

ОТЗЫВ

официального оппонента на диссертационную работу

Солдато́ва Руслана Андреевича

на тему «Методы предсказания структурных элементов РНК»,

представленную на соискание ученой степени

кандидата физико-математических наук по специальности 03.01.09 -
математическая биология, биоинформатика.

Работа Солдато́ва Руслана Андреевича представляет собой методологическое исследование подходов по предсказанию структурированных участков РНК в геномных последовательностях. В последнее время представление о роли молекул РНК в клеточных процессах претерпело значительное изменение. Так, в последние десять лет был открыт ряд классов некодирующих структурированных РНК, играющих важную роль в различных клеточных процессах. Основные базы данных структурированных РНК, такие как Rfam или miRbase, неуклонно расширяются за счёт новых аннотированных элементов РНК. С другой стороны, методы высокопроизводительного прочтения ДНК привели к появлению большого количества секвенированных геномов и транскриптомов, в связи с чем образовалась проблема, связанная как с аннотацией структурированных участков РНК в геномных последовательностях, так и с выявлением их функциональной роли. С этой точки зрения, работа Руслана Андреевича является актуальным и важным исследованием в области разработки новых стратегий к предсказанию структурированных участков РНК.

Диссертационная работа Солдато́ва Р.А. имеет традиционную структуру: содержит обзор литературы, две главы с описанием методов и результатов, а также выводы и список публикаций по теме диссертации.

В главе "Обзор литературы" достаточно полно представлены методы анализа вторичной структуры РНК, сильные и слабые стороны современных подходов. Автор делает обзор диффузионных моделей эволюции количественных характеристик.

В Главе 2 «RNASurface: эффективный алгоритм предсказания локально-оптимальных структурированных РНК» Руслан Андреевич указывает на существенный недостаток использования размера "окна" при предсказании структурированных участков РНК и вводит определение «локально-оптимального структурированного сегмента РНК». Это соображение актуально, так как в общем случае ни локализация, ни размер структурного элемента РНК неизвестны, а известные структурированные элементы РНК имеют широкий диапазон величин размеров. Автор предлагает новые алгоритмические и статистические методы для обнаружения структурированных сегментов и демонстрирует преимущества своего метода. Кроме того, автор предлагает эффективную практическую реализацию в виде программы RNASurface и описывает её применение к геному *Bacillus subtilis*.

В Главе 3 «Сравнительно-геномный метод предсказания функциональных элементов» автор предлагает метод оценки статистической значимости консервативности структурированности набора ортологичных последовательностей. Основной идеей является сравнительно-геномный анализ Z-значений и исследование консервативности Z-значений, а не просто нуклеотидных последовательностей. Безусловной новизной здесь является использование диффузионных процессов для описания эволюции Z-значений. Примечательно то, что автор показал, как эволюция Z-значений может быть описана специальным типом диффузионного уравнения (процессом Орнштейна-Уленбека). Удивляет, что этот факт не нашёл отражения в выводах работы. Руслан Андреевич продемонстрировал преимущества разработанного диффузионного подхода на примере анализа аннотированных некодирующих РНК в геномах дрожозил.

Разработанные подходы обладают новизной и могут иметь широкое практическое применение, а также показывают глубокое знание алгоритмических, эволюционных и статистических методов диссертантом. Однако, к работе есть ряд замечаний и вопросов.

В частности, использование термина «структурная РНК» стилистически неверно, при этом используется часто по всему тексту диссертации и автореферата. Правда, имеет место упоминание правильной терминологии: «структурированные участки РНК».

Цитирую на стр. 27: «имеет значение энергетическая поверхность структур» - непонятно, что автор имеет в виду; «Z-значение лучше всего и достаточно полно отражает термодинамические свойства известных структурных РНК». Автор должен дать пояснение, как Z-значение отражает термодинамические свойства известных структур РНК. На той же странице в заголовке «Поиск структурированных участков в последовательностях» не сказано о каких последовательностях идёт речь. Много предложений где пропущены слова и не стыкуются падежи, например, «В этом случае пики значения Z-значения будут соответствовать наиболее структурированным сегментам.»

Стр. 44 «Отдельный вид или клада могут иметь другой оптимальный» - Что автор имел в виду?

Подпись к рисунку 1.3.3. дана с ошибкой.

В Автореферате: «Аккуратное и быстрый метод вычисления Z-значений». Это заголовок и он написан с ошибкой. Подпись к Таблице 1 дана с ошибкой. Расположение рисунков и таблиц в Автореферате не продумано. *Vacillus subtilis* написано без курсива на стр. 11 автореферата.

Рисунки 2.2.4., 2.2.5 в диссертации и рисунки 3, 7 в Автореферате плохого качества: линии на рисунках очень тонкие, а подписи очень мелкие.

Таблица 2 в диссертации и Таблица 1 в автореферате: вместо $Z=-3$ стоит $Z=-2$. Что дано в таблице в скобках – непонятно, если это доля, то тогда не сходится с графой «Всего РНК». Там же в подписи к таблице: «стратифицированных по классам» .

Термин «либеральный порог» стр. 75 диссертации и «либеральный FPR в 20%» на стр. 69 – неуместен.

Термин «статистическая мощность программ» для данного класса задач неуместен, так как все программы работают с большими ограничениями, и, как замечает сам автор, их метод тоже имеет «слабую предсказательную мощность на масштабе генома».

Можно приводить список неточностей, стилистических и орфографических ошибок, но я ограничусь вышеперечисленными ошибками.

Может ли автор работы дать своё объяснение, почему одни структурированные элементы РНК предсказываются хорошо, а другие плохо? Следует отметить, что рРНК и тРНК, участвующие в процессе синтеза белков, предсказываются плохо, в среднем 16%. Также автор отмечает, что тРНК является мало структурированной относительно других классов РНК (стр.25). Хотя мы хорошо знаем, что вторичная структура тРНК имеет строго определённую структуру типа клеверного листа. Это можно объяснить, скорее всего, тем, что вторичная структура формируется одновременно с пространственной структурой, как это было показано в работе для тРНК^{Asp} (Wilkinson et al., RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. J. Am. Chem. Soc. 2005, 127, 4659–4667), и такие взаимодействия не учитываются при подсчёте свободной энергии тРНК.

Как изменятся результаты расчётов, если использовать другие энергетические параметры для расчёта вторичной структуры РНК? Почему программа RNASurface была применена только к одному геному *Bacillus subtilis*?

Мной не найден интернет-адрес программы RNASurface для пользователей, что дало бы возможность провести часть пробных вычислений.

Недостатки, описанные в отзыве, затрудняют понимание и интерпретацию разработанных методов. С другой стороны, они не ставят под сомнение высокий методический уровень работы, новизну полученных результатов и их практическую важность. Личный вклад соискателя несомненен, а содержание работы отражено в двух публикациях по теме диссертации. Автореферат отражает содержание диссертации. Диссертационная работа Солдатова Руслана Андреевича "Методы предсказания структурных элементов РНК" является законченным научно-квалифицированным исследованием и соответствует всем требованиям "Положения о порядке присуждения ученых степеней", утвержденного Постановлением Правительства РФ от 24 сентября 2013 г. №842. Автор заслуживает присуждения ученой степени кандидата физико-математических наук

