**ОТЗЫВ**

**члена диссертационного совета на диссертацию Алексея Александровича Гуревича «Вычислительные методы для анализа подверженных ошибкам метабологеномных данных», представленную на соискание ученой степени кандидата физико-математических наук по специальности 03.01.09 – Математическая биология, биоинформатика**

Диссертация посвящена актуальным проблемам алгоритмической биоинформатики — сборке геномов и метагеномов и масс-спектрометрии коротких рибосомных и нерибосомных пептидов, содержащих модифицированные и нестандартные пептиды и потенциально имеющих нелинейную структуру. Основные результаты изложены в четырех главах, соответствующих четырем главным публикациям, сделанным в международных журналах высокого уровня. Эти главы естественным образом можно объединить в две части, каждой из которых, в принципе, хватило бы для кандидатской диссертации. Помимо основных статей, составляющих основу рассматриваемой диссертации, А. А. Гуревич является соавтором еще пяти статей, в которых было использовано созданное им программное обеспечение.

На рассмотрение представлен текст на английском и русском языках. Выборочный анализ показывает, что эти тексты соответствуют друг другу, а английский текст, как и следует, соответствует англоязычным публикациям, на основе которых написана диссертация. В тексте имеется ремарка (*перевод с английского мой — А.Г.*). Смысл такого двуязычия от меня ускользает; видимо, это как-то связано с правилами защит в СПбГУ. Нижеследующая рецензия относится к русскому варианту.

Первые две главы посвящены описанию созданного автором конвейера для оценки качества геномных сборок. Хотя эта работа носит в значительной степени технический характер, она крайне актуальна, поскольку опубликовано множество различных программ сборки геномов (в т.ч. единичных клеток) и метагеномов, и у исследовательского сообщества имеется насущная нужда в их сравнении. Важность этой части работы подтверждается актуальной востребованностью созданного А. А. Гуревичем

вх 09/2-230 от 27.11.18

программного конвейера, который не только обширно цитируется (более 500 ссылок), но и используется как основной инструмент в международном соревновании по сборке геномов CAMI и сравнительных исследованиях типа GAGE-B.

Третья и четвертая главы описывают программные инструменты для анализа пептидов по данным масс-спектрометрии. Автором в значительной степени решена задача опознания пептида сложной структуры и состава, если в базе данных содержится похожий, но не идентичный пептид. В результате удалось найти более 19 тысяч новых вариантов, относящихся в более 2 тысяч пептидов.

Замечания к работе носят исключительно редакционный характер. Автор не до конца справился с задачей перевода английских терминов на русский язык (точнее, создания соответствующих русских терминов). В результате в тексте диссертации появились *мисассемблы*, *ложная база*, *грибковые источники* и *одноклеточное секвенирование*. Диссертация начинается словами: *Биоинформатика обрабатывает различные виды биологических данных: ДНК, РНК, белки и метаболиты. Эти вещества изучаются соответствующими направлениями исследований в молекулярной биологии...* Во-первых, ДНК и пр. — это именно *вещества*, а никак не *данные*. Во-вторых, нельзя сказать, что что-то *изучается направлением исследований*. Впрочем, следует отметить, что подобных стилистических огрехов в тексте немного.

В подписи к рис. 1.7 указан диапазон значений GC-состава *от 0% (белый) до 54% (черный)* — это выглядит несколько странно: на каком окне усреднялись значения? В разделе 2.3.2 упоминаются дефекты сборки тестового штамма относительно эталонного: идет ли речь об инверсиях или о вставках и потерях? Это было бы полезно понимать, потому что штаммы кишечной палочки могут сильно различаться по наличию фрагментов, но при этом в их геномах относительно немного инверсий. В подписи к рис. 3.3 упоминаются сплошные и пунктирные линии: я вижу красные и синие.

Я не вполне уверен, что автор к работе можно применить термин *метабологеномика*. Под ним понимается анализ метаболизма в привязке к геному, метаболическая реконструкция по (мета)геномным последовательностям, идентификация генов, отвечающих за те или иные

метаболические пути и подсистемы, в т.ч. за синтез конкретных веществ. Автором же выполнены исследования в области геномики и метаболомики отдельно; лишь в отдельных местах имеются редкие попытки синтеза (например, объяснение состава некоторых пептидов путем анализа доменной структуры ферментов, ответственных за их синтез). Я понимаю, что таким образом была сделана попытка подчеркнуть методическое единство работы, но это можно было бы сделать другими средствами, в частности, посвятив в самом тексте большее внимание общности примененных алгоритмических подходов.

Повторю, что сделанные замечания ни в коей мере не умаляют актуальности, ценности и научного уровня рассматриваемой работы. Диссертация А. А. Гуревича «Вычислительные методы для анализа подверженных ошибкам метабологеномных данных» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете». Пункт 11 указанного Порядка диссертантом не нарушен. Соискатель А. А. Гуревич заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 03.01.09. – Математическая биология, биоинформатика.

Член диссертационного совета

доктор биологических наук, профессор, заместитель директора по научным вопросам Федерального государственного бюджетного учреждения науки Институт проблем передачи информации им. А. А. Харкевича Российской академии наук

М. С. Гельфанд

11.11.2018

Подпись
Гельфанда М.С.
УДОСТОВЕРЯЮ
Зав. канцелярией