

ГЕНЕТИКА, ПРОТЕОМИКА И МЕТАБОЛОМИКА GENETICS, PROTEOMICS AND METABOLOMICS

ИССЛЕДОВАНИЕ ПОТЕНЦИАЛЬНЫХ СОБЫТИЙ РЕКОМБИНАЦИИ В БЕЛОК-КОДИРУЮЩЕЙ ОБЛАСТИ CRISPR-CAS ЛОКУСОВ В ГЕНОМАХ РАЗНЫХ СЕРОВАРИАНТОВ *SALMONELLA ENTERICA* МЕТОДАМИ *IN SILICO*

РЕЗЮМЕ

Арефьева Н.А.^{1,2,3},
Букин Ю.С.⁴,
Эрдынеев С.В.²,
Джиоев Ю.П.²,
Мирошниченко Л.А.⁵

¹ ФГБНУ «Научный центр проблем
здоровья семьи и репродукции
человека» (664003, г. Иркутск,
ул. Тимирязева, 16, Россия)

² ФГБОУ ВО «Иркутский государственный
медицинский университет»
Минздрава России (664003, г. Иркутск,
ул. Красного Восстания, 1, Россия)

³ ФГБОУ ВО «Иркутский государственный
университет» (664003, г. Иркутск,
ул. Карла Маркса, 1, Россия)

⁴ ФГБУН Лимнологический институт
СО РАН (664033, г. Иркутск,
ул. Улан-Баторская, 3, Россия)

⁵ ФГБУН Институт математики
им. С.Л. Соболева СО РАН (630090,
Новосибирск, пр. Академика Коптюга, 4,
Россия)

Автор, ответственный за переписку:

**Арефьева Надежда
Александровна,**

e-mail: arefieva.n4@gmail.com

Обоснование. Исследование процессов гомологичной рекомбинации в области *cas*-генов у *Salmonella enterica* позволит выяснить фундаментальные механизмы эволюции CRISPR-Cas систем, что важно для изучения возникновения у данного патогена резистентности к фагам.

Цель работы. Исследование процессов рекомбинации в белок-кодирующей части CRISPR-Cas локусов в геномах сероваров *Salmonella Enteritidis*, *Infantis* и *Typhimurium*, используя методы *in silico*.

Материалы и методы. Геномные последовательности серовариантов *Salmonella Enteritidis*, *Infantis* и *Typhimurium* были скачаны из базы данных NCBI GenBank. Кодированные последовательности *cas*-генов были извлечены из геномов и выровнены с учётом позиции кодона. В полученном выравнивании был выполнен поиск событий рекомбинации. Проведена верификация событий рекомбинации.

Результаты. Найдено 7683 потенциальных события рекомбинации в области *cas*-локуса в геноме *S. enterica*. Среди них верифицировано 810 (10,54 %) событий; 45 (0,59 %) событий были идентифицированы как результаты конвергентной эволюции. События рекомбинации детектируются чаще между штаммами, принадлежащими разным серовариантам, чем между штаммами, принадлежащими одному сероварианту. Все сероварианты могут рекомбинировать друг с другом, однако чаще всего рекомбинация происходит между штаммами *Enteritidis* и *Infantis* и между *Typhimurium* и *Infantis*. Не было найдено ни одного события рекомбинации между штаммами сероварианта *Enteritidis*. События конвергентной адаптивной эволюции в основном локализованы в генах эффекторного модуля: *cas5*, *cas6*, *cas7*.

Заключение. Показано, что гомологичная рекомбинация часто происходит в геноме *S. enterica* в области *cas*-генов. Биоинформатические алгоритмы находят больше событий рекомбинации между эволюционно более отдалёнными штаммами, что не согласуется с известными исследованиями, проведёнными *in vitro*.

Ключевые слова: гомологичная рекомбинация, CRISPR-Cas, *Salmonella enterica*, методы *in silico*

Для цитирования: Арефьева Н.А., Букин Ю.С., Эрдынеев С.В., Джиоев Ю.П., Мирошниченко Л.А. Исследование потенциальных событий рекомбинации в белок-кодирующей области CRISPR-Cas локусов в геномах разных серовариантов *Salmonella enterica* методами *in silico*. *Acta biomedica scientifica*. 2025; 10(1): 59-68. doi: 10.29413/ABS.2025-10.1.6

Статья поступила: 07.04.2024

Статья принята: 27.12.2024

Статья опубликована: 13.03.2025

THE STUDY OF POTENTIAL RECOMBINATION EVENTS IN THE PROTEIN-CODING REGIONS OF CRISPR-CAS LOCI IN THE GENOMES OF DIFFERENT *SALMONELLA ENTERICA* SEROVARIAN TS USING *IN SILICO* METHODS

Arefieva N.A.^{1,2,3},
Bukin Yu.S.⁴,
Erdyneev S.V.²,
Dzhioev Yu.P.²,
Miroshnichenko L.A.⁵

¹ Scientific Centre for Family Health and Human Reproduction Problems (Timiryazeva str. 16, Irkutsk 664003, Russian Federation)

² Irkutsk State Medical University (Krasnogo Vosstaniya str. 1, Irkutsk 664003, Russian Federation)

³ Irkutsk State University (Karla Marksa str. 1, Irkutsk 664004, Russian Federation)

⁴ Limnological Institute, Siberian Branch of the Russian Academy of Sciences (Ulan-Batorskaya str. 3, Irkutsk 664033, Russian Federation)

⁵ Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences (Akademika Koptyuga Ave. 4, Novosibirsk 630090, Russian Federation)

Corresponding author:
Nadezhda A. Arefieva,
e-mail: arefieva.n4@gmail.com

ABSTRACT

Background. The study of recombination processes in the CRISPR-Cas loci of *Salmonella enterica* can help investigate the fundamental evolutionary mechanisms of the CRISPR-Cas system to better understand the acquisition of phage resistance.

The aim of the study. To investigate the recombination processes in the protein-coding regions of CRISPR-Cas loci in the genomes of *Salmonella enterica* serovars Enteritidis, Infantis, and Typhimurium using *in silico* methods.

Materials and methods. The genomic sequences of the *Salmonella* serovars Enteritidis, Infantis, and Typhimurium were downloaded from the NCBI GenBank database. The coding sequences of cas genes were extracted from the genomes and aligned according to codon position. Recombination events were identified in the resulting alignment using multiple algorithms. Verification of recombination events was performed.

Results. A total of 7683 potential recombination events were identified. Among these, 810 (10.54 %) were verified, and 45 (0.59 %) were recognized as results of convergent evolution. Recombination events are detected more frequently between strains belonging to different serovariants than between those of the same serovariant. All serovariants can recombine with each other; however, recombination primarily occurs between Enteritidis and Infantis strains, as well as between Typhimurium and Infantis strains. Infantis and Typhimurium serovariants also exhibit recombination within themselves. No recombination events were found between strains of the Enteritidis serovariant. The events of convergent adaptive evolution were mainly found in the effector module genes: cas5, cas6, cas7.

Conclusion. It has been shown that homologous recombination often occurs in the *S. enterica* genome in the region of the cas genes. Bioinformatic algorithms detect more recombination events between evolutionarily more distant strains, which are inconsistent with known *in vitro* studies.

Key words: homologous recombination, CRISPR-Cas, *Salmonella enterica*, *in silico* methods

Received: 07.04.2024

Accepted: 27.12.2024

Published: 13.03.2025

For citation: Arefieva N.A., Bukin Yu.S., Erdyneev S.V., Dzhioev Yu.P., Miroshnichenko L.A. The study of potential recombination events in the protein-coding regions of CRISPR-Cas loci in the genomes of different *Salmonella enterica* serovariants using *in silico* methods. *Acta biomedica scientifica*. 2025; 10(1): 59-68. doi: 10.29413/ABS.2025-10.1.6

ВВЕДЕНИЕ

Salmonella enterica является одним из наиболее часто выявляемых кишечных патогенов как по количеству инфекций среди людей, так и по тяжести заболевания [1]. Чаще всего сальмонеллёз связывают с несоблюдением гигиенических стандартов при производстве продуктов питания птицеводческого происхождения [2]. По некоторым оценкам, инфекции, вызываемые сальмонеллой, ежегодно становятся причиной более 300 тысяч смертей – в основном в развивающихся странах [3].

В роде *Salmonella* выделяют два вида – *S. bongori* и *S. enterica*, которые включают более 2600 охарактеризованных различных серовариантов, многие из которых способны вызывать заболевания как у людей, так и у животных. При этом вид *S. enterica* имеет шесть подвидов: *enterica*, *salamae*, *arizonae*, *diarizonae*, *houtenae* и *indica* [1, 3]. Среди них доминирующим этиологическим агентом заболеваемости как в мире, так и в России является *S. enterica* subsp. *enterica* с преобладанием серовариантов Enteritidis, Typhimurium и Infantis в период 2011–2019 гг. [4].

Лечение инфекций, вызываемых этими серовариантами сальмонелл, бывает затруднено в связи с появлением у них устойчивости к антибиотикам. Одним из перспективных способов борьбы с патогенными бактериями является фаготерапия, однако бактерии также способны приобретать резистентность к фаговым препаратам посредством системы CRISPR-Cas. Для разработки и повышения эффективности фаготерапии против *S. enterica* необходимо изучить особенности функционирования и эволюции CRISPR-Cas систем. Согласно базе данных CRISPRCasdb [5], *Salmonella enterica* содержит CRISPR-Cas систему класса 1 подтипа I-E (включает белки Cas3, Cse1, Cse2, Cas7, Cas5, Cas6, Cas1, Cas2). В классификации CRISPR-Cas систем принято выделять два класса, различающихся структурно и функционально. По своей структуре CRISPR-Cas система представляет собой рибонуклеопротеиновый комплекс, состоящий из двух функциональных модулей – адаптационного и эффекторного. Первый отвечает за встраивание новых спейсеров в массив CRISPR, а второй – за распознавание и уничтожение чужеродной ДНК или РНК. Основным отличием систем первого класса является у них наличие комплексного адаптационного модуля, состоящего из нескольких Cas-белков, в то время как у систем второго класса адаптационный модуль представлен одним белком. Адаптационный модуль системы I-E стандартно состоит из белков Cas1 и Cas2. Эффекторный модуль включает белки Cse1, Cse2, Cas7, Cas5, Cas6. Разрезание ДНК-мишени осуществляет эндонуклеаза Cas3 [6, 7].

Гомологичная рекомбинация играет важную роль в эволюционной адаптации бактерий [8]. Следствием событий рекомбинации могут являться такие адаптации, как: появление новых штаммов и видов; расширение круга хозяев; повышение вирулентности; появление новых механизмов уклонения от иммунитета хозяина; возникновение резистентности к антибактериальным препаратам. Считается, что гомологичная рекомбинация оказывает влияние главным образом на гены, которые являются общими для большинства штаммов в популя-

ции, в то время как остальные геномные области приобретают дополнительный контент посредством горизонтального переноса [9]. Исследование процессов рекомбинации в области CRISPR-cas локусов у наиболее часто выделяемых у человека штаммов *S. enterica* позволит изучить особенности приобретения у *S. enterica* резистентности к фагам посредством гомологичной рекомбинации в белок-кодирующей части локусов CRISPR-Cas систем.

Таким образом, **целью** данной работы являлось проведение исследования процессов гомологичной рекомбинации в белок-кодирующей части CRISPR-Cas локусов в геномах штаммов *Salmonella enterica* серовариантов Enteritidis, Infantis и Typhimurium, используя методы *in silico*.

МАТЕРИАЛЫ И МЕТОДЫ

Подготовка набора геномных данных

В работе были использованы кодирующие последовательности *cas*-генов штаммов *Salmonella* Enteritidis (297 геномов), Infantis (87 геномов) и Typhimurium (167 геномов), скачанные из базы данных NCBI GenBank. Штаммы выбирались с учётом наличия всех *cas*-генов, необходимых для функционирования CRISPR-Cas системы подкласса I-E (*cas3*, *cse1*, *cse2*, *cas7*, *cas5*, *cas8*, *cas1*, *cas2*).

Для извлечения последовательностей *cas*-генов из полных геномов был использован алгоритм BLASTn с параметром «-outfmt 6 qseq». В качестве затравок были использованы *cas*-гены штамма *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. LT2 (номер в NCBI GenBank AE006468.2). Массивы полученных последовательностей *cas*-генов были собраны в fasta-файлы для дальнейшего анализа. В итоге было получено 8 массивов для каждого из *cas*-генов исследуемых серовариантов. К названию последовательностей генов было добавлено название сероварианта для удобства анализа.

Далее было проведено выравнивание последовательностей *cas*-генов с учётом позиции кодона (первых двух или третьей позиций соответственно) при помощи программы TranslatorX (США) [10], установленной локально. В качестве параметров было выбрано автоматическое транслирование нуклеотидных последовательностей в аминокислотные, в качестве метода выравнивания был поставлен алгоритм MAFFT. Остальные параметры были оставлены по умолчанию.

Полученные выравнивания с учётом позиции кодона были соединены в одну общую кодирующую последовательность из *cas*-генов с учётом их расположения в CRISPR-*cas* локусе: *cas3*, *cse1*, *cse2*, *cas7*, *cas5*, *cas8*, *cas1*, *cas2*. Длинные гэпы были удалены из выравнивания.

Поиск событий рекомбинации

Полученное выравнивание было проанализировано при помощи нескольких методов программы RDP5 (RDP, Geneconv, Chimaera, BootScan, SiScan, 3Seq), которая в настоящее время является одной из наиболее часто используемых [11]. Метод MaxChi не был использован в исследовании, поскольку его улучшенной модификацией

является алгоритм Chimaera. Отличия методов RDP5 заключаются главным образом в расчёте *P*-value и определении границ участка рекомбинации. Каждый из этих алгоритмов адаптирован для сканирования триплетов последовательностей (A, B, C) на наличие сигналов рекомбинации в выравнивании нуклеотидных последовательностей. Из каждого триплета в выравнивании удаляются участки, которые либо идентичны, либо отличаются во всех трёх последовательностях. Потенциальные рекомбинантные участки идентифицируются как области, в которых процент идентичности между последовательностями A и C (донор – рекомбинант) в участке рекомбинации и между B и C (акцептор – рекомбинант) в осталь-

ной части выравнивания значительно выше, чем между последовательностями A и B (донор – акцептор) (рис. 1). Далее проверяется нулевая гипотеза о том, что сходство в потенциальном рекомбинантном участке появилось случайно. Подробно принцип работы каждого из методов описан в руководстве к RDP5, доступном на официальном сайте программы [12].

Все методы программы RDP5 запускались с расслабленными параметрами, позволяющими получить большее число потенциальных событий рекомбинации. Параметры, с которыми был проведён запуск программы, представлены в таблице 1. Алгоритм данного исследования представлен на рисунке 2.

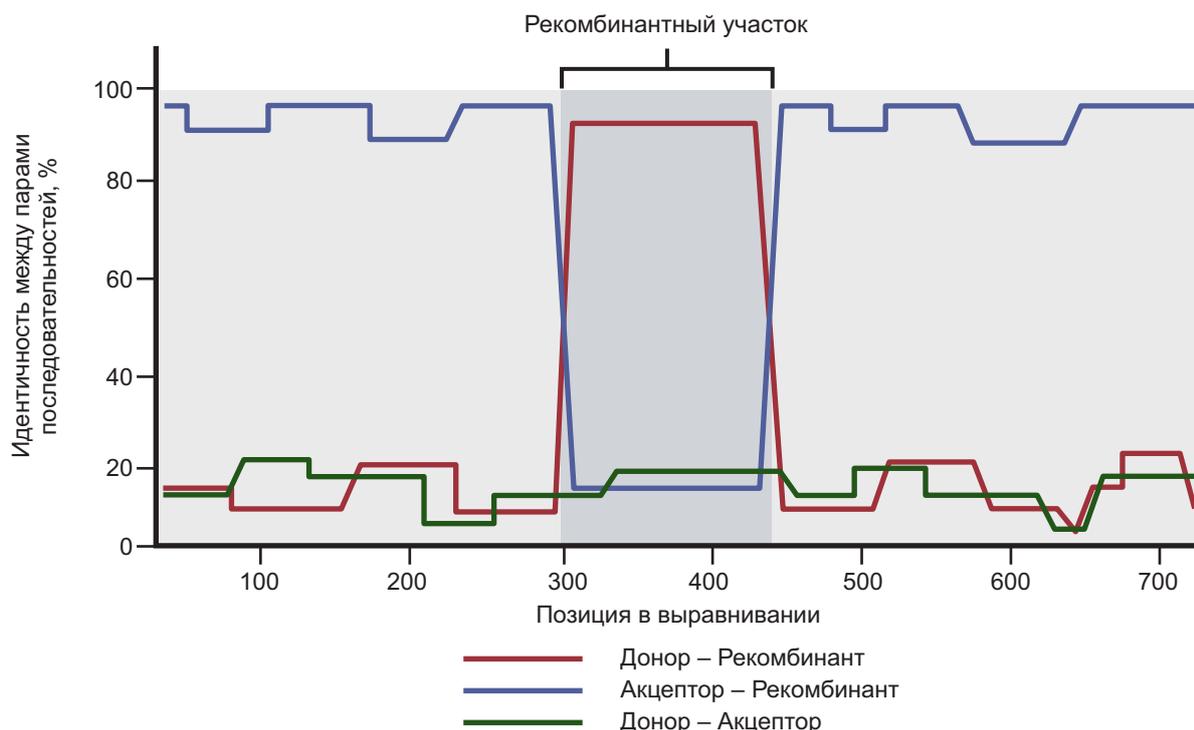


РИС. 1.
График идентичности между попарно выровненными последовательностями донора, акцептора и рекомбинанта

FIG. 1.
The identity plot for pairwise aligned sequences of donor, acceptor and recombinant

ТАБЛИЦА 1
ПАРАМЕТРЫ АЛГОРИТМОВ ПРОГРАММЫ RDP5

TABLE 1
THE PARAMETERS OF THE RDP5 PROGRAM ALGORITHMS

Алгоритмы	Параметры
Общие настройки	– Отключить поправку на множественную проверку гипотез (No multiple comparison correction) – Показывать события, найденные хотя бы одним методом (List all events) – Порог <i>P</i> -value ≤ 0,05
RDP	No reference (по умолчанию)
Geneconv	По умолчанию
Bootscan	Window size = 100, step size = 10
Chimaera	Variable site per window = 40
SiScan	Window size = 100, step size = 10
3seq	По умолчанию

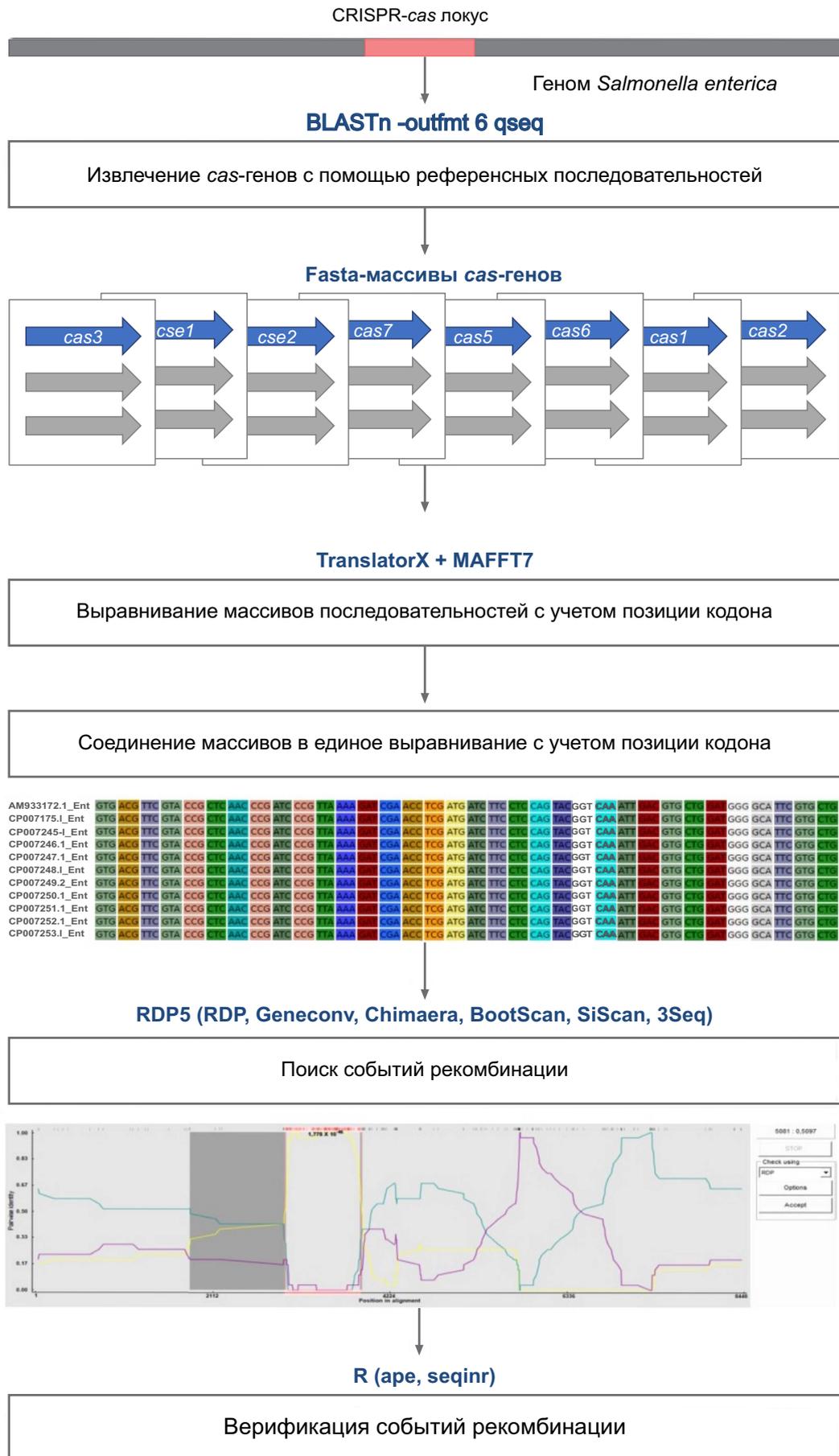


РИС. 2.
Дизайн исследования

FIG. 2.
Research design

Верификация результатов

Несмотря на то, что алгоритмы RDP5 проверяют тот факт, могла ли схожесть в потенциальном рекомбинантном участке возникнуть в результате эволюционных процессов, не связанных с рекомбинацией, каждый из них имеет свои недостатки, что приводит к появлению в выходных данных ложноположительных результатов. Таким образом, необходима дополнительная верификация с использованием иного подхода.

В белок-кодирующих последовательностях третья позиция кодона, как правило, эволюционирует гораздо быстрее, чем первая и вторая позиции. Мутации в третьей позиции часто являются синонимичными, в то время как мутации в первой и второй позициях с большей вероятностью приводят к несинонимичным заменам [13]. Адаптивная конвергентная эволюция чаще всего сопровождается несинонимичными заменами в белок-кодирующих последовательностях. Это связано с тем, что такие мутации меняют аминокислотную последовательность и, соответственно, фенотип, что может быть полезно для адаптации. При этом в третьей позиции кодона, которая склонна к синонимичным заменам, можно ожидать меньшее количество изменений. Рекомбинация сопровождается обменом протяжёнными участками, что делает различия в частотах синонимичных и несинонимичных замен в участке рекомбинации между донором и рекомбинантом менее отличимыми. На основе этих данных был разработан алгоритм верификации событий рекомбинации в белок-кодирующих последовательностях. Алгоритм написан на языке программирования R 4.3.3 с использованием пакетов *ape* 5.7-1 и *seqinr* 4.2-36.

Написанный скрипт использует для анализа следующие входные данные:

- выравнивание белок-кодирующих последовательностей, в котором был осуществлён поиск событий рекомбинации;
- координаты участка рекомбинации в выравнивании;
- номера последовательностей рекомбинанта и донора в выравнивании.

Для верификации события рекомбинации достаточно последовательностей донора и рекомбинанта, которые преобразуются алгоритмом в равные по длине вектора. Затем исходные векторы разбиваются на две пары: первая содержит нуклеотиды и гэпы первых и вторых позиций кодона последовательностей донора и рекомбинанта, а вторая – нуклеотиды и гэпы их третьих позиций. Для верификации события рекомбинации алгоритм обрабатывает эти пары векторов, рассчитывая попарную дистанцию между нуклеотидами первых и вторых позиций и между нуклеотидами третьих позиций в участке рекомбинации. Порог для рассчитанных $P\text{-value}_{1+2}$ (для первых двух позиций) и $P\text{-value}_3$ (для третьих позиций) был равен 0,05.

Для проверки гипотез алгоритм рассчитывает k -мерные дистанции для первых и вторых позиций векторных последовательностей донора и рекомбинанта, за исключением нуклеотидов рекомбинантного участ-

ка, и, аналогично, для векторов, содержащих нуклеотиды третьих позиций. Длина k -меров соответствует количеству нуклеотидов первых и вторых позиций в участке рекомбинации для первой пары векторов или количеству нуклеотидов третьих позиций для второй пары. На основе полученных данных рассчитывается статистика распределения попарных дистанций между k -мерами, равными по длине количеству нуклеотидов первых и вторых позиций (или третьих позиций) в участке рекомбинации. P -value рассчитывается как удвоенное минимальное значение долей, которые соответствуют числу случаев; попарные дистанции в случайных участках оказались больше или меньше наблюдаемой дистанции в рекомбинантном участке. На основе рассчитанных значений P -value делаются выводы о тестируемом участке:

- 1) если $P\text{-value}_{1+2} \leq 0,05$ и $P\text{-value}_3 \leq 0,05$, а также количество замен в первых, вторых и третьих позициях в тестируемом участке меньше, чем количество замен в первых, вторых и третьих позициях в среднем по геному, то тестируемый участок является рекомбинантным;
- 2) если $P\text{-value}_{1+2} \leq 0,05$ и $P\text{-value}_3 > 0,05$, а также количество замен в первых и вторых позициях в тестируемом участке меньше, чем количество замен в первых и вторых позициях в среднем по геному, то сходство между последовательностями донора и рекомбинанта в участке появилось в результате конвергентной адаптивной эволюции.

Остальные случаи считались ложноположительными результатами, либо сходство между последовательностями возникло в результате других эволюционных процессов, которые в данной работе не были исследованы.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

При помощи разработанного алгоритма найдено 7683 потенциальных события гомологичной рекомбинации в области локуса *cas*-генов. Среди них количество верифицированных с помощью разработанного алгоритма событий рекомбинации составило 810 (10,54 %). Средняя длина рекомбинантного участка – 1477 нуклеотидов, 45 (0,59 %) событий были идентифицированы как результаты конвергентной адаптивной эволюции (см. табл. 4), также приводящей к сходству соответствующих участков генома, из-за чего события конвергентной адаптивной эволюции ошибочно определяются программами как события рекомбинации.

Все найденные события рекомбинации можно разделить на две группы: происходящие между штаммами, принадлежащими разным серовариантам, и между штаммами, принадлежащими одному сероварианту (табл. 2). В первой группе рекомбинация в области *cas*-локуса чаще всего детектируется между штаммами *Enteritidis* и *Infantis* и между штаммами *Typhimurium* и *Infantis*. Штаммы серовара *Infantis* чаще, чем другие, являются донорами рекомбинантного участка, а штаммы *Typhimurium* в основном определяются как рекомбинанты. Во второй группе рекомбинация чаще обнару-

живается между серовариантами Typhimurium. Следует отметить то, что не было найдено ни одного события рекомбинации между штаммами Enteritidis, даже среди ложноположительных результатов, несмотря на значительный объём геномных последовательностей данного сероварианта в выборке (297 геномов).

В целом события рекомбинации гораздо чаще детектируются между штаммами, принадлежащими разным серовариантам, чем между штаммами, принадлежащими одному сероварианту, что не согласуется с известными исследованиями, проведёнными *in vitro*. Например, в исследовании X. Didelot и соавт. показано, что рекомбинация в геноме *S. enterica* происходит чаще между штаммами одного серовара. Авторы выдвинули гипотезу о том, что серовары сальмонеллы ограничены конкретными видами-хозяевами [14]. Также было доказано, что шансы на гомологичную рекомбинацию между донором и акцептором снижаются экспоненциально с увеличением генетического расстояния между ними вследствие различий систем репарации ДНК [14, 15].

Причина данного несоответствия заключается в том, что события рекомбинации легче детектируются программами между геномными последовательностями, обладающими большим полиморфизмом. Большинство биоинформатических программ для поиска участков рекомбинации выполняют поиск аномально схожих участков в выравнивании, используя технику скользящего окна, методы распределения замен, статистические тесты, а также филогенетические методы [8–11]. Такие схожие гомологичные участки программа будет считать рекомбинантными, что не всегда верно, поскольку сходству гомологичных участков генома могут способствовать другие эволюционные процессы.

Чувствительность подобных алгоритмов будет снижаться пропорционально увеличению сходства последовательностей. Таким образом, чем более близкородственны последовательности, тем сложнее обнаружить в них участки рекомбинации.

Большая часть найденных участков рекомбинации локализована в гене *cse1*, кодирующем белок, который в комплексе CRISPR-Cas участвует в распознавании у ДНК-мишени последовательности PAM (protospacer adjacent motif – мотив, смежный с протоспейсером), позволяющей отличить бактериальную ДНК от целевой [16]. Рекомбинации в области гена *cse1* могут привести к изменению структуры белка Cse1, что может повлиять на его способность распознавать последовательности PAM. Это может привести как к увеличению, так и к снижению эффективности защиты бактерии от фагов. В некоторых случаях рекомбинации в области данного гена могут быть выгодными, если они приводят к адаптации системы CRISPR-Cas для борьбы с новыми видами фагов. Например, изменение структуры Cse1 может привести к расширению диапазона последовательностей PAM, что позволит системе распознавать большее количество фагов. В случае неудачных рекомбинаций возможна полная потеря функции гена *cse1*, что приведёт к нефункциональности всей CRISPR-Cas системы I-E, поскольку Cse1 необходим для инициации взаимодействия с чужеродной ДНК.

Реже события рекомбинации детектируются в других генах эффекторного модуля (*cse2*, *cas6*, *cas5*, *cas6*, *cas7*). Меньше всего верифицированных участков рекомбинации найдено в гене адаптационного модуля *cas2* (рис. 3). Ген *cas2* является высококонсервативным, имеющим небольшое количество полиморфизмов. Мно-

ТАБЛИЦА 2
КОЛИЧЕСТВО СОБЫТИЙ РЕКОМБИНАЦИИ, НАЙДЕННЫХ В ГЕНОМАХ *SALMONELLA ENTERITIDIS*, *INFANTIS*, *TYPHIMURIUM*

TABLE 2
THE NUMBER OF RECOMBINATION EVENTS FOUND IN THE *SALMONELLA ENTERITIDIS*, *INFANTIS* AND *TYPHIMURIUM* GENOMES

Рекомбинант (акцептор)	Донор	Количество рекомбинаций	Всего событий найдено
События рекомбинации между штаммами разных серовариантов			
Enteritidis	Infantis	283	745
	Typhimurium	56	
Infantis	Enteritidis	5	
	Typhimurium	32	
Typhimurium	Enteritidis	113	
	Infantis	256	
События рекомбинации между штаммами одного сероварианта			
Enteritidis	Enteritidis	0	65
Infantis	Infantis	16	
Typhimurium	Typhimurium	49	

Примечание. Представлены верифицированные события рекомбинации, для которых $P\text{-value}_{1+2} \leq 0,05$ и $P\text{-value}_3 \leq 0,05$.

гие методы поиска рекомбинационных событий полагаются на наличие в последовательностях переменных участков [9, 11], что делает обнаружение в консервативных генах событий рекомбинации более сложным. Кроме того, рекомбинации в консервативных генах обычно устраняются естественным отбором. Если же рекомбинация происходит, то она может не давать явных адаптивных преимуществ.

События конвергентной адаптивной эволюции в основном обнаружены у штаммов сероварианта *Typhimurium* (табл. 3). Конвергентная адаптивная эволюция в большей степени затрагивает гены эффекторного

модуля: *cas5*, *cas6*, *cas7* (табл. 4). В генах *cse1*, *cse2* и *cas1*, *cas2* событий конвергентной эволюции найдено не было.

Некоторые ложноположительные результаты могут являться истинными событиями рекомбинации, которые из-за накопления мутаций не могут быть подтверждены нашим алгоритмом верификации. Следовательно, с помощью данного алгоритма можно верифицировать только события, которые произошли относительно недавно. В некоторых случаях высокая эволюционная консервативность генов объясняет их сходство, поэтому много ложноположительных результатов приходится на область генов адапционного комплекса, особенно на область гена *cas1*

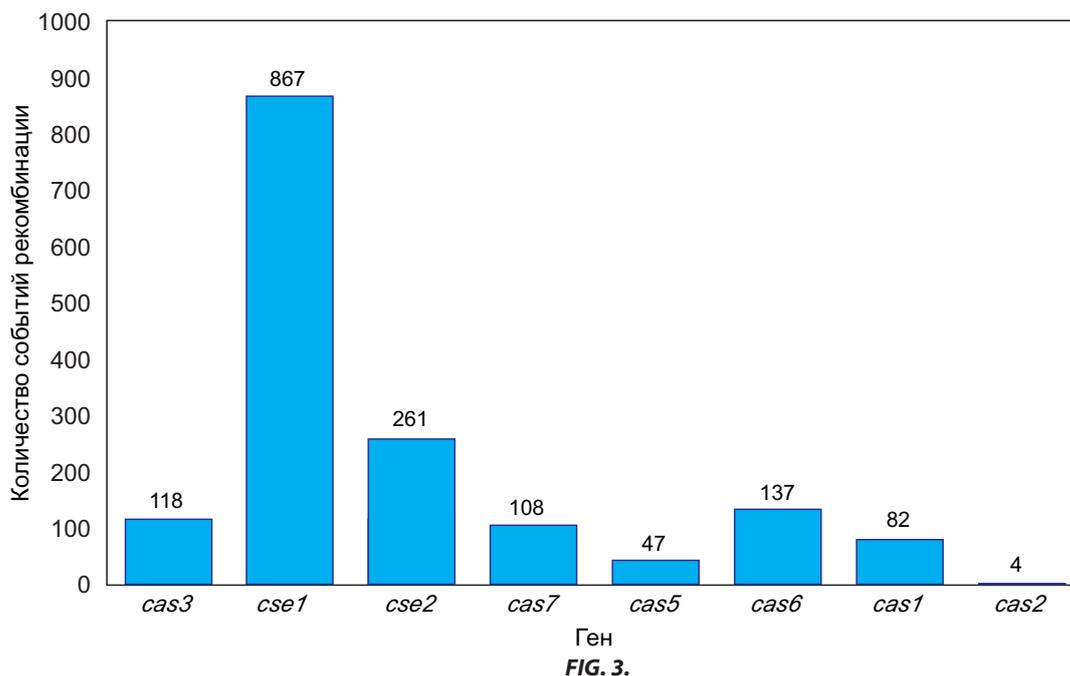


РИС. 3. *Cas*-гены, в которых расположены горячие точки рекомбинации

FIG. 3. *Cas*-genes where recombination hotspots are located

ТАБЛИЦА 3
СОБЫТИЯ КОНВЕРГЕНТНОЙ ЭВОЛЮЦИИ

TABLE 3
CONVERGENT EVOLUTION EVENTS

Штамм 1	Штамм 2	Количество событий	Всего событий для генома
Enteritidis	Enteritidis	0	
Enteritidis	Infantis	0	0
Enteritidis	Typhimurium	0	
Infantis	Infantis	0	
Infantis	Enteritidis	0	22
Infantis	Typhimurium	22	
Typhimurium	Typhimurium	22	
Typhimurium	Enteritidis	1	23
Typhimurium	Infantis	0	
Всего событий – 45			

Примечание. Представлены верифицированные события рекомбинации, для которых $P_{1+2} \leq 0,05$ и $P_3 > 0,05$.

ТАБЛИЦА 4
УЧАСТКИ С КОНВЕРГЕНТНОЙ ЭВОЛЮЦИЕЙ

Начало	Конец	Гены на участке	Количество
7006	7986	<i>cas6</i>	1
5714	6502	<i>cas7, cas5</i>	7
5223	6875	<i>cas7, cas5, cas6</i>	1
5614	6582	<i>cas7, cas5, cas6</i>	36
Всего событий – 45			

Примечание. Представлены верифицированные события рекомбинации, для которых $P_{1+2} \leq 0,05$ и $P_3 > 0,05$.

(1305 ложноположительных событий), который является самым консервативным геном CRISPR-Cas систем [5, 6].

ЗАКЛЮЧЕНИЕ

В проведённом исследовании показано, что гомологичная рекомбинация часто происходит в районе белок-кодирующей части CRISPR-Cas локуса как между штаммами разных серовариантов (Enteritidis, Infantis, Typhimurium), так и между штаммами одного сероварианта. Исключение составил серовариант Enteritidis, между штаммами которого не было идентифицировано ни одного события рекомбинации. События рекомбинации чаще детектируются программами между штаммами, принадлежащими разным серовариантам, чем между штаммами, принадлежащими одному сероварианту, что не согласуется с известными исследованиями, проведёнными *in vitro* и связано со спецификой работы программных алгоритмов поиска. Большое количество ложноположительных результатов обнаружено в области высококонсервативных генов адаптационного модуля. Такие «старые» события рекомбинации программам сложнее обнаружить вследствие высокого накопления мутаций в участке рекомбинации. Найденные горячие точки рекомбинации в области *cas*-генов указывают на участки генома, которые часто подвержены изменениям и, вероятно, влияют на адаптацию сальмонелл к бактериофагам. Это знание может быть важным при разработке новых стратегий борьбы с инфекциями, вызываемыми сальмонеллами.

Финансирование

Исследование выполнено за счёт гранта Российского научного фонда (проект № 23-25-00520).

Конфликт интересов

Авторы данной статьи сообщают об отсутствии конфликта интересов

ЛИТЕРАТУРА / REFERENCES

1. Jajere SM. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity

TABLE 4
CONVERGENT EVOLUTION REGIONS

and antimicrobial resistance including multidrug resistance. *Vet World*. 2019; 12(4): 504-521. doi: 10.14202/vetworld.2019.504-521

2. Семина А.Н. Подбор генов для идентификации *Salmonella Enteritidis*. *Международный научно-исследовательский журнал*. 2018; 10-1(76): 108-110. [Semina AN. Selecting genes for *Salmonella enteritidis* identification. *International Research Journal*. 2018; 10-1(76): 108-110. (In Russ.)]. doi: 10.23670/IRJ.2018.76.10.023

3. Galán JE. *Salmonella* Typhimurium and inflammation: A pathogen-centric affair. *Nat Rev Microbiol*. 2021; 19(11): 716-725. doi: 10.1038/s41579-021-00561-4

4. Рожнова С.Ш., Кулешов К.В., Павлова А.С., Гусева А.Н., Кожаметова Т.А., Акулова Н.К., и др. Гетерогенность изолятов нетифоидных сальмонелл из различных источников выделения в Российской Федерации в 2010–2019 гг. *Эпидемиология и инфекционные болезни*. 2020; 25(1): 26-34. [Rozhnova SS, Kuleshov KV, Pavlova AS, Guseva AN, Kozhakhmetova TA, Akulova NK, et al. Heterogeneity of *Salmonella* isolates obtained from various sources in Russia 2010–2019. *Epidemiology and Infectious Diseases*. 2020; 25(1): 26-34. (In Russ.)]. doi: 10.17816/EID35184

5. Pourcel C, Touchon M, Villeriot N, Vernadet JP, Couvin D, Toffano-Nioche C, et al. CRISPR-Casdb a successor of CRISPRdb containing CRISPR arrays and *cas* genes from complete genome sequences, and tools to download and query lists of repeats and spacers. *Nucleic Acids Res*. 2020; 48(D1): D535-D544. doi: 10.1093/nar/gkz915

6. Makarova KS, Koonin EV. Annotation and classification of CRISPR-Cas systems. *Methods Mol Biol*. 2015; 1311: 47-75. doi: 10.1007/978-1-4939-2687-9_4

7. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nat Rev Microbiol*. 2020; 18(2): 67-83. doi: 10.1038/s41579-019-0299-x

8. Shikov AE, Malovichko YV, Nizhnikov AA, Antonets KS. Current methods for recombination detection in bacteria. *Int J Mol Sci*. 2022; 23(11): 6257. doi: 10.3390/ijms23116257

9. Shikov AE, Savina IA, Nizhnikov AA, Antonets KS. Recombination in bacterial genomes: Evolutionary trends. *Toxins (Basel)*. 2023; 15(9): 568. doi: 10.3390/toxins15090568

10. Abascal F, Zardoya R, Telford MJ. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*. 2010; 38(Web Server Issue): W7-13. doi: 10.1093/nar/gkq291

11. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, et al. RDP5: A computer program for analyzing

recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* 2020; 7(1): veaa087. doi: 10.1093/ve/veaa087

12. Martin DP. RDP5 instruction manual. URL: <http://web.cbio.uct.ac.za/~darren/RDP5Manual.pdf> [date of access: 30.10.2024].

13. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 1994; 11(5): 725-736. doi: 10.1093/oxfordjournals.molbev.a040153

14. Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, et al. Recombination and population structure in *Salmonella enterica*. *PLoS Genet.* 2011; 7(7): e1002191. doi: 10.1371/journal.pgen.1002191

15. Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. *Science.* 2007; 315(5811): 476-480. doi: 10.1126/science.1127573

16. Tay M, Liu S, Yuan YA. Crystal structure of *Thermobifida fusca* Cse1 reveals target DNA binding site. *Protein Sci.* 2015; 24(2): 236-245. doi: 10.1002/pro.2609

Сведения об авторах

Арефьева Надежда Александровна – младший научный сотрудник лаборатории молекулярной эпидемиологии и генетической диагностики, ФГБНУ «Научный центр проблем здоровья семьи и репродукции человека»; лаборант-исследователь лаборатории молекулярной вирусологии и биотехнологии Научно-исследовательского института биомедицинских технологий, ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России; аспирант, ФГБОУ ВО «Иркутский государственный университет», e-mail: arefieva.n4@gmail.com, <https://orcid.org/0000-0003-2222-4518>

Букин Юрий Сергеевич – кандидат биологических наук, старший научный сотрудник, ФГБНУ Лимнологический институт СО РАН, e-mail: bukinyura@mail.ru, <https://orcid.org/0000-0002-4534-3846>

Эрдынеев Сергей Викторович – аспирант кафедры микробиологии, вирусологии и иммунологии, ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России, e-mail: orry230@yandex.ru, <https://orcid.org/0009-0006-7937-1382>

Джиоев Юрий Павлович – кандидат биологических наук, ведущий научный сотрудник, заведующий лабораторией молекулярной вирусологии и биотехнологии Научно-исследовательского института биомедицинских технологий, ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России, e-mail: alanir07@mail.ru, <https://orcid.org/0000-0001-5410-5113>

Мирошниченко Любовь Александровна – кандидат технических наук, старший научный сотрудник лаборатории анализа данных, ФГБНУ Институт математики им. С.Л. Соболева СО РАН, e-mail: luba@math.nsc.ru, <https://orcid.org/0000-0003-0127-5826>

Informationa about the authors

Nadezhda A. Arefieva – Junior Research Officer at the Laboratory of Molecular Epidemiology and Genetic Diagnostics, Scientific Centre for Family Health and Human Reproduction Problems; Research Assistant at the Laboratory of Molecular Virology and Biotechnology, Research Institute of Biomedical Sciences, Irkutsk State Medical University; Postgraduate, Irkutsk State University, e-mail: arefieva.n4@gmail.com, <https://orcid.org/0000-0003-2222-4518>

Yurij S. Bukin – Senior Research Officer, Limnological Institute, Siberian Branch of the Russian Academy of Sciences, e-mail: bukinyura@mail.ru, <https://orcid.org/0000-0002-4534-3846>

Sergey V. Erdyneev – Postgraduate at the Department of Microbiology, Virology and Immunology, Irkutsk State Medical University, e-mail: orry230@yandex.ru, <https://orcid.org/0009-0006-7937-1382>

Yuri P. Dzhioev – Cand. Sc. (Biol.), Leading Research Officer, Head of the Laboratory of Molecular Virology and Biotechnology, Research Institute of Biomedical Sciences, Irkutsk State Medical University, e-mail: alanir07@mail.ru, <https://orcid.org/0000-0001-5410-5113>

Lyubov A. Miroshnichenko – Cand. Sc. (Tech.), Senior Research Officer at the Laboratory of Data Analysis, Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, e-mail: luba@math.nsc.ru, <https://orcid.org/0000-0003-0127-5826>