



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ  
ИНСТИТУТ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ  
им. А.А. Харкевича Российской академии наук

---

Большой Каретный пер., д. 19, стр. 1, Москва, ГСП-4, 127994  
Тел.: (495) 650-42-25. Факс: (495) 650-05-79. E-mail: director@iitp.ru  
ОКПО: 02699464 ОГРН: 1037700064940 ИНН/КПП: 7707020131/770701001  
<http://www.iitp.ru>

### ОТЗЫВ

**члена диссертационного совета на диссертацию Сергея Юрьевича Нурка «Сборка геномов некультивируемых микроорганизмов по данным высокопроизводительного секвенирования», представленную на соискание ученой степени кандидата физико-математических наук по специальности 03.01.09 — Математическая биология, биоинформатика**

Диссертационная работа С. Ю. Нурка сделана в актуальной области алгоритмической биоинформатики — сборке геномов, причем в самых «горячих» ее направлениях, сборке геномов по данным секвенирования отдельных клеток и сборке геномов по данным секвенирования метагеномов. Разработанные автором программы широко используются и стали фактическим стандартом: статьи с участием автора, в которых описан SPAdes, процитированы суммарно более трех с половиной тысяч раз, а статья 2017 года про metaSPAdes уже набрала около сотни цитирований. Обе программы широко используются как отдельными исследовательскими группами, так и большими международными консорциумами.

Собственно говоря, сказанного более чем достаточно, чтобы однозначно положительно оценить представленную работу. Тем не менее, для соблюдения формальностей приведу более подробный разбор.

Диссертация состоит из трех глав, из которых первая является кратким введением, а две другие излагают основные результаты: описание алгоритмов и сопоставление результатов их работ с результатами, показанными другими аналогичными программами. При этом введение является очень общим (на уровне учебника), а содержательные аспекты и идеи, на которых базируются предложенные алгоритмы, упоминаются в начале каждого раздела. Это не вполне традиционный подход, который не создает общей картины развития области, но зато очень удобен для чтения и понимания конкретных задач и алгоритмов. На всякий случай, замечу, что атрибутирование чужих идей и результатов, как сторонних авторов, так и соавторов сделано безукоризненно, а решающий личный вклад автора очевиден.

09/2-125 от 26.04.2019



Результаты диссертации опубликованы в трех основных статьях, в двух из которых диссертант является первым автором. Еще в пяти статьях, материалы которых в основном не вошли в текст диссертации, описаны применения программ для анализа конкретных метагеномов и обсуждены некоторые частные алгоритмические проблемы. Кроме того, результаты неоднократно доложены на международных конференциях, в т.ч. дважды как устные доклады на ведущей конференции по алгоритмической биоинформатике RECOMB.

На рассмотрение представлен текст на английском и русском языках. Выборочный анализ показывает, что эти тексты в точности соответствуют друг другу. Рецензия относится к русскому варианту.

Во второй главе, посвященной сборщику SPAdes, описан ряд важных алгоритмических приемов, которые существенно улучшили качество сборки, в особенности геномов единичных клеток, и пользовательские характеристики, такие как время работы и требуемая память.

В третьей главе описан сборщик metaSPAdes, специально разработанный для анализа метагеномов.

В обеих главах приведены результаты глубокого тестирования тестирования предложенных автором алгоритмов, точнее, разработанных на их основе программ, и сравнения их с другими используемыми в литературе алгоритмами.

Структура диссертации логична, текст написан в целом хорошо, понятно и аккуратно. Замечания к диссертации носят чисто редакционный, если не сказать корректорский, характер.

Первая встреченная опечатка обнаруживается в названии организации, в которой выполнена работа: «Центр Алгоритмической Юиотехнологии» (стр. 8). Чуть далее, видимо, случилась делеция: «В работе Fang и др., 2018 [14], а также анализ композиции штаммов бактерии *E.coli* в рассматриваемых образцах» (стр. 10). Несогласованы обозначения: на стр. 17 метрики обозначены то  $NG_x/LG_x$ , то  $N(G)_x/L(G)_x$ ; на стр. 33 программы: `bbhash` и `VBhash`. Грамматически рассогласованы фразы «расстоянием вставки для типичной библиотеки парных ридов характеризуется средним значением» (стр. 19) и «...а также делает тривиальной восстановление последовательностей» (стр. 48). Пропала ссылка на стр. 37: «При работе с данными секвенирования изолятов используется вероятностная модель, схожая с использованной в работах cite» (sic!). Выражение «Теорема 1. Теорема 4 из Nurk и др., 2013 [8] (секция аппендикса B)» выглядит как химера с работой по хирургии, а фраза «к сожалению, применение описанной стратегия приводит к нежелательному отсоединению ребер, соответствующих входам и выходам из регионов,



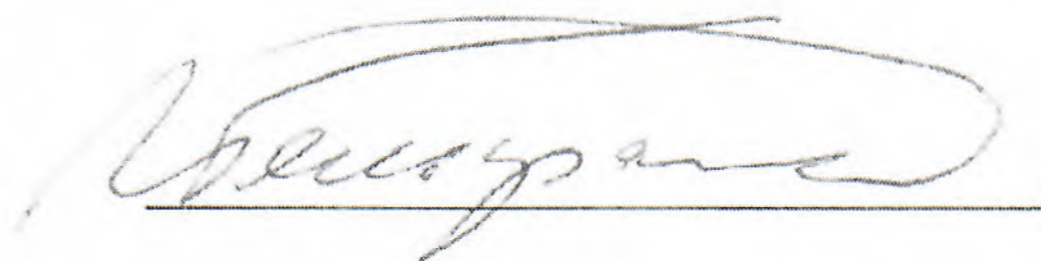
представляющих в графе повторы высокой кратности» — как недоперевод с английского

Из чуть более существенного: традиционные русские транслитерации — «Крейг Вентер» (Википедия и 27 тысяч хитов в Гугле) и «де Брёйн» (Википедия и 16 тысяч хитов), а не «Вентор» (145 хитов) и «де Брюин» (последний вариант употребляется для других De Bruin, например, голландского художника XVII в. — такая ситуация не должна удивлять, ср. Джемса Уотсона и доктора Ватсона). Впрочем, возможно, автор употребляет питерский вариант нормы, подобно генетической «аллели» и математическим «вещественным числам».

Сделанные замечания ни в коей мере не умаляют актуальности, ценности и научного уровня рассматриваемой работы. Диссертация Сергея Юрьевича Нурка «Сборка геномов некультивируемых микроорганизмов по данным высокопроизводительного секвенирования» 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 Сергей Юрьевич Нурк 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

04.04.2019



М. С. Гельфанд

