

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу **Манолова Александра Ивановича «Биоинформатический анализ изменчивости генового состава прокариот, в том числе в ассоциации с патогенностью»**, представленную на соискание ученой степени кандидата биологических наук по специальностям 1.5.8 – «математическая биология, биоинформатика»

Диссертационная работа Манолова Александра Ивановича посвящена разработке программного конвейера для идентификации высокоизменчивых областей геномов прокариот. Развитие методов высокопроизводительного секвенирования привело к тому, что расшифровка геномов прокариот стало рутинной операцией. В открытых базах данных представлено уже более 400 тысяч геномов прокариот. Анализ такого большого числа геномных данных требует развития биоинформационных методов и инструментов.

В диссертационной работе Манолова А.И. предложен оригинальный метод для оценки интенсивности изменений состава генов в определенном регионе генома. Метод основан на графовом представлении расположения ортологичных генов в разных геномах в виде графов с последующим подсчетом числа путей в графе. С помощью разработанного инструмента были проанализированы доступные геномы *Escherichia coli*. В результате, как и ожидалось, наиболее изменчивыми областями геномов являлись места встраивания мобильных элементов и фагов. Но наряду с этими событиями удалось выявить ряд высоко изменчивых регионов не содержащих следы наличия мобильных элементов. И это важный результат, который представляет интерес для дальнейших исследований механизмов горизонтального переноса генов.

Диссертация Манолова А.И. имеет традиционную структуру, которая содержит следующие разделы: введение, обзор литературы, материалы и методы, результаты, обсуждение, выводы, список литературы

и источники финансирования. Объем диссертации составляет 131 страницу, в том числе 45 рисунков и 2 таблицы. Список литературы включает 231 источник.

Обзор литературы условно можно поделить на три подраздела. Первый раздел посвящен описанию архитектуры геномов прокариот и известным механизмам горизонтального переноса генов у прокариот. Во втором разделе рассматриваются известные биоинформационные методы для исследования изменчивости геномов и визуализации представления геномных данных. И третий раздел в краткой форме описывает известные данные о связи болезни Крона и присутствия бактерий *Escherichia coli*. Стоит отметить, что в настоящее время в области исследования организации геномов накопился достаточно большой объем информации. На мой взгляд Манолову А.И. успешно удалось в достаточно краткой форме изложить основные данные, в том числе включая работы последних лет, об архитектуре геномов и биоинформационных методах их исследований. Обзор литературы, свидетельствует о том, что автор хорошо ориентируется в проблеме и способен к критическому анализу данных и их обобщению.

Вторая глава посвящена описанию используемых в работе материалов и примененных методических подходов. Достаточно подробно описан подбор используемых геномов, а также методика сборки и анализа геномных последовательностей *E. coli* от пациентов с болезнью Крона. С учетом того, что работа посвящена разработке биоинформационного инструмента по поиску высоко вариабельных геномных последовательностей, а не выделению новых изолятов, части посвященные выделению геномной ДНК и их секвенированию можно было бы сократить и представить их просто в виде таблицы с основными характеристиками полученных изолятов и их геномов. При этом вполне логично решение автора вынести описание методики по поиску вариабельных регионов в

раздел с описанием результатов.

Третья глава самая объемная, посвящена описанию полученных результатов работы, и состоит из восьми частей. В первой части описан метод представления в виде графа информации о том, как расположены гены в геномных последовательностях. Предлагается рассматривать группу близкородственных организмов, например - штаммов одного бактериального вида, строить ортогруппы и дальше переводить эту информацию в графовое представление, связывая ребрами ортогруппы расположенные по соседству. Графы предлагается строить для отдельных регионов референсного генома с тем, чтобы, подсчитав количество путей в графе, оценить количество возможных вариантов состава и расположения генов в данном регионе. Согласно этому представлению, получается, что чем более изменчива рассматриваемая область генома, тем большее количество путей присутствует в графе.

Во второй части главы рассматриваются вопросы применимости и эффективности предложенных алгоритмов. В частности, оценивается скорость и точность метода в зависимости от размера выборки геномных последовательностей и проводится сравнение результатов, полученных на полных геномах либо на драфт-геномах. Далее приводится описание изменчивости генома кишечной палочки на примере штамма, ассоциированного с болезнью Крона. Приводится описание выявленных областей повышенной изменчивости, их генный состав и ассоциация с мобильными элементами генома. Также рассматривается эволюционная консервативность расположения областей повышенной и пониженной изменчивости. Для этого были выбраны и независимо рассмотрены несколько крупных клад филогенетического дерева кишечной палочки. Отмечается устойчивая ассоциация областей повышенной изменчивости с областями расположения профагов. В четвертой части третьей главы рассматривается ассоциация уровня изменчивости и распределения χ^2 -

сайтов, а также плотности межхромосомных контактов. Далее описывается способ визуализации графового представления расположения генов в геномах и приводятся примеры построения графов для отдельных областей генома с повышенной изменчивостью, предлагаются способы интерпретации подобных визуализаций. Также приводится описание разработанного компьютерного приложения, позволяющего применять предложенные способы анализа набора геномов. В заключительной части главы приводится способ поиска генов, встречаемость которых различает две выборки геномных последовательностей на примере поиска ассоциаций различных изолятов кишечной палочки с болезнью Крона.

Четвертая глава посвящена обсуждению полученных результатов. Стоит отметить, что Манолов А.И., критично подошел к оценке полученных результатов, и достаточно подробно описывает сильные и слабые стороны применимости предложенного метода, а также возможные пути его улучшения.

Среди недостатков диссертационной работы Манолова А.И. можно отметить следующие моменты.

1. В работе уделено достаточно много внимания описанию биоинформационного анализа изменчивости геномов, а применение этого метода для идентификации генов связанных с патогенностью представлено достаточно слабо. Проведенный анализ по поиску наиболее изменчивых регионов в штаммах *E.coli* выделенных у пациентов с болезнью Крона в этом плане может быть не репрезентативным. Отбор этих генетических перестроек может быть связан с их адаптацией, вызванной именно болезнью, а не их патогенностью, поскольку они являются комменсальными штаммами.
2. Мало внимания уделено практическому применению

разработанного метода. Как он может быть применен на практике? Хотя упоминается несколько возможных применений, хотелось бы более развернутого описания.

3. В работе провели сравнения поиска изменчивости при использовании полных геномов и драфт-геномов и полученные результаты показали схожие результаты. Однако остается вопрос будут ли сохраняться эти результаты при увеличении длины анализируемой цепочки генов в окне. Ведь в этом случае, во-первых, будет иметь значение качество сборки драфт-геномов, а, во-вторых, при использовании драфт-геномов теряется информация об организации генома, в том числе о локализации интегрированных фагов и мобильных элементов.
4. Диссертационная работа Манолова А.И. посвящена разработке биоинформационного инструмента анализа геномов и для практического использования всегда интересен функциональный контекст выявленных областей. В работе хотя и называются конкретные гены, однако, роли их в метаболизме подробно не описывается. А такой анализ мог бы усилить описание полученных результатов.

Также можно отметить ещё замечания, касающиеся опечаток, неудачных фраз и оформления диссертационной работы. В качестве примера, можно привести опечатку на стр. 39 слово «поход» нужно заменить на слово «подход». В части оформления наиболее ярким примером является странная нумерация подглав, с излишним дроблением стр. 47-48 (2.0.1) или отсутствием нумерации для некоторых разделов стр. 11 - раздел «Оперон как функциональная единица архитектуры генома», стр. 24 – раздел «Плазмиды», стр. 25 – раздел «Интегративные конъюгативные элементы» и т.д.

В целом, сделанные замечания носят дискуссионный характер и не влияют на высокую оценку самой диссертационной работы Манолова А.И. Все представленные в диссертационной работе результаты получены впервые. Выносимые на защиту научные положения основаны на предложенном оригинальном методе, реализованном в виде программного продукта с открытым исходным кодом. Предложенный алгоритм верифицирован при помощи сравнения получаемых результатов с ранее опубликованными, а также при помощи методов компьютерного моделирования. Корректность методов не вызывает сомнения и подтверждается международной экспертной оценкой, что отражено в публикациях в высокорейтинговых научных журналах среди которых Scientific reports, BMC genomics, PLOS computational biology.

Заключение. Диссертационная работа Манолова Александра Ивановича «Биоинформатический анализ изменчивости генового состава прокариот, в том числе в ассоциации с патогенностью», выполненная в ФНКЦ Физико-Химической Медицины ФМБА России, представляет собой цельное и законченное исследование, в котором предложены новые способы анализа бактериальных геномов, описаны алгоритмы и программная реализация этих способов. Работа выполнена на высоком научном и методическом уровне, соответствует критериям, предъявляемым к кандидатским диссертациям, и соответствует специальности 1.5.8. – математическая биология, биоинформатика. Автореферат достаточно полно отражает объем и содержание работы.

По актуальности, научной новизне, практической значимости, диссертация Манолова Александра Ивановича соответствует требованиям п. 9 «Положения о присуждении ученых степеней», утвержденного правительством РФ от 24 сентября 2013 года № 842, предъявляемым к кандидатским диссертациям, а её автор Манолов А. И. заслуживает присуждения ученой степени кандидата биологических наук по

специальности 1.5.8. - математическая биология, биоинформатика.

Заведующий лабораторией геномики микроорганизмов и метагеномики
Федерального государственного учреждения «Федеральный
исследовательский центр «Фундаментальные основы биотехнологии»
Российской академии наук»

119071, Москва, Ленинский пр-т, д. 33, стр. 2

Тел. +7 (495) 783-32-64

www.fbras.ru

e-mail оппонента: mardanov@biengi.ac.ru

доктор биологических наук

Марданов Андрей Владимирович

16.05.2022

Подпись д.б.н. Марданова А.В. заверяю

Ученый Секретарь

Федерального государственного учреждения

«Федеральный исследовательский центр

«Фундаментальные основы биотехнологии»

Российской академии наук»

Орловский Александр Федорович

16.05.2022

