

Улучшение качества сборки генома с использованием mate-pair библиотек

Аня Малова

АУ РАН

Научный руководитель: Шон Фам

15 декабря 2012 г.

Цель

Улучшение
качества
сборки
генома с
использова-
нием
mate-pair
библиотек

Аня Малова

Улучшение качества сборки генома с использованием mate-pair библиотек

- *Staphylococcus aureus* (Золотистый стафилококк)
- Library 1: Insert size *180bp*, средняя длина чтений *101bp*
- Library 2 (mate-pair): Insert size *3500bp*, средняя длина чтений *37bp*

Цель

Улучшение
качества
сборки
генома с
использова-
нием
mate-pair
библиотек

Аня Малова

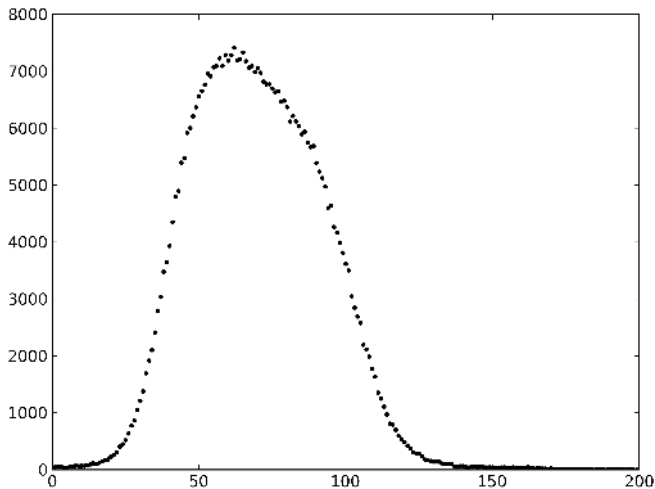
Улучшение качества сборки генома с использованием mate-pair библиотек

- *Staphylococcus aureus* (Золотистый стафилококк)
- Library 1: Insert size $180bp$, средняя длина чтений $101bp$
- Library 2 (mate-pair): Insert size $3500bp$, средняя длина чтений $37bp$

Распределение расстояния между чтениями (на геноме). Library 1.

Улучшение
качества
сборки
генома с
использова-
нием
mate-pair
библиотек

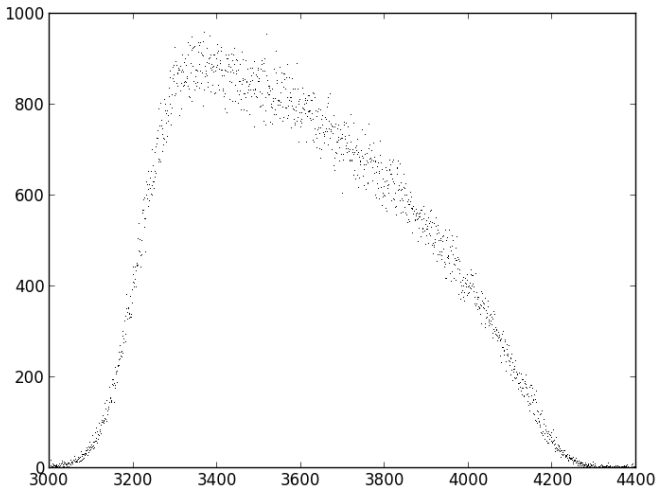
Аня Малова



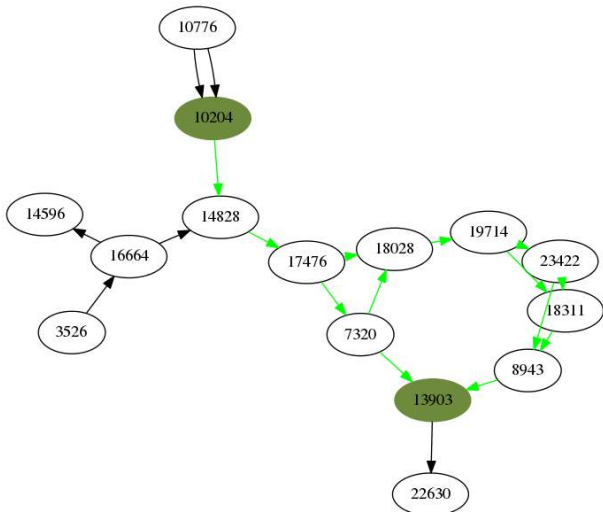
Распределение расстояния между чтениями (на геноме). Library 2.

Улучшение
качества
сборки
генома с
использова-
нием
mate-pair
библиотек

Аня Малова



Нахождение всех путей в заданном диапазоне между чтениями на графе Де-Брюина.



Дальнейшее направление

Улучшение
качества
сборки
генома с
использова-
нием
mate-pair
библиотек

Аня Малова

Модификация алгоритма Path-Extend:

- Другая инициализация:
 - объединять два последовательных длинных ребра, если это подтверждается многими парами чтений
 - Для некоторых пар вершин есть только один путь между ними. Возьмём его
- Объединять два сильно пересекающихся пути