

Sibelia

Колмогоров Михаил

Руководитель: Илья Минкин



SiBELia

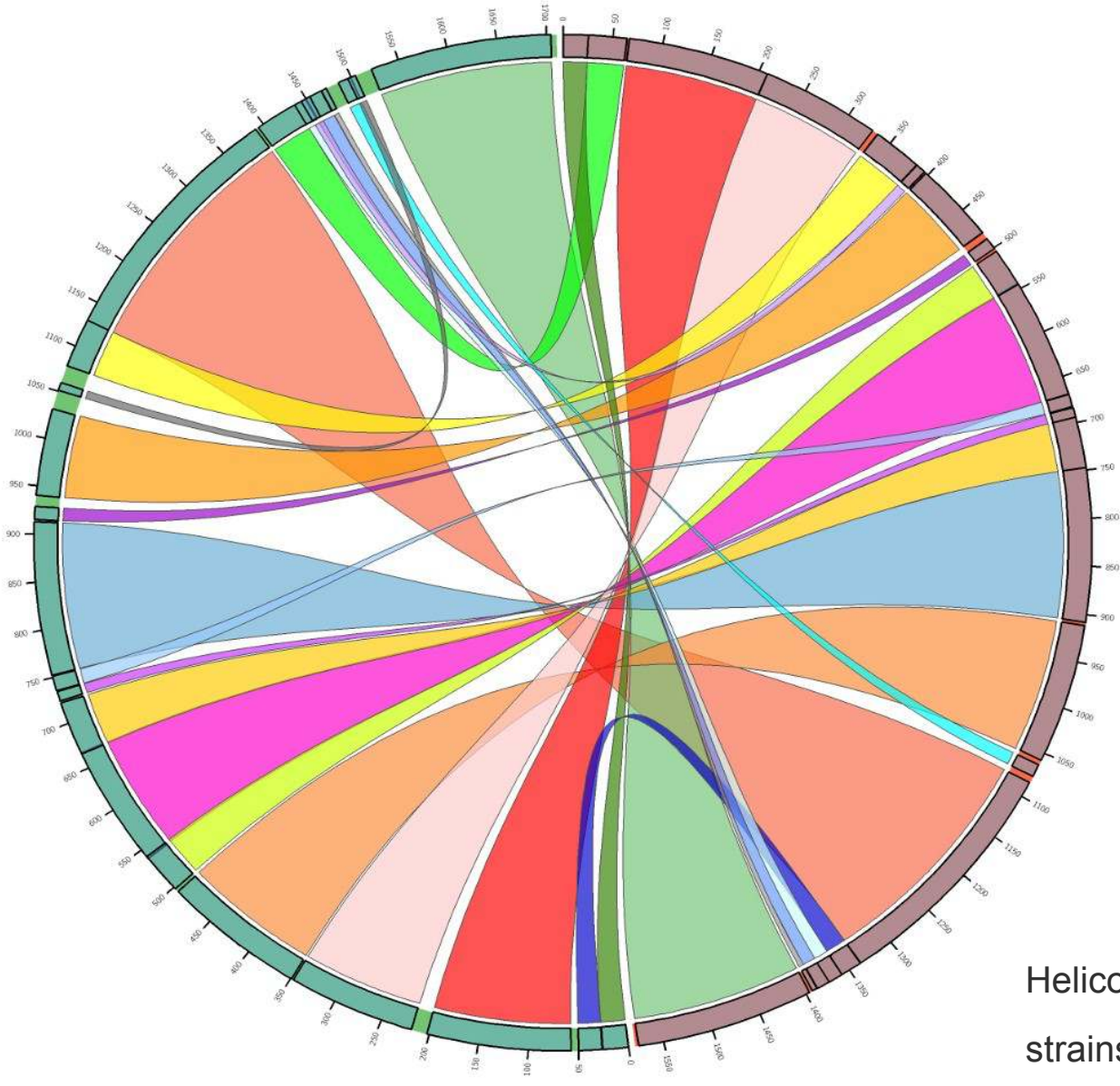
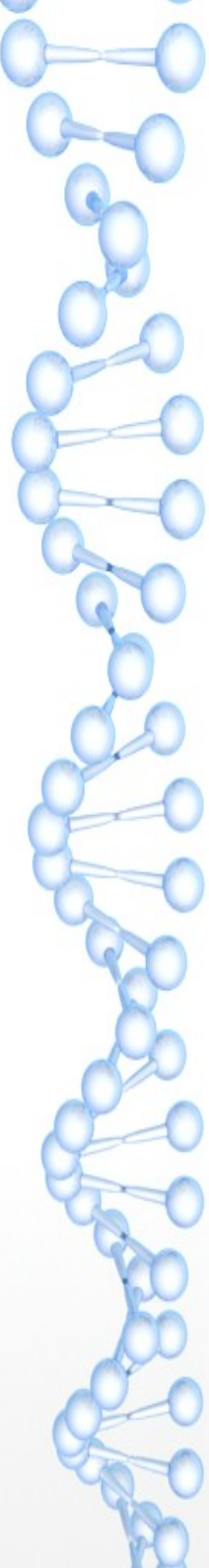
- Synteny Blocks ExpLoration tool
- Поиск Synteny Blocks в геномах, представленных последовательностями нуклеотидов
- Подход с использованием графа де Брюйна
- Synteny Blocks — схожие (обычно, не точно) участки различных геномов



Задачи

- Визуализация найденных блоков
- Структура данных для хранения сжатого графа де Брюйна
- Оптимизация алгоритма упрощения графа

Circos



Helicobacter pylori,
strains F32 and Gambia94/24.

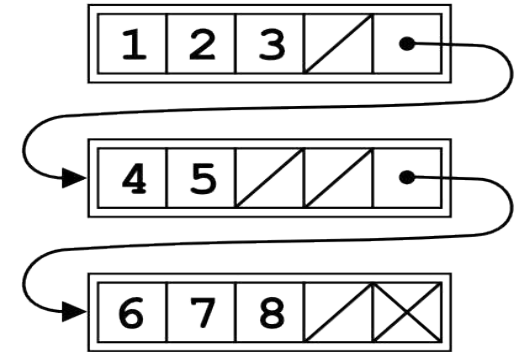


Unrolled list

- Структура для хранения последовательностей геномов
- Минимальный оверхед по памяти (к примеру, на указатели)
- Быстрые вставка/удаление
- Минимальное количество элементов, перемещаемых при вставке
- STL - совместимый

Done

- «Список массивов»
- Вставка — дописываем в массив, если надо добавляем новую ячейку списка. Иногда надо перемещать старые элементы в новую ячейку
- Удаление — помечаем как удаленный
- Функции обновления «испорченных» итераторов
- Потребление памяти уменьшилось в 2-3 раза. Выросла производительность (меньше cache-miss`ов)



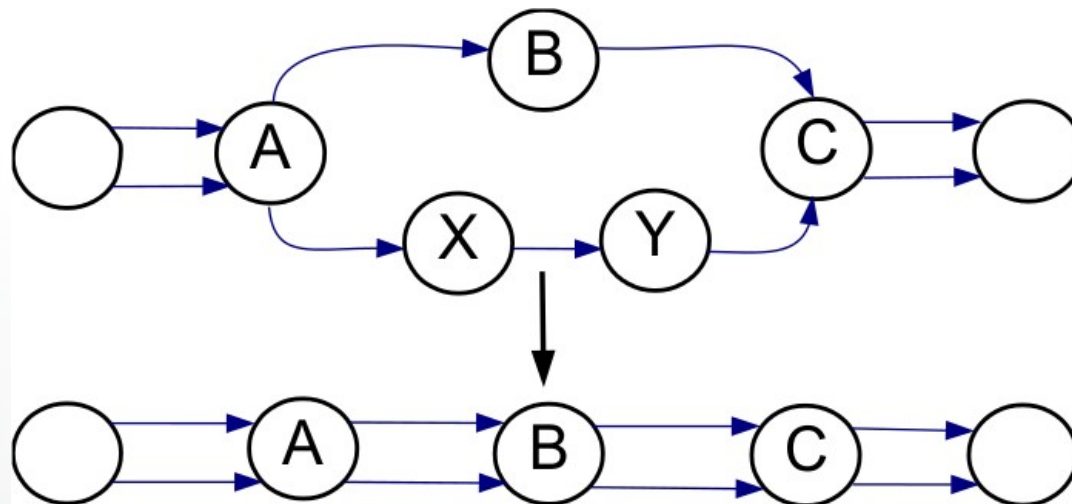


Сравнение результатов

Dataset\Version	Sibelia 1.0	Sibelia 2.0 (with unrolled list)
Helicobacter pylori	100mb, 96 seconds	55mb, 42 seconds
Staphylococcus	346mb, 485 seconds	111mb, 177 seconds

Упрощение графа

- Пузыри в графе портят длинные неразветвленные пути (соответствующие synteny block`ам)
- Пузырь удаляется заменой одного из путей на копию другого
- Для каждой пары исходящих из вершины графа путей надо найти и «схлопнуть» пузырь — квадратичная сложность





Оптимизация

- Отсечение исходящих путей, которые заведомо не образуют пузырь
- Линейно проходим по каждому из путей, выходящих из вершины, записываем все их пары, которые в дальнейшем пересекаются
- Ищем и «схлопываем» пузыри по всем парам путей, их образующих (и лишь по ним)
- На «тяжелых» входных данных производительность выросла в 14-15 раз.

Конец

<http://bioinf.spbau.ru/sibelia>

**Не понимаю, как я раньше искала
Synteny Block`и без Sibelia!**

Я тоже!

