УТВЕРЖДАЮ
ВРИО Директора Федерального государственного бюджетного учреждения науки «Федеральный исследовательский центр «Пущинский научный центр биологических исследований Российской академии наук»

_____ П.Я. Грабарник
«29» апреля 2019 г.

Отзыв ведущей организации на диссертационную работу
Бочкарёвой Ольги Олеговны
«Реконструкция эволюционной истории геномных перестроек в бактериях»,
представленную на соискание ученой степени кандидата биологических наук по специальности 03.01.09 - «математическая биология, биоинформатика»

Молекулярная филогенетика давно стала неотъемлемой частью современной биологии и важнейшим инструментом реконструкции эволюционных событий. Филогенетический анализ различных видов бактерий необходим для понимания эволюции патогенов и разработки новых подходов к терапии вызываемых ими заболеваний. Однако широко используемые методы молекулярной филогении не всегда точно учитывают такие геномные изменения как делеции, инсерции и инверсии, особенно при анализе близкородственных видов или штаммов одного вида. При этом геномные перестройки играют важнейшую роль в эволюции бактерий, в том числе, патогенных, позволяя им приобретать устойчивость к различным антибиотикам и быстро уменьшать размер генома за счет крупных делеций, что дает им существенное эволюционное преимущество. Поэтому поставленная перед Ольгой Олеговной задача интегрального анализа геномных перестроек у бактерий в контексте их эволюционной истории является, безусловно, актуальной. Важным результатом является разработка подходов, позволяющих проанализировать эволюционную историю геномных перестроек и оценивать поток генов, распространившихся в популяции в результате горизонтального переноса, на основе расположения генов на хромосоме. В ходе выполнения работы были описаны новые случаи распространения геномных островков и антигенной вариации в таких патогенах человека, как *Streptococcus pneumoniae* и *Burkholderia pseudomallei*.

Работа Ольги Олеговны содержит все необходимые для диссертационной работы разделы, в том числе: Введение, Обзор литературы, Материалы и Методы, Результаты и их обсуждение (правда, не выделенные в отдельный раздел), Заключение, Выводы и Приложение, содержащее список использованных в работе штаммов. Список процитированной литературы включает 112 ссылок. Работа изложена на 111 страницах машинописного текста. Результаты работы отражены в 37 рисунках и 5 таблицах. Кроме того, 5 таблиц включает в себя Приложение.

В разделе **«Введение»**, представленном на 8 страницах, автором обоснована актуальность проведённого исследования, а также сформулированы цель и конкретные задачи проделанной работы. В отдельных разделах сформулированы **Новизна**, а также **Практическая ценность работы** и **Положения, выносимые на защиту**. Во Введении также приведён список опубликованных работ. Основные результаты работы Ольги

Олеговны опубликованы в виде двух статей в журналах, входящих в базу данных Web of Science, и были представлены на 13 конференциях.

Раздел «Обзор литературы» состоит из трех глав, изложенных на 18 страницах. Первая глава посвящена методам изучения геномных перестроек. Довольно подробно описана история их развития, введены термины, описывающие разные виды геномных перестроек, и приведены способы их математического выражения. Отмечаются преимущества и недостатки применения каждого из методов. Стоит отметить аккуратно построенные самим автором с помощью точечных матриц примеры выравнивания хромосом в геномах *Escherichia coli* и *Shigella* (Рис. 1.1.) и цепочек перестроек (Рис. 1.4.). Кажется, что в качестве иллюстрации к последней части главы стоило бы привести и примеры филогенетических деревьев, построенных с использованием различных алгоритмов.

Вторая глава Обзора посвящена эволюции бактериальных геномов. В главе подробно описываются различные типы геномных перестроек и их влияние на эволюционные события - инверсии и их примеры, инсерции, горизонтальный перенос генов. Отмечается, что многие горизонтально перенесенные гены находятся в окруженных мобильными элементами «геномных островках», которые могут включать в себя как гены вирулентности, так и, например, гены, кодирующие ферменты биосинтеза, что может приводить к появлению у вида новых метаболических путей. Также вводится понятие пангенома и описывается его структура. В целом, глава довольно легко читается, но недостаточно структурирована. Так, в частности, в самом ее начале рассматриваются три варианта модели бактериального генома – как набора генов, определенной последовательности генов и нуклеотидной последовательности, но при этом дальше эта классификация в главе не используется – возможно, стоило бы сделать некоторую логическую связку-переход к следующему абзацу или как-то структурировать текст в соответствии с этими моделями. Не очень понятно, что имелось в виду под «геном-кодирующими областями» (Стр. 22). В подписи к Рис. 1.5. белковые продукты генов обозначены курсивом, что неверно, а в описание рисунка, вероятно, закралась опечатка – вместо FliB должен быть белок FljB.

В третьей главе приведены роды и виды микроорганизмов, геномы которых были использованы в работе. Описаны роды *Yersinia*, *Streptococcus*. *Burkholderia*, *Chlamidia*, *Escherichia* и *Salmonella*. Отмечены их особенности и перечислены вызываемые ими заболевания. На странице 31 остался непереведённым с английского союз «and».

В целом, **Обзор Литературы** в точности соответствует сути проделанной работы, информативен и достаточно легко читается, хотя, как уже было сказано ранее, можно было бы пожелать автору чуть лучше его структурировать. Кроме того, можно было бы посвятить отдельный небольшой раздел мобильным элементам, которые, безусловно, играют важную роль в геномных перестройках и являются своего рода «ограничителями» геномных островков

В разделе «**Материалы и Методы»** детально описаны методы и подходы, использованные в работе. В работе было использовано достаточно внушительное число геномов – геномы 12 штаммов *Yersinia*, 103 штаммов *Streptococcus*, 127 штаммов *Burkholderia*, 127 штаммов *Chlamydia* и 41 штамм *Escherichia coli/Salmonella enterica*. Их полный список приведен в Приложении. В работе Ольгой Олеговной были использованы такие методы, как построение ортологических рядов и филогенетических деревьев с использованием различных алгоритмов (PhyML для далеких штаммов и RAxML для близких, а также MLGO для построения деревьев по положению синтенных блоков), построение синтенных блоков (также с использованием разных алгоритмов). Реконструкция геномных перестроек проводилась с использованием алгоритма MGRA, который позволяет реконструировать такие эволюционные события, как инверсии, делеции, инсерции и транслокации. Для реконструкции межхромосомных транслокаций была использована программа PAML версии 4.6. Точки начала и конца репликации были

определены с помощью функции GC-skew, что основывается на наблюдении о перепредставленности гуанина и цитозина на лидирующей и запаздывающей цепях бактериальных хромосом, соответственно. Такая методическая база полностью отвечает поставленным задачам и свидетельствует о высоком уровне аккуратности и планирования работы, что является её безусловным достоинством.

Раздел **«Результаты и их обсуждение»** состоит из трех глав. В первой из них описаны результаты по восстановлению филогении на основе расположения синтенных блоков в геномах *Yersinia pestis* и ее предка *Yersinia pseudotuberculosis*. В качестве внешнего штамма для укоренения был использован штамм *Yersinia enterocolitica*. Автором была установлено, что в ходе эволюции *Y. pestis* произошло множество геномных перестроек, при этом у *Y. pseudotuberculosis* их частота существенно ниже – кажется, что утверждение автора, что она «сохраняется» чересчур категорично. Филогенетическое дерево, построенное по расположению общих синтенных блоков с использованием метода максимального правдоподобия, выявило две больших общих инверсии – порядка 150 тысяч пар оснований для штаммов *Y. pestis* D 106004 и Z176003 и порядка 350 тысяч пар оснований для штаммов A1122 и D182038. Обе инверсии приводят к низкому уровню бутстреп-поддержки в соответствующих узлах. Добавление в анализ синтенных блоков, найденных только в части штаммов, не привело к изменению оптимальной топологии дерева, но повлияло на значения бутстреп-поддержки некоторых узлов, что может объясняться множественными параллельными делециями и инсерциями, типичными для быстро эволюционирующих недавно образованных патогенов. Делается вывод, что оптимальной стратегией для реконструкции эволюционных событий является именно сочетание двух алгоритмов и анализ полученных расхождений в топологии. Таким образом, на примере 12 геномов *Yersinia*, для которых в целом характерна высокая частота перестроек, было показано, что построение филогенетических деревьев с учетом расположения синтенных блоков позволяет уточнить эволюционную историю и разрешить узлы в кладах, формируемых недавно образованными штаммами. Корректность первого вывода не вызывает сомнений.

Вторая глава результативного раздела посвящена реконструкции геномных перестроек в различных видах *Streptococcus* и *Burkholderia*. Было показано, что частота геномных перестроек даже у близкородственных видов может различаться на порядки, отражая их адаптацию к разным условиям среды. Для недавно образованных патогенов характерна более высокая частота перестроек по сравнению с предковыми видами, которая, в том числе, связана с присутствием большого количества копий мобильных элементов. Именно накопление мобильных элементов и высокая частота рекомбинации могли сыграть важную роль в адаптации *Burkholderia mallei* к внутриклеточному стилю жизни за счет существенной редукции генома. Было показано, что при трансклокации между хромосомами действует отбор на сохранение положения генов в перенесенных сегментах относительно лидирующей/запаздывающей цепи. Были выявлены новые геномные островки и новые случаи антигенной вариации у данных патогенов. Важным результатом работы стало установление точки рекомбинации внутри генов, кодирующих белки PhtB и PhtD в геноме *Streptococcus* (отражено в выводах 3 и 4). С учетом довольно подробного обсуждения белков PhtD и PhtB, хотелось бы увидеть чуть более подробное описание их функции и более четкого объяснения, почему PhtD рассматривается в качестве одной из кандидатных мишеней для вакцины нового поколения. В подписи к таблице 4.1. не фигурирует видовое название анализируемого организма, называемого просто «пневомокком» - лучше было бы его добавить. Слово «бреакпоинт» в подписи к Рис. 4.2. можно было бы заменить на более литературное.

В третьей главе Результатов Ольга Олеговна описывает результаты анализа горизонтального переноса генов у *Escherichia coli*, *Salmonella enterica*, *Streptococcus sp* и *Chlamidia sp*. Поскольку построение точной филогении близкородственных штаммов является сложной задачей, то диссертантом анализировались периферийные гены,

присутствующие в геномах более, чем двух видов рода и более, чем трех штаммов каждого вида. Несогласованность дерева, построенного для конкретного гена, с общим деревом видов, говорит о том, что между штаммами, принадлежащими к разным видам, произошла рекомбинация. Пример такого дерева приведен ан Рис. 5.1. Для 16 штаммов *Salmonella enterica* и 25 штаммов *Escherichia coli* было выявлено 2858 периферийных ортологичных групп, 16 из которых имеют несогласованное филогенетическое дерево с «мозаичной» структурой и найдены в различных стабильных связках общих генов. Все эти гены являются редкими для исследуемых штаммов и встречаются менее, чем у половины штаммов каждого вида. Аналогичный анализ был проведен для представителей рода *Streptococcus* – была проанализирована 1401 ортологичная группа генов, 92 из которых имели мозаичную структуру дерева. Нахождение таких генов в различных стабильных связках общих генов подтверждает горизонтальный перенос. Как и в случае с *Escherichia*, более половины этих генов оказались редкими. Для геномов *Chlamydia* было протестировано 1046 ортологичных групп, и представители только 12 из них были найдены в различных стабильных связках. Был выявлен только один ортологичный ряд, в истории которого произошел независимый перенос гена, кодирующего белок внешней мембраны, в разные штаммы одного вида. Хотелось бы увидеть чуть больше информации об этом белке. Также в самом начале раздела 5.3. на странице 77, видимо, указаны ошибочные цифры – 2858 ортологичных групп вместо 1401.

В целом, применение предложенной автором модели показало, что разные виды бактерий имеют разную частоту потока генов, при этом наиболее открытый пан-геном характерен для штаммов рода *Streptococcus*, что хорошо согласуется с гипотезой о том, что пан-геномы с большей периферией характерны для организмов с большим эффективным размером популяции и способностью к быстрой адаптации. Таким образом, выводы 5 и 6 также полностью обоснованы.

Диссертационная работа хорошо оформлена. В ней достаточно мало опечаток и стилистических неточностей – почти все они были перечислены выше, как и смысловые замечания.

К общим техническим замечаниям можно отнести следующее:

1. В работе встречается некоторое количество жаргонизмов и неточностей. Так, например, часто встречаются фразы «бактериальные данные» и «молодые патогены», что, безусловно, является жаргонизмом.
2. В подписи к Рис. 1.1. в скобках сразу стоило бы ввести сокращение *E. coli* и далее использовать уже только сокращенное название. Это же касается и всех остальных видов бактерий, которые в тексте указаны вразнобой – то полностью, то сокращенно.
3. Довольно часто названия белков выделены курсивом (например, подпись к Рис. 1.5., AcpA и TonB на стр. 54, легенды к Рис. 4.16. и 4.17), что неверно.

Это нисколько не снижает хорошего впечатления от работы, которая является цельным, хорошо спланированным и законченным в рамках поставленных задач исследованием. Достоверность результатов никаких сомнений не вызывает, а выводы отражают суть полученных данных. Автореферат соответствует содержанию диссертации, а в двух публикациях автора и тезисах конференций отражены все ключевые результаты. По актуальности темы, объему и важности проведенных исследований и сделанных выводов рецензируемая работа, несомненно, соответствует требованиям «Положения о порядке присуждения учёных степеней», утверждённым Постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а сам диссертант, Бочкарёва Ольга Олеговна заслуживает присуждения учёной степени кандидата биологических наук по специальности 03.01.09 - «математическая биология, биоинформатика».

Отзыв составлен старшим научным сотрудником Лаборатории функциональной геномики и клеточного стресса, к.б.н., Шавкуновым Константином Сергеевичем.

Отзыв был заслушан и одобрен на семинаре Лаборатории функциональной геномики и клеточного стресса Института биофизики клетки - обособленного подразделения Федерального государственного бюджетного учреждения науки «Федеральный исследовательский центр «Пущинский научный центр биологических исследований Российской академии наук» (ИБК РАН) «26» апреля 2019 г. (Протокол № 37).

Старший научный сотрудник
Лаборатории функциональной
геномики и клеточного стресса,
к.б.н. Шавкунов К. С.

Сведения о составителе отзыва: Шавкунов Константин Сергеевич, кандидат биологических наук по специальности 03.01.03 – «молекулярная биология», старший научный сотрудник Лаборатории функциональной геномики и клеточного стресса Института биофизики клетки - обособленного подразделения Федерального государственного бюджетного учреждения науки «Федеральный исследовательский центр «Пущинский научный центр биологических исследований Российской академии наук» (ИБК РАН).
Адрес: 142290, г. Пущино, ул. Институтская, д. 3,
Телефон: 8(496)773-91-40,
E-mail: shavkunovks@gmail.com

Подпись Шавкунова К.С. удостоверяю

Зав. отделом кадров ФИЦ ПНЦБИ РАН Левченко Г. Н.

Сведения о лице, утвердившем отзыв: Грабарник Павел Яковлевич, доктор физико-математических наук по специальности 05.13.18. – «Математическое моделирование, численные методы и комплексы программ», ВРИО директора ФИЦ ПНЦБИ РАН.