

УДК 575.174:599.9

“ХОЛОДНЫЕ” ТОЧКИ ГИПЕРВАРИАБЕЛЬНОГО СЕГМЕНТА 1 МИТОХОНДРИАЛЬНОЙ ДНК ЧЕЛОВЕКА

© 2008 г. Б. А. Малярчук

Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, 685000

Поступила в редакцию 06.12.2007 г.
Принята к печати 24.01.2008 г.

Анализ распределения мутаций в гипервариабельном сегменте 1 мтДНК в массиве данных более чем от 37000 человек из различных регионов мира позволил оценить интенсивность мутационных процессов и особенности распределения “холодных” точек мтДНК. Результаты анализа структурно-функциональной организации и изменчивости исследуемого участка мтДНК позволяют считать, что снижение изменчивости связано, возможно, с теми участками гипервариабельного сегмента 1, которые имеют функциональное значение. Анализ распределения “холодных” точек САТ во вторичных структурах свидетельствует об отсутствии связи между положением “холодной” точки и типом структуры участка мтДНК.

Ключевые слова: митохондриальная ДНК человека, главная некодирующая область, гипервариабельный сегмент 1, “холодные” точки, вторичная структура ДНК.

“COLD” SPOTS IN HYPERVARIABLE SEGMENT 1 OF HUMAN MITOCHONDRIAL DNA, by B. A. Malyarchuk* (Institute of Biological Problems of the North, Far-East Division, Russian Academy of Sciences, Magadan, 685000 Russia; *e-mail: malyarchuk@ibpn.ru). Analysis of distribution of mutations in hypervariable segment 1 (HVS1) of mitochondrial DNA (mtDNA) in data set of more than 37000 individuals, representing different regions of the world, allowed one to estimate the mutation processes intensity and peculiarity of mtDNA “cold” spots distribution. Results of analysis of structure-function organization and variability of the mtDNA fragment under study revealed that decrease of variability is characteristic for HVS1 sequences with probable functional importance. Analysis of distribution of “cold” spots CAT in secondary structures testified against the correlation between “cold” spot location and mtDNA structure type.

Key words: human mitochondrial DNA, major non-coding region, hypervariable segment 1, “cold” spots, DNA secondary structures.

С использованием методов филогенетического анализа установлено, что в гипервариабельных сегментах главной некодирующей области (ГВС1 и ГВС2) мтДНК существуют “горячие” точки – нуклеотидные позиции, которые характеризуются повышенной скоростью мутаций [1, 2]. Гипервариабельность нуклеотидных позиций приводит к появлению идентичных (параллельных) мутаций в филогенетически неродственных типах мтДНК.

Проведенный нами ранее анализ мутационных спектров ГВС1 мтДНК позволил выявить четыре класса нуклеотидных участков, характеризующихся различной степенью вариабельности [1]. В двух из них находятся “горячие” точки, а один класс включает очевидные “холодные” сайты, количество независимых (параллельных) мутаций в кото-

рых колеблется от 0 до 2. Контекстный анализ ГВС1 позволил выявить два типа консенсусных последовательностей “горячих” точек в положении Т и С. Они имеют вид КТNCNK и СС (где К = G или Т, а N – любой нуклеотид; “горячая” точка подчеркнута). Следует отметить, что частным случаем консенсуса КТNCNK является последовательность GTACAT, которая, как показано ранее [1, 2], может быть как гипервариабельной, так и консервативной. Между тем известно, что некоторые участки ГВС1 в мтДНК практически мономорфные (т.е. в них полностью отсутствуют мутации, в то время как класс “холодных” точек шире и включает 0–2 мутации [1]), однако результаты изучения мономорфизма в очень большой степени зависят от размеров баз данных и от того, насколько в них учтено этнорасовое разнообразие. Лишь в последние годы стали появляться данные об из-

* Эл. почта: malyarchuk@ibpn.ru

менчивости мтДНК в популяциях коренного населения Южной Азии, Австралии и Америки, что позволяет подойти к анализу распределения “холодных” точек с большей степенью надежности.

Цель настоящей работы состояла в анализе распределения мутаций в ГВС1 мтДНК в массиве данных, полученных более чем от 37000 человек из различных регионов мира.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Реконструкция мутационного спектра ГВС1 мтДНК. Для такой реконструкции недостаточно одних только данных о распределении переменных позиций вдоль участка ДНК. Наиболее точную оценку распределения “горячих” точек в ГВС1 мтДНК может дать анализ изменчивости нуклеотидных последовательностей ГВС1, объединенных в монофилетические кластеры (группы мтДНК) на основании данных о полиморфизме кодирующих участков мтДНК, темпы изменчивости которых значительно ниже, чем в ГВС1 [1–3]. При выделении групп мтДНК учитывают варианты полиморфизма ГВС1 в тех случаях, когда они являются группоспецифическими. Таким образом, в соответствии с принятой нами ранее методологией [1, 2], в пределах каждой группы мтДНК проведен поиск переменных позиций в ГВС1 и определено количество параллельных мутаций в различных группах мтДНК. Под параллельными мутациями понимают идентичные варианты полиморфизма, возникшие независимо в различных группах мтДНК.

В настоящей работе проанализировано несколько наборов популяционных данных. Изменчивость ГВС1 мтДНК проанализирована в выборке из 7482 последовательностей ГВС1 мтДНК (между позициями 16090–16365), принадлежащих 88 группам мтДНК, как описано нами ранее [2]. Эта база данных представлена населением Западной Евразии ($n = 3834$), Восточной Евразии ($n = 801$) и Африки ($n = 2847$). Состав групп мтДНК по регионам: Западная Евразия, 28 групп (H, HV*, pre-V, pre-HV, R*, T1, T*, J*, J1a, J1b, J2, K, U*, U1, U2, U3, U4, U5, U7, U8a, U8b, N1a, N1b, N1c, N*, I, W, X); Восточная Евразия, 34 группы (C, Z, M8a, D4, D5, G2, G3, G4, E, M*, M7*, M7b, M7c, M9, M10, A, N9a, N2, N*, Y, R9a, R*, F*, F1a, F1b, F1c, F2, B*, B4*, B4a, B4b, B5*, B5a, B5b); Африка, 26 групп (L0a1, L0a2, L1b, L1c*, L1c1, L1c2, L1c3, L1d, L1e, L2a*, L2a1a, L2a1b, L2b, L2c, L2d1, L2d2, L3b1, L3b2, L3d, L3e1, L3e2, L3e3, L3e4, L3f*, L3f1, L3g).

Анализ изменчивости ГВС1 мтДНК. В популяциях Южной Азии (Индия, Пакистан, Иран, Китай, Таиланд, Бангладеш) изменчивость ГВС1 мтДНК анализировали с использованием базы данных [4] ($n = 4653$); в популяциях коренного населения Северной и Южной Америки – базы данных [5]

($n = 601$); в популяциях коренного населения Северной Азии (Южная Сибирь, Монголия, Корея, Таджикистан и Иран) – базы данных [6] ($n = 1432$); у населения Австралии, Меланезии и Папуа Новой Гвинеи – базы данных mtDB [7] ($n = 1865$). Проанализирована также база данных “HV1_polymorphic” [8] ($n = 21141$). Эта база данных представлена, главным образом, “евро-американским” населением США. Таким образом, всего проанализированы 37174 нуклеотидных последовательности ГВС1. Линии мтДНК из баз данных включены в состав упомянутых 88 групп мтДНК, поскольку они полностью охватывают существующее разнообразие вариантов мтДНК человека. Некоторые редкие группы мтДНК, характерные для населения Австралии, Новой Гвинеи и Юго-Восточной Азии, включены на данном этапе исследований в состав парагрупп R*, N* и M*.

Мутации учитывали относительно L-цепи кембриджской референтной мтДНК человека [9]. Во всех случаях не учитывали точечные инсерции и делеции, а также трансверсии в нестабильном участке ГВС1, расположенном между позициями 16184–16193. Вторичную структуру участков мтДНК реконструировали с использованием компьютерной программы RNAdraw 1.1b2 (США).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

На рисунке показан реконструированный мутационный спектр ГВС1 мтДНК человека. Анализ столь обширного набора популяционных данных позволил выявить позиции, характеризующиеся низкой вариабельностью, в том числе позиции, в которых в проанализированных 88 группах мтДНК обнаружены от 0 до 2 независимых (параллельных) мутаций на сайт, и которые входят (согласно нашим предыдущим исследованиям [1]) в состав класса “холодных” точек (табл. 1). Этот класс представлен 115 позициями, что составляет 42% от размера исследованного участка ГВС1. Видно, что ГВС1 мтДНК, кроме отдельных “холодных” точек, содержат также короткие нуклеотидные последовательности, характеризующиеся низкой вариабельностью. Это участки 16096–16103, 16105–16107, 16115–16123, 16137–16139, 16155–16157, 16159–16161, 16197–16202, 16236–16238, 16251–16253, 16275–16277, 16279–16282, 16306–16308, 16313–16315, 16328–16330, 16332–16334, 16337–16341, 16345–16351. Отсутствие полиморфизма обнаружено в 29 нуклеотидных позициях, что составляет 10.5% от размера проанализированного участка мтДНК.

Консервативность некоторых участков мтДНК может быть обусловлена их функциональной значимостью, поскольку в ГВС1 находятся элементы, необходимые для инициации и регуляции процессов репликации и транскрипции митохондриального генома. Пониженная изменчивость наблюдается, например, в SP-участке (16104–16106) и в области тер-

16090	2 2 6 T A T T T C G T A c a t T A C T G C C 4 1 1 1 1 2 6 1 1 2 4 1 1 3 9 1 1 2 1 1
16120	1 1 1 9 1 2 1 5 3 2 1 7 1 8 1 2 1 7 2 6 4 8 4 5 2 A T A T T G T A C G G T A C c a t A A A T A C T T G A C C A
16150	1 1 1 3 7 2 2 1 6 1 1 1 9 5 5 2 4 2 3 0 3 7 0 7 7 4 9 1 6 5 C C T G T A G T A c a t A A A A A C C C A A T C C A c a t C
16180	7 4 1 5 1 8 7 5 9 2 2 1 6 4 1 2 3 1 1 1 2 5 1 1 3 1 1 3 A A A A C C C C C T C C C c a t G C T T A C A A G C A A G T
16210	1 4 5 6 7 9 3 8 2 4 5 7 0 2 8 2 1 5 1 1 1 1 1 4 3 5 3 1 2 A C A G C A A T C A A C C C T C A A C T A T C A C A c a t C
16240	1 1 1 1 1 1 1 1 3 1 1 4 5 4 5 8 5 2 2 5 6 7 7 1 2 2 2 1 6 A A C T G C A A A 7 5 3 1 1 A A G C C A C C C T C A C C C A
16270	1 1 2 2 3 1 3 4 1 2 7 0 0 1 2 1 2 4 2 2 2 1 3 7 7 8 9 4 7 C T A G G A T A C C A A C A A A C C T A C C C A C C C T T A
16300	1 2 7 1 1 2 2 0 4 1 1 8 3 3 0 3 4 3 1 8 7 2 3 1 A C A G T A c a t A G T A c a t A A A G C c a t T T A C C G
16330	2 3 1 3 2 2 8 9 2 1 2 1 2 1 1 1 5 3 4 0 2 2 4 5 T A c a t A G C A c a t T A C A G T C A A A T C C C T T C T
16360	1 5 2 1 4 1 1 6 C G T C C C

Спектр мутаций в ГВС1 мтДНК человека. Числа над азотистыми основаниями показывают количество независимых мутаций, произошедших в 88 филогенетических группах мтДНК. Подчеркнуты участки ДНК, входящие в стебли шпильчатых структур, согласно результатам анализа вторичной структуры ГВС1 мтДНК. САТ-участки мтДНК выделены малыми прописными буквами.

минации синтеза 7S ДНК (TAS), расположенной между позициями 16157 и 16172. Довольно консервативная лишь 5'-концевая область TAS между позициями 16155–16157 и 16159–16161. Пониженная изменчивость характерна и для контрольного элемента SE (между позициями 16194–16208). Этот участок содержит целый ряд “холодных” точек (16194, 16195, 16197–16202, 16204, 16205, 16208).

Почти половина консервативных участков ГВС1, указанных в табл. 2, имеет вид САТ (16099–16101, 16159–16161, 16306–16308, 16313–16315 и

16332–16334). Они входят в состав последовательности GTACAT, характерной для участков мтДНК, ассоциированных с терминацией транскрипции: ETAS1 (16081–16138), TAS (16157–16172) и ETAS2 (16294–16353). Следует отметить, что эта последовательность может быть как гипервариабельной, так и консервативной. Так, в участках 16096–16101, 16156–16161 и 16329–16334 последовательность GTACAT практически мономорфная, в то время как участок 16303–16315 содержит два участка GTACAT, которые в обоих случаях имеют

Таблица 1. “Холодные” точки в ГВС1 мтДНК человека

Число параллельных мутаций на сайт	Нуклеотидная позиция
0–2	16090, 16091, 16096–16103, 16105–16107, 16109, 16110, 16112, 16115–16123, 16125, 16127, 16128, 16130, 16132, 16133, 16135, 16137–16139, 16141, 16143, 16149, 16151, 16152, 16155–16157, 16159–16161, 16165, 16177, 16182, 16190, 16191, 16194, 16195, 16197–16202, 16204, 16205, 16208, 16210, 16225, 16226, 16228, 16229, 16236–16238, 16246, 16251–16253, 16267, 16268, 16272, 16273, 16275–16277, 16279–16282, 16285, 16306–16308, 16313–16315, 16323, 16326, 16328–16330, 16332–16334, 16337–16341, 16345–16351, 16361, 16363, 16364

Таблица 2. Консервативные участки в ГВС1 мтДНК человека и их расположение во вторичных структурах ДНК

Позиция	Нуклеотидная последовательность	Участок ДНК
16096–16103	GTACATTA	Стебель шпильки
16105–16107	TGC	Стебель шпильки
16115–16123	CATGAATAT	Межшпилечный участок
16137–16139	AAA	Петля шпильки
16155–16157	AGT	Стебель шпильки
16159–16161	CAT	Стебель шпильки
16197–16202	СТТАСА	Стебель шпильки
16236–16238	CAT	Межшпилечный участок
16251–16253	CAA	Петля шпильки
16275–16277	ATA	Межшпилечный участок
16279–16282	СААС	Межшпилечный участок
16306–16308	CAT	Петля шпильки
16313–16315	CAT	Межшпилечный участок
16328–16330	CGT	Стебель шпильки
16332–16334	CAT	Петля шпильки
16337–16341	CACAT	Стебель шпильки и межшпилечный участок
16345–16351	AGTCAAA	Межшпилечный участок

“горячую” точку в первом из двух остатков тимина. Можно предположить, что последовательно GTACAT принимают участие в формировании вторичных структур, поскольку они организованы в виде прямых повторов, частично инвертированы по отношению друг к другу и способны, таким образом, формировать альтернативные палиндромно-шпилечные структуры. Известно, например, что у грызунов ETAS-участки, включающие GTACAT и их производные, формируют очень стабильные вторичные структуры со стеблем из 11–17 н. и петлей из 5–9 н. [10]. Более того, у ряда видов животных повторяемость GTACAT-мономеров в главной некодирующей области мтДНК приводит к появлению разнообразных вторичных структур, различающихся как числом шпилек, так и числом мономеров в стеблях палиндромно-шпилечных структур [10]. Вполне вероятно, что подобная организация характерна и для главной некодирующей области мтДНК человека. В таком случае появление и “горячих”, и “холодных” точек в последовательностях GTACAT может определяться их положением во вторичных структурах и объясняться такими специфическими механизмами мутагенеза, как, например, коррекция несовершенных гетеродуплексов в составе палиндромно-шпилечных структур. С целью проверки этого предположения нами изучен характер расположения “холодных” точек во вторичных структурах участков ГВС1 мтДНК. С помощью программы RNAdraw 1.1b2 реконструирована организация вторичной структуры участка ГВС1, расположенного между позициями

16070 и 16368. Показано, что этот участок мтДНК содержит восемь шпилечных структур, в организации пяти из них принимают участие последовательности GTACAT. Однако тринуклеотиды CAT в формировании стеблей шпилек практически не участвуют (рисунок, табл. 2). Консервативные участки ГВС1 мтДНК человека, указанные в табл. 2, находятся как в пределах шпилечных структур (в семи случаях – в стеблях шпилек и в четырех – в петлях), так и в участках, расположенных между шпилечными структурами (семь случаев).

Таким образом, анализ большого объема данных об изменчивости нуклеотидных последовательностей, относящихся к различным группам мтДНК, позволил оценить интенсивность мутационных процессов и особенности распределения “холодных” точек в ГВС1 мтДНК. Результаты анализа структурно-функциональной организации и изменчивости этого участка мтДНК позволяют считать, что снижение изменчивости связано, возможно, с теми участками ГВС1, которые имеют функциональное значение. Между тем, результаты анализа распределения “холодных” точек CAT во вторичных структурах указывают на отсутствие связи между положением “холодной” точки и типом структуры участка мтДНК.

Автор выражает благодарность М.В. Деренко (ИБПС ДВО РАН) за помощь в проведении данного исследования и рецензенту статьи за очень конструктивные замечания.

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. 2002. Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. *Hum. Genet.* **111**, 46–53.
2. Malyarchuk B.A., Rogozin I.B. 2004. Mutagenesis by transient misalignment in human mitochondrial DNA control region. *Ann. Hum. Genet.* **68**, 324–339.
3. Bandelt H.-J., Quintana-Murci L., Salas A., Macaulay V. 2002. The fingerprint of phantom mutations in mtDNA data. *Am. J. Hum. Genet.* **71**, 1150–1160.
4. Metspalu M., Kivisild T., Metspalu E., Parik J., Hudjashov G., Kaldma K., Serk P., Karmin M., Behar D.M., Gilbert M.T., Endicott P., Mastana S., Papiha S.S., Skorecki K., Torroni A., Villems R. 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet.* **5**, e26.
5. Tamm E., Kivisild T., Reidla M., Metspalu M., Smith D.G., Mulligan C.J., Bravi C.M., Rickards O., Martinez-Labarga C., Khusnutdinova E.K., Fedorova S.A., Golubenko M.V., Stepanov V.A., Gubina M.A., Zhadanov S.I., Ossipova L.P., Damba L., Voevoda M.I., Dipierri J.E., Villems R., Malhi R.S. 2007. Beringian standstill and spread of Native American founders. *PLoS ONE*. **2**, e829.
6. Derenko M., Malyarchuk B., Grzybowski T., Denisova G.A., Dambueva I., Perkova M., Dorzhu C., Luzina F., Vanacek T., Villems R., Zakharov I. 2007. Phylogeographic analysis of mitochondrial DNA in North Asian populations. *Am. J. Hum. Genet.* **81**, 1025–1041.
7. Ingman M., Gyllensten U. 2006. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucl. Acids Res.* **34**, D749–D751.
8. Behar D.M., Rosset S., Blue-Smith J., Balanovsky O., Tzur S., Comas D., Mitchell R.J., Quintana-Murci L., Tyler-Smith C., Wells R.S., Genographic Consortium. 2007. The genographic project public participation mitochondrial DNA database. *PLoS Genet.* **3**, e104.
9. Andrews R.M., Kubacka I., Chinnery P.F., Lightowers R.N., Turnbull D.M., Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147.
10. Larizza A., Pesole G., Reyes A., Sbisa E., Saccone C. 2002. Lineage specificity of the evolutionary dynamics of the mtDNA D-loop region in rodents. *J. Mol. Evol.* **54**, 145–155.