

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное учреждение науки
Институт биологии гена Российской академии наук
(ИБГ РАН)

Вавилова ул., 34/5, Москва, 119334

Тел.: (499)135-60-89, (499)135-98-84 Факс: (499)135-41-05

e-mail: info@genebiology.ru; <http://www.genebiology.ru>

ОКПО 00244660 ОГРН 1027739618037 ИНН/ КПП 7736020369/773601001

19.05.2022 № *123А-120*

На №

от



«УТВЕРЖДАЮ»

Директор ИБГ РАН

академик РАН Георгиев П.Г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

на диссертационную работу Манолова Александра Ивановича на тему:
«Биоинформатический анализ изменчивости геномного состава
прокариот, в том числе в ассоциации с патогенностью»,
представленную на соискание ученой степени кандидата биологических наук
по специальности 1.5.8 — «Математическая биология, биоинформатика»

Актуальность темы диссертационной работы

Высокая геномная пластичность – примечательное свойство прокариот, которое позволяет им адаптироваться к существенным изменениям окружающей среды. Изучение закономерностей данного явления с помощью анализа геномных данных и с применением новых биоинформатических подходов представляет собой существенный интерес с точки зрения понимания механизмов существования бактериальной клетки и эволюции прокариот. Ввиду нарастающих объемов коллекций геномных данных, существенно важно, чтобы соответствующие алгоритмы были вычислительно эффективными и предоставляли наглядную визуализацию получаемых результатов.

В контексте биомедицины, геномная изменчивость позволяет клинически значимым микроорганизмам-возбудителям заболеваний приобретать устойчивость к широкому спектру антимикробных препаратов, что затрудняет лечение соответствующих заболеваний и таким образом представляет собой насущную проблему здравоохранения. Также существенную проблему представляет собой перенос генов, кодирующих факторы вирулентности и позволяющих микроорганизмам эффективно вносить вклад в патогенез заболеваний желудочно-кишечного тракта и других систем органов человека. В этом свете, разработка и применение алгоритмов для анализа геномной изменчивости бактерий из клинических образцов позволят в перспективе выявить возможные способы предотвращения лекарственной устойчивости, профилактики и лечения заболеваний, в которых микроорганизмы и их сообщества являются причинным фактором. Проведенное автором с помощью нового разработанного метода исследование геномов *Escherichia coli*, ассоциированных с болезнью Крона, является наглядным примером подобного анализа. Таким образом, тема диссертационной работы, несомненно, обладает высокой степенью актуальности как с фундаментальной, так и прикладной точки зрения.

Научная новизна и научно-практическая значимость работы

В работе впервые представлены алгоритмы и программное обеспечение для количественного анализа и визуализации изменчивости геномов прокариот с помощью графового подхода. Данные методы представляют собой ценные инструменты для проведения масштабных исследований эволюции бактериальных геномов. С помощью разработанных оригинальных методов проведен анализ оперонов *Escherichia coli*, специфических для кишечника пациентов с болезнью Крона. Это позволяет углубить понимание механизмов участия микробиома человека в патогенез данного тяжелого заболевания, а также служит примером эффективности подхода для геномного анализа микроорганизмов в клиническом контексте в целом. Реализация разработанных

алгоритмов в виде открыто доступного программного обеспечения представляет собой несомненную ценность для научно-исследовательского сообщества в области геномики и эволюции микроорганизмов.

Обоснованность и достоверность полученных результатов

Результаты, представленные в диссертационной работе, получены с помощью современных методов биоинформатики и молекулярной биологии, с использованием современного оборудования и программного обеспечения для анализа данных.

Диссертация оформлена по правилам, логично и последовательно изложена. При обсуждении результатов работы привлечен большой объем литературных данных.

Достоверность полученных результатов подтверждена 6 публикациями в журналах, индексируемых Web of Science и Scopus, тезисами 6 докладов, а также выступлениями на 5 научных конференциях. Достоверность предложенных алгоритмических подходов подкреплена верификацией на синтетических данных, а также сопоставлением результатов с опубликованными данными других научных групп. Для одного из методов, о достоверности его работы также положительно свидетельствуют результаты проведенной оценки границ его применимости.

Общая оценка работы

Диссертация Манолова А.И. построена по общепринятому плану и состоит из введения, обзора литературы, описания материалов и методов, результатов, обсуждения, выводов и списка литературы, а также отдельной главы с раскрытием финансирования. Работа изложена на 131 странице, содержит 45 рисунков и 2 таблицы, библиографический список содержит 231 источник.

В разделе Введение автор определяет цель и задачи исследования, обосновывает актуальность исследуемой проблемы, характеризует научную

новизну, практическую значимость полученных результатов, формулирует научные положения, выносимые на защиту, характеризует достоверность и описывает свой личный вклад.

В обзоре литературы приведены основные сведения относительно архитектуры генома прокариот, описаны известные механизмы горизонтального переноса генов, описаны биоинформатические подходы к анализу вариабельности прокариотических геномов, а также описаны графовые подходы к визуализации и анализу данного явления. Кратко описаны известные на текущий момент сведения о связи между болезнью Крона и кишечной палочкой. Обзор литературы включает большое количество актуальных опубликованных работ по теме диссертационного исследования, наглядно проиллюстрирован и полно отражает современное состояние изучаемой проблемы.

В разделе Материалы и методы кратко перечислены источники представительного набора публично доступных данных по прокариотическим геномам, а также указаны алгоритмы и программные средства, использованные в работе. Приведены описания этапов сбора и пробоподготовки образцов биоматериала и генерации новых геномных данных, с корректным указанием вклада исследователей помимо автора. Использованные методы соответствуют современному уровню исследований, адекватны для поставленной цели и задач диссертационного исследования.

В разделе Результаты четко и в сопровождении наглядных иллюстраций описаны разработанные автором оригинальные алгоритмические подходы, их программные реализации, примеры их применения к реальным данным – как опубликованным, так и новым; предложены биологические интерпретации результатов. Описано представление множества геномов и их генов в виде графа, а также алгоритм численной оценки геномной изменчивости на его основе и алгоритм анализа локальной вариабельности при помощи построения подграфов; описаны реализующие их вычислительные конвейеры. Программное обеспечение публично доступно в сети Интернет, в частности, как веб-сервис с

интуитивным интерфейсом, что предоставляет полезный инструмент для анализа геномной изменчивости широкому кругу пользователей-исследователей без навыков программирования. Детально описана проведенная автором валидация конвейера на синтетических данных; совместно с обнаруженной высокой корреляцией с другим методом оценки изменчивости по опубликованным данным *E. coli*, они подтверждают корректность разработанного подхода. Описаны результаты исследования зависимости скорости и результатов работы алгоритма от входных параметров, что важно для ресурсоемких программных решений. Примечательно, что метод позволяет работать с черновыми сборками геномов — это существенно расширяет границы его применения и повышает полезность для мирового научного сообщества. Биомедицинская релевантность подхода проиллюстрирована результатами применения подхода к реальным данным на отдельных клинических штаммах и широких филогенетических группах - как по бактерии *E. coli*, так и по обширным представительным выборкам геномов другим клинически значимых видов. О достоверности работы алгоритма свидетельствует тот факт, что профили геномной вариабельности существенно схожи между крупными филогруппами *E. coli*. В результате анализа сходства профилей вариабельности между близкородственными видами были обнаружены консервативно расположенные области повышенной изменчивости. Обнаружена корреляция найденных профилей локальной вариабельности с распределением сайтов Chi, играющих важную роль в инициации процесса гомологичной рекомбинации, а также со шкалограммой хромосомных контактов (по опубликованным данным по *E. coli* и *B. subtilis*). Новый алгоритм для анализа локальной вариабельности при помощи построения подграфов отработан на примере нескольких оперонов из геномов *E. coli*, ассоциированных с болезнью Крона; при этом проведен детальный разбор входящих в них генов и предложена интерпретация полученных наблюдений. Также автором был разработан универсальный алгоритмический подход для выявления оперонов,

наличие которых в геноме ассоциировано с определенным внешним фактором; описана его апробация и интерпретация результатов на примере *E. coli* в связи с болезнью Крона.

В разделе Обсуждение результатов полученные в работе алгоритмические разработки и результаты их работы обобщены и сопоставлены с предыдущими исследованиями по теме диссертационной работы, на основании в последующем разделе Выводы сделаны корректные и достоверные выводы, соответствующие целям и задачам исследования.

Вопросы и замечания

По работе имеются следующие вопросы и замечания:

- 1) В разделе Обзор литературы на стр. 18 упомянуто, что: «бактериальная хромосома разделена на домены взаимодействия хромосом (chromosome interaction domains, CID), которые аналогичны топологически ассоциированным доменам (topologically associating domains, TAD) у эукариот». Стоит уточнить, что исходя из имеющихся на момент данных в Hi-C формате для бактерий различных видов домены CID все же не очень устойчиво определены, между ними могут быть промежутки, и набор CID в конкретном геноме может разительно варьировать в зависимости от выбора алгоритма поиска.
- 2) Имеется вопрос относительно поиска оперонов *E. coli*, которые значимо чаще встречаются в изолятах от пациентов с болезнью Крона. Из описания пациентов в разделе 2.1 не очевидно, являлись ли пациенты первично диагностированными, не подвергавшимися еще лечению от данного заболевания. Если лечение все-таки предшествовало сбору образцов, насколько вероятно, что найденные опероны ассоциированы не с болезнью Крона как таковой, а представляют собой особенности подвидов *E. coli*, отобравшихся в результате лекарственной терапии данного заболевания?

- 3) Хотя в литературном обзоре описана важность плазмид как участников горизонтального переноса генов, в работе не описаны анализы плазмидной составляющей генома с помощью разработанных алгоритмов геномной изменчивости. Применимы ли данные графовые методы для плазмид? Если да, проводился ли такой анализ? Было бы интересно увидеть их – например, для тех же *E. coli*, ассоциированных с болезнью Крона, являющихся одним из центральных объектов исследования в работе.
- 4) В тексте присутствуют небольшое количество языковых ошибок и опечаток, которых, как представляется, легко можно было избежать, применив программу для поиска и исправления ошибок.

Впрочем, указанные замечания и комментарии не носят принципиального характера и не снижают научной и практической значимости выполненной автором работы.

Заключение

В представленной соискателем диссертационной работе успешно решены заявленные цель и задачи. Диссертационная работа Манолова А.И. выполнена на высоком научно-методическом уровне и заслуживает положительной оценки.

Диссертационная работа диссертации Манолова Александра Ивановича «Биоинформатический анализ изменчивости генового состава прокариот, в том числе в ассоциации с патогенностью» является оригинальной научно-квалификационной работой, выполненной на высоком методическом уровне. По содержанию, актуальности, новизне, научному и методическому уровню, практической и теоретической значимости полученных результатов, она полностью соответствует требованиям «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства РФ от 24 сентября 2013 г. №842, представляемым к диссертациям на соискание ученой степени кандидата биологических наук, а ее автор, Манолова Александр

Иванович, заслуживает присуждения степени кандидата биологических наук по специальности 1.5.8 — «Математическая биология, биоинформатика».

Отзыв заслушан и утвержден на заседании межлабораторного семинара Федерального государственного бюджетного учреждения Институт биологии гена Российской академии наук (протокол заседания № 2 от 12 мая 2022 года).

Отзыв составил

Кандидат биологических наук,

научный сотрудник

Лаборатории молекулярной генетики микроорганизмов

Федерального государственного бюджетного учреждения науки

Института биологии гена Российской академии наук

Куликовский Алексей Дмитриевич

E-mail: al.kulikovsky@gmail.com

Федеральное государственное бюджетное учреждение науки

Институт биологии гена Российской академии наук

Адрес: 119334, г. Москва, ул. Вавилова, д. 34/5

E-mail: info@genebiology.ru

Веб-сайт: <http://www.genebiology.ru>

Телефон: +7 (499) 135-60-89, +7 (499) 135-41-05



ПОДПИСЬ *Куликовского А.Д.*
С В Е Р Я Ю
Ученый секретарь ИБГ РАН Набирочкина Е.Н.