A large group of graduates in blue gowns and caps are celebrating in front of a modern university building. Many are throwing their caps into the air. The scene is festive and celebratory.

Свободное Программное Обеспечение в Передовой Инженерной Школе

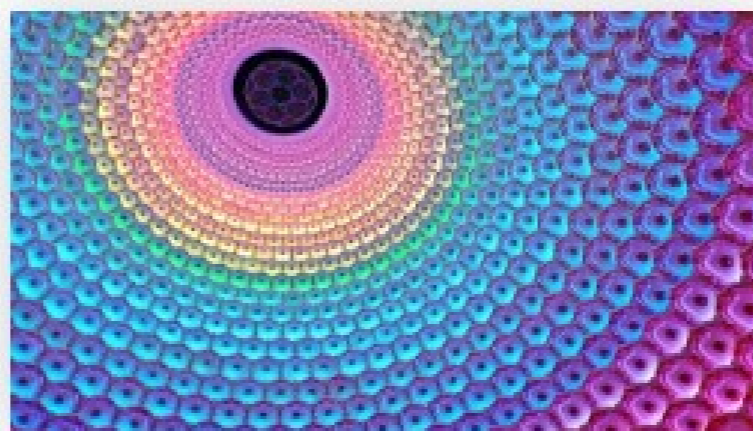
**Академический Университет им. Жореса Алфёрова РАН,
Санкт-Петербург**

**к.ф.-м.н. Сергей Александрович Симонов, Юрий
Александрович Андреев**

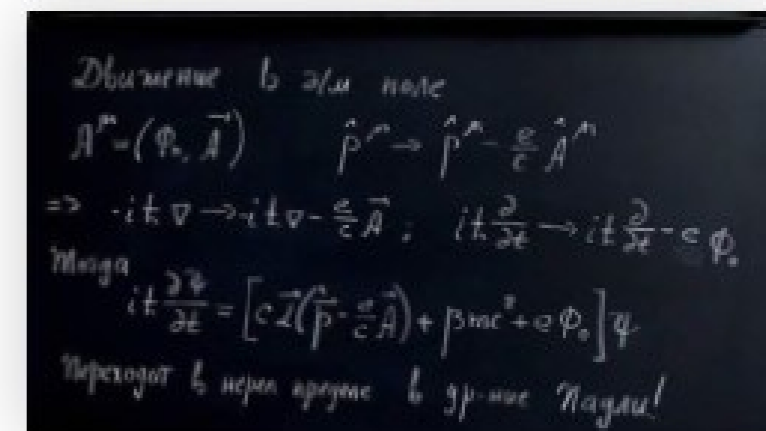
Профили обучения (Выпускающие кафедры)



Биоинформатика



Нанотехнологии



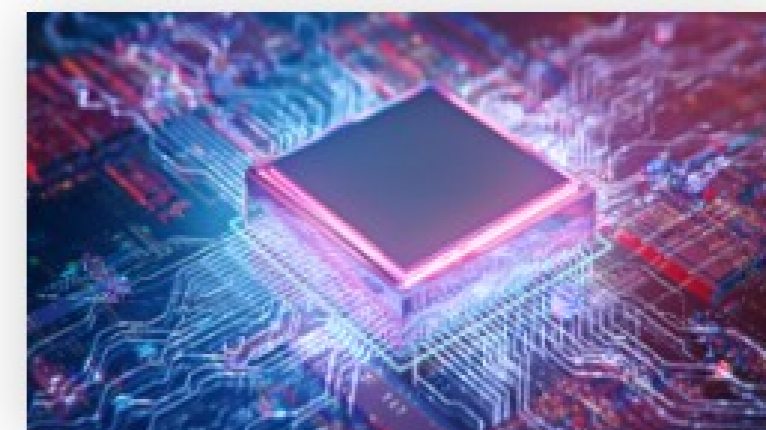
Теоретическая физика



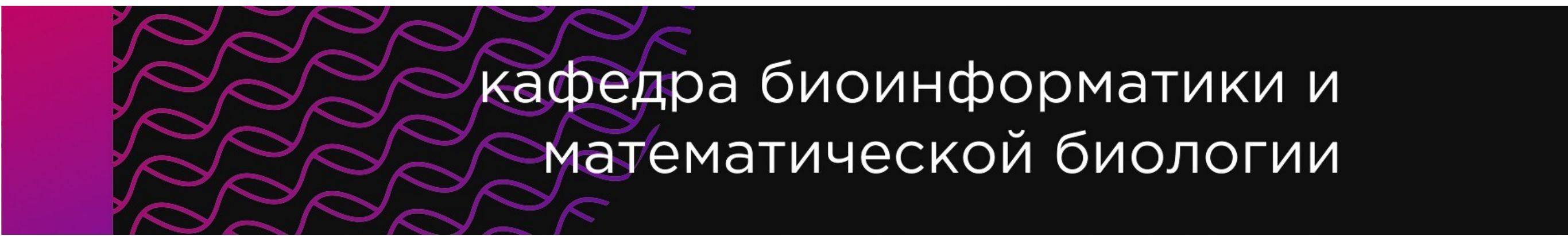
НаноБИОтехнологии



Физика космоса



**Полупроводниковые
нанотехнологии**



кафедра биоинформатики и математической биологии

Кафедра осуществляет подготовку студентов по общеобразовательным курсам разделов математики, а также подготовку студентов по профилям обучения «Биоинформатика» и «Цифровая физика».

Биоинформатика – активно развивающееся междисциплинарное направление, объединяющая биологию, генетику, химию, компьютерные науки, математику и статистику и специализирующееся на крупномасштабных биологических проблемах, требующих анализа больших объемов данных.



Лабораторные работы

- Обработка данных NGS (New Generation Sequencing)
- Полигеномный поиск ассоциаций
- Вычислительная транскриптомика
- Вычислительная геномика
- Биоинформатика рака
- Алгоритмы и структуры данных в биоинформатике
- Биостатистика



Обработка данных NGS

Обработка данных NGS – “выравнивание сырых ридов” – по сути задача поиска наилучшего размещения подстроки на длинной строке.

Для человека:

- длина строки $3,2 * 10^{**9}$;
- количество подстрок длины 200 – $(6-10) * 10^{**9}$
примерно терабайт данных.



Полногеномный поиск ассоциаций – GWAS

У человека до 1,5 миллионов СНИПов – мест, в которых один геном отличается от другого.

Задача:

найти статистическую значимость влияния каждого из СНИПов на тот или иной фенотипический признак



Трудности

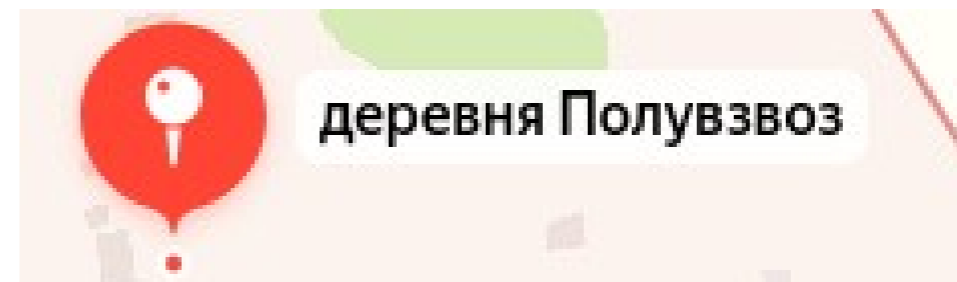
- Очевидно такие задачи не могут решаться на персональном компьютере. Поэтому решаются на примерах очень маленьких организмов – бактерий и вирусов.
- Кроме того, хотя почти всё программное обеспечение и вычислительные процедуры есть в свободном доступе, санкционные ограничения делают многие тулы недоступными.



Трудности

- Многие алгоритмы используют цепочку (пайпы) из инструментов. И надо следить, чтобы все тулы в пайпах были доступны. Либо прорабатывать альтернативные пайпы.

пропустишь приставку —
не туда приедешь!





Решение

При поддержке администрации АУ было решено выделить **сервер** для решения следующих задач.

- В качестве репозитория для геномных данных и средств их обработки.
- Для проведения вычислений студентами “в облаке”.



Комментарии

Модель образования

Работа проводится группой студентов 2-5 курса под руководством сотрудников АУ, интегрирована с дисциплинами:

- Разработка ПО
- Технологии программирования
- Летняя практика



Комментарии

Контроль результата

Используется:

- В учебном процессе
- Для работы над ВКР...



Спасибо за внимание!

**Сергей Александрович Симонов,
Юрий Александрович Андреев**

andreev.yurij@gmail.com

t.me/mathsalon