Университет ИТМО

Казаков Сергей Владимирович

Автоматизация сборки генома и сравнительного анализа метагеномов для обучения геномной биоинформатике

Диссертация на соискание ученой степени кандидата технических наук

Специальность 05.13.06 – Автоматизация и управление технологическими процессами и производствами (образование)

Научный руководитель – доктор технических наук, профессор
Шалыто Анатолий Абрамович

22.12.2016

Введение

За последние 40 лет **основным методом** получения информации о клетке и процессах, протекающих в ней, стало *секвенирование дезоксирибонуклеиновой кислоты (ДНК)* или *рибонуклеиновой кислоты (РНК)*.

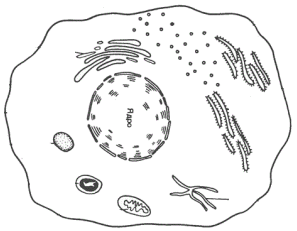
Дезоксирибонуклеиновая кислота (*ДНК*) и её более простая форма *рибонуклеиновая кислота* (*РНК*) – биологический полимер, обеспечивающий хранение и передачу из поколения в поколение биологической информации.

Геном – совокупность наследственного материала, заключенного в клетке организма.

Секвенирование – это процесс получения по физической субстанции ДНК или РНК ее нуклеотидной последовательности.

Секвенирование состоит из двух частей – физико-химической и компьютерной. Вторая часть обычно называется «*сборка генома*».

Клетка



Секвенирование ДНК / РНК

Данные

01011000011…

Актуальность (предметная) (1)

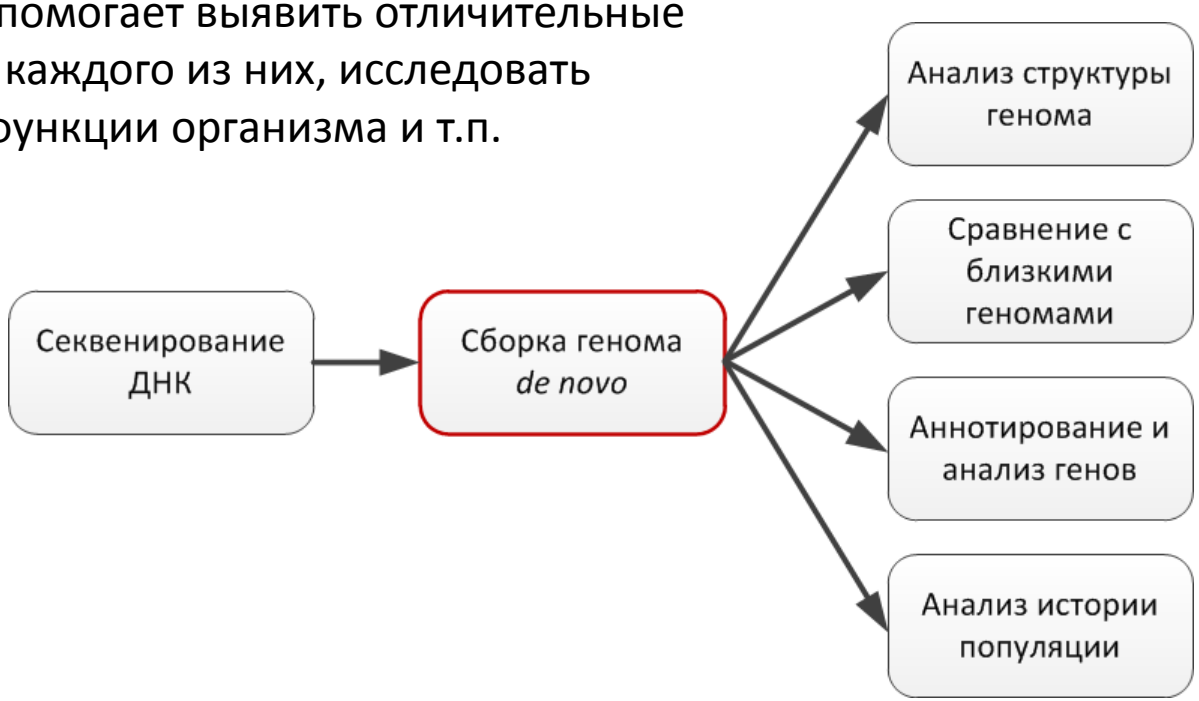
В диссертации рассматриваются две задачи:

1. ***de novo* сборка генома**

- *Сборка генома* – задача восстановления исходного генома клетки из небольших фрагментов (чтений), которые получаются при секвенировании.
- Задача *сборка генома de novo* – задача сборки ранее неизвестного генома.

Задача *de novo* сборки генома **используется при проведении научных исследований** при изучении геномов вирусов, бактерий, растений и животных.

Их изучение помогает выявить отличительные особенности каждого из них, исследовать возможные функции организма и т.п.

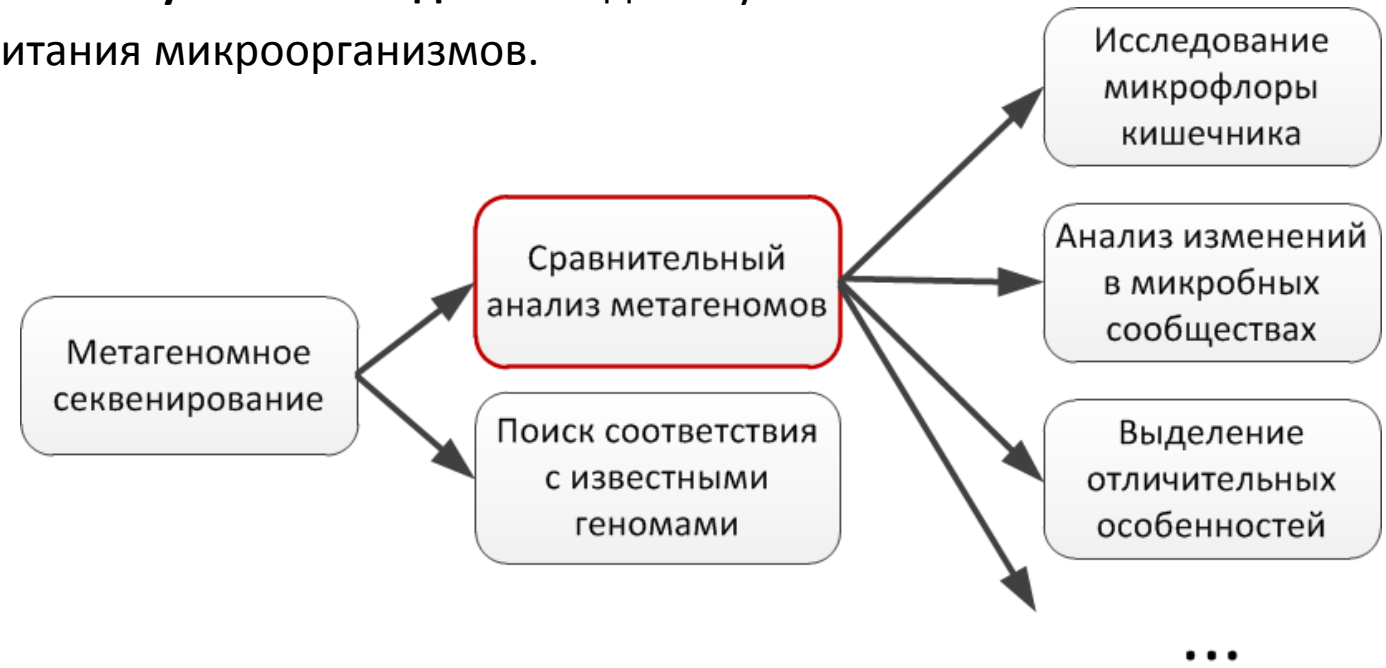


Актуальность (предметная) (2)

2. Сравнительный анализ метагеномов

- *Метагеном* – совокупность геномов микроорганизмов (бактерий, архей, вирусов) из одной среды обитания (почва, водные ресурсы, кишечник человека).
- *Сравнительный анализ метагеновов* – задача сравнения нескольких метагеномов между собой с целью выявления характерных признаков каждого из них.

Методы и алгоритмы сравнительного анализа метагеномов **используются при проведении научных исследований** для изучения сред обитания микроорганизмов.



Актуальность (предметная) (3)

Сборка генома *de novo* и сравнительный анализ метагеномов обычно осуществляются:

- **Специалистами** узкой направленности (в основном биоинформатиками).
- Выполняются на серверах и кластерах.
- Используют **десятки гигабайт** оперативной памяти для работы.
- **Работают только на ОС *Linux*.**
- Сложны в установке и использовании.

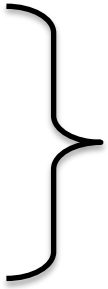
Указанные требования могут быть обеспечены в ведущих геномных центрах мира при научных исследованиях, но для целей обучения решению этих задач описанные требования обычно невыполнимы!

Актуальность (образовательная) (1)

Однако обучение рассматриваемым задач необходимо выполнять!

Оно включает в себя:

1. Проведение лекционных занятий.
2. Проведение практических занятий (работа с учителем).
3. Выполнение домашних заданий (самостоятельная работа).



Работа с данными **на ЭВМ**

Для проведения практических занятий **наиболее доступны персональные компьютеры** обучающихся.

Цели диссертационной работы

Цель диссертационной работы – автоматизация процессов обучения сборке генома *de novo* и сравнительному анализу метагеномов.

При этом вместо чисто лекционного обучения обеспечивается **практико-ориентированное обучение** на персональных компьютерах.

В соответствии с паспортом специальности 05.13.06 – «Автоматизация и управление технологическими процессами и производствами (образование)» диссертация относится к следующей **области исследований**:

20. Разработка автоматизированных систем научных исследований в области образования.

В соответствии с **формулой специальности** диссертационное исследование направлено на обеспечение «интеллектуальной поддержки процессов обучения»

Задачи диссертационной работы

Задачи работы:

1. Произвести анализ существующих методов сборки генома *de novo* и сравнительного анализа метагеномов по возможности их применения в образовательном процессе. **Решается в главе 1.**
2. Разработать метод сборки генома *de novo,* предназначенный для целей обучения. **Решается в главе 2.**
3. Разработать метод сравнительного анализа метагеномов, предназначенный для целей обучения. **Решается в главе 3.**
4. Обеспечить автоматизацию процессов обучения сборке генома *de novo* и сравнительному анализу метагеномов. **Решается в главе 4.**

Выносимые на защиту положения, научная новизна

1. Автоматизированный метод сборки генома *de novo* на основе совместного применения графа де Брейна и графа перекрытий.
    - Программа на основе предложенного метода позволяет производить сборку малых и средних по размеру геномов на персональных компьютерах под управлением трех самых распространенных операционных систем (*Windows*, *macOS/OS X*, *Linux*), что отличает ее от существующих программ.

2. Автоматизированный методы сравнительного анализа метагеномов, основанный на анализе компонент связности в графе де Брейна.
    - Разработанный метод отличается от существующих тем, что он выполняет «упрощенную» сборку метагеномов вместо стандартной, позволяя значительно сократить требуемые вычислительные ресурсы.

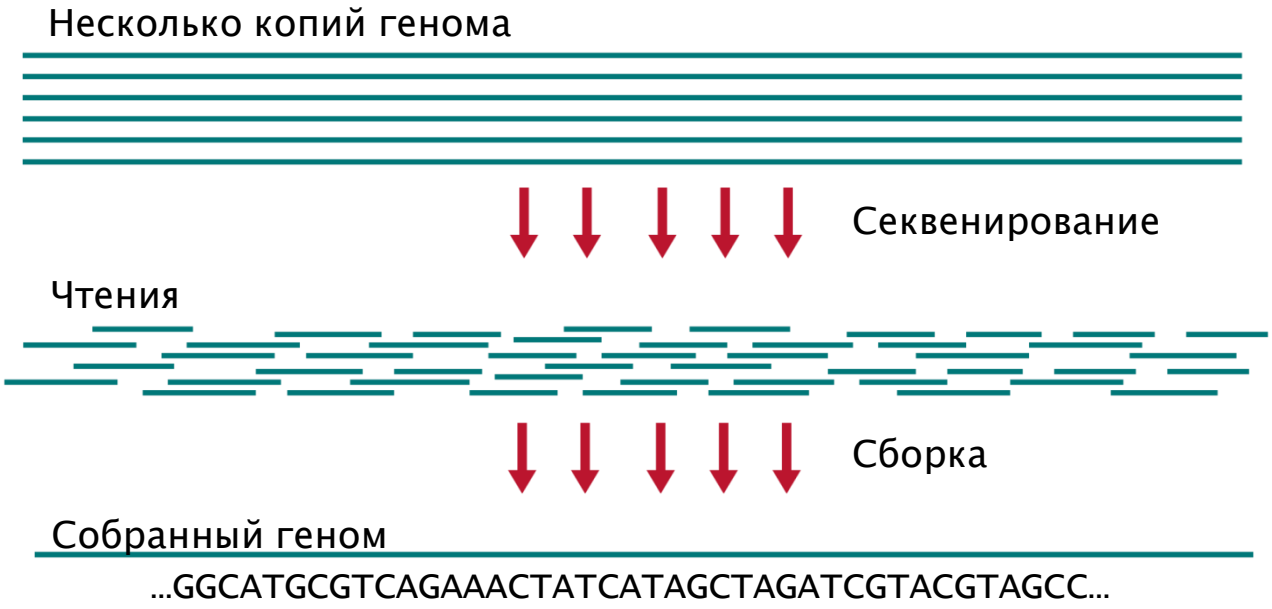
1.1. Задача сборки генома *de novo*

Имеется набор чтений $S = \{s_i\},\ s_i \in \Sigma^*$.

- *Чтения* (англ. *reads*) – небольшие фрагменты генома, получаемые при секвенировании

Задача сборки генома состоит в восстановлении исходного генома из чтений:

- Где должно находится каждое из чтений – неизвестно.
- Чтения покрывают исходный геном несколько десятков раз.



1.2. Сравнение известных программных продуктов *de novo* сборки генома для обучения биоинформатике (1)

Сравнение по критериям:

1. Возможность свободного использования.
2. Возможность работы на персональном компьютере (при небольшом объеме оперативной памяти).
3. Кроссплатформенность (работа под ОС Windows, Mac OS, Linux).
4. Наличие графического интерфейса пользователя.
5. Простота запуска.

Таблица 1

Название сборщика	1	2	3	4	5
ABySS	+	–	–	–	+
Allpaths-LG	+	–	–	–	–
CLC Genomics Workbench	–	+	+	+	+
MaSuRCA	+	–	–	–	–
Minia	+	+	–	–	–
Newbler	±	–	–	+	–
SPAdes (АУ, Санкт-Петербург)	+	–	–	–	+
Sparse Assembler	+	+	–	–	–
Velvet	+	–	–	–	–

Наиболее подходящим для целей обучения является *CLC Genomics Workbench*, который, однако, не распространяется свободно, **является платным** и дорогим!

1.2. Сравнение известных программных продуктов *de novo* сборки генома для обучения биоинформатике (2)

Анализ работы сборщика *CLC Genomics Workbench*:

В связи с отсутствием доступа к продукту у автора, анализ структуры и качества этого продукта не проводился.

Однако в статье сотрудников Академического Университета (АУ) [1] было проведено сравнение работы сборщика *CLC Genomics Workbench* с другими сборщиками, и его преимуществ по качеству сборки генома не обнаружено.

Исходя из изложенного можно утверждать, что для целей образования отсутствует подходящее программное обеспечение, которое может быть также использовано и в научных исследованиях.

[1] *Nurk S., Bankevich A., Antipov D., Gurevich A.A., Korobeynikov A., Lapidus A., Prjibelski A.D., Pyshkin A., Sirotkin A., Sirotkin Y., Stepanauskas R.* Assembling single-cell genomes and mini-metagenomes from chimeric MDA products // Journal of Computational Biology, 2013. № 20(10), pp.714-737.

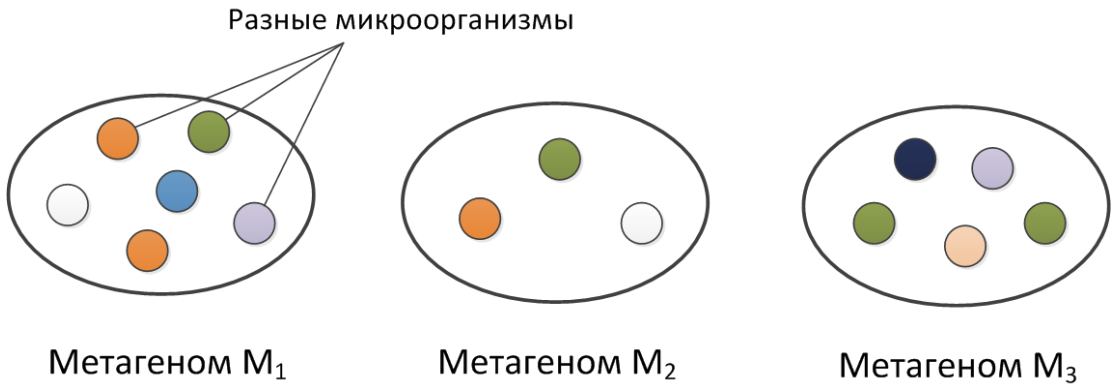
1.3. Задача сравнительного анализа метагеномов

Задано *M* метагеномов, каждый метагеном содержит K_i геномов.

Каждый метагеном из *M* отдельно секвенирован – имеется *M* наборов чтений.

Необходимо произвести сравнительный анализ данного набора из *M* метагеномов:

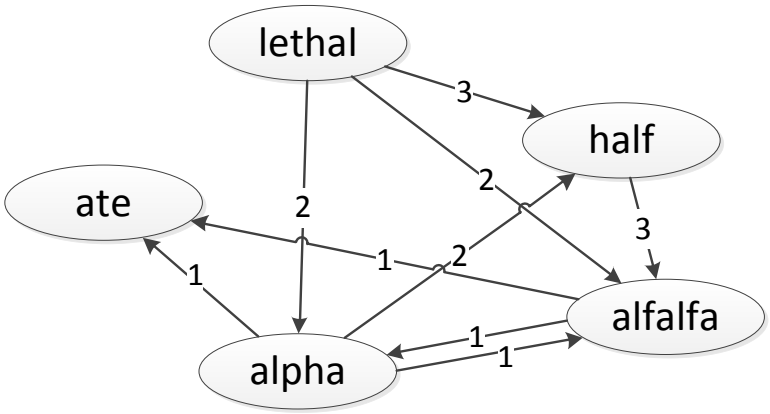
- Извлечь информативные признаки F_k, которые отличают один метагеном от другого.
- Построить матрицу расстояния D, где $D_{i,j}$ – степень сходства метагеномов M_i и M_j.



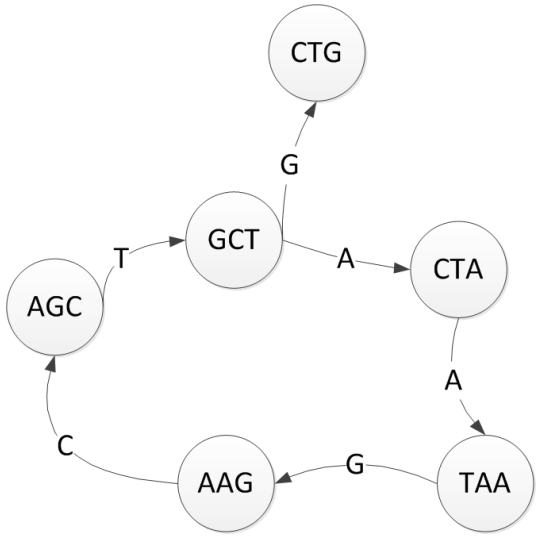
1.4. Граф перекрытий и граф де Брейна

Граф перекрытий – это взвешенный ориентированный граф $G = \langle V, E, \omega \rangle$, вершины которого – чтения генома s_i, а ребра обозначают перекрытия соответствующих строк.

Граф де Брейна – это ориентированный граф $G = \langle V, E \rangle$, в котором вершины – k-меры (строки длины k), а ребра соединяют два k-мера, которые перекрываются на k-1 нуклеотид.



Граф перекрытий



Граф де Брейна

Именно эти графы используются при сборке генома и сравнительном анализе метагеномов виду удобности их применения для решения описанных задач.

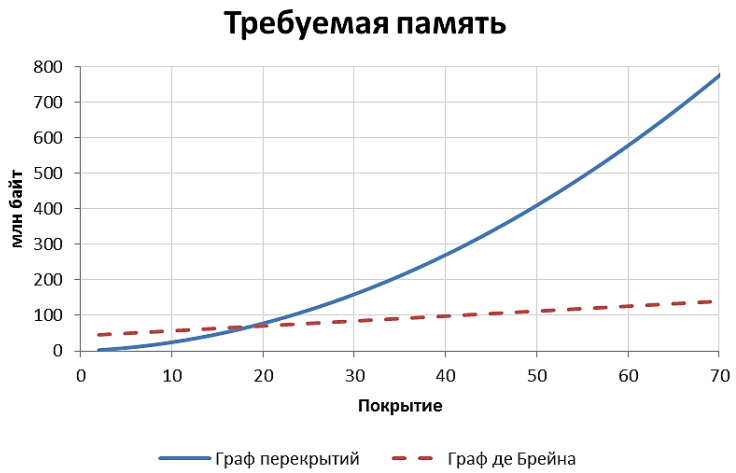
2.1. Теоретическое исследование. Анализ требуемых ресурсов для хранения графов

Автором получены следующие зависимости объема требуемой памяти в байтах для хранения графа перекрытий (*ov-graph*) и графа де Брейна (*db-graph*):

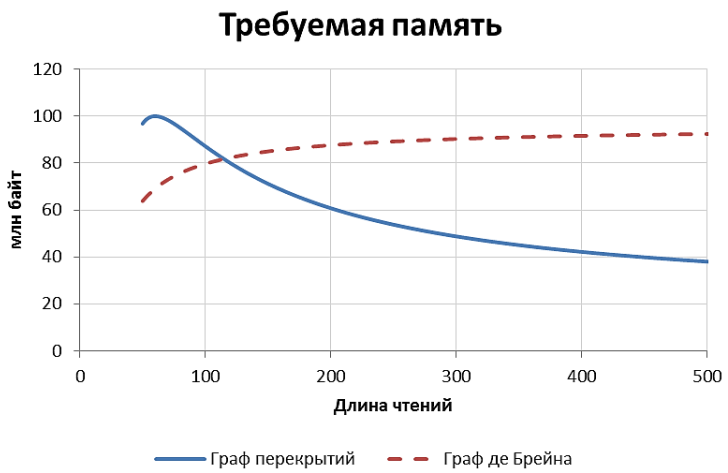
$$M_{ov-graph} = \frac{1}{4} \cdot N \cdot L + 6 \cdot N \cdot \left\lfloor \frac{N(L - L_{ov})}{G} \right\rfloor,$$

$$M_{db-graph} = 10\frac{2}{3} \cdot \left(K_g + \lfloor N \cdot L \cdot p_{er} \rfloor \cdot \frac{k(L - k + 1)}{L} \right).$$

где N – число чтений, L – средняя длина чтения, G – длина генома, L_{ov} – длина перекрытия, k – размер k-мера, K_g – число безошибочных k-меров, p_{er} – вероятность ошибки в одной позиции чтения.



При фиксированных параметрах: L = 100 нукл.
Меняются параметры: N, покрытие $c = \frac{N \cdot L}{G}$

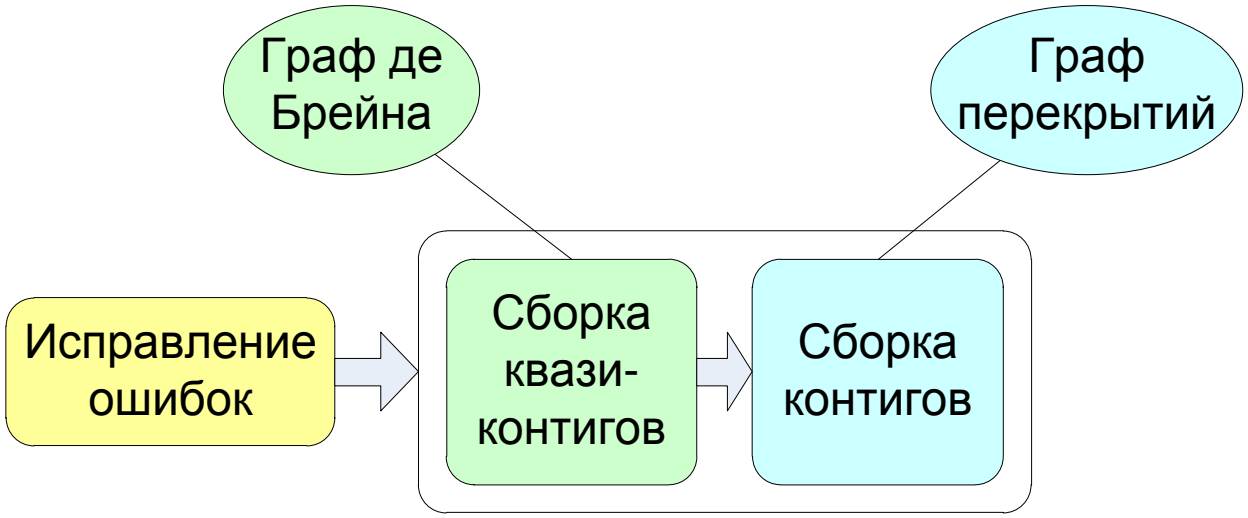


При фиксированных параметрах: c = 20
Меняются параметры: L, N

2.2. Предлагаемый метод сборки генома *de novo* (1)

Предлагаемый метод ориентирован на минимизацию объема используемой оперативной памяти и состоит из трех этапов:

- исправление ошибок в наборе чтений (основан на анализе *k*-мерного спектра);
- сборка квазиконтигов (восстановление фрагментов по парным чтениям, выполняется на графе де Брейна);
- сборка контигов (расширение исходных квазиконтигов, выполняется на графе перекрытий).

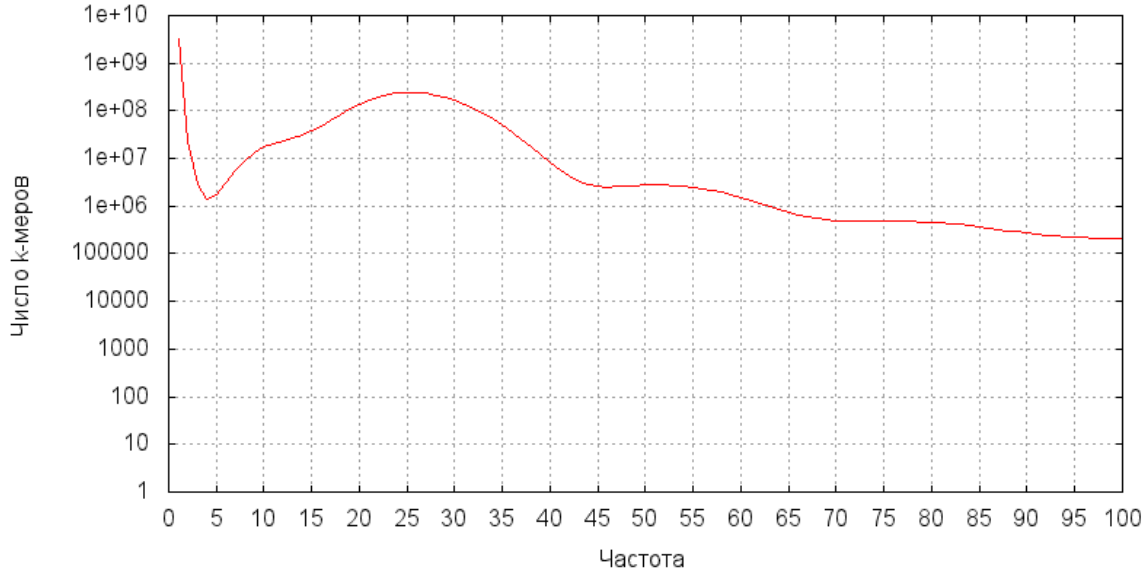


2.2. Предлагаемый метод сборки генома *de novo* (2)

Этап исправления ошибок состоит из следующих шагов:

- Подсчет числа *k*-меров (подстрок длины *k*), содержащихся в чтениях.
- Определение *ошибочных k-меров* (k-меры, которые встречаются меньше некоторого порога *f*) и *достоверных k-меров*.
- Исправление ошибочных k-меров на достоверные путем единичных замен нуклеотидов на отдельных позициях.

Новизна: возможность разбивать шаги на несколько шагов для уменьшения используемой памяти.



2.2. Предлагаемый метод сборки генома *de novo* (3)

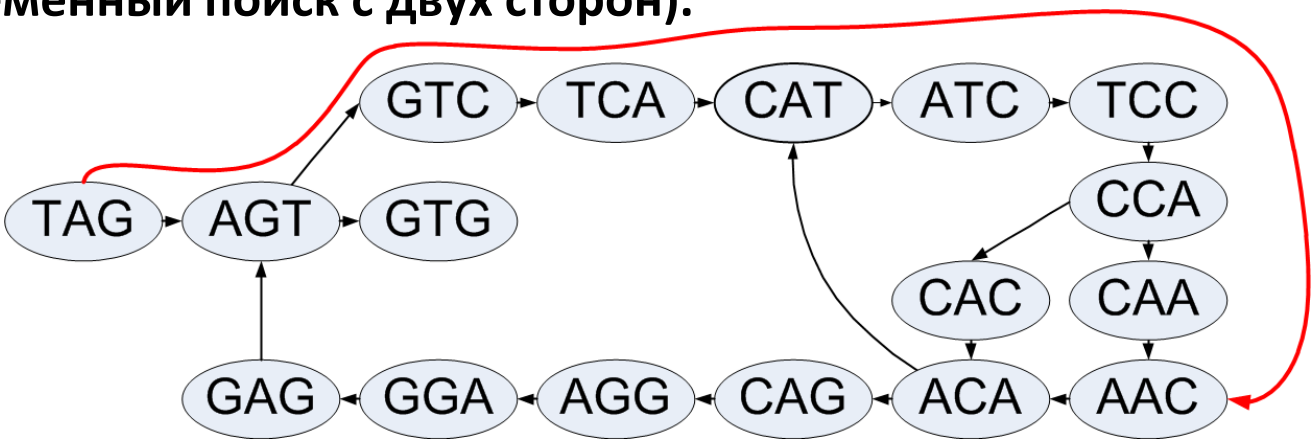
Этап сборки квазиконтигов состоит из следующих шагов:

- Построение графа де Брейна из достоверных *k*-меров.
- Для каждой пары чтений производится поиск пути в графе, который по длине укладывается в границы длин фрагментов.
- При единственном решении путь преобразуется в квазиконтиг.

Новизна:

1. Сохраняем только вершины графа, без сохранения дополнительной информации в вершинах и на ребрах.

2. Для поиска подходящего пути используем подход *meet-in-the-middle* (одновременный поиск с двух сторон).

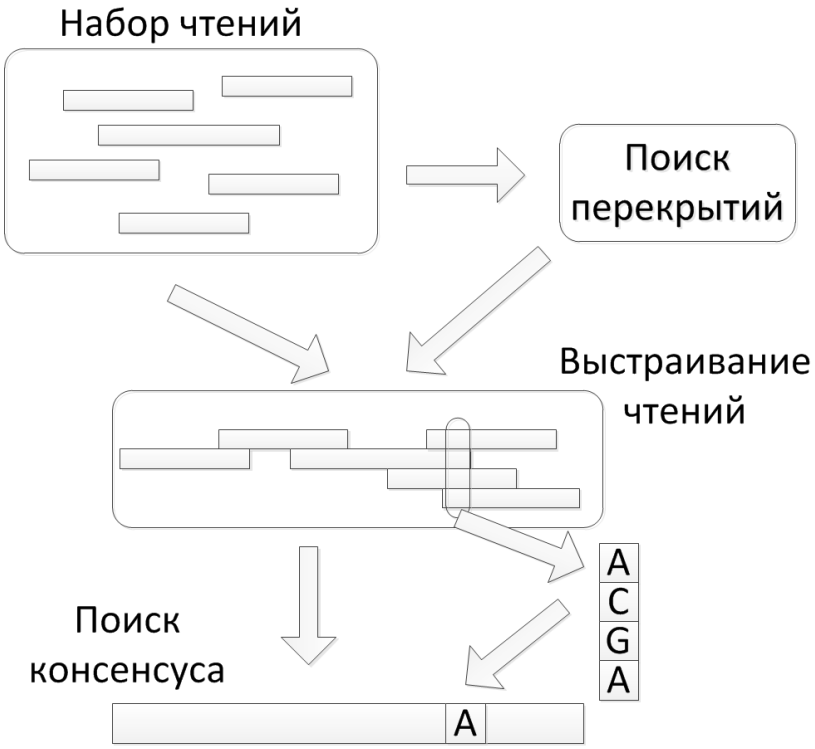


2.2. Предлагаемый метод сборки генома *de novo* (4)

Этап сборки контигов использует подход O*verlap-Layout-Consensus* (*OLC*). Он состоит из следующих шагов:

- Поиск перекрытий между квазиконтигами (*overlap stage*).
- Построение графа перекрытий.
- Упрощение графа перекрытий.
- Поиск неветвящихся путей (*layout stage*).
- Вывод консенсуса для найденных путей (*consensus stage*);

Новизна: возможность разбивать шаги на несколько шагов для уменьшения используемой памяти.



2.3. Реализация предложенного метода

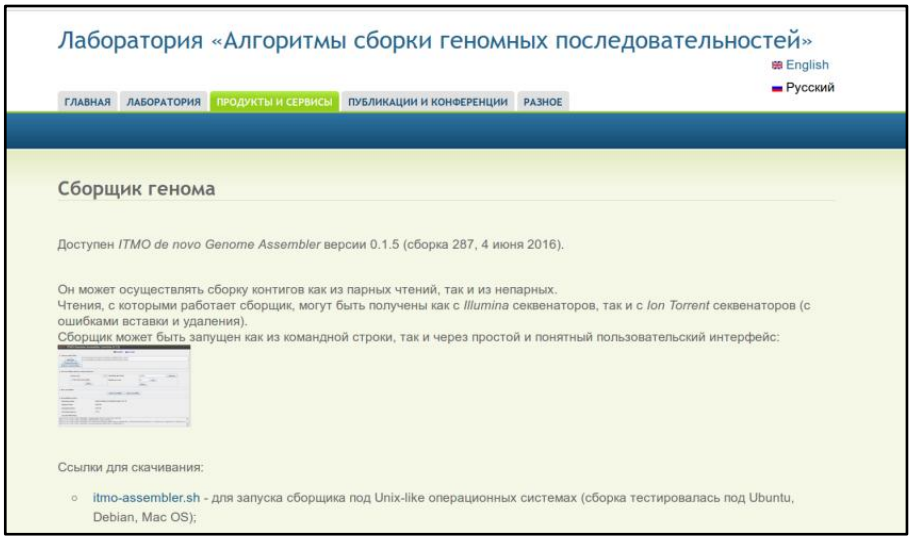
Предложенный метод был реализован на языке программирования *Java*

- Сборщик **ITMO Genome Assembler**
- Может работать при малом объеме ОЗУ, а также имеет возможность задавать объем памяти для использования, что не позволяют сделать другие сборщики.
- Свободно и бесплатно распространяется:

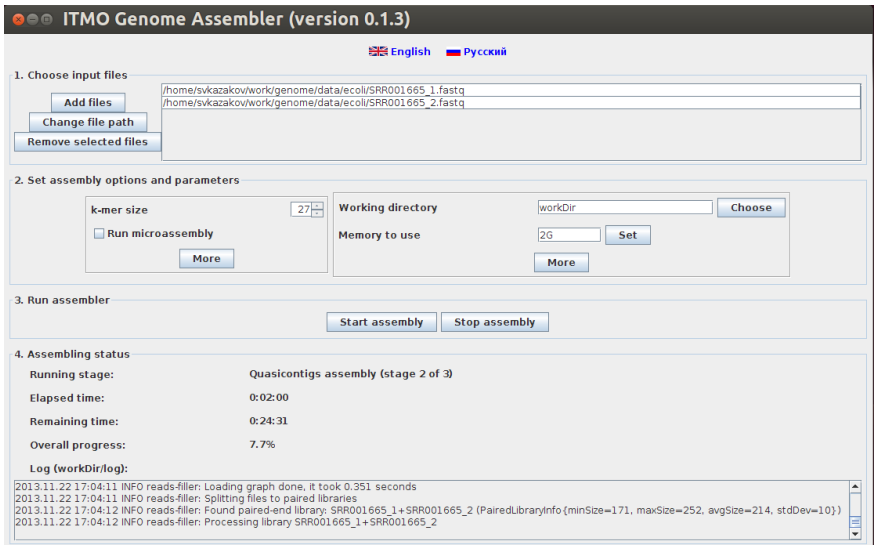
`http://genome.ifmo.ru/ru/assembler`

- **Кроссплатформенность** обеспечивается за счет использования языка *Java*: работает на *Linux, macOS/OS X, Windows.*
- Графический интерфейс пользователя для простоты запуска.
- Не требует других программ для работы.

Сайт лаборатории



Графический интерфейс пользователя



2.4. Экспериментальные исследования

Было проведено экспериментальное исследование по сравнению разработанного программного средства с существующими, указанными в таблице 1.

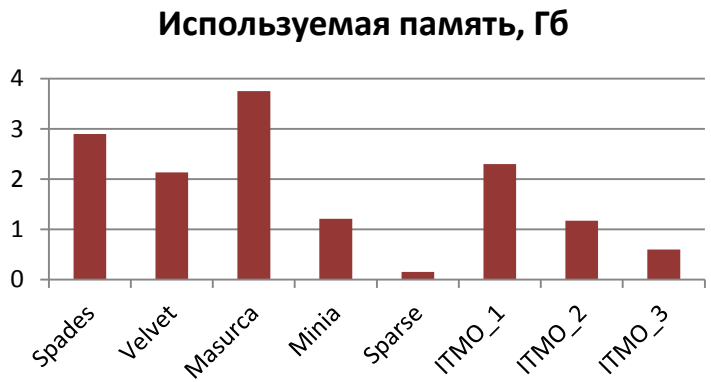
Использовались три набора данных:

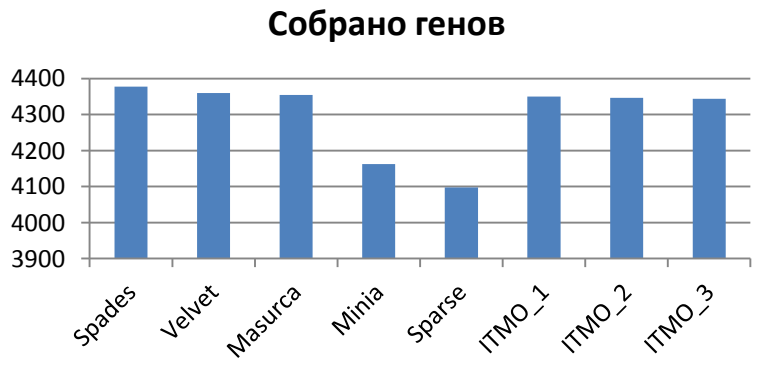
1. Малый по размеру геном (бактерия *E.coli*), стандартное покрытие.
2. Малый по размеру геном (бактерия *E.coli*), большое покрытие.
3. Средний по размеру геном (14-я хромосома человека), стандартное покрытие.

Сравнение выполнялось по затратам (память, время) и по оценке качества полученного генома.

Предлагаемое решение:

- Использует меньше памяти в среднем (средний выигрыш в **3.0–4.1** раз, на **1.4–7.9 Гб**).
- Работает в среднем быстрее (в **1.3–5** раз).
- Число собранных генов в сборке незначительно ниже (разница на **0.1–0.6%**) при сравнении с наилучшим показателем, но лучше по сравнению с другими сборщиками (разница до **81.7%**).





Выводы по главе 2

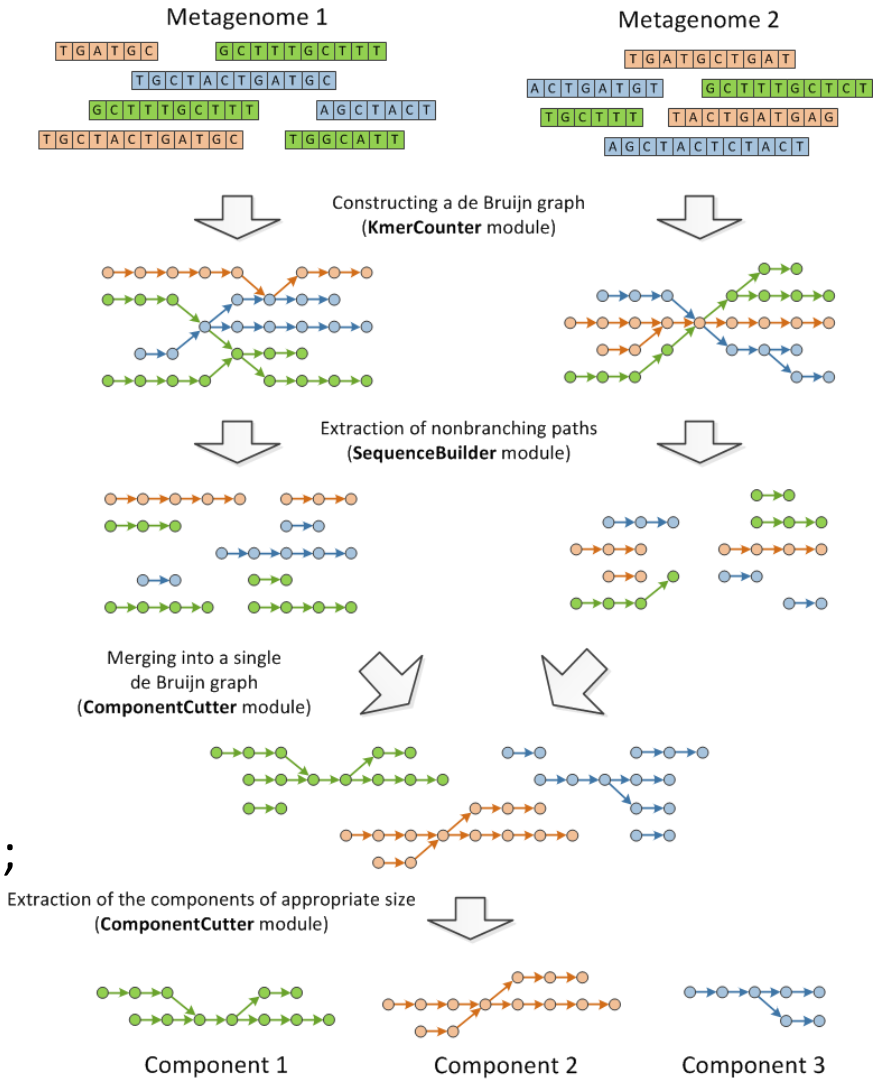
1. Разработан автоматизированный метод сборки генома *de novo,* предназначенный для целей обучения.
2. Проведены экспериментальные исследования по сравнению разработанного программного средства с существующими и показана его эффективность по объему используемой оперативной памяти.
3. Разработанное инструментальное средство *ITMO Genome Assembler* далее используется в автоматизации процесса обучения сборке генома.

3.1. Предлагаемый метод сравнительного анализа метагеномов

Задача сравнительного анализа метагеномов поставлена на слайде 13.

Предлагаемый метод *MetaFast* состоит из трех этапов:

- выполнение «упрощенной» сборки для каждого метагенома отдельно (выполняется на графе де Брейна);
- выделение компонент связности в обобщенном графе де Брейна (каждая компонента далее используется как единичный признак);
- вычисление матрицы расстояния на основе выделенных признаков (индекс Брея-Кертиса).



3.2. Новизна предлагаемого метода

Новизна предлагаемого метода состоит в применении **«частичной» (не полной) сборки** метагеномов вместо стандартной.

При «частичной» сборке выполняется только два первых шага (построение графа де Брейна и выделение неветвящихся путей) по сравнению со стандартной сборкой, при которой выполняется пять «сложных» шагов (слайд 19).

Благодаря таким изменениям:

- Получаются «короткие» последовательностей вместо длинных.
- Уменьшаются необходимые вычислительные ресурсы для анализа.
- Сохраняется разнообразие в геномных последовательностях, присутствующих в среде образцов.

Уменьшение длин собранных последовательностей **положительно сказывается** на последующем анализе, сохраняя разнообразие в них и значительно сокращая требуемые вычислительные ресурсы.

3.3. Сравнение предлагаемого метода с известными

Основанные на выравнивании на каталог известных геномов (*reference-based*)

- *Kraken*, *CLARK*, *FOCUS*, *MetaPhlAn2,* и др.
- **Необходима репрезентативная база геномов!**

Основанные на совместной сборке (*assembly-based methods*):

- *MetaVelvet*, *Meta-IDBA*, *Ray*, *MetAMOS*, *crAss*, и др.
- Получают информативные признаки.
- **Необходимы большие вычислительные ресурсы!**

Основанные на абстрактом разложении (*composition-based methods*):

- Анализ k-мерного спектра, нейронные сети, Марковские модели и др.
- *AbundanceBin*, *CompostBin*, *MaxBin*, *MetaCluster*, и др.
- Работают очень быстро.
- **Получаемые признаки неинформативны!**

Сравнение показало, что предлагаемый метод обладает преимуществами:

- Не используется база референсных геномов!
- Извлекаемые признаки информативны!
- Требует небольших вычислительных ресурсов!

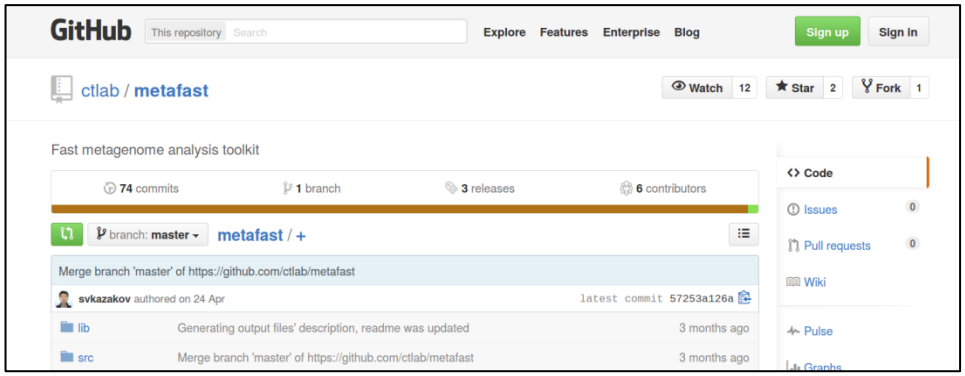
3.4. Реализация предложенного метода

Предложенный метод был реализован на языке программирования *Java:*

- Программное средство **MetaFast**.
- Может работать при малом объеме ОЗУ, имеет возможность задавать объем памяти для использования.
- Свободно и бесплатно распространяется:

`https://github.com/ctlab/metafast`

- **Кроссплатформенность** обеспечивается за счет использования языка *Java*: работает на *Linux, macOS/OS X, Windows.*
- Графический интерфейс пользователя для простоты запуска.
- Не требует других программ для работы.



3.5. Экспериментальные исследования

Были проведены экспериментальные исследования по сравнению разработанного программного средства с существующими.

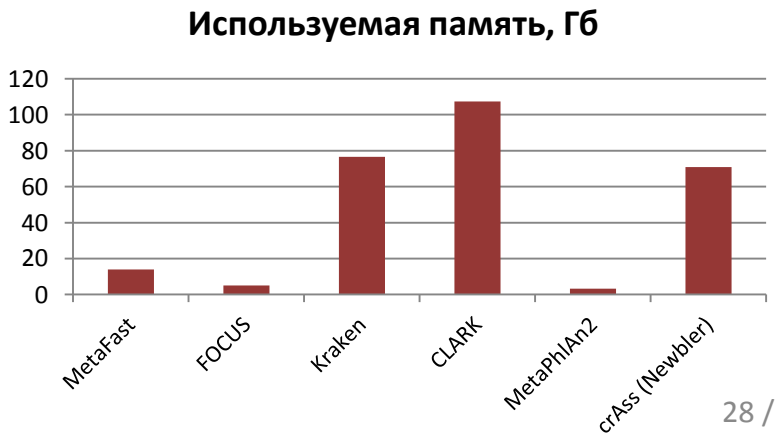
Использовались четыре набора данных:

1. Искусственные метагеномы.
2. Стандартный бактериальный набор.
3. Большой набор (157 метагеномов).
4. Набор из плохо изученной среды.

Сравнение по затратам (память, время) и по оценке качества получаемых результатов.

По результатам были сделаны следующие выводы:

- Предложенный метод не зависит от репрезентативности базы референсных геномов.
- Предложенный метод работает **на порядок быстрее** и **требует меньше памяти**, чем программы, основанные на совместной сборке.
- Метод получает информативные признаки, которые могут быть в дальнейшем использованы для их анализа.



Выводы по главе 3

1. Разработан автоматизированный метод сравнительного анализа метагеномов, предназначенный как для целей обучения, так и для научных исследований.
2. Проведено экспериментальное исследование по сравнению предложенного метода *MetaFast* с существующими решениями, которое подтвердило преимущества *MetaFast* над существующими решениями по требуемым для работы вычислительным ресурсам и возможности работы на данных из ранее неисследованных сообществ.
3. Разработанное инструментальное средство *MetaFast* далее используется в автоматизации процесса обучения сравнительному анализу метагеномов.

4.1. Внедрение в практико-ориентированное образование (1). Имеется акт внедрения

Политехнический университет:

- Проведение лекционных и практических занятий в рамках магистерской программы «Прикладная математика и информатика. Биоинформатика».
- Лабораторное задание по *de novo* сборке генома **двух бактерий** (искусственные и реальные данные).
- Лабораторная работа была выполнена **восемью магистрантами**.

Разработанные программы **впервые позволили провести практическое занятие** по теме *de novo* сборки генома.

Анализ изменений в обучении:

Таблица 2

Критерий	Без использования предложен. программ	С использованием предложен. программ
Проведение практических занятий с учителем	Не проводились	Проводились
Скорость понимания лекционного материала с учетом практических занятий	Средняя	Высокая
Возможность повторения практического занятия дома	Нет	Есть

4.2. Внедрение в научно-образовательный процесс. Имеется акт внедрения

Разработанные методы были внедрены в научно-образовательный процесс по анализу геномов бактерий в Казанском (Приволжском) федеральном университете.

Разработанная программа **впервые позволила провести полный процесс анализа** геномов бактерий на персональном компьютере с ОС *Windows*.

- Лаборатория масс-спектрометрии Института фундаментальной медицины и биологии КФУ.
- Анализ геномов **шести малоизученных бактерий**.
- Последующий анализ полученных геномов: сравнение с близкими видами, анализ присутствующих генов, выделение отличительных особенностей и сопоставление их с имеющейся информацией о выполняемых функциях.
- По результатам этих работ было **опубликовано три статьи** в англоязычных журналах *Genome Announcements* и *Standards in Genomic Sciences* (2014–2015 гг.).
- По результатам этих работ **три генома** были добавлены в Национальный центр биотехнологической информации (*NCBI,* США), **один** также добавлен в Базу геномов *GOLD* (США).

4.3. Внедрение в практико-ориентированное образование (2). Имеется акт внедрения

Университет ИТМО:

- Проведение лекционных и практических занятий на тему «Сборка генома *de novo* и сравнительный анализ метагеномов» на кафедре «Компьютерные технологии» в рамках курса лекций по биоинформатике.
- Лабораторное задание по анализу **двух метагеномных наборов** (один искусственный и один реальный набор данных).
- Лабораторная работа была выполнена **двенадцатью студентами**.

Разработанные программы **впервые позволили провести практические занятия** по темам сборки генома *de novo* и сравнительному анализу метагеномов.

Анализ изменений в обучении:

Таблица 3

Критерий	Без использования предложен. программ	С использованием предложен. программ
Проведение практических занятий с учителем	Не проводились	Проводились
Выполнение домашних заданий на пройденный материал	Не использовались	Использовались

Выводы по главе 4

1. Обеспечена автоматизация процессов обучения сборки генома *de novo* и сравнительного анализа метагеномов, который базируется на использовании инструментальных средств сборки генома *ITMO Genome Assembler* и сравнительного анализа метагеномов *MetaFast*.
2. Результаты диссертационной работы внедрены в образовательный процесс в Санкт-Петербургском политехническом университете Петра Великого и в Университете ИТМО при проведении практических занятий и лабораторных работ.
3. Результаты диссертационной работы внедрены в научно-образовательный процесс в Казанском (Приволжском) Федеральном Университете при выполнении по анализу геномов бактерий.

Заключение. Результаты работы (1)

Результаты диссертационной работы состоят в следующем:

1. Предложен автоматизированный метод сборки генома *de novo*, предназначенный как для целей обучения, так и для научных исследований.
   Показано, что метод в среднем требует меньше памяти по сравнению с существующими средствами и может быть использован для обучения геномной биоинформатике.

2. Предложен автоматизированный метод сравнительного анализа метагеномов, предназначенный как для целей обучения, так и для научных исследований.
   Показано, что метод работает эффективнее по вычислительным ресурсам, чем известные решения, а также имеет преимущество в независимости от базы референсных геномов.

Разработанные инструментальные средства обеспечивают автоматизацию процессов обучения сборки генома и сравнительного анализа метагеномов.

Внедрение результатов работы было осуществлено:

- в образовательный процесс **в Политехническом университете**, имеется акт внедрения;
- в образовательный процесс **в Университете ИТМО**, имеется акт внедрения;
- в научно-образовательный процесс **в Казанском (Приволжском) федеральном университете**, имеется акт внедрения;

Заключение. Результаты работы (2)

19 публикаций, из них:

- четыре статьи в журналах, рекомендованных ВАК:
  - Александров А. В., Казаков С. В., Мельников С. В., Сергушичев А. А., Царев Ф. Н., Шалыто А. 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.
  - Александров А. В., Казаков С. В., Мельников С. В., Сергушичев А. А., Царев Ф. Н. 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.
  - Сергушичев А. А., Александров А. В., Казаков С. В., Царев Ф. Н., Шалыто А. А. Совместное применение графа де Брейна, графа перекрытий и микросборки для *de novo* сборки генома // Известия Саратовского университета. Новая серия. Серия Математика. Механика. Информатика. 2013. Т. 13, вып. 2, ч. 2. С. 51–57.
  - Казаков С. В., Шалыто А. А. Анализ геномных и метагеномных данных в образовательных целях // Компьютерные инструменты в образовании. 2016. № 3, с. 5–15.

Заключение. Результаты работы (3)

- четыре публикации в изданиях, индексируемых *Web of Science* или *Scopus*
  - *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 the 12th annual conference in bioinformatics "Bioinformatics 2012". Stockholm, Sweden. 2012. P. 72.
  - *Bradnam K., Fass J., Alexandrov A., Baranay P., Bechner M., Kazakov S. et al.* Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species // GigaScience. 2013. № 2(10), pp. 1–31.
  - *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). 2014. Pp. 222–226.
  - *Ulyantsev V., Kazakov S., Dubinkina V., Tyakht A., Alexeev D.* MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data // Bioinformatics. 2016. № 32 (18), pp. 2760–2767.

Заключение. Результаты работы (4)

Сделано 15 докладов на конференциях

- 12 докладов на российских конференциях
  - Всероссийская межвузовская конференция молодых ученых (Санкт-Петербург, 2011 г.)
  - Международная конференция «Постгеномные методы анализа в биологии, лабораторной и клинической медицине: геномика, протеомика, биоинформатика» (Новосибирск, 2011 г.)
  - Всероссийская конференция «Телематика'2012» (Санкт-Петербург, 2012 г.)
  - Международная конференция «Постгеномные методы анализа в биологии, лабораторной и клинической медицине» (Казань, 2012, 2014 гг.)
  - Всероссийская конференция по проблемам информатики СПИСОК (Санкт-Петербург, 2012, 2016 гг.)
  - Всероссийский конгресс молодых ученых (Санкт-Петербург, 2012 г.)
  - Международная школа-конференция студентов, аспирантов и молодых ученых «Биомедицина, материалы и технологии XXI века» (Казань, 2015 г.)
  - Moscow Conference on Computational Molecular Biology (Москва, 2015 г.)
  - Летняя школа по биоинформатике (Москва, 2015 г.)
  - Международная конференция «Компьютерные науки и информационные технологии» (Саратов, 2016 г.)
- Три доклада на зарубежных конференциях
  - *de novo* Genome Assembly Assessment Project workshop (dnGASP) (Испания, Барселона, 2011 г.)
  - «Bioinformatics 2012» Conference (Швеция, Стокгольм, 2012 г.)
  - International Conference on Intelligent Systems and Agents (Португалия, Лиссабон, 2014 г.)

Заключение. Результаты работы (5)

Получено пять свидетельств о регистрации программы для ЭВМ:

- № 2013660881 от 21 ноября 2013 года «Программное средство, реализующее алгоритм упрощения графа перекрытий при сборке геномных последовательностей»;
- № 2013619155 от 26 сентября 2013 года «Программное средство, реализующее запуск этапов сборки генома через графический интерфейс пользователя»;
- № 2013616471 от 09 июля 2013 года «Программное средство, реализующее алгоритм поиска перекрытий между квазиконтигами»;
- № 2012616774 от 27 июля 2012 года «Программное средство для сборки квазиконтигов из парных чтений»;
- № 2011614454 от 06 июня 2011 года «Программное средство для удаления ошибок из набора чтений нуклеотидной последовательности».

Часть результатов получено при выполнении государственных контрактов:

- «Разработка методов сборки генома, сборки транскриптома и динамического анализа протеома» (Государственный контракт № 14.B37.21.0562, 2012–2013 гг.)
- «Разработка метода сборки геномных последовательностей на основе восстановления фрагментов по парным чтениям» (Государственный контракт № 16.740.11.0495, 2011–2013 гг.)