

РЕЦЕНЗИЯ

на выпускную квалификационную работу студента Д.А. Бибы «Поиск скомпенсированных сдвигов рамки считывания и определение их роли в формировании белковых последовательностей»

Как возникают новые белок-кодирующие последовательности в эволюции? Это вопрос, который интересует многих эволюционных биологов. С обилием геномных и транскриптомных данных ответ на этот вопрос постепенно перемещается из плоскости чисто теоретической и даже порой спекулятивной в плоскость эмпирическую, где высказанные предположения можно проверить на полногеномном уровне. В выпускной квалификационной работе Дмитрия Андреевича Бибы рассматривается один из важных аспектов этого вопроса: как эволюционируют белковые последовательности после вставок/делеций (инделов), вызвавших сдвиг рамки считывания, при этом скомпенсированных расположенными рядом инделами? Как следствие, в части нуклеотидной последовательности меняется рамка считывания, в результате на белковом уровне мы имеем дело с совершенно другой последовательностью. Открытым остаётся вопрос, насколько часты такие события в эволюции и к каким последствиям на уровне белковой последовательности они приводят, а также как указанные события связаны с нуклеотидными заменами, происходящими неподалёку. Автор систематически исследовал случаи скомпенсированных сдвигов рамки считывания на дереве позвоночных и дереве двукрылых (100 и 124 полных геномов соответственно).

Цель и основные задачи ясно сформулированы, актуальность темы не вызывает сомнений. Литературный обзор весьма подробный, его было интересно читать, однако большую его часть занимает описание моделей эволюции дублицированных генов, что не является непосредственной темой исследования, эту часть на мой взгляд можно было бы сократить. В разделе «Материалы и методы» содержится описание оригинального алгоритма автора по поиску скомпенсированных сдвигов рамки считывания, что само по себе является большим методологическим достижением. Работа в целом выполнена на высоком научном уровне, последовательно изложена. Стоит отметить местами весьма неформальный стиль изложения автора, что хоть и придаёт живости повествованию, иногда сбивает с толку (см. ниже).

К работе есть несколько замечаний и вопросов.

В разделе, описывающем взаимодействие скомпенсированных инделов с однонуклеотидными заменами, рассматриваются две крайние ситуации: однонуклеотидные мутации произошли после инделов, либо до инделов во времени. Однако, в реальности возможна ситуация, когда часть из мутаций произошла до инделов, а часть позже, а может и между ними. Не может ли это повлиять на результаты, представленные на рис. 17 и 18? Вероятно, более точно было бы рассмотреть все возможные последовательности замен и инделов?

В литературном обзоре на стр. 12 читаем: «кодирование двух белков одним участком может быть полезно, если оба эти белка вовлечены в разные пути одного процесса. Тогда путем модификации иРНК можно быстро переключать эти пути, синтезируя разные белки (Szklarczyk et al., 2007).» Из сказанного лично мне не

понятно, в чем состоит польза двойного кодирования, по сравнению скажем с двумя отдельными генами в данном случае?

В описании алгоритма (стр.19, п.15) говорится, что длинные инделы, обладающие некоторыми характеристиками, выбрасываются, т.к. предположительно являются артефактами выравнивания. Во-первых, непонятно зачем нужен этот шаг, когда в дальнейшем рассматривается только короткие инделы. Во-вторых, не ясно, что значат термины “базальное животное” и “родительский узел животного”?

В «Результатах» на стр. 23 упоминается программа PAML, используемая для оценки отбора, однако не приводятся параметры её запуска, этой информации также не содержится и в разделе “Материалы и методы”.

На стр. 25 выдвигается гипотеза: “сдвиг рамки скорее произойдет в консервативных белках, чем в неконсервативных”. На чем основано это предположение? С первого взгляда кажется, что ситуация должна быть обратной.

На стр. 25 утверждается: “Видно, что, хотя белки со скомпенсированным сдвигом рамки находятся не вблизи экстремальных значений консервативности, они все же чаще менее консервативны, чем в среднем”. Из рис. 12 следует, что это верно для 7 из 11 генов, что близко к случайному совпадению.

В «Обсуждении» Дмитрий Андреевич анализирует результаты эффекта мутаций и прибегает, по его собственному выражению, к наивной нулевой гипотезе: “мутации с одинаковой вероятностью могут как уменьшать расстояние до предкового состояния, так и увеличивать его”. Действительно, хотелось бы понять, насколько наблюдаемые паттерны мутаций не случайны. Наверное, стоит в будущем подумать о более аккуратном контроле, учитывающем локальную нуклеотидную последовательность и свойства генетического кода, хотя это конечно, не является претензией, а скорее предложением дальнейшего развития работы.

Также в работе встречаются мелкие опечатки и неточности. Например ген *Diar6* вместо *Diar1* (стр. 23). На подписи к рис. 17: “красное” расстояние > “зеленое”

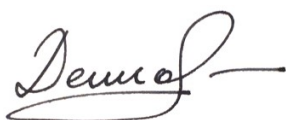
расстояние ⇒ мутации компенсаторны”. Видимо под “зеленым” расстоянием имеется в

виду “синее” расстояние.

Высказанные замечания не умаляют высокого качества работы, адекватности постановки задачи и важности исследования. Отдельно хотелось бы подчеркнуть нестандартность работы и то, что для ее выполнения использовался разработанный автором алгоритм, реализованный на языке Python.

Дмитрий Андреевич Бибабезусловно заслуживает присуждения степени бакалавра, а его работа заслуживает оценки “отлично”.

Научный сотрудник
Манчестерского университета
к.б.н.



С.В. Денисов

Division of Evolution & Genomic Sciences
Faculty of Biology Medicine and Health Sciences
The University of Manchester
email: stepan.denisov@manchester.ac.uk
tel: +44 7947 154046