

doi 10.18699/vjgb-26-10

Развитие метода филогенетического футпринтинга для распознавания сайтов связывания транскрипционных факторов на основе использования bootstrap-испытаний для анализа больших бактериальных геномных данных

А.М. Мухин ^{1,2,3} , Т.М. Хлебодарова^{1,2}, Д.Ю. Ощепков ^{1,2}¹ Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия² Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия³ Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия mukhin@bionet.nsc.ru

Аннотация. Активное развитие технологий высокопроизводительного секвенирования привело к взрывообразному накоплению высококачественных данных по последовательностям бактериальных геномов – их число приближается к трем миллионам, и дальнейший рост продолжается. Это, в свою очередь, дает дополнительный импульс развитию технологий для более эффективной их аннотации аналитическими методами с прицелом на применение таких больших геномных данных, а также получение нового качества аннотаций. Одним из таких аналитических подходов стал метод филогенетического футпринтинга, направленный на выявление мотивов, соответствующих сайтам связывания транскрипционных факторов в промоторных областях бактериальных геномов путем сравнения соответствующих выборок регуляторных последовательностей генов-ортологов для родственных организмов. Дальнейшее накопление геномных данных стало стимулом для развития подхода. Так, было обнаружено, что избыточное число последовательностей в выборке, анализируемой с использованием филогенетического футпринтинга, лишь ухудшает точность метода, тогда как включение этапа отбора последовательностей в анализируемую выборку с учетом данных о взаимных эволюционных расстояниях повышает качество работы метода. В настоящей статье нами предложен и реализован следующий шаг развития метода филогенетического футпринтинга, основанный на множественном запуске описанного выше этапа отбора для формирования различающихся подвыборок, последующего запуска конвейера для каждой из подвыборок и на статистическом анализе получаемых результатов множественных запусков конвейера. Предложенный подход, реализованный в методе MotifsOnFly, позволяет повысить устойчивость получаемых результатов распознавания мотивов, выявляемых в многократных запусках конвейера. Эффективность метода MotifsOnFly продемонстрирована на примере анализа хорошо аннотированного промотора гена *OmpW Escherichia coli*.

Ключевые слова: филогенетический футпринтинг; бактериальный геном; сайты связывания транскрипционных факторов; мотивы; бутстреп; Python

Для цитирования: Мухин А.М., Хлебодарова Т.М., Ощепков Д.Ю. Развитие метода филогенетического футпринтинга для распознавания сайтов связывания транскрипционных факторов на основе использования bootstrap-испытаний для анализа больших бактериальных геномных данных. *Вавиловский журнал генетики и селекции*. 2026;30(1):15-26. doi 10.18699/vjgb-26-10

Финансирование. Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (Федеральная научно-техническая программа развития генетических технологий на 2019–2030 годы, соглашение № 075-15-2025-516).

Благодарности. Авторы выражают благодарность Центру коллективного пользования (ЦКП) «Биоинформатика» за предоставление вычислительных ресурсов.

Improvement of a phylogenetic footprinting method for transcription factor binding sites recognition based on the use of bootstrap trials for the analysis of large bacterial genomic data

А.М. Mukhin ^{1,2,3} , Т.М. Khlebodarova^{1,2}, D.Yu. Oshchepkov ^{1,2}¹ Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia² Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia³ Novosibirsk State University, Novosibirsk, Russia mukhin@bionet.nsc.ru

Abstract. The rapid development of high-throughput sequencing technologies has led to an explosive accumulation of high-quality bacterial genome sequence data – their number is approaching three million, and this growth continues. This, in turn, provides additional impetus for the development of technologies for more efficient annotation using analytical methods designed to utilize such large-scale genomic data, as well as for achieving new levels of annotation quality. One such analytical approach is phylogenetic footprinting, which aims to identify motifs corresponding to transcription factor binding sites in the promoter regions of bacterial genomes by comparing corresponding sets of regulatory sequences of orthologous genes in related organisms. The continued accumulation of genomic data has served as the basis for further development of this approach. It has been found that an excessive number of sequences in a set analyzed using phylogenetic footprinting only reduces the accuracy of the method, whereas the inclusion of a sequence selection step in the analyzed set based on data on mutual evolutionary distances improves the method's performance. In this paper, we propose and implement a further step in the development of the phylogenetic footprinting method. This step involves multiple runs of the selection step described above to generate distinct subsamples, subsequent pipeline runs for each subsample, and statistical analysis of the results obtained from multiple pipeline runs. The proposed approach, implemented in the MotifsOnFly method, improves the robustness of motif recognition results obtained from multiple pipeline runs. The effectiveness of the MotifsOnFly method is demonstrated using the analysis of the well-annotated promoter of the *Escherichia coli OmpW* gene.

Key words: phylogenetic footprinting; bacterial genome; transcription factor binding sites; motifs; bootstrap; Python

For citation: Mukhin A.M., Khlebodarova T.M., Oshchepkov D.Yu. Improvement of a phylogenetic footprinting method for transcription factor binding sites recognition based on the use of bootstrap trials for the analysis of large bacterial genomic data. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed.* 2026;30(1):15-26. doi 10.18699/vjgb-26-10

Введение

Стремительное развитие и массовое применение технологий высокопроизводительного секвенирования в молекулярной генетике стимулировало развитие не только биотехнологий, обеспечив возможность массовой сборки бактериальных геномов для их анализа, модификации и дальнейшего использования штаммов бактерий при решении биотехнологических задач, но и методов биоинформатики для все более качественной их аннотации. Аннотация бактериальных геномов сайтами связывания транскрипционных факторов (ССТФ) является одним из важнейших этапов в задачах биотехнологии и микробиологии, поскольку через связывание транскрипционных факторов со своими сайтами в промоторах генов обеспечивается один из основных механизмов регуляции их экспрессии в бактериях (Browning, Busby, 2004).

Стратегия филогенетического футпринтинга, направленная на выявление ССТФ и впервые предложенная в 1988 г. (Tagle et al., 1988; Katara et al., 2012), оказалась плодотворной для поиска мотивов *de novo* с учетом выросшего на тот момент числа секвенированных геномов. Эта стратегия основана на общем принципе, что регуляторные элементы в промоторах, такие как ССТФ, в подавляющем большинстве случаев эволюционируют с меньшей скоростью и более консервативны на уровне последовательности ДНК, чем окружающие их нефункциональные последовательности (Levy et al., 2001). Дальнейший рост числа секвенированных бактериальных геномов делал применение филогенетического футпринтинга все более эффективным (Blanchette, Tompa, 2002), дав стимул к развитию множества алгоритмов, наиболее результативных для имеющихся к тому времени секвенированных геномов, таких как MotifSuite, FootPrinter, AlignACE, BioProspector, CONSENSUS, MDscan, MEME, CUBIC и BoBro (Hertz, Stormo, 1999; Liu X. et al., 2001, 2002; Blanchette, Tompa, 2003; Olman et al., 2003; Chen et al., 2008; Bailey et al., 2009; Li et al., 2011a; Claeys et al., 2012).

Впоследствии было обнаружено, что ограниченное число правильным образом отобранных референтных про-

моторов может быть достаточным для идентификации ССТФ в гене (McCue et al., 2002), поскольку сходные последовательности, имеющие слишком малые эволюционные расстояния, оказываются малоинформативными для филогенетического футпринтинга в силу малого числа накопленных мутаций в окружающих ССТФ нефункциональных последовательностях. Поэтому дальнейшее накопление геномных данных для бактерий позволило улучшить метод филогенетического футпринтинга путем применения этапа отбора последовательностей в анализируемую выборку с учетом данных о взаимных эволюционных расстояниях между ними. Это обеспечивает получение более информативных наборов ортологичных промоторов для распознавания функциональных мотивов. Описанный подход, реализованный в работе (Liu B. et al., 2016), превосходил по качеству распознавания ССТФ популярные методы поиска мотивов, разработанные ранее и перечисленные выше.

Используя в основе подход, предложенный В. Liu с коллегами (2016), мы ранее разработали вычислительный конвейер для поиска ССТФ в бактериальных геномах (Мухин и др., 2024). Он интегрирует исчерпывающий набор необходимых баз данных и алгоритмов для работы с ними, обеспечивая быстрое применение этих ресурсов для аннотации выбранных бактериальных геномов сайтами связывания транскрипционных факторов. Однако в ходе его использования для поиска ССТФ в геноме *Geobacillus icigianus* (Peltek et al., 2024) мы выявили зависимость результата распознавания от порядка, состава и способа отбора подаваемой на вход программы выборки промоторных районов, не позволяющую сделать однозначный вывод о наиболее вероятном расположении сайта связывания и наиболее вероятном транскрипционном факторе, взаимодействующем с этим сайтом.

В настоящей работе на базе ранее разработанного конвейера (Мухин и др., 2024) нами предложена и реализована модификация метода, названного нами MotifsOnFly. В основе этой модификации лежат: а) множественный запуск этапа конвейера для формирования подвыборок

промоторов из полной выборки с учетом данных о взаимных эволюционных расстояниях для формирования различающихся подвыборок; б) запуск конвейера для каждой подвыборки для поиска перепредставленных мотивов *de novo* в каждой из них; в) статистический анализ получаемых результатов. Такой бутстреп-подход (с возвращением) обеспечивает гораздо более полное использование исходных данных и повышает устойчивость и достоверность получаемых результатов за счет статистического анализа результатов распознавания мотивов, выявляемых в многократных запусках. На примере анализа хорошо аннотированного промотора гена *ompW Escherichia coli* показаны преимущества разработанного метода MotifsOnFly, ставшие результатом предложенного развития подхода филогенетического футпринтинга.

Материалы и методы

Вычислительный конвейер. Для распознавания потенциальных сайтов связывания транскрипционных факторов на основе выявления *de novo* мотивов в бактериальных геномах с помощью реализованного в работе вычислительно-компьютерного конвейера использовались аннотации оперонов для 3850 бактериальных геномов, доступные в базе данных (БД) DOOR2 (Mao et al., 2014), последовательности которых были извлечены из БД NCBI (Sayers et al., 2021).

В конвейере использован интегративный подход для реализации метода филогенетического футпринтинга, предложенный ранее В. Liu с коллегами (2016). В ходе выполнения этапов работы конвейера для исследуемого (целевого) гена производится выявление ортологичных генов среди всех генов, имеющихся в базе (3850) аннотированных геномов. Поиск осуществляется путем выявления сходства белковых последовательностей, для этого применяется программный модуль GOST (Li et al., 2011b). Далее для каждого из выявленных генов-ортологов, исходя из известной оперонной структуры соответствующего генома, извлекается его промоторная последовательность, в результате чего формируется полная выборка промоторных последовательностей генов-ортологов исследуемого целевого гена. Оценка расстояний между каждой парой ортологичных промоторных последовательностей, входящих в анализируемую полную выборку, проводится на основании построенного филогенетического дерева с помощью программы ClustalW2 (Larkin et al., 2007). Формирование различающихся подвыборок последовательностей промоторов путем их отбора в зависимости от их взаимных эволюционных расстояний проводили согласно принципам, эффективность которых была показана ранее (Liu B. et al., 2016). Далее, по аналогии с подходом (Liu B. et al., 2016), для выявления мотивов *de novo* в каждой из подвыборок промоторов применяли принцип голосования между результатами, получаемыми набором базовых поисковых систем, в качестве которых были использованы ранее разработанные алгоритмы AlignACE, BioProspector, CONSENSUS, MDscan, MEME, CUBIC и BoBro (Hertz, Stormo, 1999; Liu X. et al., 2001, 2002; Olman et al., 2003; Chen et al., 2008; Bailey et al., 2009; Li et al., 2011a).

Найденные *de novo* мотивы сравнивались с помощью алгоритма Tomtom (Gupta et al., 2007) с известными ССТФ, содержащимися в двух базах данных – SwissRegulon (Pachkov et al., 2013) для *E. coli* и PRODORIC для бактерий (Dudek, Jahn, 2022). На основе рассчитываемой в Tomtom статистической меры сходства определялась выборка ССТФ, наиболее похожая на выявленный *de novo* мотив, таким образом определялся наиболее вероятный ТФ, взаимодействующий с найденным мотивом.

Выбор двух баз данных обусловлен возможностью дополнительного сравнения выявленных *de novo* мотивов в промоторах *E. coli* как с выборками известных ССТФ именно этого организма, представленных в БД SwissRegulon, так и с обобщенными выборками известных ССТФ бактерий, представленных в БД PRODORIC. Результаты множественного запуска конвейера подвергались статистическому анализу и визуализации.

Аналитические скрипты были реализованы на Python 3.12, 3.6 и 2.7 с использованием платформы Anaconda (<https://anaconda.com>) для управления программными окружениями и различными версиями программ в них. Для хранения необходимых данных и структурированной метаинформации о генах, последовательностях и оперонах применялась система управления базами данных PostgreSQL (<https://www.postgresql.org/>), которая была развернута на инфраструктуре Института цитологии и генетики СО РАН силами ЦКП «Биоинформатика».

Визуализация и анализ результатов. Данные по известным ССТФ в геноме *E. coli* были извлечены из БД RegulonDB, содержащей наиболее полный массив знаний о регуляции инициации транскрипции *Escherichia coli* K-12 (Salgado et al., 2024). При расчетах использована аннотация генома NC_000913.3, извлеченная из БД NCBI.

Для визуализации как данных по аннотации генома из NCBI (последовательность генома) и RegulonDB (экспериментально выявленные сайты связывания транскрипционных факторов), так и проекции результатов работы конвейера на целевой промотор применялся модульный геномный браузер JBrowse2 (Diesh et al., 2023), способный визуализировать множество распространенных форматов файлов с геномными данными. Визуализация дерева, отражающего таксономическую принадлежность анализируемых геномов, представленную в NCBI (Sayers et al., 2021), осуществлялась с помощью программного модуля ETE3 (Huerta-Cepas et al., 2016). Для визуализации результатов множественных запусков конвейера и их статистического анализа использован программный пакет Matplotlib (Hunter, 2007).

Результаты

Модификация подхода филогенетического футпринтинга, реализованного в методе MotifsOnFly

В основе реализованной в этой работе модификации подхода лежат: множественный запуск этапа конвейера, обеспечивающего формирование различающихся подвыборок последовательностей промоторов с сохранением принципов отбора, эффективность которых была показана



Рис. 1. Схема конвейерного поиска *de novo* функциональных мотивов, соответствующим сайтам связывания транскрипционных факторов, основанная на подходе филогенетического футпринтинга и реализованная в методе MotifsOnFly.

Фиолетовой рамкой (1) обозначены данные от пользователя, синими цветами – программные инструменты (2, 7) и внешние данные (3), использованные в конвейере. Зеленым цветом (4, 6) показаны промежуточные данные, генерируемые в процессе работы конвейера. Красной рамкой отмечены этапы, разработанные в рамках метода MotifsOnFly (5, 8 и 9). Этап, включающий отбор с формированием подвыборок промоторов исходя из данных о взаимных эволюционных расстояниях между ними (5), позволяет проводить множественный запуск конвейера для поиска мотивов *de novo* (7) на этих подвыборках. Получаемые в результате данные о выявляемых *de novo* мотивах (8) и статистике их распределения в промоторах генов-ортологов для каждой позиции выравнивания анализируемого промотора позволяют получить гистограмму частоты найденных *de novo* мотивов, отображаемую в геномном браузере JBrowse2 (9). Сравнение полученных в каждом запуске *de novo* мотивов с помощью инструмента Tomtom с базами данных известных CCTФ SwissRegulon и PRODORIC (8) и дальнейший анализ статистики для этих сравнений позволяют предположить наиболее вероятные транскрипционные факторы-регуляторы исследуемого промотора.

ранее (Liu B. et al., 2016); накопление и статистический анализ результатов множественного запуска алгоритма на этих подвыборках. Принципиальная схема анализа, реализованного в разработанном конвейере метода MotifsOnFly, представлена на рис. 1. В соответствии с подробно описанными в разделе «Материалы и методы» этапами работы конвейера, исходя из белковой последовательности исследуемого гена, подаваемой на вход конвейера (1), происходит выявление ортологичных генов программным модулем GOST (2) для имеющихся в базе (3) аннотированных геномов. Далее, используя эти же геномные последовательности и разметку оперонной структуры (3), извлекаются все промоторные последовательности генов-ортологов (4).

Базовым для разработанной модификации является этап создания множественных подвыборок (5) с сохранением условий отбора промоторов в анализируемую подвыборку исходя из данных об их взаимных эволюционных расстояниях. Эффективность использования таких условий отбора для реализации метода филогенетического футпринтинга была показана ранее (Liu B. et al., 2016). Сформированная полная выборка промоторов (4) по рассчитанным с помощью ClustalW2 эволюционным расстояниям от целево-

го промотора разделялась на три подгруппы, расстояния которых до целевого промотора составляли от 0.05 до 0.31, от 0.31 до 0.55 и от 0.55 до 0.73 соответственно. Далее на основании размеров этих подгрупп определялось максимальное число подвыборок M , способных вместить все промоторы полной выборки в подвыборки по 12 штук с пропорцией промоторов из соответствующих подгрупп 3:3:6. На последнем этапе формировались $N = 2M$ подвыборок путем случайного отбора, исходя из пропорции 3:3:6 и дополнительных принципов отбора, которые учитывают расстояние между входящими в итоговую выборку последовательностями, не допуская попадания в нее слишком схожих промоторов. Такой бутстреп-подход (с возвращением) обеспечивал максимальное, но не полное вхождение промоторов исходной выборки в анализ, поскольку сохранение принципов отбора, ограничивающих вхождение промоторов с малым информационным вкладом для каждой подвыборки, является приоритетным для эффективности метода (McCue et al., 2002).

Сформированные описанным выше способом подвыборки (5), дополненные целевым промотором, используются для многократного запуска этапа конвейера для поиска мотивов *de novo* (7) (см. раздел «Материалы и методы»).

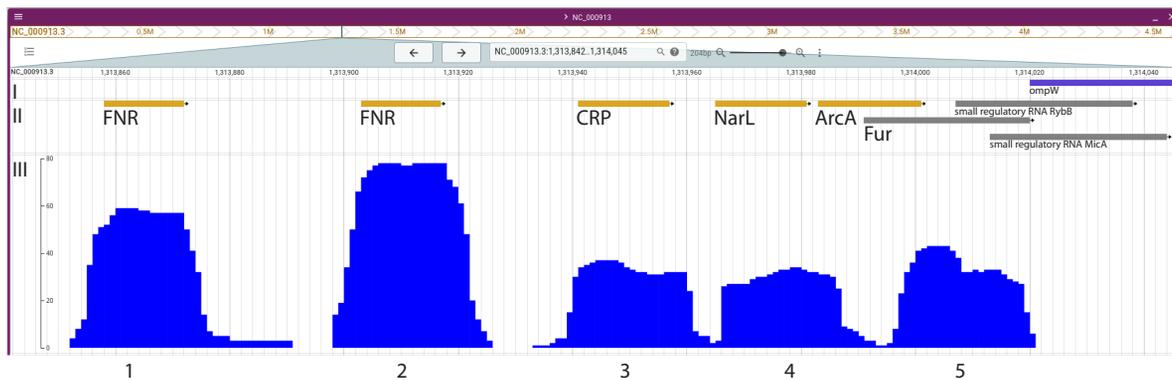


Рис. 2. Скриншот геномного браузера JBrowse2.

I – аннотация генома в области промотора гена *ompW* согласно данным NCBI; II – экспериментально подтвержденная локализация ССТФ из БД RegulonDB; III – результаты работы конвейера: гистограмма частоты найденных *de novo* мотивов в промоторах генов-ортологов для каждой позиции анализируемого промотора. Гистограмма формирует пики, которые пронумерованы в соответствии с их упоминанием в тексте.

В результате таких множественных запусков алгоритма (7) на подвыборках собирается статистика (8), суммирующая, в каких позициях целевого промотора какие *de novo* мотивы были выявлены, а также в каких позициях каких промоторов генов-ортологов они одновременно были обнаружены. Собранная таким образом гистограмма частоты найденных *de novo* мотивов в промоторах генов-ортологов для каждой позиции анализируемого промотора позволяет оценить надежность получаемого *de novo* мотива на позицию (Tompa et al., 2005). Данные гистограммы сохраняются в файл формата BigWig для дальнейшего отображения в геномный браузер JBrowse2 (9).

Анализ промотора гена *ompW* *Escherichia coli* K-12 на основе разработанного метода MotifsOnFly

Наиболее хорошо изученным и полно аннотированным бактериальным геномом на текущий момент остается *E. coli* K-12 (Salgado et al., 2024). Для демонстрации возможностей разработанного нами метода MotifsOnFly, расширяющего потенциал подхода, реализованного В. Liu с соавторами (2016), был выбран ген *ompW* *E. coli*, в контроле экспрессии которого участвуют шесть транскрипционных факторов с пятью известными сайтами связывания. Ген *ompW* кодирует белок наружной мембраны и обладает широким спектром физиологических функций, включая устойчивость бактерий к различным антибиотикам, гербицидам, осмотическому стрессу и поддержку роста бактерий в суровых условиях окружающей среды, таких как гипоксия и повышенная температура (Zhang et al., 2020). Характерной особенностью промотора этого оперона, содержащего только ген *ompW*, является наличие как отдельно расположенных сайтов связывания ТФ, так и групп пересекающихся сайтов разных ТФ. В качестве входных данных конвейера использовались белковая и промоторная последовательности гена, как описано в разделе «Материалы и методы».

На рис. 2 представлены результаты работы конвейера, визуализированные с помощью JBrowse2, включая: аннотацию генома в области гена *ompW* согласно данным

NCBI (I), экспериментально подтвержденную локализацию ССТФ из БД RegulonDB (II) и гистограмму частоты найденных *de novo* мотивов в промоторах генов-ортологов для каждой позиции анализируемого промотора (III). Гистограмма формирует отчетливые пики (см. рис. 2), указывающие на то, что в соответствующих позициях промотора гена *ompW* в значительной доле запусков были найдены мотивы *de novo*. При этом пики 1 и 2 с хорошей точностью совпадают с экспериментально выявленными сайтами связывания транскрипционного фактора FNR, а пик 3 – с известной локализацией сайта связывания транскрипционного фактора CRP. Область промотора, в которой локализован пик 4, пересекается с экспериментально выявленным сайтом связывания ТФ NarL. Что же касается пика 5, то он пересекается с локализацией как сайта связывания ТФ ArcA, так и сайта связывания ТФ Fur, что говорит о сложности исследования регуляторных сигналов вблизи старта транскрипции и требует отдельного рассмотрения. Также легко заметить «хвост» справа от пика 1, свидетельствующий о том, что для этого района в некоторых запусках конвейера были идентифицированы мотивы *de novo*. Последнее подтверждает неустойчивость результатов, получаемых в однократных запусках.

Сравнение выявленных *de novo* мотивов с помощью инструмента Tomtom с базами данных известных ССТФ SwissRegulon и PRODORIC позволяет предположить наиболее вероятные транскрипционные факторы-регуляторы исследуемого промотора. Для каждого такого сравнения рассчитывается значимость E-value сходства выявленного мотива с каждым из известных ССТФ. Нами были рассчитаны статистики значений E-value для всех возможных кандидатов всех пиков (рис. 3). Как видно из диаграмм, полученная статистика позволяет однозначно определить ТФ FNR в качестве наиболее вероятного кандидата на связывание для пиков 1 (см. рис. 3, а, б) и 2 (см. рис. 3, в, г), а ТФ CRP – для пика 3 (см. рис. 3, д, е) при использовании любой из баз данных.

Для пика 4 значимые совпадения обнаружены только с сайтами из БД SwissRegulon (рис. 4) для ТФ NanR и NarL.

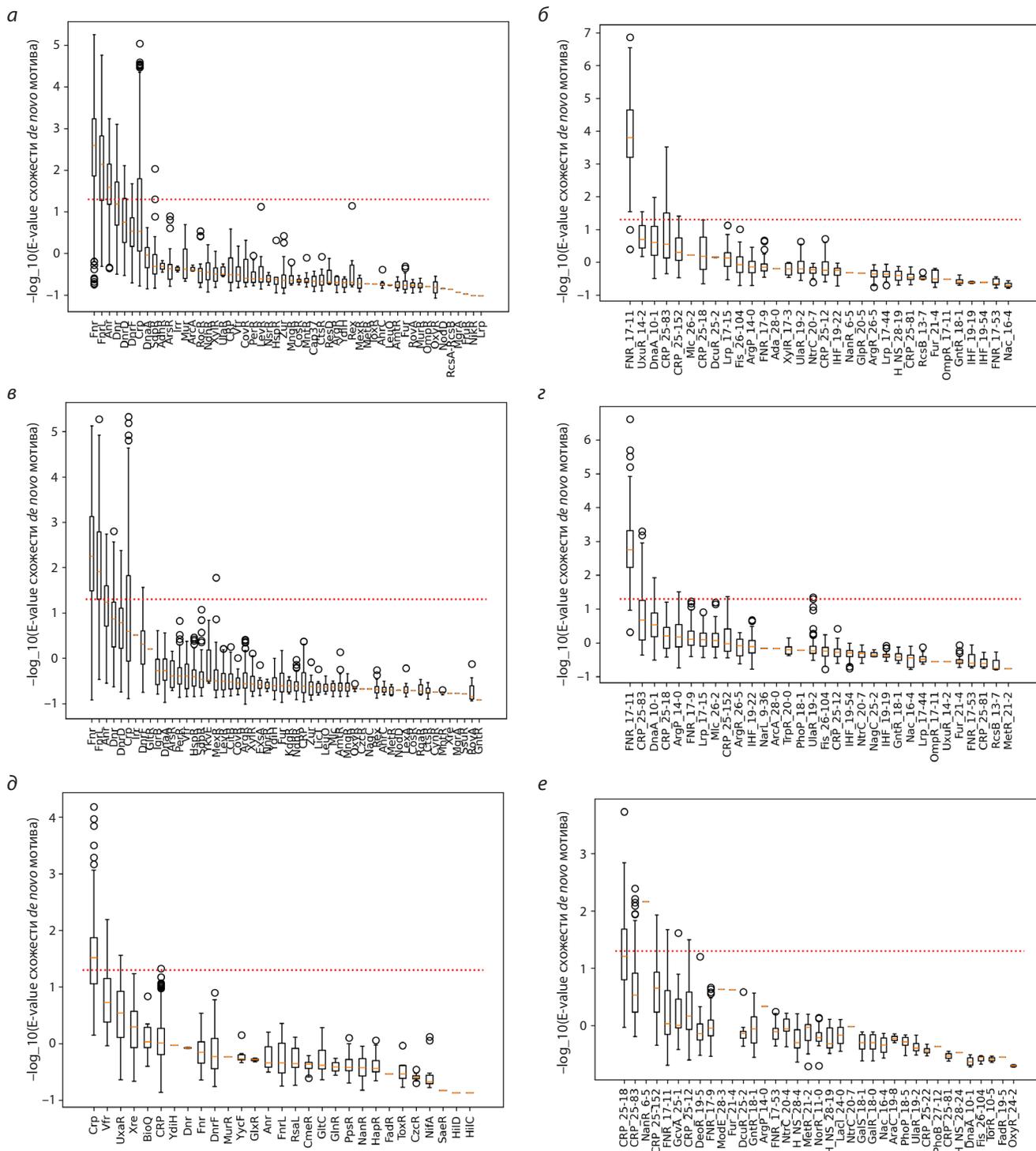


Рис. 3. Диаграммы типа «ящик с усами» для статистики сравнения выявленных мотивов для пика 1 (а, б), пика 2 (в, г), пика 3 (д, е) с базами данных известных ССТФ PRODORIC (а, в, д) и SwissRegulon (б, г, е).

Красным пунктиром обозначен порог значимости 5 %. Для визуализации результатов статистического анализа множественных запусков был применен программный пакет Matplotlib.

Для пика 5 такой расчет не дает значимых совпадений с известными ССТФ из обеих баз данных, что предполагает необходимость дополнительного анализа.

Следует отметить, что для пика 1 в случае однократного запуска конвейера лишь в 72 % случаев при использовании

БД SwissRegulon и в 43 % случаев при использовании БД PRODORIC приоритетным результатом был бы именно FNR. Аналогичные оценки для пика 2 дают 90 и 65 % случаев правильного распознавания FNR, а для пика 3 – 49 и 54 % случаев правильного распознавания CRP со-

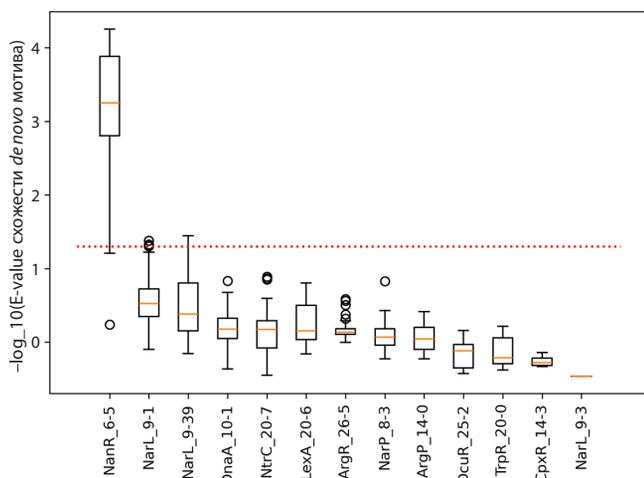


Рис. 4. Диаграмма типа «ящик с усами» для статистики сравнения выявленных мотивов для пика 4 с базой данных известных ССТФ SwissRegulon.

Красным пунктиром обозначен порог значимости 5%. Для визуализации результатов статистического анализа множественных запусков был применен программный пакет Matplotlib.

ответственно. Предложенный нами модифицированный подход позволяет преодолеть эту неустойчивость работы алгоритма филогенетического футпринтинга при однократном запуске конвейера и однозначно расставить приоритеты для полученных результатов в полном согласии с экспериментальными данными.

Метод также дает возможность проследить наличие *de novo* мотивов, соответствующих каждому пику, в каждом из промоторов генов-ортологов и отобразить эти данные на таксономическом дереве геномов, рассмотренных в процессе анализа. В случае, если в промоторе гена-ортолога определенного вида бактерий при выравнивании с целевым промотором в позиции пика также найден сходный мотив, это отображается красной точкой на узле дерева, соответствующем данному виду. На рис. 5 в качестве примера показано наличие сходных мотивов в исследованных промоторах генов-ортологов для позиций, соответствующих пику 1 выравнивания с целевым промотором.

Информация, дающая возможность отследить наличие сходных участков по всем промоторам генов-ортологов (см. рис. 5) в пределах одного пика, также позволяет объединить выравненные участки только тех промоторов генов-ортологов, для которых в любом из запусков конвейера одновременно с целевым промотором был выявлен сходный мотив. Благодаря этому можно получить объединенное выравнивание всех сходных консервативных участков ДНК в пределах одного пика, что обеспечивает максимально полное их представительство в такой выравненной выборке. Границы информативных участков выравнивания, очевидно, определяются шириной каждого из пиков, что и стало критерием для выбора длины объединенного выравнивания и позволило использовать также информацию, содержащуюся на флангах сайтов,

для более качественного сравнения с выборками известных ССТФ. Собранные таким образом объединенные выравнивания для каждого пика были дополнительно проанализированы в программе Tomtom. В результате сравнения полученных объединенных выравниваний для пиков 1–3 с известными ССТФ максимальный уровень достоверности сходства был достигнут при использовании БД SwissRegulon (рис. 6). При этом наиболее вероятными кандидатами для связывания, в полном соответствии с экспериментальными данными, снова оказались ТФ FNR для пиков 1 (см. рис. 6, а) и 2 (см. рис. 6, б) и ТФ CRP для пика 3 (см. рис. 6, в) с уровнем значимости (E-value) 3.13×10^{-8} , 9.17×10^{-8} и 1.80×10^{-4} соответственно. Это согласуется с результатами, приведенными выше для анализа статистики сравнения выявленных мотивов (см. рис. 3), и показывает при этом значительно более высокую достоверность сходства мотивов с соответствующими выборками ССТФ для полученных объединенных выравниваний.

Анализ программой Tomtom объединенного выравнивания для пика 4 дает результаты лишь при использовании БД SwissRegulon – выявлено его значимое сходство с ССТФ NanR (рис. 7, а) и незначимое сходство с выборкой ССТФ NarL (см. рис. 7, б). Других ТФ, имеющих сходные с пиком 4 сайты связывания, выявлено не было.

Объединенное выравнивание для участка пика 5 при анализе программой Tomtom дает результаты лишь при сравнении с БД PRODORIC. При этом ни один ССТФ не показывает значимого сходства с этой выборкой. Тем не менее было выявлено слабое сходство объединенного выравнивания с ССТФ ArcA и Fur (рис. 8), сайты связывания которых имеют значительное пересечение.

Обсуждение

Нами разработан вычислительно-компьютерный конвейер для поиска функциональных мотивов, соответствующих ССТФ, основанный на модификации метода филогенетического футпринтинга, разработанного ранее В. Liu с коллегами (2016). Ключевая идея модификации состоит в множественном запуске эффективной процедуры отбора промоторов, учитывающей их взаимные эволюционные расстояния, для формирования различающихся подвыборок промоторов, последующих запусков конвейера для каждой из таких подвыборок и в статистическом анализе результатов, получаемых при множественных запусках конвейера (см. рис. 1). В предложенном подходе, реализованном в методе MotifsOnFly, применена схема бутстреп-подхода с возвращением, что позволяет использовать максимальное число имеющихся на текущий момент геномных данных, одновременно учитывая статистику результатов на различающихся подвыборках промоторов.

Эффективность работы реализованного конвейера продемонстрирована на примере анализа наиболее полно аннотированного промотора гена *ompW*, кодирующего белок наружной мембраны *Escherichia coli* K-12. Для него известно шесть транскрипционных факторов-регуляторов, у которых экспериментально определены пять сайтов связывания (Salgado et al., 2024). Подробный анализ регуляторных взаимодействий этого промотора был пред-

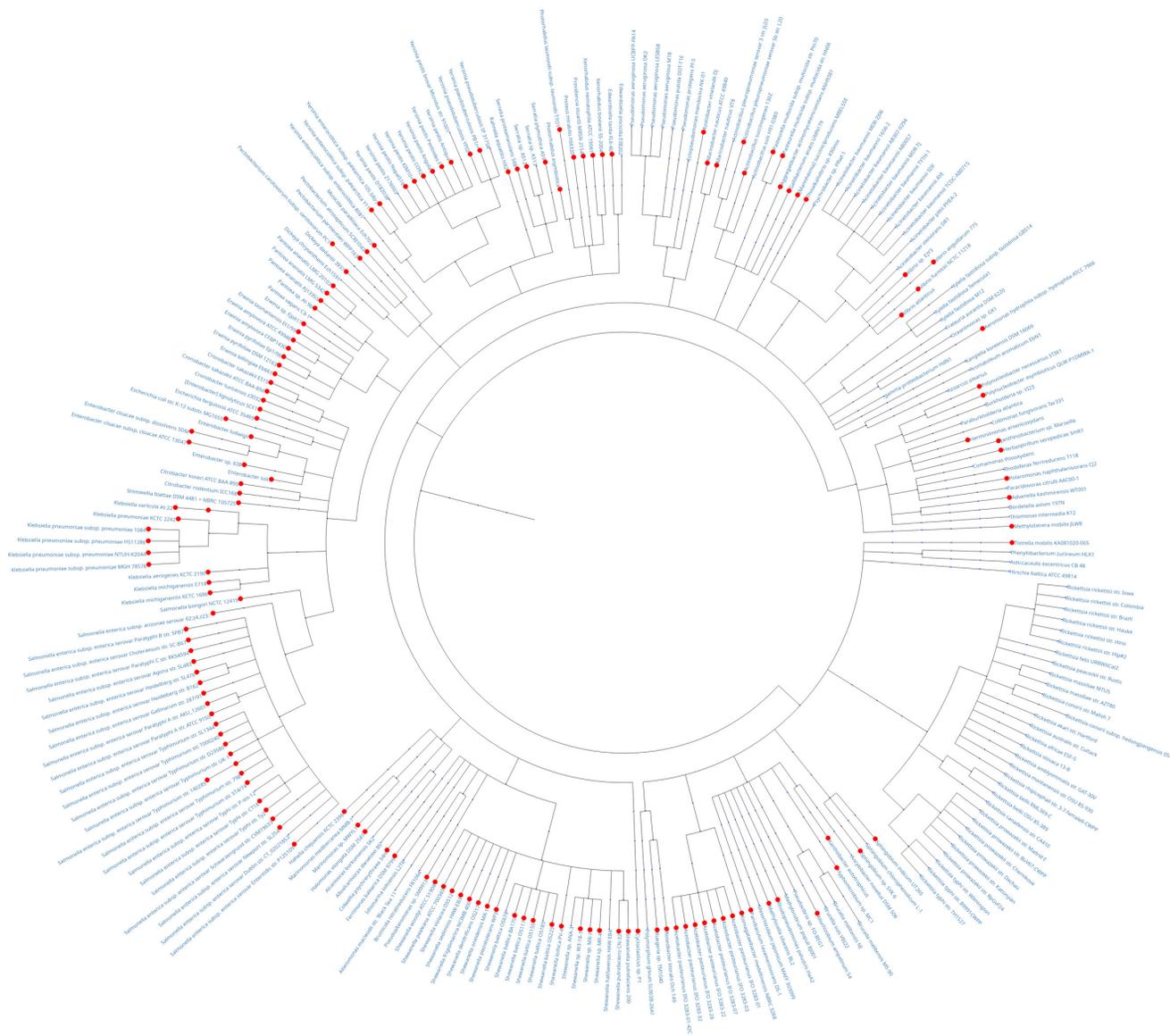


Рис. 5. Отображение наличия сходных мотивов (отмечены красными точками) в исследованных промоторах генов-ортологов разных видов бактерий для позиций, соответствующих пику 1 выравнивания с целевым промотором.

Дерево построено в соответствии с данными по таксономической принадлежности анализируемых геномов, представленными в NCBI. Для визуализации дерева использован пакет ETE3 (Huerta-Cepas et al., 2016).

ставлен ранее (Xiao et al., 2016). Статистика, посчитанная в результате работы конвейера, позволила получить гистограмму частоты найденных *de novo* мотивов в промоторах генов-ортологов для каждой позиции анализируемого промотора, что, в свою очередь, обеспечило надежную идентификацию консервативных участков промотора, обогащенных регуляторными мотивами и формирующих характерные пики (см. рис. 2), которые, как показал приведенный анализ, точно совпадают с экспериментально доказанными. Таким образом, предложенный нами подход наглядно иллюстрирует суть филогенетического футпринтинга, выявляя наиболее консервативные участки выравниваний промоторов генов-ортологов.

Проведенный статистический анализ результатов сравнения *de novo* мотивов, выявленных во множественных запусках, с известными выборками ССТФ из БД SwissRegulon и БД PRODORIC позволил точно определить наиболее вероятные регуляторные факторы, связывающиеся с соответствующими участками анализируемого промотора. Предсказания метода MotifsOnFly полностью совпали с экспериментальными данными для пиков 1, 2, 3, определив ТФ FNR (Constantinidou et al., 2006; Myers et al., 2013; Xiao et al., 2016) как наиболее вероятный белок-регулятор для пиков 1 и 2 и ТФ CRP (Gaston et al., 1990; Ushida, Aiba, 1990; Xiao et al., 2016) – для пика 3 (см. рис. 3). Одновременно этот анализ показал, что не

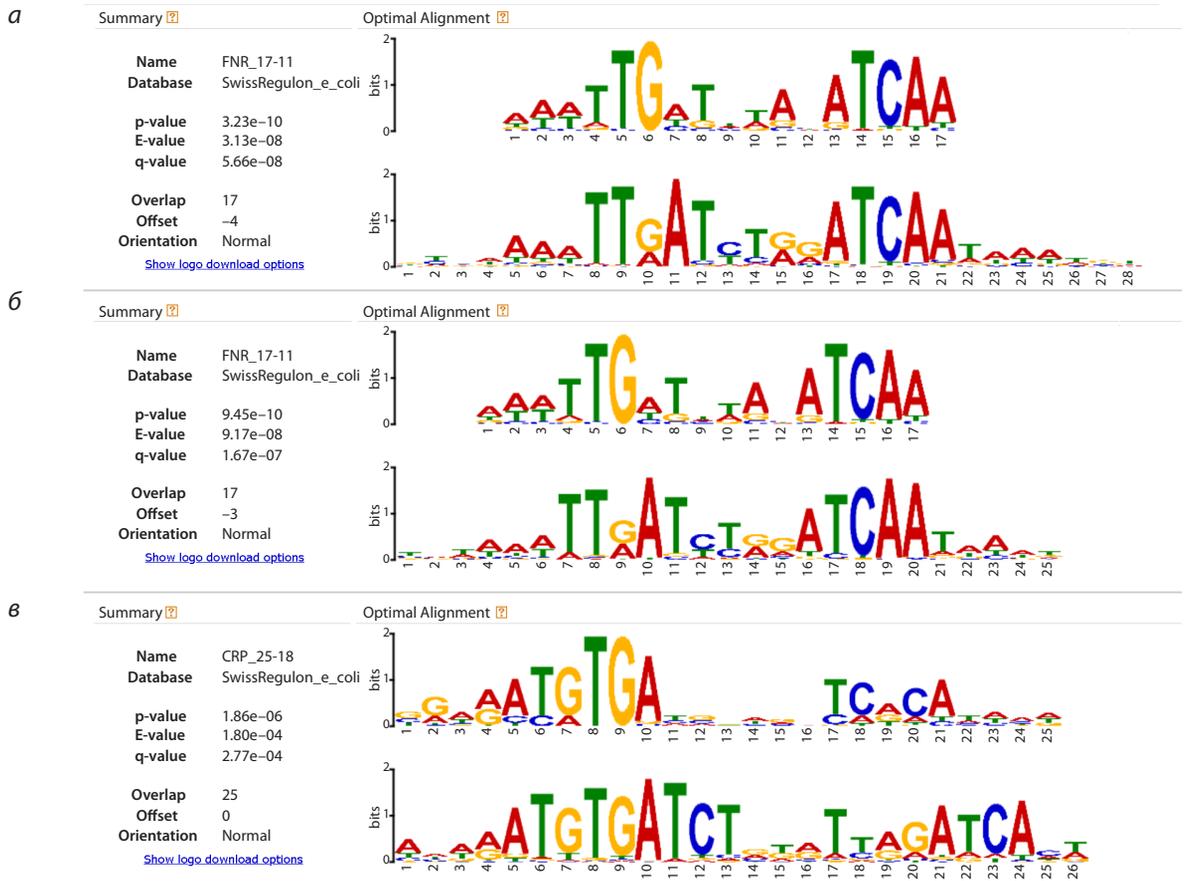


Рис. 6. Результаты сравнения выявленных объединенных выравниваний участков пиков 1–3. Для каждого пика приведен скриншот лучшего совпадения с данными БД SwissRegulon из выдачи программы Tomtom. Показано значимое сходство этих выравниваний с ССТФ FNR для пиков 1 (а, E-value=3.13e-08) и 2 (б, E-value = 9.17e-08) и с ССТФ CRP для пика 3 (в, E-value = 1.80e-04), что полностью совпадает с экспериментальными данными.

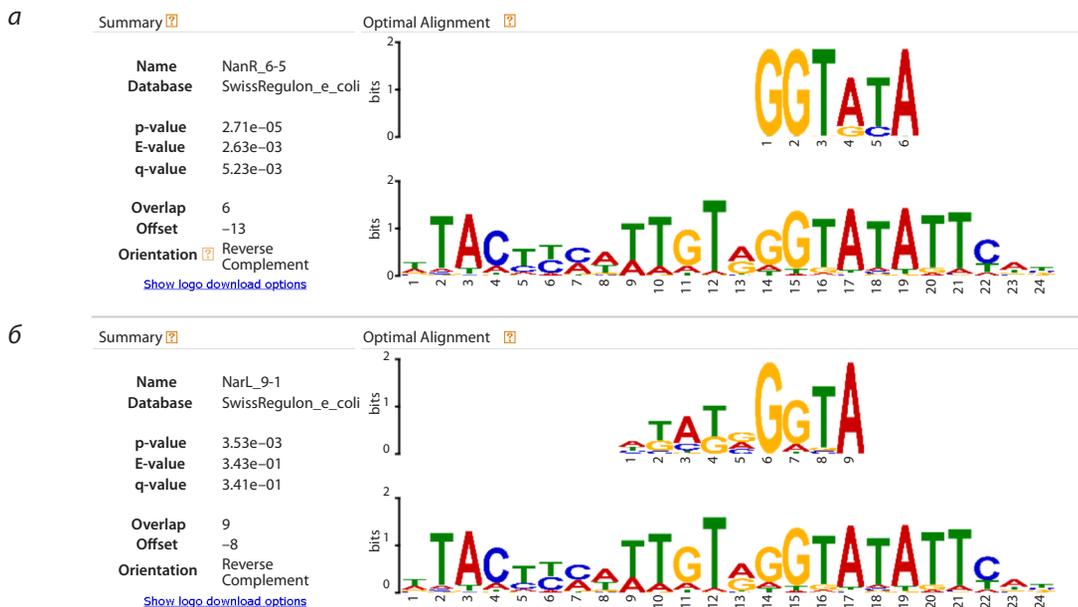


Рис. 7. Результат сравнения выявленного объединенного выравнивания для участка пика 4. Приведены скриншоты сравнения с данными БД SwissRegulon из выдачи программы Tomtom. Показано значимое сходство с выборкой ССТФ NanR (а) и незначимое сходство с выборкой ССТФ NarL (б).

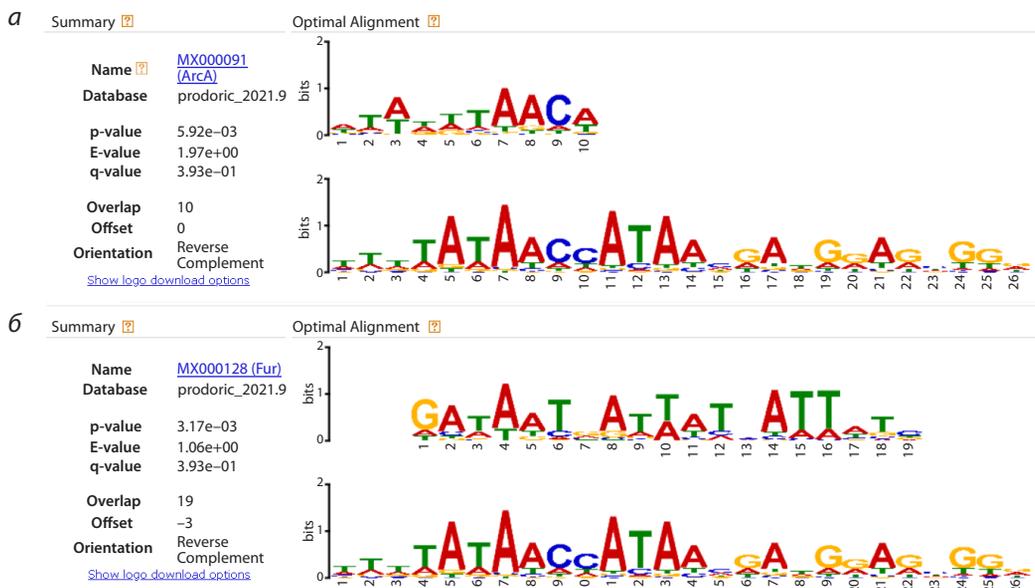


Рис. 8. Результат сравнения выявленного объединенного выравнивания для участка пика 5.

Приведены скриншоты сравнения с данными БД PRODORIC из выдачи программы Tomtom. Показано незначительное сходство частей объединенного выравнивания для пика 5 с ССТФ ArcA (а) и Fur (б). Видно, что расположение выявленных сайтов выравнивания значительно пересекается.

всегда и далеко не в каждом однократном запуске имеется возможность правильной идентификации наиболее вероятного транскрипционного фактора, совпадающего с экспериментальными данными (см. рис. 3). Описанное выше продемонстрировало неустойчивость результатов при однократном запуске конвейера, тогда как многократный запуск, напротив, позволяет надежно выявлять наиболее вероятный ТФ не только с помощью анализа статистики многократных запусков (см. рис. 3), но и с помощью анализа объединенного выравнивания для соответствующего пика с последующим его сравнением с выбранными базами данных известных ССТФ (см. рис. 6). Полученная оценка значимости сходства объединенных выравниваний с эталонными выборками из БД SwissRegulon и PRODORIC, достигающая величины E-value ~3.1e-08, открывает возможность генерации качественных выборок для распознавания некоторых ССТФ, недостаточно представленных в базах данных. В рамках конвейера реализована процедура, позволяющая проследить наличие *de novo* мотивов, соответствующих каждому пику, в каждом из проанализированных промоторов генов-ортологов и отобразить эти данные на таксономическом дереве геномов, рассмотренных в процессе анализа (см. рис. 5). Эта функция может быть полезна при анализе эволюции и представленности отдельных регуляторных сигналов.

Интересен результат, полученный с помощью метода MotifsOnFly для пика 4 (см. рис. 4 и 7). Экспериментальные данные указывают на наличие в этом районе сайта связывания ТФ NarL (Tyson et al., 1993, 1994; Xiao et al., 2016). Однако анализ статистики множественных запусков конвейера (см. рис. 4) и объединенного выравнивания сходных консервативных участков ДНК (см. рис. 7) вы-

являет NanR как наиболее вероятный ТФ для связывания с данным сайтом. При этом ТФ NarL лишь второй по значимости кандидат на связывание, несмотря на сходство с консенсусной последовательностью для его сайта связывания “ТАСУУМТ” (где Y = C или T, а M = A или C) (Tyson et al., 1993) с объединенным выравниванием для пика 4 (см. рис. 7). В то же время, рассматривая этот результат, следует учесть, что последовательность GGTATA, выявленная нами как высокодостоверный сайт связывания ТФ NanR (см. рис. 7), может быть просто недостаточной для связывания активной формы этого фактора, который является димером. Более того, ранее экспериментально показано, что высокое сродство связывания ТФ NanR с ДНК обеспечивается кооперативным связыванием трех димеров ТФ NanR с тремя рядом расположенными повторами GGTATA (Kalivoda et al., 2013; Horne et al., 2021). Наличие только одной последовательности GGTATA в позиции пика 4 ставит под сомнение возможность участия NanR в регуляции экспрессии гена *ompW*. Кроме того, в настоящее время роль NanR в регуляции экспрессии генов *E. coli* считается минорной и ограничивается его участием в регуляции только четырех оперонов, кодирующих ферменты катаболизма сиаловой кислоты (Kalivoda et al., 2013; Shimada et al., 2018). Тем не менее возможность регуляции гена *ompW* или некоторого числа генов-ортологов через найденный нами ССТФ NanR заслуживает внимания и требует дополнительной экспериментальной проверки.

Что касается результатов анализа участка промотора, соответствующего пику 5, то в первую очередь следует отметить, что этот пик полностью совпадает с районом посадки РНК-полимеразы $\sigma 70$, включающим старт транскрипции и специфичные -35/-10 элементы. Этот участок промотора перед самым стартом транскрипции гена *ompW*

насыщен регуляторными сигналами. Так, согласно RegulonDB (Salgado et al., 2024), здесь сосредоточены сайты связывания ArcA (Park et al., 2013; Xiao et al., 2016) и Fur (Zhang et al., 2020). Известная пластичность архитектуры сайта связывания TF ArcA (Park et al., 2013) обеспечивает эффективный способ кодирования сайтов ArcA вместе с сайтами других факторов транскрипции в одном и том же узком пространстве последовательностей. Тот факт, что сайт связывания TF Fur в районе пика 5 оценивается как слабый (Zhang et al., 2020), может объясняться особенностями его конкурентных отношений с TF SoxS (Graham et al., 2012), обладающим способностью взаимодействовать с элементом -35 (Taliaferro et al., 2012) в условиях окислительного стресса, провоцируемого активностью Fur в присутствии ионов железа. Таким образом, область промотора гена *OmpW* в районе пика 5 обладает значительным регуляторным потенциалом, однако по данным нашего анализа не содержит ни одного сайта со значительным сходством с известными ССТФ. Это можно интерпретировать как отсутствие сильных сайтов связывания каких-либо транскрипционных факторов, которые препятствовали бы посадке РНК-полимеразы $\sigma 70$. Одновременно сравнимое по значениям выявляемой достоверности сходство объединенного выравнивания для пика 5 (E-value) с большим числом известных ССТФ из БД PRODORIC можно интерпретировать как наличие множества слабых сайтов различных TF, включая пересекающиеся сайты ArcA и Fur (см. рис. 8), что находится в соответствии с приведенными выше литературными данными.

Заключение

Предложенная в нашей работе модификация подхода филогенетического футпринтинга, реализованная в методе MotifsOnFly, является естественным развитием подхода В. Liu с коллегами (2016) в ответ на непрерывно увеличивающийся объем широкомасштабных геномных данных. MotifsOnFly позволяет производить множественный запуск конвейера на различающихся подвыборках промоторов генов-ортологов, получать уточненные данные по локализации ССТФ и проводить статистический анализ. Варианты такого статистического анализа не ограничиваются описанными в этой статье и могут быть расширены в соответствии с целями и задачами конкретных исследований. Метод MotifsOnFly, позволяющий получать, как продемонстрировано здесь, надежные и устойчивые результаты выявления ССТФ, может быть востребован широким кругом исследователей при аннотации и анализе регуляторных последовательностей бактериальных геномов.

Список литературы / References

Мухин А.М., Ощепков Д.Ю., Лашин С.А. Вычислительный конвейер по распознаванию сайтов связывания транскрипционных факторов в бактериальных геномах *de novo*. *Проблемы информатики*. 2024;4(65):69-83. doi 10.24412/2073-0667-2024-4-69-83 [Mukhin A., Oschepkov D., Lashin S. A computational pipeline for *de novo* recognition of transcription factor binding sites in bacterial genomes. *Problems of Informatics*. 2024;4(65):69-83. doi 10.24412/2073-0667-2024-4-69-83 (in Russian)]

- Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W., Noble W.S. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(W):W202-W208. doi 10.1093/nar/gkp335
- Blanchette M., Tompa M. Discovery of regulatory elements by a computational method for phylogenetic footprinting. *Genome Res.* 2002; 12(5):739-748. doi 10.1101/gr.6902
- Blanchette M., Tompa M. FootPrinter: a program designed for phylogenetic footprinting. *Nucleic Acids Res.* 2003;31(13):3840-3842. doi 10.1093/nar/gkg606
- Browning D.F., Busby S.J. The regulation of bacterial transcription initiation. *Nat Rev Microbiol.* 2004;2(1):57-65. doi 10.1038/nrmicro787
- Chen X., Guo L., Fan Z., Jiang T. W-AlignACE: an improved Gibbs sampling algorithm based on more accurate position weight matrices learned from sequence and gene expression/ChIP-chip data. *Bioinformatics.* 2008;24(9):1121-1128. doi 10.1093/bioinformatics/btn088
- Claeys M., Storms V., Sun H., Michoel T., Marchal K. MotifSuite: workflow for probabilistic motif detection and assessment. *Bioinformatics.* 2012;28(14):1931-1932. doi 10.1093/bioinformatics/bts293
- Constantinidou C., Hobman J.L., Griffiths L., Patel M.D., Penn C.W., Cole J.A., Overton T.W. A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. *J Biol Chem.* 2006;281(8):4802-4815. doi 10.1074/jbc.M512312200
- Diesh C., Stevens G.J., Xie P., De Jesus Martinez T., Hershsberg E.A., Leung A., Guo E., ... Haw R., Cain S., Buels R.M., Stein L.D., Holmes I.H. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol.* 2023;24(1):74. doi 10.1186/s13059-023-02914-z
- Dudek C.-A., Jahn D. PRODORIC: state-of-the-art database of prokaryotic gene regulation. *Nucleic Acids Res.* 2022;50(D1):D295-D302. doi 10.1093/nar/gkab1110
- Gaston K., Bell A., Kolb A., Buc H., Busby S. Stringent spacing requirements for transcription activation by CRP. *Cell.* 1990;62(4):733-743. doi 10.1016/0092-8674(90)90118-x
- Graham A.I., Sanguinetti G., Bramall N., McLeod C.W., Poole R.K. Dynamics of a starvation-to-surfeit shift: a transcriptomic and modelling analysis of the bacterial response to zinc reveals transient behaviour of the Fur and SoxS regulators. *Microbiology (Reading)*. 2012;158(Pt.1):284-292. doi 10.1099/mic.0.053843-0
- Gupta S., Stamatoyannopoulos J.A., Bailey T.L., Noble W.S. Quantifying similarity between motifs. *Genome Biol.* 2007;8(2):R24. doi 10.1186/gb-2007-8-2-r24
- Hertz G.Z., Stormo G.D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics.* 1999;15(7-8):563-577. doi 10.1093/bioinformatics/15.7.563
- Horne C.R., Venugopal H., Panjikar S., Wood D.M., Henrickson A., Brookes E., North R.A., Murphy J.M., Friemann R., Griffin M.D.W., Ramm G., Demeler B., Dobson R.C.J. Mechanism of NanR gene repression and allosteric induction of bacterial sialic acid metabolism. *Nat Commun.* 2021;12(1):1988. doi 10.1038/s41467-021-22253-6
- Huerta-Cepas J., Serra F., Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 2016;33(6): 1635-1638. doi 10.1093/molbev/msw046
- Hunter J.D. Matplotlib: a 2D graphics environment. *Comput Sci Eng.* 2007;9(3):90-95. doi 10.1109/MCSE.2007.55
- Kalivoda K.A., Steenbergen S.M., Vimr E.R. Control of the *Escherichia coli* sialoregulon by transcriptional repressor NanR. *J Bacteriol.* 2013;195(20):4689-4701. doi 10.1128/JB.00692-13
- Katara P., Grover A., Sharma V. Phylogenetic footprinting: a boost for microbial regulatory genomics. *Protoplasma.* 2012;249(4):901-907. doi 10.1007/s00709-011-0351-9
- Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lo-

- pez R., Thompson J.D., Gibson T.J., Higgins D.G. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947-2948. doi 10.1093/bioinformatics/btm404
- Levy S., Hannenhalli S., Workman C. Enrichment of regulatory signals in conserved non-coding genomic sequence. *Bioinformatics*. 2001; 17(10):871-877. doi 10.1093/bioinformatics/17.10.871
- Li G., Liu B., Ma Q., Xu Y. A new framework for identifying cis-regulatory motifs in prokaryotes. *Nucleic Acids Res*. 2011a;39(7):e42. doi 10.1093/nar/gkq948
- Li G., Ma Q., Mao X., Yin Y., Zhu X., Xu Y. Integration of sequence-similarity and functional association information can overcome intrinsic problems in orthology mapping across bacterial genomes. *Nucleic Acids Res*. 2011b;39(22):e150. doi 10.1093/nar/gkr766
- Liu B., Zhang H., Zhou C., Li G., Fennell A., Wang G., Kang Y., Liu Q., Ma Q. An integrative and applicable phylogenetic footprinting framework for cis-regulatory motifs identification in prokaryotic genomes. *BMC Genomics*. 2016;17:578. doi 10.1186/s12864-016-2982-x
- Liu X., Brutlag D.L., Liu J.S. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput*. 2001;6:127-138
- Liu X.S., Brutlag D.L., Liu J.S. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol*. 2002;20(8):835-839. doi 10.1038/nbt717
- Mao X., Ma Q., Zhou C., Chen X., Zhang H., Yang J., Mao F., Lai W., Xu Y. DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res*. 2014;42(D1): D654-D659. doi 10.1093/nar/gkt1048
- McCue L.A., Thompson W., Carmack C.S., Lawrence C.E. Factors influencing the identification of transcription factor binding sites by cross-species comparison. *Genome Res*. 2002;12(10):1523-1532. doi 10.1101/gr.323602
- Myers K.S., Yan H., Ong I.M., Chung D., Liang K., Tran F., Keleş S., Landick R., Kiley P.J. Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. *PLoS Genet*. 2013;9(6):e1003565. doi 10.1371/journal.pgen.1003565
- Olman V., Xu D., Xu Y. CUBIC: identification of regulatory binding sites through data clustering. *J Bioinform Comput Biol*. 2003;1(1): 21-40. doi 10.1142/s02197200030000162
- Pachkov M., Balwierz P.J., Arnold P., Ozonov E., van Nimwegen E. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. *Nucleic Acids Res*. 2013;41(D1):D214-D220. doi 10.1093/nar/gks1145
- Park D.M., Akhtar M.S., Ansari A.Z., Landick R., Kiley P.J. The bacterial response regulator ArcA uses a diverse binding site architecture to regulate carbon oxidation globally. *PLoS Genet*. 2013;9(10): e1003839. doi 10.1371/journal.pgen.1003839
- Peltek S., Bannikova S., Khlebodarova T.M., Uvarova Y., Mukhin A.M., Vasiliev G., Scheglov M., Shipova A., Vasilieva A., Oshchepkov D., Bryanskaya A., Popik V. The transcriptomic response of cells of the thermophilic bacterium *Geobacillus icigianus* to terahertz irradiation. *Int J Mol Sci*. 2024;25(22):12059. doi 10.3390/ijms252212059
- Salgado H., Gama-Castro S., Lara P., Mejia-Almonte C., Alarcón-Carranza G., López-Almazo A.G., Betancourt-Figueroa F., ... Hernández-Alvarez A.J., Santos-Zavaleta A., Capella-Gutiérrez S., Gelpi J.L., Collado-Vides J. RegulonDB v12.0: a comprehensive resource of transcriptional regulation in *E. coli* K-12. *Nucleic Acids Res*. 2024;52(D1):D255-D264. doi 10.1093/nar/gkad1072
- Sayers E.W., Beck J., Bolton E.E., Bourexis D., Brister J.R., Canese K., Comeau D.C., ... Wang J., Ye J., Trawick B.W., Pruitt K.D., Sherry S.T. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2021;49(D1):D10-D17. doi 10.1093/nar/gkaa892
- Shimada T., Ogasawara H., Ishihama A. Single-target regulators form a minor group of transcription factors in *Escherichia coli* K-12. *Nucleic Acids Res*. 2018;46(8):3921-3936. doi 10.1093/nar/gky138
- Tagle D.A., Koop B.F., Goodman M., Slightom J.L., Hess D.L., Jones R.T. Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol*. 1988;203(2):439-455. doi 10.1016/0022-2836(88)90011-3
- Taliaferro L.P., Keen E.F., Sanchez-Alberola N., Wolf R.E. Transcription activation by *Escherichia coli* Rob at class II promoters: protein-protein interactions between Rob's N-terminal domain and the σ^{70} subunit of RNA polymerase. *J Mol Biol*. 2012;419(3-4):139-157. doi 10.1016/j.jmb.2012.03.019
- Tompa M., Li N., Bailey T.L., Church G.M., De Moor B., Eskin E., Favorov A.V., ... Vandenbogaert M., Weng Z., Workman C., Ye C., Zhu Z. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*. 2005;23(1):137-144. doi 10.1038/nbt1053
- Tyson K.L., Bell A.I., Cole J.A., Busby S.J. Definition of nitrite and nitrate response elements at the anaerobically inducible *Escherichia coli* *nirB* promoter: interactions between FNR and NarL. *Mol Microbiol*. 1993;7(1):151-157. doi 10.1111/j.1365-2958.1993.tb01106.x
- Tyson K.L., Cole J.A., Busby S.J. Nitrite and nitrate regulation at the promoters of two *Escherichia coli* operons encoding nitrite reductase: identification of common target heptamers for both NarP- and NarL-dependent regulation. *Mol Microbiol*. 1994;13(6):1045-1055. doi 10.1111/j.1365-2958.1994.tb00495.x
- Ushida C., Aiba H. Helical phase dependent action of CRP: effect of the distance between the CRP site and the -35 region on promoter activity. *Nucleic Acids Res*. 1990;18(21):6325-6330. doi 10.1093/nar/18.21.6325
- Xiao M., Lai Y., Sun J., Chen G., Yan A. Transcriptional regulation of the outer membrane porin gene *ompW* reveals its physiological role during the transition from the aerobic to the anaerobic lifestyle of *Escherichia coli*. *Front Microbiol*. 2016;7:799. doi 10.3389/fmicb.2016.00799
- Zhang P., Ye Z., Ye C., Zou H., Gao Z., Pan J. OmpW is positively regulated by iron via Fur, and negatively regulated by SoxS contribution to oxidative stress resistance in *Escherichia coli*. *Microb Pathog*. 2020;138:103808. doi 10.1016/j.micpath.2019.103808

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 15.11.2025. После доработки 09.12.2025. Принята к публикации 10.12.2025.