Отзыв официального оппонента на диссертацию Бочкарёвой Ольги Олеговны «Реконструкция эволюционной истории геномных перестроек в бактериях», представленную на соискание ученой степени кандидата биологических наук по научной специальности 03.01.09 «Математическая биология, биоинформатика»

Диссертация Ольги Олеговны Бочкарёвой посвящена интереснейшей теме: исследованию геномных перестроек в бактериях. Известно, что перестройки являются одним из мощнейших движущих факторов эволюции, но в классических исследованиях при построении филогенетических деревьев роль перестроек обычно учитывается очень слабо, в частности это связано с недостаточной разработанностью математического аппарата и сложной структурой и историей перестроек в различных организмах. Поэтому затронутая тема является очень актуальной и позволяет добавить «новое измерение» в традиционные филогенетические реконструкции.

Работа изложена на 111 страницах и довольно четко структурирована. Диссертация содержит все необходимые разделы - «Введение», «Обзор литературы», «Материалы и методы», «Результаты» (логично разделенные на три смысловые группы) и «Выводы». Список литературы насчитывает 112 пунктов. По материалам диссертации опубликованы две статьи в международных высокорейтинговых журналах (причем в обеих статьях Ольга Олеговна является первым автором). Также результаты были представлены на 13 российских и международных конференциях.

Во введении простым и четким языком обозначена основная проблема, решаемая в данной работе, и у читателя не остается сомнений в актуальности выбранной темы. Говоря о практической ценности работы, автор утверждает, что «Работа имеет преимущественно теоретический характер», но подобные реконструкции могут иметь важное применение в медицине и эпидемиологии - известны примеры, когда инверсии приводили к образованию патогенных штаммов.

Обзор литературы представлен на 18 страницах. В первой части обзора автор делает интересный экскурс в историю изучения феномена геномных перестроек и описывает различные подходы к их анализу. Правда, остается непонятным, какие методы являются наиболее предпочтительными - видимо, байесовские, но из текста это напрямую не следует.

Вторая часть литобзора посвящена эволюции бактериальных геномов. В этой части подробно описано влияние геномных перестроек на адаптивность организмов. Приведены интересные примеры сайт-специфических инверсий, а также характеристики этих инверсий в геномном контексте. Также отдельно обсуждаются механизмы горизонтального переноса генов - пожалуй, наиболее функционально значимого процесса, связанного с изменением структуры генома. Затем автор вводит понятие кор- и пан-генома, и приводит данные о функциональных свойствах разных фракций генов.

В финальной части литобзора описаны пять видов бактерий, выбранных в качестве объектов изучения. Показательно, что каждый из них имеет свои особенности, связанные с геномными перестройками.

В целом обзор очень информативен. Правда, создается впечатление, что в работе решаются две принципиально различные задачи - исследование истории перестроек внутри вида и анализ потока генов из-за горизонтального переноса. Возможно, часть про горизонтальный перенос стоило вынести в отдельный раздел.

Раздел «Материалы и методы» написан достаточно подробно - для всех исходных данных приведены ссылки на базу данных NCBI. Для использованных программ указаны их версии и параметры, что позволяет заинтересованному читателю с легкостью воспроизвести приведенные результаты. Правда, почему-то сделано исключение для пакетов Interproscan и PhyML. Если актуальная версия PhyML 3.0 не менялась с 2010 года, то в случае Interproscan базы обновляются постоянно, и это может повлиять на полученный результат. В разделе «Построение синтенных блоков» не очень понятно, как определялись «виды с низким уровнем сходства генного состава» - в каких случаях использовалась Sibelia, а в каких DRIMM-Synteny. В разделе «Определение точек начала и конца репликации» из контекста непонятно, для чего в дальнейшей работе это будет использовано. «*Значения функции были вычислены с помощью библиотек на BioPython*» - возможно, имело бы смысл привести используемые скрипты в приложении (если, конечно, это подразумевает формат диссертации). Также стоит отметить грамотный подход к выдвижению и тестированию статистических гипотез. В целом описанные методы отвечают поставленным задачам, и автор демонстрирует знание наиболее актуальных подходов к заявленной теме.

Первая глава раздела «Результаты» посвящена восстановлению филогении на основании положения синтенных блоков. Здесь в качестве модели выступают штаммы *Y. pestis*, отличающиеся высокой частотой рекомбинаций. Автор делает наглядное сравнение двух подходов к восстановлению филогении. Методически вопросов не возникает, все параметры и аутгруппа выбраны корректно.

Автор демонстрирует границы применимости метода на примере группы молодых штаммов *Y. pestis*. Восстановление филогенетической сети позволило ей разрешить узлы с низкой бутстреп-поддержкой на дереве, полученном по выравниванию генов. Тем не менее, уже на этом примере показана параллельная инверсия (350 т.п.н) в штаммах *Y. pestis* A1122 и *Y. pestis* D182038. Эта инверсия значительно снижает бутстреп в соответствующих узлах в дереве, построенном по истории инверсий. Автор совершенно справедливо замечает, что независимые инверсии в некоторых случаях могут давать ложный сигнал, и поэтому данные об истории перестроек должны использоваться совместно с данными о нуклеотидных выравниваниях, с «ручным» анализом различий в топологии.

Приятно отметить, что Ольга Олеговна учла влияние выбранного порогового значения на результат предсказания синтенных блоков и провела реконструкцию для блоков разной длины. Это привело к обнаружению новой параллельной инверсии в штаммах *Y. pestis* KIM и *Y. pestis* Nepal. В целом, корректность выводов в этом разделе не вызывает сомнений. Из замечаний - возможно, имело бы смысл для восстановления филогении по ортологичным генам использовать не просто конкатенат выравниваний, а отдельные модели для каждого гена.

Вторая глава раздела «Результаты» посвящена реконструкции истории перестроек в различных видах *Streptococcus* и *Burkholderia*, каждый из которых представляет в этом плане особый интерес. У исследуемых видов количество перестроек недостаточно для восстановления филогении, поэтому автор рассматривает перестройки в контексте дерева, полученного на основе выравнивания ортологов. Автор реконструирует значительное количество инверсий, в т.ч. параллельных. Для параллельных инверсий автор выполнила филогенетический анализ генов внутри инверсий и опровергла гипотезу об их горизонтальном переносе.

Интересен показанный пример инверсии с вариацией внутри генов PhtD и PhtB - в данном случае в качестве повторяющихся элементов на границах инверсий выступают гены-паралоги. Было бы интересно рассмотреть другие примеры подобных фазовых вариаций. Для стрептококков ожидаемо и наглядно подтвержден отрицательный отбор на инверсии внутри одной реплихоры с корректным статистическим обоснованием.

Для видов *Burkholderia* с несколькими хромосомами автор продемонстрировала несколько случаев независимого переноса генов между хромосомами. Для *B. mallei* продемонстрированы множественные делеции во вторых хромосомах относительно *B. pseudomallei*, что связано с переходом к внутриклеточному образу жизни. Также продемонстрирован отрицательный отбор на инверсии внутри реплихор. Для *B. cepacia* в ходе анализа синтенных блоков выявлено два геномных островка, переносящихся горизонтально и несущих наборы функциональных генов. Из замечаний, в подписи к рисунку 4.15 фраза «Цветными треугольниками помечены штаммы, в которых детектированы одинаковые инверсии», видимо, должна относиться к рисунку 4.14.

Финальная глава результатов посвящена анализу потока генов между бактериальными геномами. Ольга Олеговна подробно описывает разработанную ею модель детектирования горизонтального переноса генов. Модель была корректно протестирована на видах *Escherichia coli* и *Salmonella enterica* и использована для анализа видов *Streptococcus* spp. и *Chlamydia* spp. - с открытым и закрытым пан-геномом соответственно. Было ожидаемо продемонстрировано, что в случае открытого пан-генома более 10% периферийных генов распространяются в популяции горизонтально, позволяя адаптироваться к различным эволюционными нишам. В то же время для закрытого пангенома горизонтальный перенос практически отсутствует, что соответствует изолированному внутриклеточному образу жизни.

Выводы диссертации хорошо и четко сформулированы и соответствуют поставленным задачам. В целом диссертация внушает уважение объемом проделанной работы, выдвинутых и проверенных гипотез, разработанных методов и новых результатов. Читать её легко и приятно, она написана понятным языком, выдержана в едином стиле и очень хорошо оформлена. Большая часть полученных результатов сопровождается наглядными иллюстрациями.

Существенных замечаний и претензий к научной составляющей работы у меня нет. Все замечания, рекомендации и претензии по оформлению были приведены выше и нисколько не снижают хорошего впечатления от работы.

Автореферат соответствует содержанию диссертации, все ключевые результаты отражены в публикациях и тезисах конференций. По актуальности темы, объему и

важности проведенных исследований и сделанных выводов рецензируемая работа, несомненно, соответствует требованиям «Положения о порядке присуждения ученых степеней», утвержденным Постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а сам диссертант заслуживает присуждения ученой степени кандидата биологических наук по специальности 03.01.09 - «математическая биология, биоинформатика».

Райко Михаил Петрович,
к.б.н.,
старший научный сотрудник,
лаборатория «Центр Алгоритмической Биотехнологии»,
Федеральное государственное бюджетное образовательное учреждение
высшего образования «Санкт-Петербургский государственный университет»

29.04.2019

ЛИЧНУЮ ПОДПИСЬ
Райко М. П.
ЗАВЕРЯЮ

ЗАМЕСТИТЕЛЬ НАЧАЛЬНИКА
УПРАВЛЕНИЯ КАДРОВ СПбГУ
Н. К. КОРЕЛЬСКАЯ