

Сборка оптических карт: доработка ассемблера и моделирование данных оптического картирования

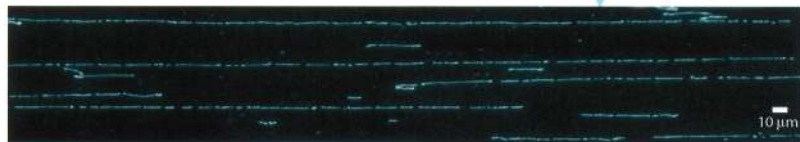
Тетюхин Артём
Шиляев Никита

Институт биоинформатики

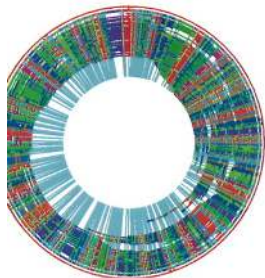
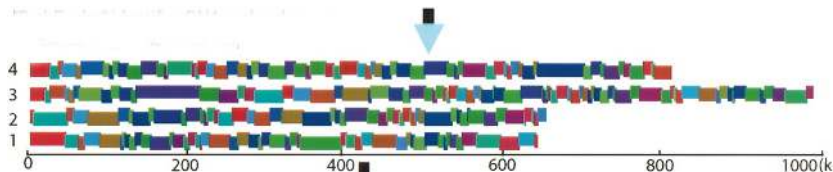
Научный руководитель: Святослав Сидоров
Санкт-Петербургский Академический Университет

Санкт-Петербург, 2014

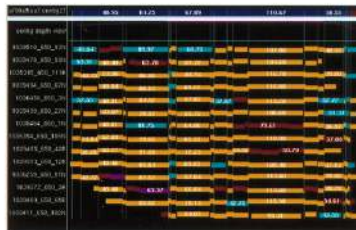
Оптическое картирование



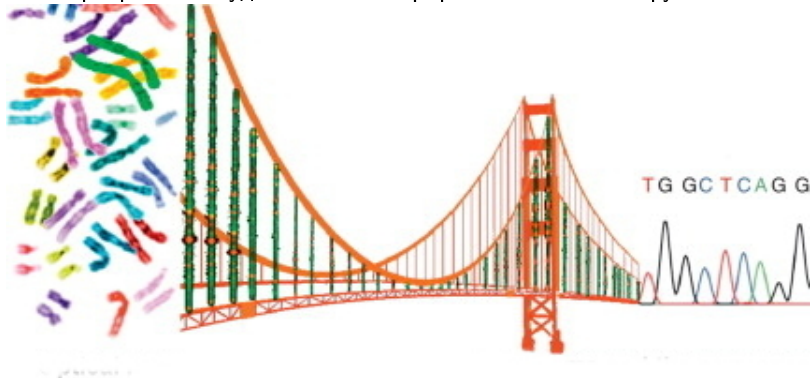
Оптическое картирование



Сборка



Доработка программы сборки оптических карт, разработанной А. Валуевым (PhD University of Southern California). Превращение этой программы в удобный биоинформатический инструмент.

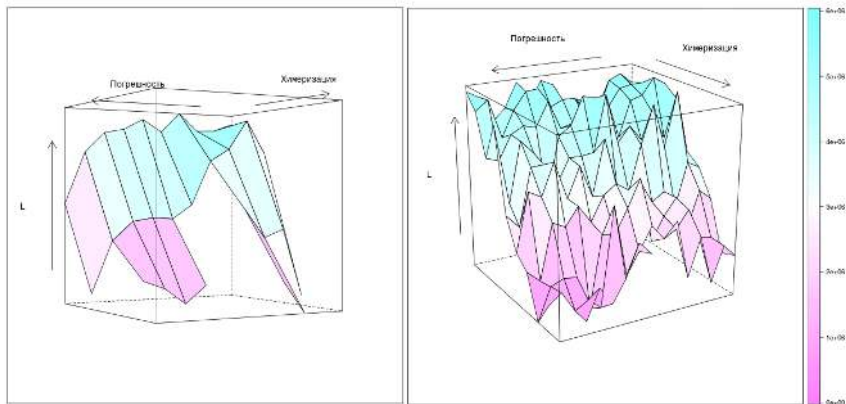


- 1 Написан скрипт для симуляции ошибок оптического картирования.
- 2 Написан скрипт для автоматического перебора значений параметров (из некоторого диапазона), задающих свойства ошибок.
- 3 На датасетах, получаемых с помощью скрипта из п. 2, проведено тестирование сборщика в условиях двух моделей симуляции ошибок.

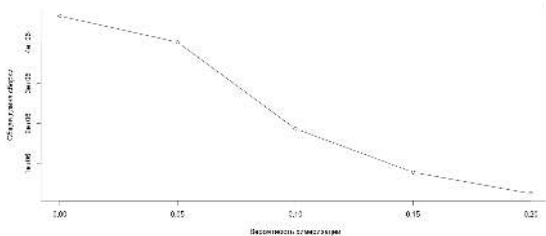
- 1 **Мультипликативная модель.** Длина фрагмента умножается на $1 - \alpha$, где $\alpha \sim \mathcal{N}(0, \sigma)$ для некоторого σ .
- 2 **Смешанная модель.** Пусть l_0 – отсечка по длине фрагмента. Тогда, если длина фрагмента l меньше l_0 , то для него действует мультипликативная модель. Если же $l \geq l_0$, то длина фрагмента становится равной $l + 1000\alpha$, где α было определено выше.

Профили сборки:

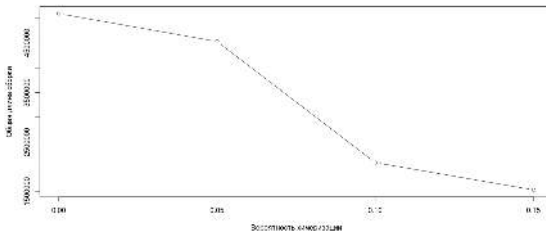
Мультипликативная модель < — > Смешанная модель



Мультипликативная модель



Смешанная модель

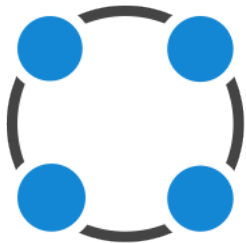


- 1 Рефакторинг кода проекта, то есть реорганизация кода. Рефакторинг исходного кода необходим для удобства его анализа и последующей доработки, а также для обеспечения содержательного журналирования работы программы. Код был приведен к единому стилю с использованием нового стандарта C++ (C++ 11).
- 2 Написан скрипт для кроссплатформенной сборки проекта (smake). Проект можно скомпилировать в любой операционной системе (Unix-like, Mac OS, Windows).
- 3 Оформление и правка описания проекта.

Данный проект находится на Github среди других репозиториях Dobzhansky Center for Genome Bioinformatics и пока закрытый, то есть находится в приватном репозитории.

<https://github.com/DobzhanskyCenter>

Таким образом, удалось подготовить проект к следующей стадии: анализу параметров выравнивания оптических рядов и анализу процедур построения и коррекции графа сборки.



Bioinformatics Institute

Спасибо за внимание!