



УДК 004.94:004.738.5:577

Выявление потенциальных событий рекомбинации в нуклеотидных последовательностях вируса клещевого энцефалита с помощью компьютерных программ

А. И. Парамонов^{1,2}, Ю. П. Джиев¹, Т. В. Дёмина¹, Ю. С. Букин³,
И. В. Козлова¹, А. А. Приставка², В. П. Саловарова²

¹Институт эпидемиологии и микробиологии, Научный центр проблем здоровья семьи и репродукции человека СО РАМН, Иркутск

²Иркутский государственный университет, Иркутск

³Лимнологический институт СО РАН, Иркутск

E-mail: Paramonov_a.i@mail.ru

Аннотация. Охарактеризован ряд компьютерных методов для детекции рекомбинации у РНК геномных вирусов. Показано наличие рекомбинации у вируса клещевого энцефалита и дана оценка применимости различных методов обнаружения рекомбинации к этому объекту.

Ключевые слова: компьютерные программы для определения рекомбинации, геномы флавивирусов, вирус клещевого энцефалита, рекомбинация.

Введение

Рекомбинация – процесс, который влияет на биологическую эволюцию на самых разных уровнях. Рекомбинация разрушает сцепленное наследование, что существенно для генного картирования, создаёт значительное генетическое разнообразие в природных популяциях путём перетасовки существующих вариантов геномов, её наличие необходимо учитывать при реконструкции филогений. Кроме того, рекомбинация играет важную роль в эволюции некоторых человеческих патогенов, что отражается на методах их лечения и профилактики [12]. Рекомбинация имеет существенное значение в эволюции вирусов, являясь способом обмена генами между вирусными частицами, а также, наряду с дрейфом генов, источником генетического разнообразия. Она отмечена во всех группах ДНК-содержащих вирусов и у многих РНК-содержащих вирусов [2; 5]. Рекомбинация отмечена и у представителей рода *Flavivirus* [7], в который входит возбудитель такого опасного заболевания, как клещевой энцефалит. Вопрос о наличии рекомбинации у ВКЭ обсуждается, однако литературные данные по этому вопросу противоречивы [3].

Геном вируса клещевого энцефалита (ВКЭ) представлен одной молекулой однонитчатой РНК положительной полярности общей длиной около 11 тыс. нуклеотидов [1]. Скорость нуклеотидных замен достаточно велика и состав-

ляет в среднем по геному $2,6 \times 10^{-4}$ на сайт в год [4]. Поскольку всестороннее изучение эволюции ВКЭ может иметь большое значение для диагностики, предупреждения и лечения вызываемого им заболевания, вопрос о возможности рекомбинации должен быть разрешён.

Компьютерные программы для обнаружения рекомбинации дают ответ на следующие вопросы: присутствуют ли в выборке рекомбинантные последовательности, где именно расположены сайты рекомбинации и какие последовательности вовлечены в рекомбинацию [8]. Если с ответом на первые два вопроса справляются практически все программы, то ответ на третий вопрос часто вызывает трудности и требует дополнительной ручной работы, такой как анализ филогенетических деревьев.

Методы обнаружения рекомбинации можно разделить на несколько групп, в зависимости от подхода, применяемого при её обнаружении.

Методы расстояний, или методы анализа шаблонов последовательностей ищут изменения в расстояниях между последовательностями. Обычно они используют скользящее окно и какую-либо статистику, основанную на генетических расстояниях. К этой группе относятся использованные в данной работе методы Maximum χ^2 , Chimaera, SiScan, GENECONW.

Филогенетические методы фиксируют наличие рекомбинации, когда филогенетические деревья, построенные на разных частях генома,

имеют различные топологии. Если сравнение двух соседних участков последовательностей даёт разный порядок ветвления, это позволяет предположить, что между ними произошло рекомбинационное событие. Эта группа методов наиболее широко упоминается в литературе, в неё входят методы TOPAL и RDP.

Методы совместимости анализируют последовательности сайт за сайтом в поисках филогенетических несовместимостей, для этого не требуется филогенетическая схема целых последовательностей. В эту группу входит метод BootScan.

Материалы и методы

Исследовались 29 полногеномных нуклеотидных последовательностей вируса клещевого энцефалита. В качестве аутгруппы использовалась полногеномная нуклеотидная последовательность вируса Омской геморрагической лихорадки. В работе применялись следующие компьютерные методы обнаружения рекомбинации: TOPAL [10], реализованный в составе пакета программ TOPALi v. 2.5 [15]; RDP [9]; GENECONW [11]; BootScan [6]; Максимум χ^2 [14]; Chimaera [12]; SiScan [8], реализованные в пакете программ RDP v. 3.34 [13]. Настройка программ производилась на рекомбинантных штаммах вируса гепатита В (штамм HPBADW1, номер в GenBank-№ D00329; HPBADW2-№ D00330; HPBADWZCG-№M57663; HPBADRC-№D00630), а также на химерных флавивирусах – искусственных гибридах ВКЭ и вируса Денге. Все последовательности взяты из базы данных GenBank.

Для всех программ была использована модель нуклеотидных замен Джукса – Кантора. Достижимый уровень значимости был принят за 0,05. В ходе работы с программами применялись следующие настройки: TOPAL – длина окна 500, шаг 10, 1 проход; RDP – внешние и внутренние ссылки, размер окна 30; GENECONW – g-scale = 1; BootScan. Длина окна 200, шаг 20, использовать UPGMA дерево, число повторов бутстрепа 100; Максимум χ^2 – заданный размер окна, 70 переменных сайтов на окно, не использовать пропуски; Chimaera – заданный размер окна, 70 переменных сайтов на окно; SiScan – размер окна 200, шаг 20, не использовать пропуски, использовать все позиции. Все прочие настройки выставлены по умолчанию.

Результаты и обсуждение

На предварительном этапе были произведены подбор и настройка компьютерных про-

грамм обнаружения рекомбинации на рекомбинантных штаммах вируса гепатита В и химерных флавивирусах. Вирус гепатита В – ДНК-содержащий вирус, явление рекомбинации в котором хорошо изучено. Для работы был использован один заведомо рекомбинантный штамм вируса гепатита; два штамма, являющиеся родительскими для рекомбинанта; один штамм выступал как аутгруппа. На этой выборке последовательностей была проверена способность компьютерных программ обнаруживать рекомбинацию, а также их склонность к ложноположительным определениям. Программы, показавшие лучшие результаты, были использованы в дальнейшей работе.

Во вторую тестовую выборку входили геномы двух химерных флавивирусов – генно-инженерные гибриды между ВКЭ и вирусом Денге и последовательности родительских вирусов. На этой выборке была произведена настройка параметров программ, наиболее подходящая для детекции рекомбинации у флавивирусов, с учётом высокой вариации частот замен между сайтами.

Затем было проведено определение наличия рекомбинации в исследуемой выборке штаммов, определено положение точек рекомбинации. В таблице представлены список использованных в работе штаммов вирусов, их номера в GenBank, обнаруженные точки рекомбинации и методы, которыми они были обнаружены.

Метод BootScan обнаружил 1 рекомбинационное событие в штамме SN205. Метод Максимум χ^2 – 4 рекомбинационных события в штаммах: Senjang, MDJ-01, Kavalerovo, а также в вирусе OHF. Метод Chimaera – 7 рекомбинационных событий в штаммах Senjang, MDJ-01, Neudoerfl, SN 263, K23, Hupr, Kavalerovo. Также зафиксирован ряд очень коротких рекомбинантных участков, предположительно являющихся ложноположительными определениями. Метод SiScan показал склонность к ложноположительным определениям: установлены более 200 рекомбинационных событий во всех штаммах. Методом TOPAL выявлено около 40 точек рекомбинации, следовательно, около 20 рекомбинационных событий. К сожалению, этот метод не позволяет определить, какие именно последовательности рекомбинантны. Метод RDP установил 4 рекомбинационных события в штаммах EK-328, Vasilichenko, Zausaev, Kolarovo-2008. Метод GENECONW показал отсутствие рекомбинации среди изучаемых последовательностей.

Обнаруженные точки рекомбинации генов в исследуемой выборке штаммов флавивирусов

Штамм	Номер доступа	Положение точки рекомбинации	Метод
Primorye-90	FJ997899	–	–
Sofjin	–	–	–
Sofjin-HO	AB062064	–	–
Oshima 5-10	AB062063	–	–
SN205	DQ989336	380–030	BootScan
Senzhang	AY182009	2746–3882 2728–3882	Chimaera Максимум χ^2
MDJ-01	AY217093	2745–3867 2715–3867	Chimaera Максимум χ^2
Glubinnoe	DQ862460	–	–
178-79	EF469661	–	–
Neudoerfl	U27495	9647–10242	Chimaera
263	DQ153877	9648–10242	Chimaera
K23	AF091010	9648–10242	Chimaera
Hypr	U39292	9326–10021	Chimaera
886-84	EF469662	–	–
EK-328	DQ486861	8903–10122	RDP
Vasilchenko	AF069066	9780–10122	RDP
Zausaev	AF527415	9746–10099	RDP
P-69	EU816453	–	–
P-94	EU816454	–	–
P-253	EU816451	–	–
P-270	EU816452	–	–
P-86	EU816455	–	–
P-212	EU816450	–	–
P-322	AY169390	–	–
P-89	FJ906622	–	–
Kavalerovo	FJ402885	2760–3867 2762–3867	Chimaera Максимум χ^2
Dalnegorsk	FJ402886	–	–
Kolarovo-2008	FJ968751	9743–10122	RDP
Salem	FJ572210	–	–
Omsk hemorrhagic fever	NC_005062	380–891	Максимум χ^2

В ходе работы было произведено определение наличия рекомбинации в выборке штаммов флавивирусов, а также сделаны выводы о возможности применения различных программ обнаружения рекомбинации к данному объекту. Метод SiScan был отвергнут на предварительном этапе, так как показал большое количество ложноположительных определений на тестовой выборке. Вероятно, это связано с высокой скоростью нуклеотидных замен в данных последовательностях. Методы Bootscan и GENECONV показали высокую чувствительность на тестовой выборке, однако в исследуемом наборе последовательностей обнаружили лишь небольшое количество рекомбинантных участков. Вероятно, это связано с величиной

выборки и входящими в неё штаммами – оба метода требуют наличия в ней родительских последовательностей.

Методы RDP, Chimaera, Максимум χ^2 и TOPAL показали высокую чувствительность и малое количество ложноположительных определений на тестовой выборке вируса гепатита В. Определение рекомбинации этими методами показало наличие значительного количества рекомбинантных участков в исследуемом наборе последовательностей ВКЭ, что позволяет сделать вывод о возможности рекомбинации у данного вируса.

Заключение

Таким образом, посредством проведенного скрининга и применения компьютерных про-

грамм для обнаружения рекомбинации у вирусов показано, что наиболее подходящими методами для выявления рекомбинационных событий в популяциях вирусов с одноцепочечным несегментированным РНК-геномом являются программные методы RDP, Chimaera, Максимум χ^2 и TOPAL. Результаты, полученные с применением этих программ на полномочной выборке штаммов ВКЭ достоверно показывают возможность детекции у них рекомбинационных событий.

Литература

1. Амосов А. Д. Клещевой энцефалит : Информационно-методическое пособие / А. Д. Амосов. – Кольцово : Ин-т средств мед. диагностики ЗАО «Вектор-Бест», 2006. – 115 с.
2. Белоусов Е. В. Некоторые аспекты нерепликативной рекомбинации между фрагментами геномной РНК вируса полиомиелита : автореф. дис. ... канд. биол. наук / Е. В. Белоусов. – М., 2002. – 167 с.
3. Карганова Г. Г. Механизмы микроэволюции вируса клещевого энцефалита : дис. ... канд. биол. наук / Г. Г. Карганова. – М., 2009. – 389 с.
4. Молекулярная эпидемиология клещевого энцефалита / В. И. Злобин [и др.]. – Иркутск : РИО ВСНЦ СО РАМН, 2003. – 272 с.
5. Филдс Б. Вирусология / Б. Филдс, Д. Найн, Ф. Мэрфи. – М. : Мир, 1989. – Т. 1. – 496 с.
6. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints / D. P. Martin // AIDS Res. Hum. Retroviruses. – 2005. – P. 98–102.

7. Edward C. H. Phylogenetic Evidence for Recombination in Dengue Virus / C. H. Edward, M. Worobey, A. Rambaut // Mol. Biol. Evol. – 1999. – Vol. 16, N 3. – P. 405–409.

8. Gibbs M. J. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences / M. J. Gibbs, J. S. Armstrong, A. J. Gibbs // Bioinformatics. – 2000. – Vol. 16. – P. 573–582.

9. Martin D. RDP: detection of recombination amongst aligned sequences / D. Martin, E. Rybicki // Bioinformatics. – 2000. – Vol. 16. – P. 562–563.

10. McGuire G. TOPAL 2.0: Improved detection of mosaic sequences within multiple alignments / G. McGuire, F. Wright // Bioinformatics. – 2000. – Vol. 16. – P. 130–134.

11. Padidam M. Possible emergence of new geminiviruses by frequent recombination // M. Padidam, S. Sawyer, C. M. Fauquet // Virology. – 1999. – Vol. 265. – P. 218–225.

12. Posada D. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations / D. Posada, K. A. Crandall // Proc. Natl. Acad. Sci. – 2001. – Vol. 98. – P. 13757–13762.

13. Recombination patterns in aphthoviruses mirror those found in other picornaviruses / L. Heath [et al.] // J. Virol. – 2006. – Vol. 80. – P. 11827–11832.

14. Smith M. J. Analyzing the mosaic structure of genes / M. J. Smith // J. Mol. Evol. – 1992. – Vol. 34. – P. 126–129.

15. TOPALi: Software for Automatic Identification of Recombinant Sequences within DNA Multiple Alignments // I. Milne [et al.] // Bioinformatics. – 2004. – Vol. 20, N 11. – P. 1806–1807.

Identification of potential recombination events in nucleotide sequences of tick-borne encephalitis virus by computer software

A. I. Paramonov^{1,2}, Yu. P. Dzhioev¹, T. V. Demina¹, Yu. S. Bukin³, I. V. Kozlova¹, A. A. Pristavka², V. P. Salovarova²

¹ Institute of Epidemiology and Microbiology, SC PFHHR SB RAMS, Irkutsk

² Irkutsk State University, Irkutsk

³ Limnological Institute SB RAS, Irkutsk

Abstract. A number of computer methods for recombination detection are characterized. The presence of recombination at a tick-borne encephalitis virus is shown and an estimation of applicability of various methods to this object is made.

Keywords: computer programs for recombination detection, the genomes of flaviviruses, tick-borne encephalitis virus, recombination

Парамонов Алексей Игоревич
Институт эпидемиологии и микробиологии
НЦ ПЗС РЧ РАМН
664025, г. Иркутск, ул. К. Маркса, 3
младший научный сотрудник
тел: (3952) 33–39–52
E-mail: Paramonov_a.i@mail.ru

Paramonov Aleksey Igorevitch
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

Джиоев Юрий Павлович
Институт эпидемиологии и микробиологии НЦ ПЗС
РЧ СО РАМН
664025, г. Иркутск, ул. К. Маркса, 3
кандидат биологических наук, старший научный
сотрудник
тел. (3952) 33–39–52
E-mail: alanir07@mail.ru

Dhizioev Yuri Pavlovitch
Institute of Epidemiology and Microbiology,
Scientific Centre of Family Health and Human Repro-
duction 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

Букин Юрий Сергеевич
Лимнологический институт СО РАН
664033, Иркутск, ул. Улан-Баторская, 3
кандидат биологических наук
старший научный сотрудник
тел. (3952) 42–29–23
E-mail: bukinyura@mail.ru

Bukin Yury Sergeevitch
Limnological Institute SB RAS
3 Ulan-Batorskaya St., Irkutsk, 664033
Ph.D. in Biology, senior research scientist
phone: (3952)42–29–23
E-mail: bukinyura@mail.ru

Козлова Ирина Валерьевна
Институт эпидемиологии и микробиологии НЦ
ПЗС РЧ РАМН
664025, г. Иркутск, ул. К. Маркса, 3
доктор медицинских наук,
заведующий лабораторией
тел: (3952) 33–39–52

Kozlova Irina Valerjevna
Institute of Epidemiology and Microbiology,
Scientific Centre of Family Health and Human
Reproduction Problems SB RAMS
3 K. Marks St., Irkutsk, 664025
D. Sc. of Medicine, Head of Laboratory
phone: (3952) 33–39–52

Приставка Алексей Александрович
Иркутский государственный университет
664003 г. Иркутск, ул. Сухэ-Батора, 5
кандидат биологических наук, доцент
тел. (3952) 24–18–55

Pristavka Aleksey Aleksandrovitch
Irkutsk State University
5 Sukhe-Bator St., Irkutsk, 664003
Ph. D. of Biology, ass. prof.
phone: (3952) 24–18–55

Саловарова Валентина Петровна
Иркутский государственный университет
664003 г. Иркутск, ул. Сухэ-Батора, 5
зав. кафедрой физико-химической биологии
доктор биологических наук, профессор.
E-mail: vsalovarova@rambler.ru
тел. (3952) 24–18–55

Salovarova Valentina Petrovna
Irkutsk State University
5 Sukhe-Bator St., Irkutsk, 664003
D. Sc. of Biology, Prof., Head of Department
of Physical and Chemical Biology
E-mail: vsalovarova@rambler.ru
phone: (3952) 24–18–55