Математическая биология и биоинформатика
2024. Т. 19. № 2. С. 565–578. doi: 10.17537/2024.19.565

========== **БИОИНФОРМАТИКА** ==========

Анализ профаговой нагрузки и *cas*-генов в геномах *Salmonella enterica* сероваров Enteritidis, Typhimurium и Infantis

Эрдынеев С.В.[1,3], Арефьева Н.А.[1,2,3,4], Джиоев Ю.П.[1], Мирошниченко Л.А.[5]

[1]*Иркутский государственный медицинский университет, Иркутск, Россия*
[2]*Иркутский государственный университет, Иркутск, Россия*
[3]*Иркутский научно-исследовательский противочумный институт Сибири и Дальнего Востока, Иркутск, Россия*
[4]*Научный центр проблем здоровья семьи и репродукции человека, Иркутск, Россия*
[5]*Институт математики имени С.Л. Соболева, Новосибирск, Россия*

Аннотация. *Salmonella enterica* – грамотрицательная, факультативно-анаэробная бактерия, являющаяся возбудителем кишечных инфекций у человека и животных. Вид *S. enterica* включает более 2600 сероваров, среди которых Typhimurium, Enteritidis и Infantis широко распространены на территории Российской Федерации и часто демонстрируют множественную лекарственную устойчивость к антибиотикам. Одним из перспективных подходов к борьбе с такими патогенами является таргетная терапия с использованием литических фагов. Однако механизмы взаимодействия фагов и бактерий на геномном уровне остаются недостаточно изученными. Как и у большинства бактерий в геномах сальмонелл обнаруживаются профаги. Исследование профаговой нагрузки в геномах *S. enterica* может быть важно для разработки таргетной фаготерапии против данного патогена. В данной работе проведён биоинформатический анализ профаговой нагрузки в геномах трёх сероваров *S. enterica* подвид *enterica*: Enteritidis (n=50), Infantis (n=50) и Typhimurium (n=50) с использованием алгоритма DBSCAN-SWA. Выявлено 805 профаговых последовательностей, включая 274 у Enteritidis, 212 у Infantis и 319 у Typhimurium. Все последовательности классифицированы как принадлежащие родам *Salmonella*, *Escherichia*, *Enterobacter*, *Cronobacter*, *Shigella*, *Klebsiella* и *Moraxella*. Дополнительно в геномах всех исследованных штаммов были обнаружены *cas*-гены, относящиеся к CRISPR-Cas системе класса 1, подтипу I-E. Эти штаммы были разделены на две группы: первую группу составили 23 штамма, содержащие мутации в *cas*-генах, вторую – 127 штаммов без мутаций. Установлено, что геномы серовара Typhimurium содержат больше профагов по сравнению с другими сероварами. Анализ также показал, что штаммы с мутациями в *cas*-генах характеризуются повышенной профаговой нагрузкой, что подчёркивает роль CRISPR-Cas систем в регуляции профаговой динамики и адаптации бактериальных популяций. Полученные результаты могут быть использованы для разработки более эффективных подходов к фаготерапии.

Ключевые слова: *Salmonella enterica, S. Typhimurium, S. Enteritidis, S. Infantis, профаговая нагрузка, CRISPR-Cas система, cas-гены, таргетная фаготерапия.*

ВВЕДЕНИЕ

Ежегодно по всему миру регистрируются многочисленные случаи сальмонеллёза, вызванного бактериями вида *Salmonella enterica*, которые включают около 2600 различных сероваров, классифицируемых на основе поверхностных антигенов [1, 2]. Согласно данным

Российского референс-центра по мониторингу за сальмонеллёзами за период 2011–2020 гг., наиболее распространёнными этиологическими агентами заболеваемости у людей являются серовары Enteritidis, Typhimurium и Infantis. Инфекции, вызванные этими сероварами, остаются одной из ключевых причин поражений внутренних органов, особенно пищеварительной системы. Это подчёркивает их значимость как объектов лабораторной диагностики и научных исследований [3].

На уровне ДНК сероварианты *Salmonella enterica* подвида *enterica* очень похожи, обычно на 95-99% идентичны в консервативных областях. Однако сероварианты сальмонелл различаются большими вставками и/или делециями, которые могут иметь длину в 500-600 тысяч нуклеотидных пар. И примерно 60% этих хромосомных вставок представляют собой гены, полученные из мобильных генетических элементов, в первую очередь из умеренных фагов [4]. Умеренные бактериофаги способны проходить как литический, так и лизогенный жизненный цикл. Во время лизогенной фазы бактериофаги способны интегрировать свой геном в геном бактерии-хозяина, образуя профаг. Если бактериальный хозяин испытывает физиологический стресс, вызванным, например, повреждением ДНК или антимикробным лечением, может произойти активация профага, что приводит к переходу профага из лизогенного состояния в литический цикл. В результате фаговые частицы начинают реплицироваться, вызывая лизис бактериальной клетки и высвобождение новых вирусных частиц [5, 6]

В геномах сальмонелл, как и у большинства других бактерий, обнаруживаются профаги – интегрированные вирусные элементы, которые могут составлять до 5% бактериального генома. Как мобильные генетические элементы, профаги играют важную роль в эволюции и адаптации бактерий, так как способны переносить гены, кодирующие токсины, факторы вирулентности и антибиотикорезистентности [7, 8]. Кроме того, профаги обеспечивают механизм горизонтального переноса генов между различными бактериальными клетками посредством псевдолизогении или трансдукции. При псевдолизогении бактериофаг присутствует внутри клетки бактерии, но его геном не интегрируется в бактериальный геном. Трансдукция – это перенос генетического материала от одной бактериальной клетки к другой с помощью бактериофагов. Различают два типа трансдукции: генерализованную и специализированную, которые отличаются механизмом, диапазоном переносимых генов и участием различных типов бактериофагов. Генерализованная трансдукция происходит, когда бактериофаг случайно захватывает фрагмент бактериальной ДНК во время литического цикла вместо своего генетического материала. Такой фаг становится дефектным и способен переносить эту ДНК в другую бактерию. Специализированная трансдукция происходит, когда умеренный бактериофаг переносит только определённые участки бактериальной ДНК, расположенные рядом с сайтом интеграции профага в геноме бактерии. Такое влияние на геном бактерий делает профаги значимым объектом изучения для понимания патогенеза инфекций и разработки новых подходов к диагностике или фаговой терапии [7, 9].

На степень фаговой нагрузки может влиять эффективность работы CRISPR-Cas систем – адаптивных иммунных систем бактерий. Генетический CRISPR-*c*as локус имеет общую структуру для большинства CRISPR-Cas систем и состоит из CRISPR-массива и кластера *cas*-генов, которые разделены лидерной последовательностью. *Cas*-гены кодируют два функциональных белковых модуля: адаптационный и эффекторный. Эффекторный модуль обеспечивают защиту от фагов, путем распознавания и разрезания ДНК-мишени (в некоторых случаях РНК-мишени). Адаптационный модуль обеспечивают приобретение устойчивости к новым фагам, посредством вырезания из ДНК фагов коротких участков ДНК, которые затем будут встроены в CRISPR-массив. Эти короткие участки в составе CRISPR-массива называются спейсерами. Между ними расположены палиндромные повторы примерно такой же длины. Лидерная последовательность содержит промотор, необходимый для транскрипции CRISPR-массива в длинную молекулу предшественника CRISPR-РНК (пре-crРНК). С этой молекулой связываются комплексы Cas-белков в районе

шпилек, образуемых палиндромными повторами, и разрезают ее на короткие направляющие crРНК. Образовавшийся рибонуклеопротеиновый комплекс crРНК и Cas-белков осуществляет поиск, распознавание и разрезание ДНК-/РНК-мишени, а также приобретение новых спейсеров [10].

На основании структуры белковых модулей и по типу разрезаемой нуклеиновой кислоты, системы CRISPR-Cas разделяют на 2 класса, 6 типов, 33 подтипа. Основное отличие между системами класса 1 и класса 2 заключается в структуре эффекторного модуля. У систем класса 1 он состоит из нескольких белков, а у систем 2 класса из 1 белка [11]. Согласно базе данных CRISPRCasdb [12], CRISPR-Cas система *Salmonella enterica* относится к классу 1 подтипу I-E [11] и присутствует практически во всех геномах *S. enterica*, присутствующих в базе данных NCBI GenBank. Генетический локус CRISPR-Cas I-E включает массив CRISPR и гены *cas3, cse1, cse2, cas7, cas5, cas6, cas1, cas2*.

Таким образом, оценка профаговой нагрузки в геномах *S. enterica* сероваринтов Typhimuriun, Enteritidis и Infantis позволит оценить эффективность работы CRISPR-Cas систем, что важно для лучшего понимания процессов антагонизма между бактериями и фагами. Полученные знания могут помочь в разработке таргетной фаговой терапии против данного патогена с использованием литических фагов.

МЕТОДЫ ИССЛЕДОВАНИЯ

Формирование выборки

Из базы данных GenBank было получено 150 геномов *S. enterica* сероваров Enteritidis, Typhimurium и Infantis. Исследуемая выборка включала по 50 геномов каждого сероварианта, отобранных у различных подателей NCBI, чтобы исключить близкородственные и клонированные сальмонеллы с общим источником выделения или происхождения.

Реконструкция филогенетического дерева

Поиск и выравнивание однонуклеотидных полиморфизмов проводили с помощью конвейера PhaME (версия 1.0.2) [13], принимая за референсную последовательность геном *S.* Typhimurium LT2 (GCA_000006945.2). Молекулярно-филогенетический анализ полученного выравнивания проводили с помощью программы IQ-TREE (версия 2.3.4) [14]. Наилучшая модель нуклеотидных замен TVM+F+ASC+G4 была определена с помощью метода максимального правдоподобия (Maximum Likelihood). Надежность филогенетического дерева оценивали с помощью ультрабыстрого бутстрепа с 1000 итерациями [15]. Полученное дерево визуализировали в веб-приложении iTOL (версия 6.7) [16].

Поиск генов CRISPR-Cas систем

Поиск элементов CRISPR-Cas систем в геномах сальмонелл осуществлялся с помощью пользовательской базы данных *cas*-генов и алгоритма BLASTn [17]. Если в геноме сальмонеллы обнаруживались гены CRISPR-Cas системы, то они транслировались в белки. Если белки всех генов были полными, то такая CRISPR-Cas система считалась функциональной.

Определение профаговых областей

Геномы сальмонелл, представленные на уровне контигов и скаффолдов в количестве более шести последовательностей, были собраны до хромосом с помощью инструмента RagTag [18] с опцией «scaffold». Для *S.* Enteritidis и *S.* Infantis использовались референсные геномы GCF_029866625 и GCF_029592045 соответственно. Геномы серовара Typhimurium были доступны на уровне полных сборок. Из собранных хромосом удалили участки, состоящие из символов N, которые обозначали места соединения и упорядочивания

контигов и скаффолдов. Это было сделано, чтобы исключить неизвестные фрагменты последовательностей из дальнейшего анализа.

Аннотацию геномов сальмонелл проводили с помощью Prokka [19]. Для поиска профаговых областей в аннотированных геномах применяли алгоритм DBSCAN-SWA [20]. К алгоритму в функцию «diamond_blastp» был установлен параметр, отвечающий за минимальную идентичность выравнивания, на уровне 60%, чтобы повысить чувствительность обнаружения фаговых белков. Статистическая обработка данных, идентифицированных профагов, была выполнена с помощью языка программирования R версии 4.4.1, применяя пакеты tidyverse и boot [21, 22].

РЕЗУЛЬТАТЫ

Результаты филогенетического анализа

Полногеномное выравнивание содержало 53 844 SNP, которые использовались для построения филогенетического дерева. Анализ SNP-замен позволил выявить четкое разделение штаммов на три крупные клады сальмонелл, соответствующие сероварианта Enteritidis, Infantis и Typhimurium (рис. 1). Такое разделение указывает на значительную генетическую дифференциацию между этими сероварианта, что подтверждает их филогенетическую обособленность.

При детальном анализе филогенетического дерева было обнаружено, что внешние ветви внутри каждой клады сероварианта образуют структуры, напоминающие «гребенку» с разной длиной ветвей (рисунки 2–4). Такая структура дерева указывает на то, что большинство исследованных штаммов сальмонелл не являются прямыми клонами. Это свидетельствует о генетической вариабельности внутри сероварианта и отсутствии доминирующих клональных линий.

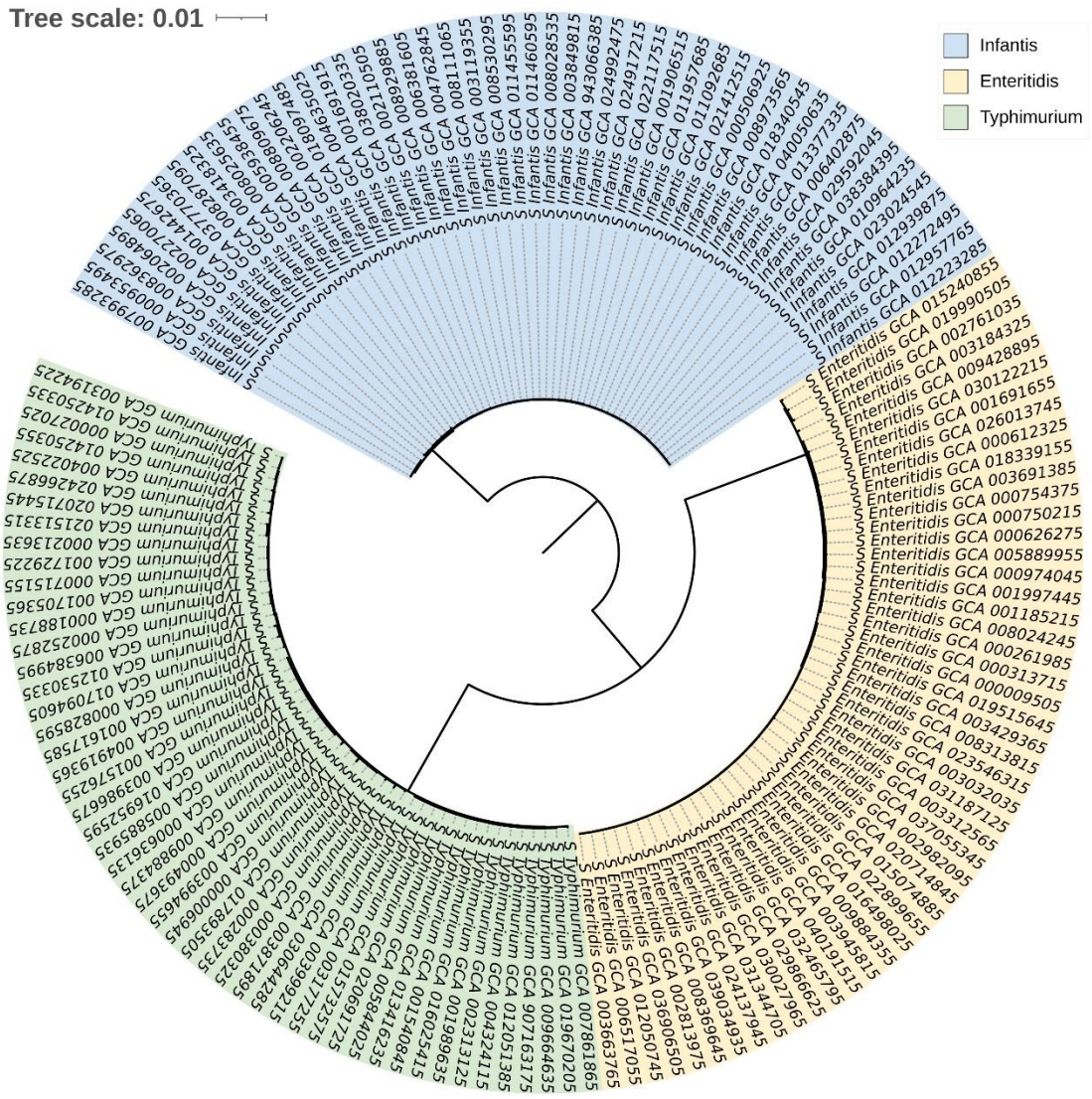

Рис. 1. Филогенетическое дерево *Salmonella enterica* подвид *enterica* сероваров Typhimurium, Enteritidis, Infantis укорененное в средней точке.

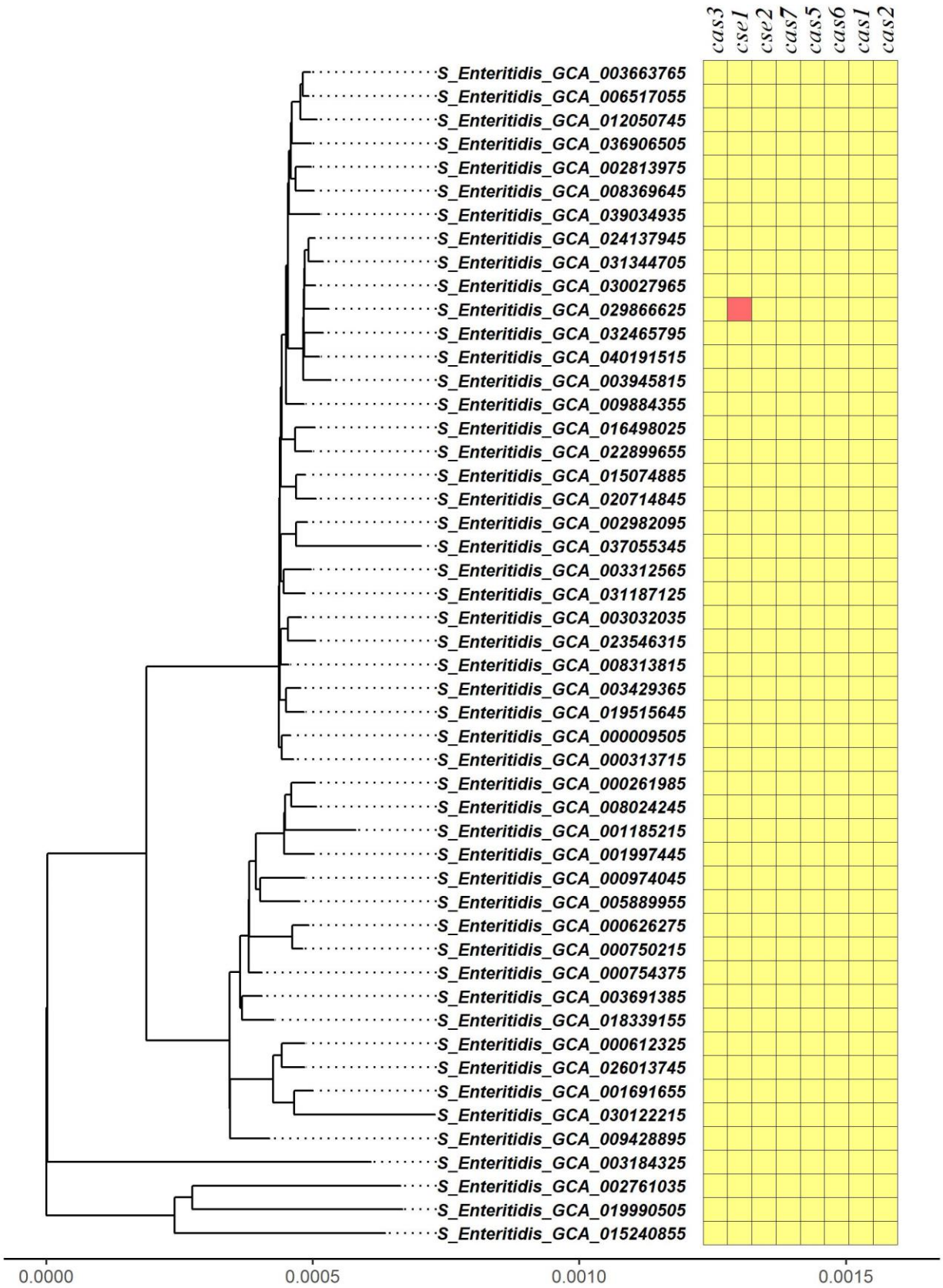

Рис. 2. Фрагмент филогенетического дерева, включающий штаммы серовара Enteritidis. Желтым цветом обозначено наличие полной белковой последовательности соответствующего *cas*-гена, красным – наличие гена с возможным отсутствием синтеза белка вследствие делеции.

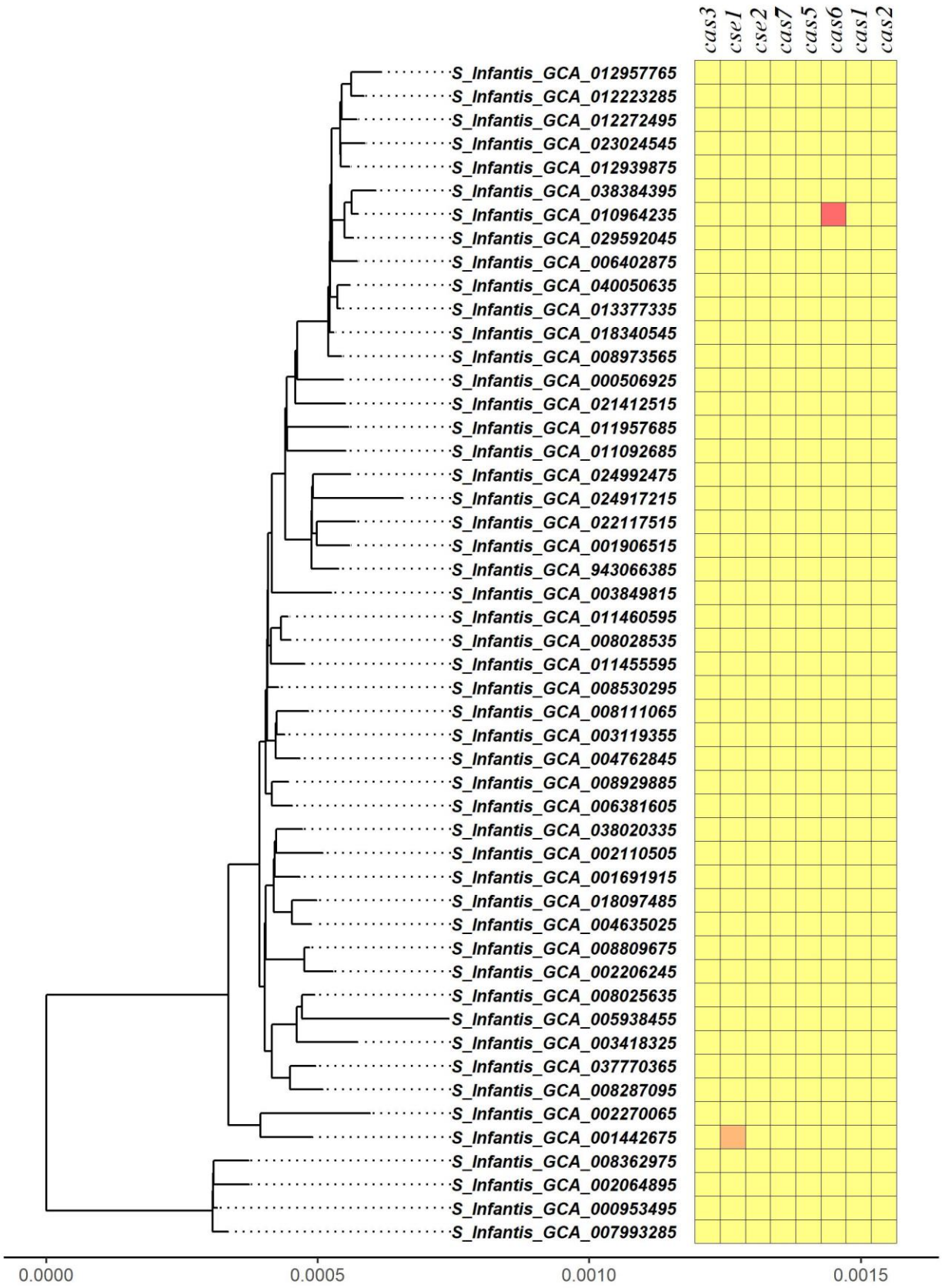

Рис. 3. Фрагмент филогенетического дерева, включающий штаммы серовара Infantis. Желтым обозначено наличие полной белковой последовательности соответствующего *cas*-гена, красным – наличие гена с возможным отсутствием синтеза белка вследствие делеции, оранжевым – наличие гена с возможным отсутствием синтеза белка вследствие вставки.

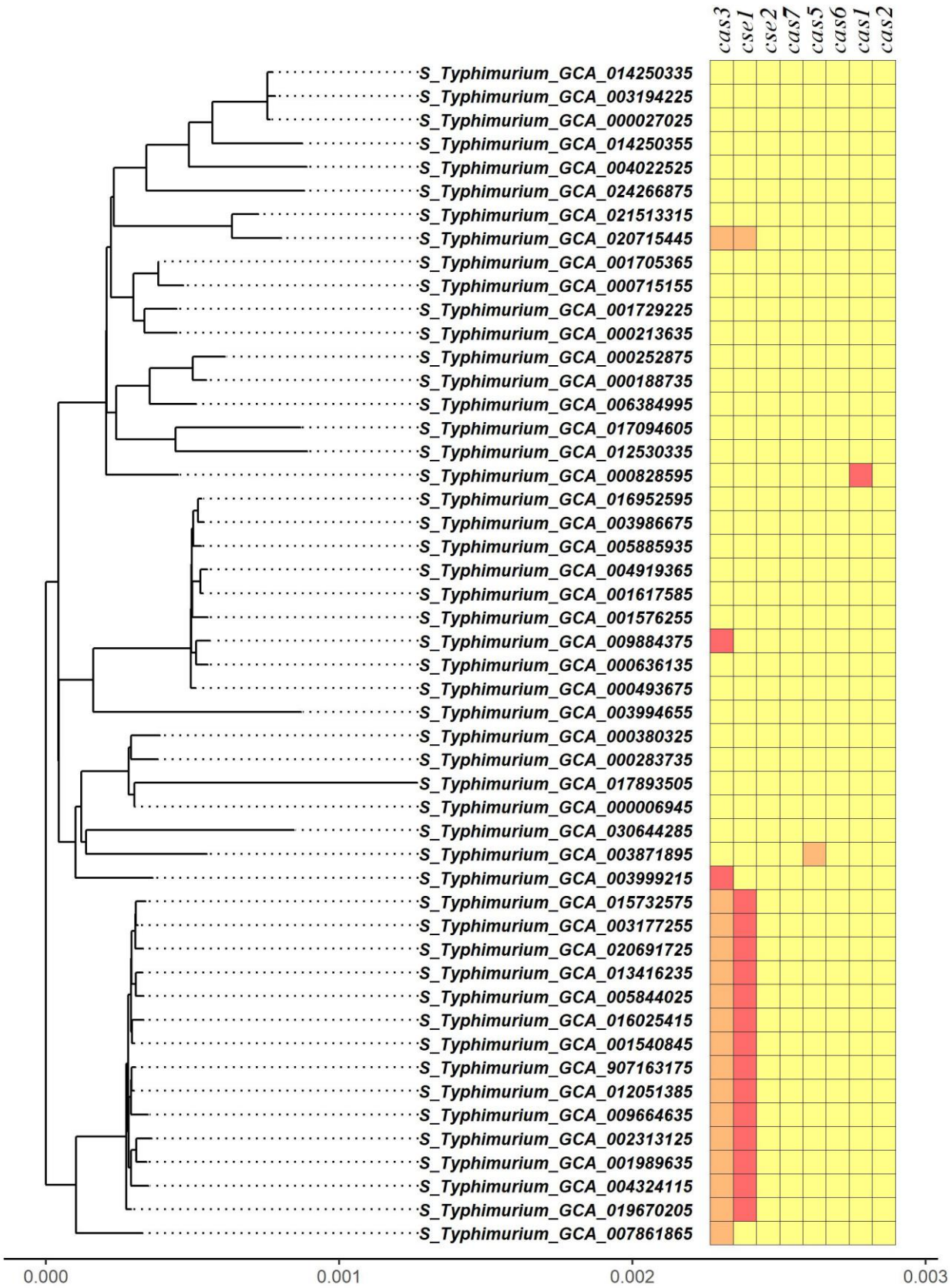

Рис. 4. Фрагмент филогенетического дерева, включающий штаммы серовара Typhimurium. Желтым обозначено наличие полной белковой последовательности соответствующего *cas*-гена, красным – наличие гена с возможным отсутствием синтеза белка вследствие делеции, оранжевым – наличие гена с возможным отсутствием синтеза белка вследствие мутации.

Идентифицированные профаговые последовательности

Всего было обнаружено 805 профаговых последовательностей, распределённых между сероварами Enteritidis, Infantis и Typhimurium. Наибольшее количество профагов выявлено у серовара Typhimurium – 319 последовательностей, тогда как у Enteritidis и Infantis обнаружено 274 и 212 последовательностей соответственно (рис. 5). Также у серовара Typhimurium выявлено больше мутаций в генах *cas3* и *cse1* (рис. 4). Ген *cas3* кодирует эндонуклеазу/хеликазу, осуществляющую разрезание двуцепочечной ДНК фага [22]. *Cse1*, кодирует белок, который в комплексе CRISPR-Cas участвует в распознавании у ДНК-мишени последовательности PAM (Protospacer adjacent motif – мотив, смежный с протоспейсером), которая позволяет отличить бактериальную ДНК от целевой [23].

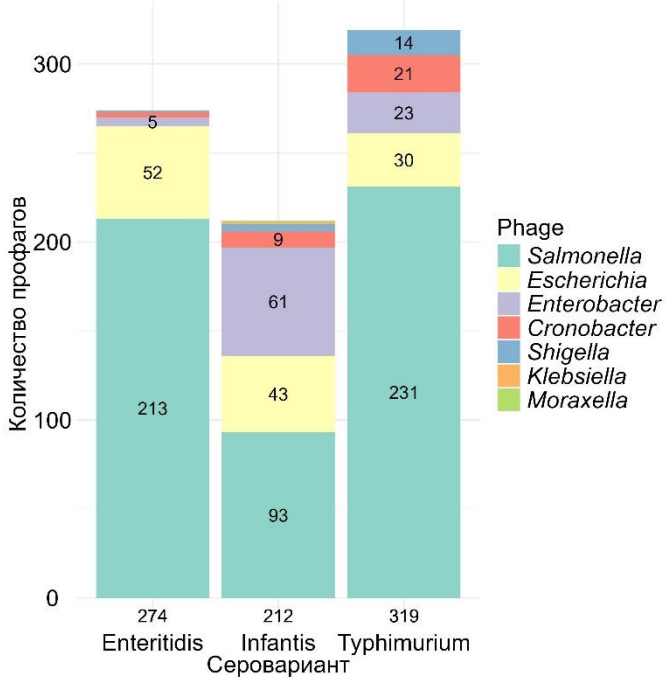

Рис. 5. Идентифицированные профаговые последовательности, узкие полоски соответствуют значениям меньше 5.

Все профаговые последовательности были отнесены к следующим таксономическим группам: *Salmonella*, *Escherichia*, *Enterobacter*, *Cronobacter*, *Shigella*, *Klebsiella* и *Moraxella*. Профаги родов *Klebsiella* и *Moraxella* идентифицированы исключительно у сальмонелл серовара Infantis. При этом доминирующими профагами у всех трёх сероваров стали *Salmonella*, *Escherichia* и *Enterobacter*. Примечательно, что профаги, отнесённые к группе *Enterobacter*, содержали гены, характерные как для *Salmonella* и *Escherichia*, так и для других фагов, что свидетельствует о возможной рекомбинации и участии в горизонтальном переносе генов.

В геномах сальмонелл серовара Enteritidis содержалось от 5 до 9 профаговых последовательностей. Большинство геномов (70 %, или 35 штаммов) имели 5 профагов, что является наиболее частым числом. В группе Infantis количество профаговых последовательностей варьировалось от 1 до 8, при этом чаще всего встречалось 4 профага (32 %), затем 5 (20 %) и 6 (12 %). Для сальмонелл серовара Typhimurium наибольшее количество профагов было зафиксировано в геномах с 6 профагами (34 %), затем следуют штаммы с 5, 7 и 8 профагами (по 18 % и 16 % соответственно). Для более точной оценки числа профаговых последовательностей был проведен бутстреп-анализ. Он показал, что распределение среднего количества профагов значительно варьируется между тремя сероварами *S. enterica*. Группа Typhimurium продемонстрировало самое высокое среднее количество профагов, с 95 % доверительным интервалом от 6.0 до 6.8. Этот результат значительно превышает показатели сероваров Enteritidis и Infantis, чьи доверительные

интервалы составляют от 5,3 до 5,7 и от 3,8 до 4,6 соответственно. Примечательно, что доверительный интервал для сероваров Typhimurium и Infantis шире, чем для Enteritidis, при этом среднее количество профагов у Typhimurium превышает показатели других сероваров. Это указывает на значимые различия в профаговой нагрузке между сероварами сальмонелл, что может отражать их разную эволюционную историю и адаптивные механизмы (рис. 6).

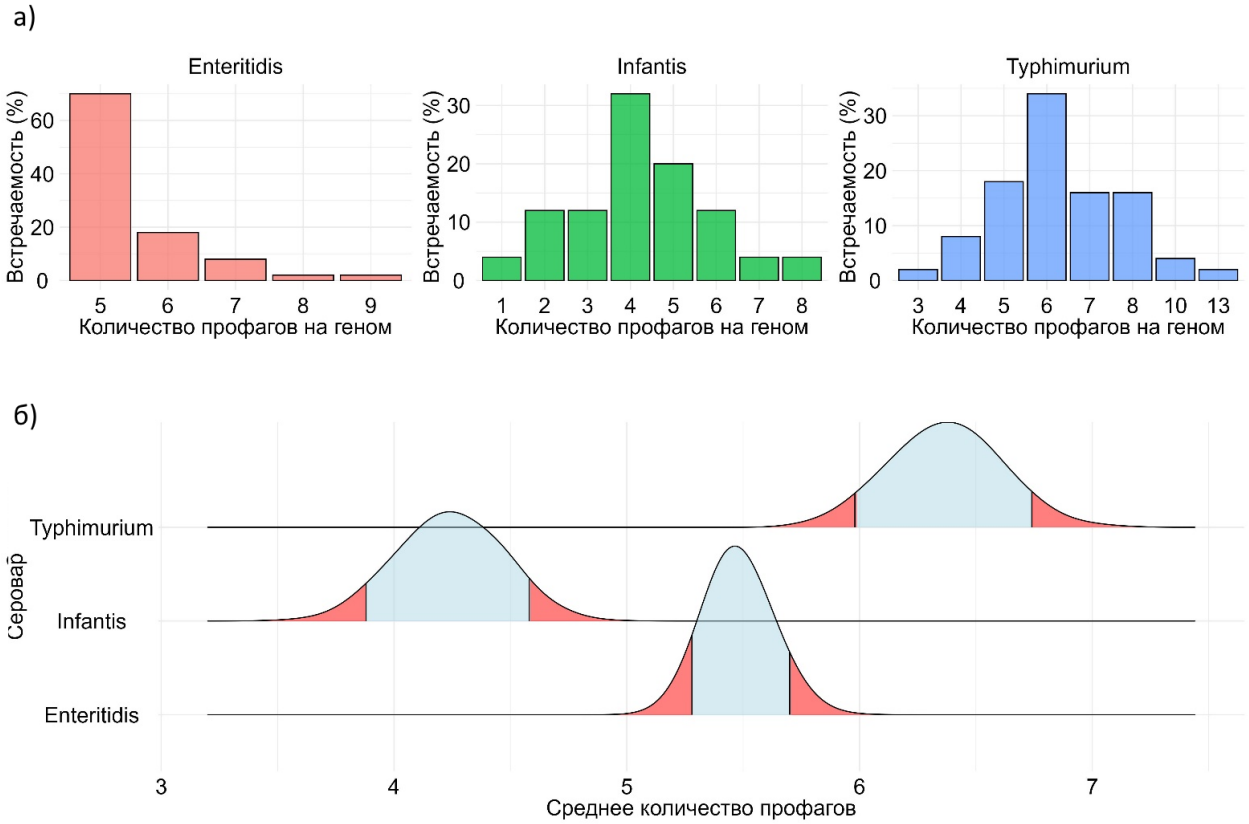

Рис. 6. Распределение среднего количества профагов на геном: а) Количество профагов на геном; б) Диаграмма плотности распределения среднего количества профагов среди сероваров Enteritidis, Infantis и Typhimurium.

Обнаруженные гены CRISPR-Cas систем

На основе анализа геномов *S. enterica* во всех исследованных штаммах были обнаружены *cas*-гены, относящиеся к системе CRISPR-Cas класса 1, подтипу I-E. Однако в ряде случаев выявлены мутации, потенциально влияющие на функциональность этих генов, с различной степенью воздействия на системы CRISPR-Cas у разных сероваров.

Среди сальмонелл серовара Enteritidis у одного штамма (GCA_029866625) выявлена делеция цитозина в позиции 297 гена *cse1* (рис. 2). Эта мутация приводит к нарушению трансляции гена, что делает невозможным синтез функционального белка.

В сальмонеллах серовара Infantis также обнаружены изменения, так у штамма (GCA_010964235) идентифицирована делеция длиной 10 нуклеотидов в гене *cas6* (позиции 62-72), которая, вероятно, нарушает его функцию. У другого штамма из этой группы (GCA_001442675) в гене *cse1* зафиксированы четыре точечных мутации (позиции 292, 294, 297 и 300), а также вставка аденина в позицию 336, что вызвало сдвиг рамки считывания и образование преждевременного стоп-кодона, нарушая способность кодировать белок.

Штаммы сальмонелл серовара Typhimurium также демонстрируют разнообразие мутаций в *cas*-генах. У штамма (GCA_000828595) в гене *cas1* выявлена делеция 20 нуклеотидов (позиции 608–628), что, вероятно, нарушает структуру и функцию белка. У другого штамма (GCA_003871895) обнаружена замена гуанина на аденин в позиции 674 гена *cas5*, что приводит к образованию преждевременного стоп-кодона TAG.

Особое внимание привлекает группа из 14 штаммов Typhimurium (GCA_001540845, GCA_001989635, GCA_002313125, GCA_004324115, GCA_005844025, GCA_012051385, GCA_015732575, GCA_016025415, GCA_019670205, GCA_020691725, GCA_907163175, GCA_003177255, GCA_009664635, GCA_013416235), принадлежащих к одной сублинии филогенетического дерева. В их гене *cas3* выявлены две замены: с цитозина на тимин в позиции 2194 и с цитозина на аденин в позиции 2196. Эти изменения приводят к замене кодона CGC (аргинин) на стоп-кодон TGA, нарушая экспрессию белка. Кроме того, у этих же штаммов в гене *cse1* зафиксирована делеция тимина в 13-й позиции, вызвавшая сдвиг рамки считывания и образование стоп-кодона TAA.

У отдельных штаммов серовара Typhimurium наблюдаются дополнительные изменения. Например, у штамма (GCA_020715445) в гене *cse1* обнаружены мутации, аналогичные тем, что зафиксированы у Infantis (GCA_001442675), включая вставку гуанина в позицию 750, что также приводит к сдвигу рамки считывания и образованию стоп-кодона. У двух других штаммов (GCA_003999215 и GCA_009884375) в гене *cas3* в позиции 156 выявлена делеция аденина, тогда как у штамма (GCA_020715445) в том же гене произошла вставка аденина в позицию 157. Эти мутации вызывают сдвиг рамки считывания и формирование преждевременного стоп-кодона, нарушая экспрессию гена *cas3*.

Учитывая выявленную генетическую неоднородность в анализе *cas*-генов, интересно отметить различия в количестве профаговых областей в геномах сальмонелл в зависимости от мутаций.

Количество профагов в геноме сальмонеллы в зависимости от наличия мутаций в cas-генах

Все исследованные штаммы *S. enterica* были разделены на две группы в зависимости от наличия мутаций в *cas*-генах. Первая группа, включающая 23 штамма, содержала мутации в этих генах, тогда как во вторую группу вошли 127 штаммов без мутаций.

Для выявления влияния мутаций в *cas*-генах на количество профагов в геноме был проведен статистический анализ, включающий проверку нормальности распределения и сравнение двух групп. Тест Шапиро–Уилка продемонстрировал, что распределение числа профагов не соответствует нормальному ($W = 0.92548$, $p < 0.05$). В связи с этим для оценки различий был применен непараметрический критерий Краскела–Уоллиса, который подтвердил статистически значимые различия в медианах числа профагов между группами ($p < 0.05$). Штаммы с мутациями демонстрируют более высокую профаговую нагрузку, что указывает на возможную роль *cas*-генов в регуляции профаговой динамики (рис. 7).

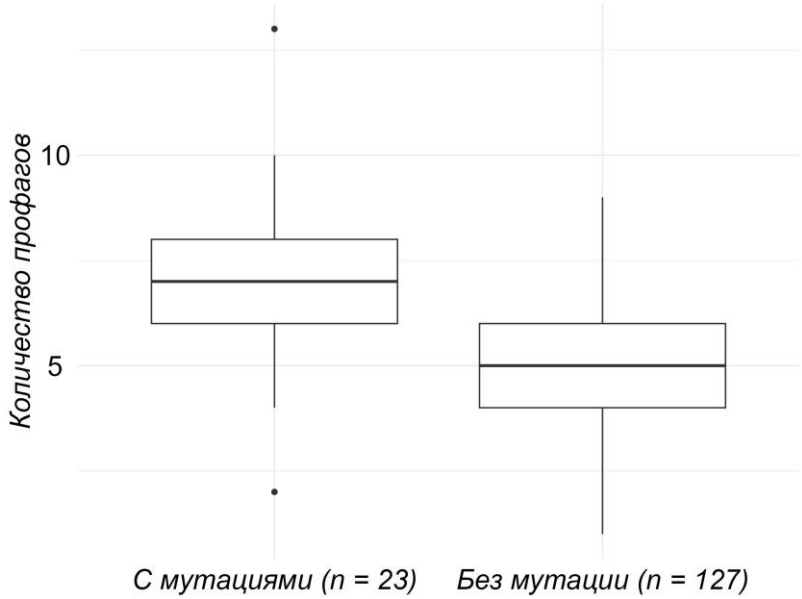

Рис. 7. Сравнение профаговой нагрузки у штаммов *S. enterica* с мутациями в *cas*-генах и без них.

Результаты бутстреп-анализа доверительных интервалов среднего числа профагов показали, что в геномах штаммов с мутациями в *cas*-генах среднее значение варьирует в диапазоне от 6,26 до 7,74 с 95 % вероятностью. Для геномов штаммов без мутаций в *cas*-генах этот интервал составляет от 4,86 до 5,28. Эти данные указывают на более высокую профаговую нагрузку у штаммов с мутациями в *cas*-генах.

ЗАКЛЮЧЕНИЕ

Проведённое исследование позволило выявить значимые различия в профаговой нагрузке в геномах *S. enterica* между сероварами Enteritidis, Infantis и Typhimurium. Серовар Typhimurium показал наибольшее количество профагов и высокую среднюю профаговую нагрузку, что может быть связано с его широкой экологической адаптацией и активным горизонтальным переносом генов. У сероваров Infantis и Enteritidis профаговый состав был менее разнообразным, но также содержал значительное число профагов. Также в геномах сероваров Typhimurium чаще чем в геномах сероваров Enteritidis и Infantis выявляются мутации в генах *cas3* и *cse1*, что может привести к нефункциональности всей CRISPR-Cas системы. Результаты данной работы свидетельствуют о наличии взаимосвязи между состоянием *cas*-генов и профаговой нагрузкой генома. Это подчеркивает важность дальнейших исследований для выяснения биологических механизмов данной зависимости, включая изучение роли CRISPR-Cas систем в регуляции профаговой динамики и адаптивных стратегий бактериальных популяций. Полученные данные могут помочь в разработке таргетной фаговой терапии против инфекций, вызываемых *Salmonella enterica*.

Исследование выполнено за счет гранта Российского научного фонда № 23-25-00520, https://rscf.ru/project/23-25-00520/.

СПИСОК ЛИТЕРАТУРЫ

1. Jajere S.M. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Vet World*. 2019. V. 12. № 4. P. 504-521. doi: 10.14202/vetworld.2019.504-521
2. El-Saadony M.T., Salem H.M., El-Tahan A.M., Abd El-Mageed T.A., Soliman S.M., Khafaga A.F., Swelum A.A., Ahmed A.E., Alshammari F.A., Abd El-Hack M.E. The control of poultry salmonellosis using organic agents: an updated overview. *Poult Sci*. 2022. V. 101. № 4. P. 101716. doi: 10.1016/j.psj.2022.101716
3. Rozhnova S.Sh., Kuleshov K.V., Pavlova A.S., Guseva A.N., Kozhakhmetova T.A., Akulova N.K., Podkolzin A.T. Heterogeneity of *Salmonella* isolates obtained from various sources in Russia 2010–2019. *Epidemiology and infectious diseases.* 2020. V. 25. № 1. P. 26-34. (In Russ)]. doi: 10.17816/EID35184
4. Edwards R.A., Olsen G.J., Maloy S.R. Comparative genomics of closely related salmonellae. *Trends Microbiol.* 2002. V. 10. № 2. P. 94-99. doi: 10.1016/S0966-842X(01)02293-4
5. Owen S.V., Wenner N., Dulberger C.L., Rodwell E.V., Bowers-Barnard A., Quinones-Olvera N., Rigden D.J., Rubin E.J., Garner E.C., Baym M., Hinton J.C.D. Prophages encode phage-defense systems with cognate self-immunity. *Cell Host Microbe*. 2021. V. 29. № 11. P. 1620-1633.e8. doi: 10.1016/j.chom.2021.09.002
6. Wahl A., Battesti A., Ansaldi M. Prophages in *Salmonella enterica*: a driving force in reshaping the genome and physiology of their bacterial host? *Mol Microbiol*. 2019. V. 111. № 2. P. 303-316. doi: 10.1111/mmi.14167
7. Trofeit L., Sattler E., Künz J., Hilbert F. *Salmonella* Prophages, Their Propagation, Host Specificity and Antimicrobial Resistance Gene Transduction. *Antibiotics (Basel)*. 2023. V. 12. № 3. P. 595. doi: 10.3390/antibiotics12030595
8. Goh S. Phage Transduction. *Methods Mol Biol*. 2016. V. 1476. P. 177-185. doi: 10.1007/978-1-4939-6361-4_13

9. Hille F., Richter H., Wong S.P., Bratovič M., Ressel S., Charpentier E. The Biology of CRISPR-Cas: Backward and Forward. *Cell*. 2018. V. 172. № 6. P. 1239-1259. doi: 10.1016/j.cell.2017.11.032
10. Makarova K.S., Wolf Y.I., Iranzo J., Shmakov S.A., Alkhnbashi O.S., Brouns S.J.J., Charpentier E., Cheng D., Haft D.H., Horvath P., et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 2020. V. 18. № 2. P. 67-83. doi: 10.1038/s41579-019-0299-x
11. Pourcel C., Touchon M., Villeriot N., Vernadet J.P., Couvin D., Toffano-Nioche C., Vergnaud G. CRISPRCasdb 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 V. 48. № D1. P. D535-D544. doi: 10.1093/nar/gkz915
12. Shakya M., Ahmed S.A., Davenport K.W., Flynn M.C., Lo C., Chain PS.G. Standardized phylogenetic and molecular evolutionary analysis applied to species across the microbial tree of life. *Sci Rep.* 2020. V. 10. № 1723. doi: 10.1038/s41598-020-58356-1
13. Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Haeseler A., Lanfear R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*. 2020. V. 37. № 5. P. 1530-1534. doi: 10.1093/molbev/msaa015
14. Hoang D.T., Chernomor O., Haeseler A., Minh B.Q., Vinh L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*. 2018. V. 35. № 2. P. 518-522. doi: 10.1093/molbev/msx28
15. Letunic I., Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*, 2024. V. 52. № W1. P. W78-W82. doi: 10.1093/nar/gkae268
16. BLAST® Command Line Applications User Manual. Bethesda: NCBI. 2008. URL: https://www.ncbi.nlm.nih.gov/books/NBK279690/ (accessed 24.11.2024)
17. Alonge M., Lebeigle L., Kirsche M., Jenike K., Ou S., Aganezov S., Wang X., Lippman Z.B., Schatz M.C., Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol*. 2022. V. 23. № 258. doi: 10.1186/s13059-022-02823-7
18. Seemann T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics*. 2014. V. 30. № 14. P. 2068-2069. doi: 10.1093/bioinformatics/btu153
19. Gan R., Zhou F., Si Y., Yang H., Chen C., Ren C., Wu J., Zhang F. DBSCAN-SWA: An Integrated Tool for Rapid Prophage Detection and Annotation. *Front. Genet*. 2022. V. 13. P. 885048. doi: 10.3389/fgene.2022.885048
20. Canty A., Ripley B.D., Brazzale A.R. boot: Bootstrap R (S-Plus) Functions. 2024. URL: https://cran.r-project.org/web/packages/boot/index.html (accessed 24.11.2024)
21. He L, St John James M, Radovcic M, Ivancic-Bace I, Bolt EL. Cas3 Protein-A Review of a Multi-Tasking Machine. Genes (Basel). 2020 Feb 18;11(2):208. doi: 10.3390/genes11020208
22. Tay M, Liu S, Yuan YA. Crystal structure of *Thermobifida fusca* Cse1 reveals target DNA binding site. *Protein Sc*i. 2015. V. 24. № 2. P. 236-245. doi: 10.1002/pro.2609

Рукопись поступила в редакцию 06.11.2024, переработанный вариант поступил 10.12.2024.
Дата опубликования 29.12.2024.

Analysis of Prophage Load and *cas* Genes in the Genomes of *Salmonella enterica* Serovars Enteritidis, Typhimurium, and Infantis

Erdyneev S.V.[1,3], Arefieva N.A.[1,2,3,4], Dzhioev Yu.P.[1], Miroshnichenko L.A.[5]

[1]*Irkutsk State Medical University, Irkutsk, Russia*
[2]*Irkutsk State University, Irkutsk, Russia*
[3]*Irkutsk Antiplague Research Institute of Siberia and Far East, Irkutsk, Russia*
[4]*Scientific Centre for Family Health and Human Reproduction Problems, Irkutsk, Russia*
[5]*The Sobolev Institute of Mathematics, Novosibirsk, Russia*

Abstract. *Salmonella enterica* is a Gram-negative, facultatively anaerobic bacterium that causes intestinal infections in humans and animals. The species *S. enterica* includes over 2,600 serovars, among which Typhimurium, Enteritidis, and Infantis are widely distributed in the Russian Federation and often exhibit multidrug resistance to antibiotics. One promising approach to combatting such pathogens is targeted therapy using lytic phages. However, the genomic mechanisms underlying phage-bacterium interactions remain insufficiently understood. Like most bacteria, *Salmonella* genomes harbor prophages. Investigating prophage abundance in *S. enterica* genomes may be crucial for developing targeted phage therapy against this pathogen. This study conducted a bioinformatic analysis of prophage content in the genomes of three *S. enterica* subsp. *enterica* serovars: Enteritidis (n=50), Infantis (n=50), and Typhimurium (n=50) using the DBSCAN-SWA algorithm. A total of 805 prophage sequences were identified, including 274 in Enteritidis, 212 in Infantis, and 319 in Typhimurium. These sequences were classified into genera including *Salmonella, Escherichia, Enterobacter, Cronobacter, Shigella, Klebsiella*, and *Moraxella*. Additionally, *cas* genes associated with class 1, subtype I-E CRISPR-Cas systems were detected in all analyzed genomes. The strains were divided into two groups: 23 strains with mutations in *cas* genes and 127 strains without mutations. It was found that Typhimurium genomes contain more prophages compared to the other serovars. Furthermore, strains with mutations in *cas* genes exhibited higher prophage abundance, highlighting the role of CRISPR-Cas systems in regulating prophage dynamics and bacterial population adaptation. These findings may contribute to the development of more effective phage therapy strategies.

Key words: *Salmonella enterica, S. Typhimurium, S. Enteritidis, S. Infantis, prophage load, CRISPR-Cas system, cas-genes, targeted phage therapy.*