	C	C	C	C	
9677H	3	G	G	T	T	G	t	T	T	T	C	T	
14798	3	A	C	T	C	A	t	T	C	A	T	C	
5442	4	C	C	C	C	A	t	T	C	C	T	C	
10084	4	A	A	T	A	A	t	C	A	A	C	A	
						R	t	Y					
4216	4	T	A	T	G	A	t	A	T	G	T	C	
6671	3	T	C	C	C	A	t	A	T	T	G	T	
3394	3	T	A	G	G	C	t	A	T	A	T	A	
10598H	4	G	T	A	G	C	t	A	T	A	A	T	
6827	3	T	C	C	G	C	t	A	C	C	A	T	
4561	3	C	T	G	A	G	t	A	G	G	C	C	
8705	3	A	A	C	C	A	t	A	C	A	C	A	
14180	3	C	A	A	T	A	t	A	T	A	C	A	
14182	3	A	T	A	T	A	t	A	C	A	C	C	
15670	4	C	T	C	C	A	t	A	T	A	T	C	
							t	A					
12414	4	A	A	C	C	C	t	A	A	C	A	A	d
6152	3	C	T	A	G	T	t	C	C	C	C	T	d
6221	5	T	T	A	C	C	t	C	C	C	T	C	d
8618	3	A	T	T	G	A	t	C	C	C	C	A	d
14308	3	C	A	G	C	T	t	C	C	T	A	C	d
14470	6	G	T	A	T	A	t	C	C	A	A	A	d
7055H	3	A	C	A	G	C	t	C	C	T	A	T	d
9545H	5	T	G	C	C	C	t	C	C	T	A	A	d
11719H	5	G	T	A	A	G	t	C	C	G	T	G	d
6680	3	G	T	A	A	C	t	T	A	C	T	A	d
15784	8	T	A	C	C	C	t	T	T	T	A	C	d

ти;  $p = 0.03$ ). Таким образом, участки мутирования Т-позиций могут быть описаны двумя типами консенсусов RtY и tA (табл. 2). В данном случае тоже очевидно, что наибольшее влияние на мутагенез в Т-позициях генов мтДНК оказывают основания в ближайших  $-1$  и  $+1$  положениях.

Распределение мутаций в участках белок-кодирующих генов в значительной степени зависит от такого фактора эволюции как отбор, обусловленный различной функциональной значимостью нуклеотидных замен в различных позициях кодонов. Проведенные ранее исследования показали, что 32% мутаций в белок-кодирующих генах мтДНК являются несинонимичными, а из параллельных мутаций в генах лишь их четверть приводит к аминокислотным заменам [18]. Анализ участков мутирования, представленных в настоящей работе, также показал, что параллельные мутации в 30.3% проанализированных позиций приводят к аминокислотным заменам. Из них наиболее частыми являются мутации в позициях 3316, 13708 и 5460 (6, 7 и 8 независимых замен аланина на треонин, соответственно), 14766 (шесть замен изолейцина на треонин) и 14470 (шесть замен серина на пролин). Роль множественных аминокислотных замен в процессе эволюции митохондриального генома до конца не ясна, хотя некоторые из подобного рода мутаций ассоциируются с наследственными заболеваниями. Например, из указанных выше “быстрых” позиций лишь мутация в позиции 13708 относится к числу вторичных патологических мутаций, ассоциирующихся с болезнью Лебера [19].

Полученные в настоящей работе данные свидетельствуют о том, что изменчивость в генах митохондриального генома, как и его главной некодирующей области, обусловлена влиянием нуклеотидного контекста на возникновение мутаций. Консенсусные участки переменных позиций мтДНК являются достаточно гетерогенными, что указывает на разнообразие механизмов спонтанного мутагенеза в митохондриях человека. Один из них – механизм дислокационного мутагенеза, приводящий к появлению неспаренных оснований на участках смещения праймерной или матричной цепи мтДНК в процессе репликации, способен объяснить происхождение 21% нестабильных позиций в генах мтДНК. Полученное значение сопоставимо с результатами аналогичного анализа “горячих точек” гипервариабельных участков главной некодирующей области [1]. Важно отметить также, что полученные в настоящей работе данные свидетельствуют о том, что наибольшее влияние на возникновение мутаций в генах мтДНК имеет ближайшее нуклеотидное окружение (хотя нельзя все же исключить возможное влияние удаленных более чем на пять оснований от позиции мутирования, но такого рода анализ в данной работе не проводился). Следует

также отметить, что относительно высокая доля позиций, подверженных влиянию дислокационного мутагенеза, также обусловлена их непосредственной близостью или входжением в состав коротких монотонных повторов, что, в целом, свидетельствует о важной роли контекст-зависимых механизмов мутагенеза в генерировании мутаций в митохондриальном геноме человека.

Работа выполнена при финансовой поддержке Дальневосточного отделения Российской академии наук (№ 03-3-А-06-096).

## СПИСОК ЛИТЕРАТУРЫ

1. *Malyarchuk B.A., Rogozin I.B., Berikov V.B., Derenko M.V.* Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region // *Hum. Genet.* 2002. V. 111. P. 46–53.
2. *Cooper D.N., Youssoufian H.* The CpG dinucleotide and human genetic disease // *Hum. Genet.* 1988. V. 78. P. 151–155.
3. *Krawczak M., Ball E.V., Cooper D.N.* Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes // *Am. J. Hum. Genet.* 1998. V. 63. P. 474–488.
4. *Rogozin I.B., Pavlov Y.I.* Theoretical analysis of mutation hotspots and their DNA sequence context specificity // *Mutat. Res.* 2003. V. 544. P. 65–85.
5. *Morton B.R., Oberholzer V.M., Clegg M.T.* The influence of specific neighboring bases on substitution bias in noncoding regions of the plant chloroplast genome // *Mol. Evol.* 1997. V. 45. P. 227–231.
6. *Рогозин И.Б., Бериков В.Б., Васюнина Е.А., Синицина О.И.* Исследование влияния особенностей первичной структуры ДНК на возникновение мутаций, индуцированных алкилирующими агентами // *Генетика.* 2001. Т. 37. № 6. С. 854–861. (*Rogozin I.B., Berikov V.B., Vasunina E.A., Sinitina O.I.* The effect of the primary structure of DNA on induction of mutations by alkylating agents // *Rus. J. Genet.* 2001. V. 37. № 6. P. 704–710.)
7. *Broschard T.H., Koffel-Schwartz N.* Sequence-dependent modulation of frameshift mutagenesis at NARI-derived mutation hot spots // *J. Mol. Biol.* 1999. V. 288. P. 191–199.
8. *Shibutani S., Suzuki N., Tan X. et al.* Influence of the flanking sequence context on the mutagenicity of acetylaminofluorene-derived DNA adducts in mammalian cells // *Biochemistry.* 2001. V. 40. P. 3717–3722.
9. *Zavolan M., Kepler T.B.* Statistical inference of sequence-dependent mutation rates // *Curr. Opin. Genet. Dev.* 2001. V. 11. P. 612–615.
10. *Timsit Y.* DNA structure and polymerase fidelity // *J. Mol. Biol.* 1999. V. 293. P. 835–853.
11. *Kunkel T.A., Soni A.* Mutagenesis by transient misalignment // *J. Biol. Chem.* 1988. V. 263. P. 14784–14789.
12. *Herrnstadt C., Elson J.L., Fahy E. et al.* Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian and European haplogroups // *Am. J. Hum. Genet.* 2002. V. 70. P. 1152–1171.

13. *Finnila S., Lehtonen M.S., Majamaa K.* Phylogenetic network for European mtDNA // *Am. J. Hum. Genet.* 2001. V. 68. P. 1475–1484.
14. *Maca-Meyer N., Gonzalez A.M., Larruga J.M. et al.* Major genomic mitochondrial lineages delineate early human expansions // *BMC Genetics.* 2001. V. 2. P. 13.
15. *Andrews R.M., Kubacka I., Chinnery P.F. et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA // *Nat. Genet.* 1999. V. 23. P. 147.
16. *Bandelt H.-J., Forster P., Sykes B.C., Richards M.B.* Mitochondrial portraits of human populations using median networks // *Genetics.* 1995. V. 141. P. 743–753.
17. *Cornish-Bowden A.* Nomenclature for incompletely specified bases in nucleic acid sequences: Recommendation // *Nucl. Acids Res.* 1985. V. 13. P. 3021–3030.
18. *Moilanen J.S., Majamaa K.* Phylogenetic network and physicochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA // *Mol. Biol. Evol.* 2003. V. 20. P. 1195–1210.
19. *Wallace D.C.* Mitochondrial diseases in man and mouse // *Science.* 1999. V. 283. P. 1482–1488.