

Prof. Dr. Olga V. Kalinina
Head of research group for
Drug Bioinformatics

Phone +49 681 98806-3600
Fax +49 681 98806-3009

olga.kalinina@helmholtz-
hips.de

**Helmholtz-Institute for Pharmaceutical
Research Saarland**
Campus E8.1
66123 Saarbrücken
Germany
www.helmholtz-hips.de

Registered Office:
Helmholtz-Zentrum
für Infektionsforschung GmbH
Inhoffenstraße 7
38124 Braunschweig

Chair of Supervisory Board:
MinDir'in Prof. Dr. Veronika von Messling,
Federal Ministry of Education and
Research

Scientific Director:
Prof. Dr. Dirk Heinz

Administrative Director:
Silke Tannapfel

Registry Court:
Amtsgericht Braunschweig HRB 477
VAT Reg. No DE 11 48 15 244

St.-Nr. 13/200/24006

ОТЗЫВ

**на автореферат диссертации Казнадзей Анны Денисовны
“Геномная ко-локализация генов углеводного метаболизма бактерий”,
представленной на соискание ученой степени кандидата биологических
наук по специальности 03.01.09 — математическая биология,
биоинформатика**

Диссертация Казнадзей А.Д. посвящена проведенным диссертанткой исследованиям в области изучения функциональной значимости ко-локализации генов у бактерий с упором на гены углеводного метаболизма. Эта тема, с одной стороны, очень давно изучается методами сравнительной геномики, на с другой стороны, по-видимому, еще может преподнести немало сюрпризов, как это было продемонстрировано в настоящей диссертации.

Гипотезы о том, что гены, связанные функционально, имеют тенденцию образовывать консервативные кассеты (набор генов, расположенные относительно близко в бактериальном геноме, причем это расположение эволюционно консервативно), неоднократно выдвигались в эволюционной биологии. Для некоторых функций этот признак является диагностическим и используется для предсказания (например, для генов синтеза вторичных метаболитов). Представляемая к защите работа является, по-видимому, самым масштабным исследованием подобного рода для генов углеводного метаболизма. Углеводный метаболизм занимает центральное место в метаболизме бактерий, поэтому объем проанализированных данных поистине огромен. Кроме того, к достоинствам работы следует отнести разработку нового оригинального алгоритма поиска сходства последовательностей, которые превосходит другие используемые в этой области методы как по точности, так и по производительности, что и позволило проанализировать такой большой объем данных.

Диссертация состоит из четырех глав, которые кратко изложены в автореферате.

Глава 1 содержит обзор литературы, разделенный на три части: обзор методов для поиска сходства нуклеотидных последовательностей; обзор предшествовавших исследований связи ко-локализации генов с их функцией; и обзор метаболических путей утилизации лактозы в *E. coli*, которые были

An institution of

HZI HELMHOLTZ
Centre for Infection Research

In cooperation with



исследованы в качестве примера приложения методов в диссертации. Следует заметить, что, тогда как для методов поиска сходства нуклеотидных последовательностей в автореферате приведены конкретные примеры, практически не упомянуты родственные методы кластеризации последовательностей, которые, так же как и представленный в диссертации метод NSimScan, бывают основаны на фильтрации на основании хэш-таблиц вхождений коротких слов (например, CD-HIT, Li and Godzik, Bioinformatics, 2006, или более поздний метод Linclust, Steinegger and Söding, Nat Commun, 2018, который не мог быть использован диссертанткой и соавторами для сравнения в оригинальной статье, так как был опубликован позднее их оригинальной статьи).

Во второй главе описан алгоритм NSimScan, разработанный и реализованный диссертанткой. В качестве несущественного недостатка в этой части я могу отметить, что не упомянут использованный язык программирования, хотя косвенные признаки позволяют предположить, что это C/C++. Также наблюдается непоследовательность в приведенных здесь оптимальных диапазонах применимости метода: на стр. 5 и 26 это «последовательности, различающиеся на 60-90%», на стр. 12 – «последовательности, отличающиеся ... на 10-40 процентов». Представлены результаты сравнения точности и производительности разработанного метода с другими используемыми поиска сходства нуклеотидных последовательностей и показано улучшение как того, так и другого.

Глава 3 излагает результаты анализа ко-локализации генов углеводного метаболизма. Было проанализировано 148 тысяч генов из 665 геномов. К сожалению, в автореферате не описана процедура выбора этих генов, а также процедура разделения их на глобальные функциональные классы. Вводится мера склонности к образованию кассеты как доля генов определенной группы или класса, входящих в состав кассет, и показано, что эта мера варьируется в широких пределах между разными функциональными классами и таксонами. Также проведено сравнение со случайной моделью.

В четвертой главе приведен пример исследования конкретной кассеты *yih* из *E. coli*. Показано, как с помощью разработанных методов удалось предсказать новую функцию этих генов в катаболизме лактозы с помощью сравнения с родственной аннотированной кассеты из *Bacilli*. Это предсказание блестяще подтвердилось экспериментально. На мой взгляд, это прекрасный пример результативного сотрудничества теоретических и экспериментальных лабораторий, где биоинформатика была движущей силой проекта. С помощью биоинформатических методов была исследована сложная регуляция этой кассеты двумя транскрипционными факторами в условиях роста бактерий на разных сахарах.

К недостаткам автореферата можно отнести некоторые опечатки и небрежности, например, на стр. 16 на рис. 9 ось ординат размечена от 0 до 80%, тогда как в сопровождающей этот рисунок тексте описывается вариация этого значения (склонности к образованию кассет) от 0 до 100%. Эти незначительные недостатки никоим образом не умаляют качества работы.

Работа выполнена на высоком методическом уровне, результаты не вызывают сомнений, выводы работы обоснованы полученными результатами. По теме диссертации опубликовано 3 статьи в журналах ВАК, а также сделано 8 докладов на российских и международных конференциях, опубликованных в сборниках

тезисов этих конференций. Квалификационная работа «Геномная ко-локализация генов углеводного метаболизма бактерий» удовлетворяет требованиям ВАК, предъявляемым к диссертационным работам. Ее автор, Казнадзей Анна Денисовна, несомненно, заслуживает присуждения ей ученой степени кандидата биологических наук по специальности 03.01.09 – Математическая биология, биоинформатика.

Кандидат физико-математических наук по специальности 03.01.03 – молекулярная биология, профессор Университета Земли Саар

Калинина Ольга Вячеславовна

Саарбрюкен, 2.09.2019

