

О работах по биоинформатике в лаборатории «Компьютерные технологии» Университета ИТМО

Работы по направлению «Биоинформатика» проводятся в Лаборатории «Компьютерные технологии» Университета ИТМО с января 2010 г. В декабре 2011 г. была создана лаборатория «Алгоритмы сборки геномных последовательностей» (<http://genome.ifmo.ru>). Инициатор создания лаборатории – докт. техн. наук, профессор Анатолий Шалыто, руководители лаборатории: с 2010 по 2014 гг. – чемпион мира по программированию 2008 г. Федор Царев, с 2014 г. – Алексей Сергушичев.

Работы начинались совместно с центром «Биоинженерия» РАН (руководитель – академик РАН К.Г. Скрыбин). Консультант работ – докт. биол. наук Е. Б. Прохорчук.

В настоящее время (май 2015 г.) исследования проводятся совместно со следующими организациями (лабораториями):

- *Washington University St. Louis (WashU)*, лаборатория Максима Артемова (Сент-Луис, США);
- *Научно-исследовательский институт акушерства, гинекологии и репродуктологии им. Д.О. Отта*, лаборатория пренатальной диагностики наследственных и врожденных заболеваний под руководством Андрея Глотова (Санкт-Петербург);
- *Научно-исследовательский институт физико-химической медицины (НИИ ФХМ)*, лаборатория биоинформатики под руководством Дмитрия Алексеева (Москва);
- *Санкт-Петербургского Академический университет (АУ)*, лаборатория нанобиотехнологий под руководством Михаила Дубины (Санкт-Петербург);
- *Казанский государственный университет (КФУ)*, Институт фундаментальной медицины и биологии, отдел микробиологии. Работы проводятся совместно с Анастасией Тойменцевой (Казань).

Начаты работы с *Федеральным центром сердца, крови и эндокринологии им. В. А. Алмазова*, Институт молекулярной биологии и генетики под руководством Анны Костаревой (Санкт-Петербург);

За время работы выполнялись/выполняются следующие проекты:

- разработка алгоритмов сборки данных секвенирования:
 - сборка генома на основе совмещения подходов использования графов де Брёйна и графа перекрытий. В дальнейшем алгоритмы были доработаны для поддержки чтений *Ion Torrent*;
 - сборка транскриптома;

- сборка скэффолдов.

Создан уникальный сборщик, позволяющий проводить сборку генома на **персональных компьютерах под управлением любой ОС**.

- анализ вариаций и поиск маркеров ассоциированных с кардиомиопатией (совместно с институтом Отта);
- анализ данных экспрессии генов и построение генетической подписи конкретного типа рака (совместно с АУ);
- совместный анализ данных секвенирования и экспрессии, а также различных эпигенетических данных определенного типа рака (совместно с WashU);
- исследование иммуногенности раковых эпитопов для создания персонализированных противораковых вакцин (совместно с WashU);
- разработка алгоритмов анализа транскрипционных и метаболических данных в контексте метаболических сетей реакций (совместно с WashU);
- разработка алгоритмов сравнительного анализа метагеномных данных (совместно с НИИ ФХМ);
- сборка геномов редких бактерий (совместно с КФУ).

С участием нашей лаборатории были организованы/организуются следующие мероприятия:

- 21-24 мая 2015 г. Выездной семинар по системной биологии (<http://bioinformaticsinstitute.ru/sbw2015>). Организовано совместно с Институтом биоинформатики (<http://bioinformaticsinstitute.ru>);
- 15-18 мая 2014 г. Первый практический семинар по системной биологии (<http://bioinformaticsinstitute.ru/sbw2014>), который провели приглашенные лекторы из Гарвардского университета и Университета Вашингтона. В семинаре приняло участие 25 человек. Организовано совместно с Институтом биоинформатики;
- 18-20 февраля 2014 г. Практическая школа по обработке данных секвенирования нового поколения. В школе приняло участие 14 человек. Организаторы: лаборатория «Компьютерные технологии» Университета ИТМО, «Life Technologies», ООО «Компания Хеликон».

За время работы было получено шесть свидетельств о регистрации программ по разработанным алгоритмам биоинформатики (<http://is.ifmo.ru/certificates/>).

Было выиграно и выполнено три государственных контракта:

- «Разработка методов сборки генома, сборки транскриптома и динамического анализа протеома» на 2012-2013 годы (№ 14.В37.21.0562, заключен в рамках Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России на 2009-2013 годы»).

- «Разработка алгоритмов сборки геномных последовательностей для вычислительных систем экзафлопсного уровня производительности» на 2011-2012 годы (№ 07.514.11.4010, заключен в рамках Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы»).
- «Разработка метода сборки геномных последовательностей на основе восстановления фрагментов по парным чтениям» на 2011-2013 годы (№ 16.740.11.0495, заключен в рамках Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России на 2009-2013 годы») (<http://is.ifmo.ru/genom/>).

В рамках работы лаборатории были **опубликованы** следующие работы в российских и зарубежных реферируемых журналах:

- *Jha A., Huang S., Sergushichev A., Lampropoulou V., Ivanova J., Loginicheva E., Chmielewski K., Stewart K., Ashall J., Everts B., Pearce E., Driggers E., Artyomov M.* **Network integration of parallel metabolomic-transcriptional data reveals novel metabolic modules regulating divergent macrophage polarization** // *Immunity*. 2015. 42 (3): 419-430. <http://dx.doi.org/10.1016/j.immuni.2015.02.005>.
- *Glotov A., Kazakov S., Zhukova E., Alexandrov A., Glotov O., Pakin V., Danilova M., Tarkovskaya I., Niyazova S., Chakova N., Komissarova S., Kurnikova E., Sarana A., Sherbak S., Sergushichev A., Shalyto A., Baranov V.* **Targeted next-generation sequencing (NGS) of nine candidate genes with custom AmpliSeq in patients and a cardiomyopathy risk group** // *Clinica Chimica Acta*. 2015. 446: 132-140. <http://www.sciencedirect.com/science/article/pii/S0009898115002053>.
- *Mardanov A., Toymentseva A., Gilyazeva A., Kazakov S., Shagimardanov E., Khaitlina S., Sharipova M.* **Draft Genome Sequence of *Serratia grimesii* Strain A2** // *Genome Announcements*. 2014. 2(5): e00937-14. <http://genomea.asm.org/content/2/5/e00937-14>.
- *Bradnam K., Fass J., Alexandrov A., Kazakov S., Sergushichev A., Melnikov S., Fedotov P., Tsarev F. et al.* **Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species** // *GigaScience*. 2013. Vol. 2, No. 1, pp. 1-31. <http://link.springer.com/article/10.1186/2047-217X-2-10>.
- *Сергушичев А., Александров А., Казаков С., Царев Ф., Шалыто А.* **Совместное применение графа де Брёйна, графа перекрытий и микросборки для de novo сборки генома** // *Известия Саратовского университета. Новая серия. Серия Математика. Механика. Информатика*. 2013. № 2, ч. 2, с. 51-57.
- *Александров А., Казаков С., Мельников С., Сергушичев А., Царев Ф.* **Метод сборки контигов геномных последовательностей на основе совместного применения графов де Брюина и графов перекрытий** // *Научно-технический*

вестник информационных технологий, механики и оптики. 2012. № 6 (82), с. 93-98.

- Александров А., Казаков С., Мельников С., Сергушичев А., Царев Ф., Шалыто А. **Метод исправления ошибок в наборе чтений нуклеотидной последовательности** // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2011. № 5 (75), с. 81-84.

Были выполнены **доклады** на следующих международных конференциях:

- *Sergushichev A.* **Network integration of parallel metabolomic-transcriptional data reveals novel metabolic modules regulating divergent macrophage polarization** / Cold Spring Harbor Laboratory Meeting. Systems Biology: Networks. 2015.
- *Kazakov S., Shalyto A.* **Overlap graph simplification using edge reliability calculation** / Proceedings of the 8th International Conference on Intelligent Systems and Agents 2014 (ISA 2014). Lisbon, Portugal. 2014. pp. 222-226.
- *Sergushichev A.* **GAM: a web-service for integrated transcriptional and metabolic network analysis** / Metabolism & Immunity: A Rediscovered Frontier 2014.
- *Akhi A., Sergushichev A., Tsarev F.* **Maximum Likelihood Scaffold Assembly** / RECOMB 2013. Book of Abstracts. Tsinghua University. 2013, p. 156 (<http://genome.ifmo.ru/files/conferences/recomb-2013-poster.pdf>).
- *Alexandrov A., Kazakov S., Melnikov S., Sergushichev A., Shalyto A., Tsarev F.* **Combining de Bruijn graph, overlap graph and microassembly for de novo genome assembly** / Proceedings of «Bioinformatics 2012». Stockholm, Sweden. 2012. p. 72.
- *Sergushichev A., Alexandrov A., Kazakov S., Melnikov S., Isenbaev V., Tsarev F.* **Combining «overlap-layout-consensus» and de Bruijn graph approaches for de novo genome assembly** / Sequence Mapping and Assembly Assessment Project workshop. Barcelona, April 5th, 2011.

Е. Капуном и Ф. Царевым получены выдающиеся результаты по оценке сложности сборки генома при использовании графов де Брёйна:

- *Kapun E., Tsarev F.* **On NP-Hardness of the Paired de Bruijn Sound Cycle Problem** / WABI 2013. Algorithms in Bioinformatics. Lecture Notes in Computer Science Vol. 8126, pp. 59-69. <http://arxiv.org/pdf/1307.7806.pdf>
http://link.springer.com/chapter/10.1007/978-3-642-40453-5_6#
- *Kapun E., Tsarev F.* **De Bruijn Superwalk with Multiplicities Problem is NP-hard** / BMC Bioinformatics. 2013, 14 (Suppl.5). Proceedings of the Third Annual RECOMB

Satellite Workshop on Massively Parallel Sequencing (RECOMB-seq 2013): S7.
<http://www.biomedcentral.com/1471-2105/14/S5/S7>.

16.11.2012 г. коллективом лаборатории был представлен доклад «Метод сборки генома на основе совместного применения графов де Брёйна и графов перекрытий» на семинаре в лаборатории «Алгоритмической биологии» Академического университета, который был высоко оценен руководителем лаборатории Павлом Певзнером (<http://genome.ifmo.ru/ru/node/21>).

В 2011 г. коллектив принимал участие в проекте *de novo Genome Assembly Assessment Project* (dnGASP, <http://cnag.bsc.es>, организован Национальным центром геномного анализа, Барселона, Испания). Целью участия НИУ ИТМО в этом проекте было проведение предварительных исследований новой архитектуры сборщика генома, в которой этап сборки контигов был разбит на два подэтапа: сборка квази-контигов (фрагментов генома длиной порядка 500 нуклеотидов) и сборка контигов.

В 2012 г. коллектив принимал участие в схожем проекте – *Assemblathon 2*.

Работы лаборатории начинались в 2010 г. с защит двух бакалаврских работ:

- Исенбаев В.В. Разработка системы секвенирования ДНК с использованием *paired-end* данных (http://is.ifmo.ru/genom/_isenbaev_thesis.pdf).
- Капун Е.Д. Разработка метода сравнения нуклеотидных последовательностей путем разбиения на фрагменты (http://is.ifmo.ru/genom/_kapun_thesis.pdf).

и магистерской диссертации:

- Дворкин М.Э. Методы минимизации необходимого числа цепей для секвенирования ДНК (http://is.ifmo.ru/diploma-theses/_dvorkin_genom.pdf).

Дополнительная ссылка:

- Биоинформатика в лаборатории «Компьютерные технологии»
http://irc.ifmo.ru/ru/87845/res_dir/1283/bioinformatika.htm.