



ИНСТИТУТ
БИОИНФОРМАТИКИ

Использование неравномерности покрытия бактерий в фазе роста для разрешения повторов при сборке *de novo*

Руководитель:

Антипов Дмитрий, центр алгоритмической
биотехнологии СПбГУ

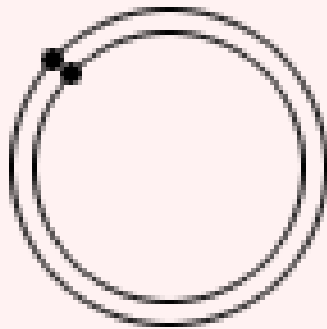
Студент:

Нестеренко Максим

Цель проекта

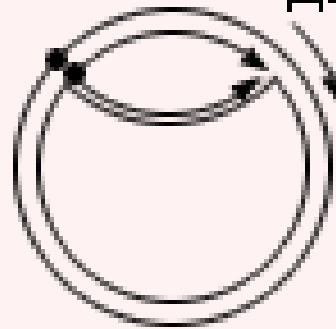
Улучшение сборки бактериальных геномов в специфическом случае, когда колония бактерий находится в фазе роста

Начало репликации



Исходный момент

Движение
вилки



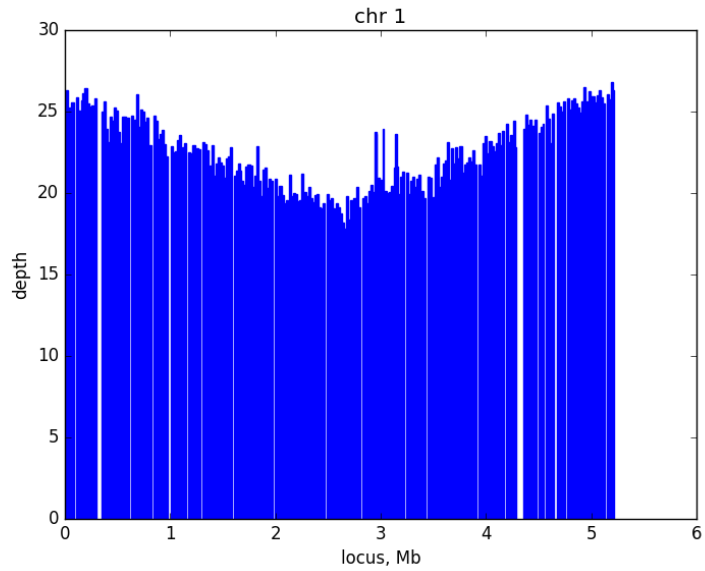
В ходе репликации

Фаза роста, как это понять?

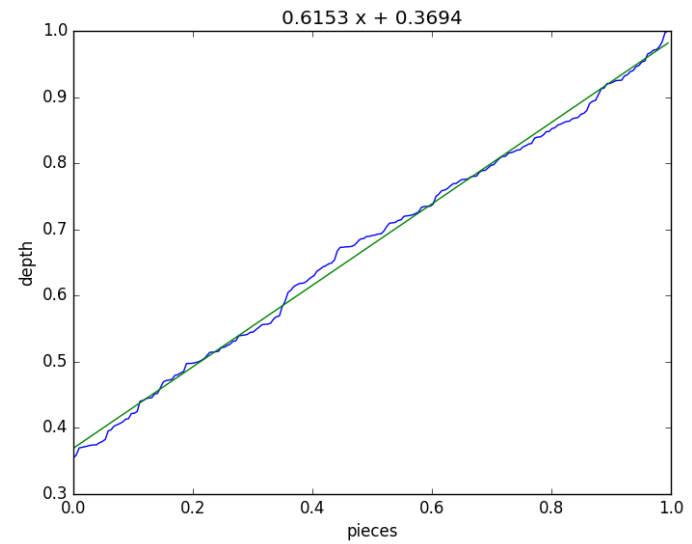
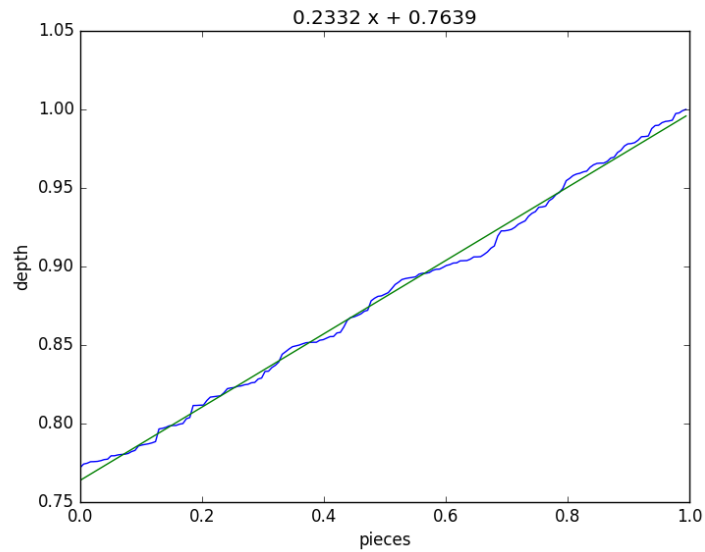
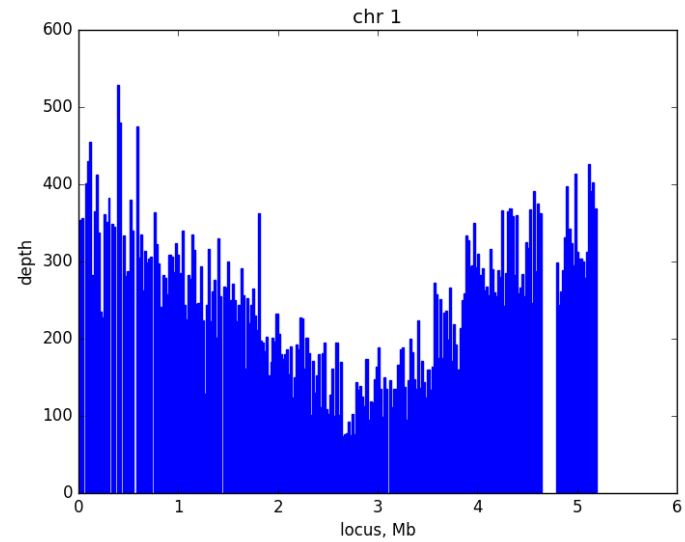
- Собираем контиги (Spades)
- Прикладываем к контигам риды и определяем покрытие в каждой позиции контига (Samtools)
- Разбиваем контиги на относительно небольшие кусочки (например 1000 bp), считаем среднее покрытие кусочка
- Сортируем кусочки по покрытию и строим зависимость: покрытие (номер куска). По углу наклона пытаемся определить, находилась ли бактерия в фазе роста

Примеры

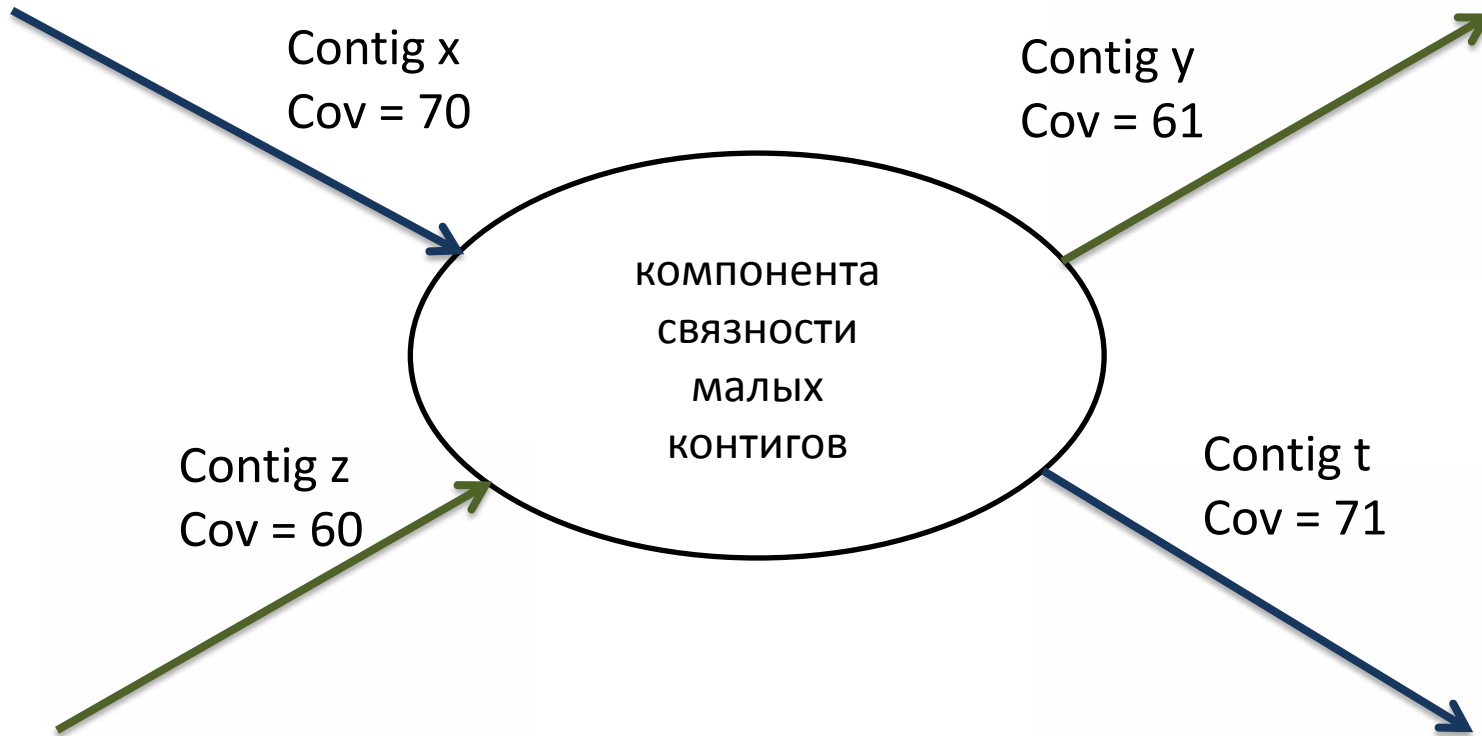
B.cereus



B.anthraxis



Разрешение повторов, простой пример

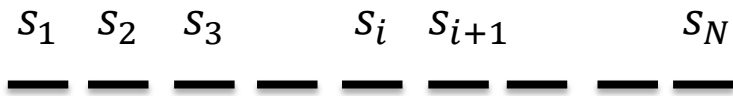


Из рисунка видно, что контиги (x, t) и (z, y) обладают схожим покрытием, что позволяет разрешить возникающий повтор и выяснить порядок в котором идут контиги

Возникшие трудности

- Сильная “зашумленность” может помешать, при разрешении повторов, чтобы оценить, насколько сильно, посчитаем sd и сравним с H

s_1 s_2 s_3 s_i s_{i+1} s_N

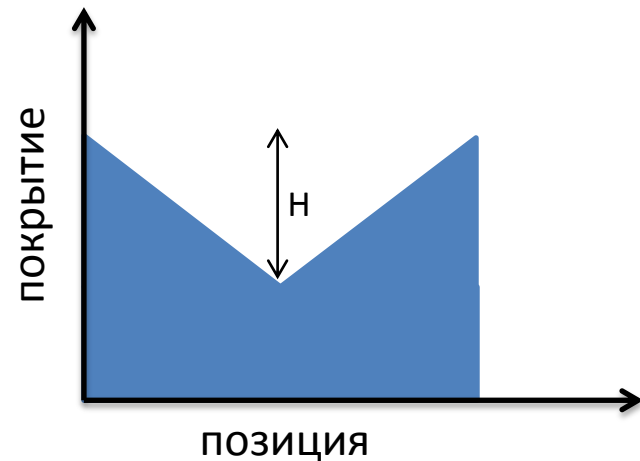


s_i - покрытие куска контига

d_i - разница покрытий соседних кусков

$$d_i = s_i - s_{i+1}$$

$$sd = \sqrt{\frac{(d_i - \bar{d})^2}{N - 1}} \sim H$$



H – изменение покрытия, обусловленное фазой роста

Результаты и выводы

bacteria	<i>sd</i>	H
V.anthraxis	110,72	352,96
V.cereus	4,23	8,88
P.stuartii	64,83	176,31
C.freundii	15,51	38,78

- Высокое значение *sd* покрытий соседних кусков в большинстве случаев не позволяет использовать неравномерность покрытия для разрешения повторов, однако, это становится возможным при достаточно низких значениях *sd*

Спасибо за внимание!