



УДК 578.56

Выявление интеграции генетического материала вируса клещевого энцефалита в геномы человека и других млекопитающих

А. И. Парамонов, Ю. П. Джигоев

Научный Центр проблем здоровья семьи и репродукции человека СО РАМН, Иркутск
E-mail: Paramonov_a.i@mail.ru

Аннотация. В ходе поиска участков интеграции генома вируса клещевого энцефалита (ВКЭ) в геномы ряда видов млекопитающих с использованием компьютерного алгоритма BLAST установлен ряд гомологий между геномами ВКЭ и геномами домашней мыши, шимпанзе и человека. Высказано предположение о возможном значении подобных интеграций для животных, восприимчивых к этому вирусу.

Ключевые слова: интеграция участков генома, вирус клещевого энцефалита, геном человека.

Введение

Известно, что геномы позвоночных содержат многочисленные копии ретровирусных последовательностей, приобретённые в течение эволюции. До недавнего времени ретровирусы считались единственным типом РНК-содержащих вирусов, представленных таким образом. Однако недавние исследования показали наличие в геномах растений, насекомых и позвоночных большого количества копий неретровирусных последовательностей [2; 3; 7]. Это доказывает, что источники генетического наследования в геномах позвоночных значительно более разнообразны, чем ранее предполагалось [9; 10]. Сохранение и экспрессия некоторых из этих эндогенных последовательностей, потенциальные взаимосвязи между их присутствием и сопротивляемостью вида вызванным вирусами заболеваниями дают основания предполагать, что их наличие предоставляет важное селективное преимущество популяциям хозяев. Известно, что виды, содержащие в своём геноме фрагменты вирусного генетического материала, более устойчивы к соответствующему патогену, чем не содержащие их [8]. Это верно и в случае, когда интегрированная последовательность не экспрессируется, что, вероятно, связано с подавлением развития вируса путём РНК-интерференции. Помимо этого, участки вирусных последовательностей, содержащиеся в геноме, могут играть регуляторную роль, изменяя уровень экспрессии генов хозяина и/или вируса.

Вирус клещевого энцефалита (ВКЭ), относящийся к роду *Flavivirus*, является возбуди-

лем одного из опаснейших природно-очаговых заболеваний на территории России и Евразии. В последние несколько десятков лет вирус активно расширяет свой ареал, захватывая всё новые территории [1]. Геном вируса представлен одонитевой положительной молекулой РНК длиной около 11 тыс. нуклеотидов. Потенциальная возможность интеграции такого рода вирусов в геном позвоночных и, следовательно, человека, является доказанной [9].

Целью данной работы является поиск возможных участков интеграции генома ВКЭ в геномы человека и других млекопитающих. Изучение этих участков даст ценную информацию для понимания механизмов устойчивости к вирусу, а также установления путей коэволюции вируса и хозяина.

Материалы и методы

Поиск интеграций генома был произведён при помощи онлайн-интерфейса группы программ BLAST [4; 6], служащих для поиска гомологов белков или нуклеиновых кислот, для которых известна первичная последовательность. С помощью BLAST можно сравнить имеющиеся в распоряжении исследователя последовательности с содержащимися в базе данных, и установить предполагаемые гомологи. В данной работе была использована программа Nucleotide Blast, алгоритм blastn, оптимизированный для поиска гомологий в разнородных нуклеотидных последовательностях, не обладающих каким-либо сходством.

Поиск интеграций генома был произведён с использованием полногеномных нуклеотидных

последовательностей прототипных штаммов трёх генотипов ВКЭ (Найдорф, Васильченко, Софьин), а также последовательностей штаммов 886-84 и 178-79, доступных в базе данных GenBank.

Был произведён скрининг против находящихся в публичном доступе геномных последовательностей человека (human genomic plus transcript, human G+T) и домовой мыши (mouse genomic plus transcript, mouse G+T), а также против базы референсных геномных последовательностей (reference genomic sequences, refseq_genomic) с исключением последовательностей вирусов (Taxid:10239).

Результаты

Проведённый поиск показал наличие 34 достоверно совпадающих участков (см. табл. с. 6–8). Длина этих фрагментов выравнивания не превышает 68 нуклеотидов, однако доля совпадающих остатков очень велика (не менее 79 %). Доля пробелов в выравниваниях не превышала 5 %. 27 из всех обнаруженных фрагментов находятся в геноме шимпанзе. Большая часть из них имеют идентичную нуклеотидную структуру и рассеяна в седьмом геномном скаффолде. В геноме человека найдены три, а в геноме мыши – четыре фрагмента. Все найденные фрагменты лежат вне известных генов, хотя иногда поблизости с их границами.

В геномах всех трёх позвоночных найдено большое количество участков малой протяжённости, показывающих высокую гомологию, однако имеющих низкую достоверность ($p > 0,05$). Несмотря на то что они также могут рассматриваться как результат интеграции участков вирусного генома в геном хозяина, они не были включены в результаты.

В ходе предварительного поиска в геноме шимпанзе с высоким значением достоверности ($p \ll 0,001$) были обнаружены гомологичные вирусу КЭ фрагменты большой (до 1500 нуклеотидов) протяжённости. Однако в результате проверки обнаружилось, что эти фрагменты выявляются лишь при анализе геномных последовательностей из базы RefSeq и отсутствуют в полногеномных последовательностях генома шимпанзе, содержащихся в других базах. Они также не были включены в таблицу результатов.

Заключение

Проведённый анализ позволяет с большой долей вероятности утверждать, что геномы исследованных организмов содержат участки,

гомологичные участкам генома ВКЭ или другого флавивируса, и, по всей вероятности, являются результатом интеграции вирусного генетического материала. Количество участков вирусных последовательностей в геноме невелико и в их расположении не наблюдается каких-либо закономерностей. Участки, показавшие высокую достоверность, обладают рядом общих свойств (небольшая длина, высокая гомология и малое число пробелов в выравниваниях). Обнаружено также частое повторение одной и той же вирусной последовательности в разных частях генома хозяина.

Участки генома вируса, гомологичные фрагментам генома позвоночных, лежат как в генах, кодирующих структурные белки, так и в неструктурных генах. В их расположении нами не обнаружено каких-либо закономерностей.

Опубликованы данные о том, что интеграция участков генома РНК-вирусов в большинстве случаев производится путём обратной транскрипции с участием LINE-подобных элементов, однако длина обнаруженных нами участков в основном слишком мала для интеграции таким способом [5].

Обнаруженные участки геномов у разных организмов не совпадают между собой, что позволяет предположить их независимую и сравнительно недавнюю интеграцию. Это предположение подтверждается сравнительно большой величиной фрагментов вирусного генома, найденных в составе генома шимпанзе в одной из баз данных. Можно предположить, что вирус интегрировался в геном того конкретного животного, чья последовательность генома помещена в базу данных, либо его предка недавно, много позже становления шимпанзе как вида.

Абсолютное преобладание количества интегрированных фрагментов в геноме шимпанзе – вида, обитающего на территориях массового распространения флавивирусов, по сравнению с мышью и человеком, может свидетельствовать о том, что подобные интеграции повышают сопротивляемость организма хозяина к соответствующим вирусам.

Литература

1. Молекулярная эпидемиология клещевого энцефалита европейской части России и некоторых стран Балтии, Восточной и Юго-Восточной Европы / Р. В. Адельшин [и др.] // Эпидемиология и вакцинопрофилактика. – 2006. – № 2. – С. 27–34.
2. Basic local alignment search tool / S. F. Altschul [et al.] // J. Mol. Biol. – 1990. – Vol. 215 (3). – P. 403–410.

3. Basic Local Alignment Search Tool. [Электронный ресурс] : сайт. – <http://blast.ncbi.nlm.nih.gov>

4. Belyi V. A. Sequences from Ancestral Single-Stranded DNA Viruses in Vertebrate Genomes: the *Parvoviridae* and *Circoviridae* Are More than 40 to 50 Million Years Old / V. A. Belyi, J. A. Levine, A. M. Skalka // *J. Virol.* – 2010. – December; 84(23). – P. 12458–12462.

5. Belyi V. A. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes / V. A. Belyi, J. A. Levine, A. M. Skalka // *PLoS Pathog.* – 2010. – Jul 29, 6(7): e1001030.

6. Endogenous non-retroviral RNA virus elements in mammalian genomes / M. Horie [et al.] // *Nature.* – 2010. – Vol. 463. – P. 84–87.

7. Isolation and characterization of Israeli acute paralysis virus, a dicistrovirus affecting honeybees in

Israel: evidence for diversity due to intra- and interspecies recombination / E. Maori [et al.] // *J. Gen. Virol.* – 2007. – Vol. 88. – 3428–3438.

8. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration / M. B. Geuking [et al.] // *Science.* – 2009. – Vol. 323. – P. 393–396.

9. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes / S. Crochu [et al.] // *J. Gen. Virol.* – 2004. – Vol. 85. – P. 1971–1980.

10. Tanne E. Occurrence of a DNA sequence of a non-retro RNA virus in a host plant genome and its expression: evidence for recombination between viral and host RNAs / E. Tanne, I. Sela // *Virology.* – 2005. – Vol. 332. – P. 614–622.

Identification of the integration of tick-borne encephalitis virus genetic material in the genomes of humans and other mammals

A. I. Paramonov, Yu. P. Dzhioev

Research Center of Family Health Problems and Human Reproduction SB RAMS, Irkutsk.

Abstract. The aim of this work was to search for the presence of viral genome integration sites in the genomes of humans and other animals. With the computer algorithm BLAST homology was found between the genomes of tick-borne encephalitis virus and some mammals. We suggested the possible significance of such integrations for the animals susceptible to this virus.

Keywords: integration of genome fragments, tick-borne encephalitis virus, human genome.

*Парамонов Алексей Игоревич
Институт эпидемиологии и микробиологии
НЦ ПЗС РЧ РАМН*

*664025, Иркутск, ул. К. Маркса, 3
младший научный сотрудник
тел: (3952) 33–39–52
E-mail: Paramonov_a.i@mail.ru*

*Джзиев Юрий Павлович
Институт эпидемиологии и микробиологии
НЦ ПЗС РЧ СО РАМН*

*664025, Иркутск, ул. К. Маркса, 3
кандидат биологических наук,
старший научный сотрудник
тел. (3952) 33–39–52
E-mail: alanir07@mail.ru*

*Paramonov Aleksey Igorevich
Institute of Epidemiology and Microbiology Scientific
Centre of Family Health and Human Reproduction
Problems SO RAMS
3 K. Marks St., Irkutsk, 664025
junior research scientist
phone: (3952) 33–39–52
E-mail: Paramonov_a.i@mail.ru*

*Dzhioev Yuri Pavlovich
Institute of Epidemiology and Microbiology,
Scientific Centre of Family Health and Human
Reproduction Problems SO RAMS
3 K. Marx St., Irkutsk, 664025
Ph. D. of Biology, senior research scientist
phone: (3952) 33–39–51,
E-mail: alanir07@mail.ru*

Таблица

Участки интеграции фрагментов генома вируса клещевого энцефалита в геномы позвоночных, установленные с использованием программы BLAST

Идентификатор целевой последовательности в Genbank	Описание целевой последовательности	Счёт выравнивания	E-значение	Длина выравнивания, пар нуклеотидов	Доля идентичных остатков, %	Доля пробелов, %	Штамм вируса в запросе	Позиция в запросе	Позиция в цели	Особенности в этой части последовательности
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	54	83	0	Sofjin	1135-1188	709230-709177	16169 bp at 5' side: myosin-Ig isoform 6; 5887 bp at 3' side: malcavernin isoform 3
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	54	83	0	Sofjin	1135-1188	719913-719860	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	54	83	0	Sofjin	1135-1188	730597-730544	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	54	83	0	Sofjin	1135-1188	741277-741224	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	Sofjin	5691-5731	747319-747279	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	Sofjin	5691-5731	757862-757822	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	Sofjin	5691-5731	768404-768364	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	Sofjin	5691-5731	778939-778899	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	Sofjin	5691-5731	789480-789440	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	Sofjin	5691-5731	800019-799979	NAC-alpha domain-containing protein 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	Sofjin	5691-5731	810559-810519	4275 bp at 5' side: NAC-alpha domain-containing protein 1; 7110 bp at 3' side: protein TBRG4 isoform 7
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	Sofjin	5691-5731	821099-821059	11682 bp at 5' side: myosin-Ig isoform 6-10383 bp at 3' side: malcavernin isoform 3
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	55,4	0,033	45	87	0	Sofjin	5691-5735	715417-715373	malcavernin isoform 3; malcavernin isoform 1
ref NW_003457177.1	Pan troglodytes геномный скаффолд хромосомы 7	55,4	0,033	45	87	0	Sofjin	5691-5735	736781-736737	malcavernin isoform 3; malcavernin isoform 1

Идентификатор целевой последовательности в Genbank	Описание целевой последовательности	Счёт выравнивания	E-значение	Длина выравнивания, пар нуклеотидов	Доля идентичных остатков, %	Доля пробелов, %	Штамм вируса в запросе	Позиция в запросе	Позиция в цели	Особенности в этой части последовательности
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	178-79	5691-5731	747319-747279	malcavernin isoform 3; malcavernin isoform 1
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	178-80	5691-5731	757862-757822	malcavernin isoform 3; malcavernin isoform 1
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	178-81	5691-5731	768404-768364	malcavernin isoform 3; malcavernin isoform 1
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	178-82	5691-5731	778939-778899	malcavernin isoform 3; malcavernin isoform 1
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	178-83	5691-5731	789480-789440	malcavernin isoform 3; malcavernin isoform 1
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	178-84	5691-5731	800019-799979	NAC-alpha domain-containing protein 1
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	178-85	5691-5731	810559-810519	4275 bp at 5' side: NAC-alpha domain-containing protein 1; 7110 bp at 3' side: protein TBRG4 isoform 7
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	41	90	0	178-86	5691-5731	821099-821059	11682 bp at 5' side: myosin-Ig isoform 6-10383 bp at 3' side: malcavernin isoform 3
ref[NT_010274.17]	Homo sapiens геномный скаффолд хромосомы 7	51,8	0,016	42	90	5	886-84	5327-5366	15612254-15612295	-
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	64,4	0,00006	50	88	0	Vasilchenko	7753-7802	829722-829673	5488 bp at 5' side: myosin-Ig isoform 6; 16566 bp at 3' side: malcavernin isoform 3
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	56	82	0	Vasilchenko	1131-1186	709234-709179	16171 bp at 5' side: myosin-Ig isoform 6; 5883 bp at 3' side: malcavernin isoform 3
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	56	82	0	Vasilchenko	1131-1186	719917-719862	malcavernin isoform 3; malcavernin isoform 1
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	56	82	0	Vasilchenko	1131-1186	730601-730546	malcavernin isoform 3; malcavernin isoform 1
ref[NW_003457177.1]	Pan troglodytes геномный скаффолд хромосомы 7	57,2	0,009	56	82	0	Vasilchenko	1131-1186	741281-741226	malcavernin isoform 3; malcavernin isoform 1

Окончание табл.

Идентификатор целевой последовательности в Genbank	Описание целевой последовательности	Счёт выравнивания	E-значение	Длина выравнивания, пар нуклеотидов	Доля идентичных остатков, %	Доля пробелов, %	Шагм вируса в запросе	Позиция в запросе	Позиция в цели	Особенности в этой части последовательности
ref[NT_004487.19]	Homo sapiens геномный контиг хромосомы 1	50,0	0,05	68	79	4	Neudoerfl	4142–4208	33172931–33172996	–
ref[NW_001838533.2]	Homo sapiens геномный контиг хромосомы 1, альтернативная сборка	50,0	0,05	68	79	4	Neudoerfl	4142–4208	24189345–24189287	–
ref[NT_096135.5]	Mus musculus штамм C57BL/6J геномный контиг хромосомы 11	51,8	0,014	36	92	0	Neudoerfl	8047–8082	31566006–31566041	protein shisa-6 homolog precursor
ref[NT_096135.5]	Mus musculus штамм C57BL/6J геномный контиг хромосомы 11	51,8	0,014	36	92	0	Neudoerfl	8047–8082	4262664–4262699	–
ref[NT_039649.7]	Mus musculus штамм C57BL/6J геномный контиг хромосомы 17	50,0	0,049	46	85	4	Neudoerfl	2815–2860	56607589–56607632	DNA-binding protein REX2 isoform 2; DNA-binding protein REX2 isoform 1
ref[NW_001030622.1]	Mus musculus штамм C57BL/6J геномный контиг хромосомы 17	50,0	0,049	46	85	4	Neudoerfl	2815–2861	13707753–13707796	DNA-binding protein REX2 isoform 2; DNA-binding protein REX2 isoform 1

Примечание: столбцы 1 и 2 содержат информацию о нуклеотидной последовательности позвоночного животного, в которой были обнаружены совпадения с последовательностью вируса. Столбцы 3–7 содержат критерии, позволяющие оценить качество полученных программой BLAST совпадений. Столбец 8 – название штамма вируса, использованного при анализе. Столбцы 9 и 10 отражают положение предполагаемого участка интеграции в геноме вируса и позвоночного соответственно. В столбце 11 указаны ближайшие к предполагаемому участку интеграции гены из генома позвоночного, а также расстояние до них в парах нуклеотидов (если известно)