

## О Т З Ы В

официального оппонента на диссертационную работу  
Сеплярского Владимира Борисовича  
на тему «Паттерны и молекулярные механизмы мутагенеза у эукариот»,  
представленную на соискание ученой степени  
кандидата биологических наук  
по специальности 03.01.09 – математическая биология, биоинформатика.

Работа Сеплярского В.Б. находится на стыке эволюционной биологии, геномики и биоинформатики и посвящена исследованию статистических особенностей мутационных процессов у высших эукариот. Понимание механизмов и особенностей мутагенеза открывает широкие перспективы для функциональной аннотации и предсказания потенциально болезнетворных мутаций, способствующих прогрессии различных заболеваний, в том числе, опухолевой трансформации клеток. Особый интерес представляет изучение распределения и особенностей горячих точек мутагенеза в геноме, что и является объектом исследования представленной диссертационной работы.

В первую очередь, хочется отметить широкий круг вопросов, освещаемых в диссертации. Автором (глава 1 диссертации) исследована локальная неравномерность скорости мутагенеза у высших эукариот с точностью до отдельных нуклеотидов, в частности, для *Drosophila* такой анализ проведен впервые. Более того, в главе 2 проведен более глубокий анализ, а именно оценка относительных частот транзиций и трансверсий, с учетом информации о локальной неравномерности мутагенеза. Выяснилось, что частота транзиций по отношению к трансверсиям в целом снижена в ближайшей локальной окрестности активно мутируемых геномных сайтов, и особенно снижена в окрестности трансверсий. Это говорит о том, что события трансверсий, замены пуриновых оснований на пиримидиновые и наоборот, склонны кластеризоваться; причем, например, для *D. melanogaster* такие мутагенные участки достигают малых десятков нуклеотидов. В главах 3-4 проведен глубокий анализ мультинуклеотидных замен и особенностей работы полимеразы дзета. Здесь хочется особенно отметить анализ конкретных случаев, когда динуклеотидные мутации ассоциированы с развитием заболеваний: синдрома Костелло и амиотрофического латерального склероза. Этот



раздел изящно сводит воедино результаты всей диссертационной работы, поскольку на конкретном примере обсуждает и роль полимеразы дзета, и взаимосвязь динуклеотидных мутаций и горячих точек мутагенеза, вызванных GC-парами.

На протяжении всего текста автор демонстрирует остроумное применение методов сравнительной геномики и уверенное владение сопутствующими статистическими методами. Полученная в ходе работы информация о локализации различных типов горячих точек мутагенеза сама по себе имеет прямую ценность и должна позволить детальное исследование особенностей мутационных контекстов и потенциально сделать возможной декомпозицию «мутационной подписи» по конкретным молекулярным механизмам, вносящим конкретные типы мутаций. Наконец, автором сделано замечательное наблюдение о схожих глобальных свойствах мутагенеза между грибами и высшими эукариотами (*H. sapiens* и *S. commune*), и факт, что результаты автора об активности «горячих точек» подтверждаются на столь удаленных организмах, кажется крайне важным.

Тем не менее, непосредственно текст диссертации до определенной степени контрастирует с высоким научным уровнем работы. В первую очередь, по диссертации возникает ряд фактических вопросов, не нашедших в тексте прямого отражения. Например, на странице 4 автор пишет «крупнейшая база данных ... TCGA ... включает 10247 полных экзомов». **О каком именно релизе TCGA идет речь? Имеются ли в виду пары экзотов (норма-опухоль) или полное число записей в базе данных?**

По тексту работы в подписях к рисункам многократно используется фраза «доверительные интервалы по 1000 испытаний». **Можно предполагать, что авторы сэмпировали изучаемые выборки нуклеотидных замен и оценивали стабильность наблюдаемых отклонений в исследуемых частотах, но было бы грамотно дать четкое описание используемой процедуры.** На странице 20 «Материалов и методов» автор пишет, что из анализа исключался ряд регионов, в том числе нетранслируемые районы РНК. **Не очень понятно, почему проксимальные промоторные области генов не были в свою очередь исключены.**

Окончание подраздела 1 раздела «Материалы и методы» вызывает массу вопросов. Автор пишет: «По аннотации референтного генома мы нашли 7,987 генов. Мы перевыравнивали эти гены программой MACSE [74]. Для анализа мутирования у *S. commune* мы использовали 4-



кратно вырожденные синонимичные сайты». **Что значит «по аннотации генома мы нашли гены»? Для какой цели было проведено «перевыравнивание» программой MACSE (и что такое «перевыравнивание»?).** Почему для анализа мутагенеза выбраны сайты в кодирующих областях, если основная работа на *H. sapiens* и *D. melanogaster* проведена на некодирующих областях?

Автор существенно злоупотребляет жаргоном, текст богат на стилистические ошибки, вот лишь краткий список авторских находок: «время репликации было взято из статьи», «скорость однонуклеотидных мутаций», «наиболее прямые данные», «мутации часто недостаточно высокого качества». Налицо путаница с англоязычной и русскоязычной терминологией, наряду с «базой данных по раковым сиквенсам» автор говорит о «механизмах удвоения или починки ДНК». На странице 18 упоминается «подверженная ошибкам» полимеразы, а в следующем абзаце речь идет уже о «неточной (low fidelity)» полимеразе. Стоило явным образом рассказать об именуемых определениях и терминах и выбрать один вариант для сквозного использования в тексте работы.

Встречаются и совершенно замечательные опечатки, так, на странице 12 читаем: «Для всех известных эукариот, с измеренной  $k$ ,  $k > 1$ , кроме *Podisma pedestris*, для которого  $k=1.13$ » (в этом случае, к счастью, из сопутствующего текста понятно, что речь идет не просто об отличии от единицы, а о статистически достоверном отличии). По всему тексту «дезаминирование» у автора многократно упоминается как «деаминирование» (что, вероятно, допустимо, но в русскоязычной литературе употребляется значительно реже, чем «дезаминирование»), а иногда и как «деаминирвание»; рисунок 1 согласно подписи содержит «схематическое изображение транзиций», случающихся (согласно рисунку) с «пирамидами»; в подписи к рисунку 12 читаем про «повреждения, накопившиеся из-за транскрипции».

Многократно по тексту работы используются фразы вида «данные были загружены с UCSC». UCSC – университет в Санта Круз, Калифорния, поддерживающий одну из наиболее крупных геномных баз данных UCSC Genome Browser, говорить о «скачивании» данных с университета по меньшей мере некорректно. Гораздо более наглядно и грамотно было бы всякий раз говорить о конкретном геноме и версии использованной геномной сборки и аннотации. Аналогичная проблема возникает и с авторской формулировкой о данных



«загруженных из Galaxy bioinformatics». Galaxy – система организации веб-интерфейсов и последовательной обработки данных для задач биоинформатики. Она может быть установлена на любом подходящем сервере и адаптирована для широкого круга задач. **Необходимо всегда уточнять, какой именно сервер под управлением Galaxy был использован в каждом конкретном случае.**

Структура текста в ряде случаев не выдерживает критики. Так, на странице 5 читаем: «Наше исследование сосредоточено на хорошо изученных модельных организмах – *Drosophila melanogaster* и *Homo sapiens* с использованием геномов ближайших к ним видов». Но продвинувшись в «материалы и методы» на 22 странице внезапно появляется анализ мутагенеза у Щелелистника обыкновенного (*S. commune*) и обсуждаются технические детали секвенирования и сборки генома. **Из текста трудно понять, какое отношение технические подробности получения геномной сборки имеют к последующему анализу мутагенеза.** На странице 40, где ожидаются результаты, дан достаточно пространственный фрагмент методического текста, который было бы логично видеть в соответствующем разделе.

К сожалению, налицо и множество простых небрежностей в работе над текстом. Например, в тексте отсутствует сквозная нумерация разделов, диссертация (но не автореферат) лишена отдельного раздела, формулирующего «цели и задачи исследования» и подчеркивающего «актуальность темы». Глава 3 начинается с туманной фразы, которая должна объяснять вклад автора в решение конкретной задачи; то, что вклад автора явным образом обозначен, хорошо; хуже, что в этой вводной фразе сразу две опечатки. Есть в тексте и более грубые оплошности, так, **выводы, приведенные в автореферате и диссертации, имеют одинаковый смысл, но по неизвестной причине дословно не совпадают.** Например, «связана со временем репликации» (вывод 5, автореферат) и «связана со временем репликации, данного сегмента ДНК» (вывод 5, диссертация); а также «полимераза зета делает мутации» (вывод 7, автореферат) и «полимераза зета приводит к мутациям» (вывод 7, диссертация). Трудно сказать какая из версий является более удачной, но формулировка «делает мутации» выглядит вульгарно. Вообще, было бы грамотно что-то конкретное сказать о типе «связи». Вывод 6 говорит об «интенсивности» работы полимеразы зета, но слово «интенсивность» нигде ранее в работе не упоминается. Там же в выводах читаем о «скорости мутирования на уровне однонуклеотидных позиций», что кажется просто неудачной калькой с

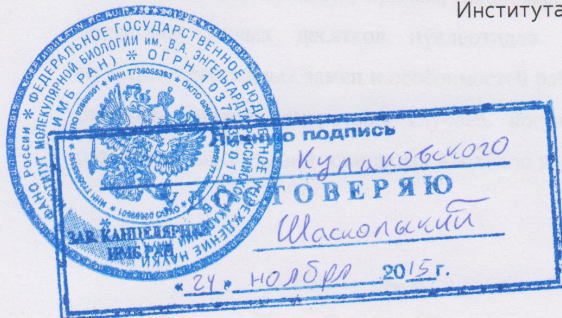


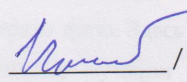
английского. Некоторые формулировки в принципе не позволяют однозначно интерпретировать смысл сказанного без детальной консультации с текстом диссертации, например, вывод 1 «SNP попадает в сайт, содержащий другой SNP» и вывод 2 «в человеке доля трансверсий растёт втрое для сайтов, содержащих другую трансверсию» страдают от отсутствия прямого определения понятия «сайта» (как может мутация попадать в то место, где уже «содержится» другая мутация»). В силу слабых формулировок выводов особенно ощущается нехватка раздела «закключение», который суммировал бы результаты, представленные в диссертации.

**Наконец, список публикаций по теме диссертации лишен номеров выпусков и страниц в журналах, идентификаторов doi и прочих формальных данных. Статьи автора, безусловно, находятся поиском в системах индексирования PubMed и Web of Science, но отсутствие прямых выходных данных выглядит неуважительно.**

Обсуждаемые выше недостатки, конечно, снижают общее впечатление от представленной диссертационной работы. Тем не менее, они никоим образом не позволяют поставить под сомнение ее высокий научный уровень. На мой взгляд, результаты работы представляют большой интерес и хорошо обоснованы, неоспорим личный вклад соискателя, содержание диссертации достойно представлено в высокорейтинговых научных публикациях по теме. Автореферат отражает содержание диссертации. Таким образом, работа Сеплярского Владимира Борисовича соответствует требованиям «Положения о присуждении ученых степеней», утвержденным Постановлением № 842 Правительства РФ от 24 сентября 2013 г. В свою очередь, автор, Сеплярский В.Б., определенно заслуживает присуждения ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

кандидат физико-математических наук, старший научный сотрудник  
Лаборатории вычислительных методов системной биологии  
Федерального государственного бюджетного учреждения науки  
Института молекулярной биологии имени В.А. Энгельгардта  
Российской академии наук (ИМБ РАН)



 / И.В. Кулаковский /  
24 ноября 2015 года