



Выполнил: Адвеев Павел

Научный руководитель: Михаил Колмогоров

Обзор

1. Построение последовательности блоков синтении (Sibelia and multiple alignment).
2. Итеративная сборка:
 - 2.1. Строим граф брейкпоинтов для всех геномов
 - 2.2. Считаем вес для каждого ребра (с помощью процедуры парсимонии)
 - 2.3. Находим совершенное паросочетание минимального веса
 - 2.4. Объединяем контиги с учетом найденных смежностей
3. Уточнение с помощью графа перекрытий:
 - 3.1. Строим граф перекрытий
 - 3.2. Находим, короткие контиги, которые нужно вставить между смежными контигами в текущем скаффолде

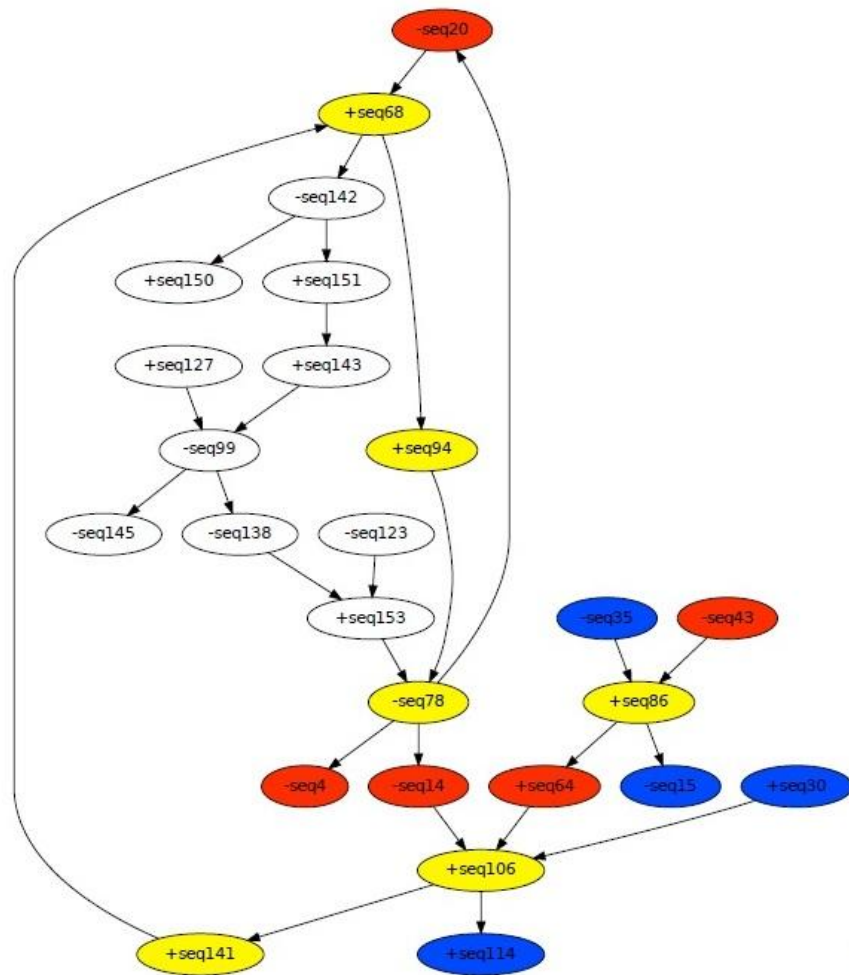
Граф перекрытий

Вершины:

Контиги

Ориентированные ребра:

Две вершины, соединяются ребром, если конец одного контига, пересекается с началом другого контига, на размер равный значению `overlap`.



Как строить граф перекрытий?

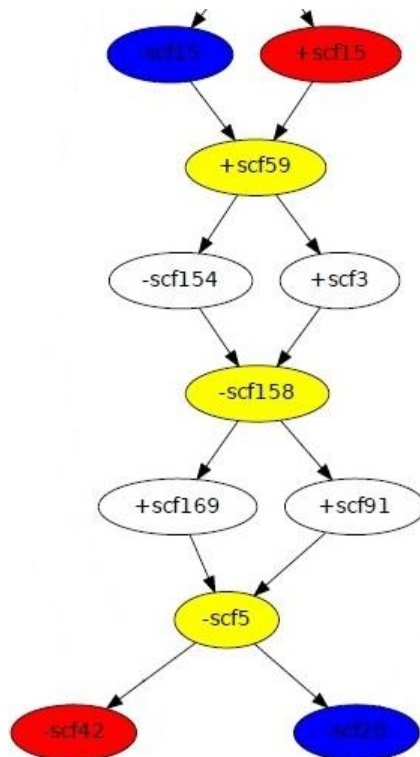
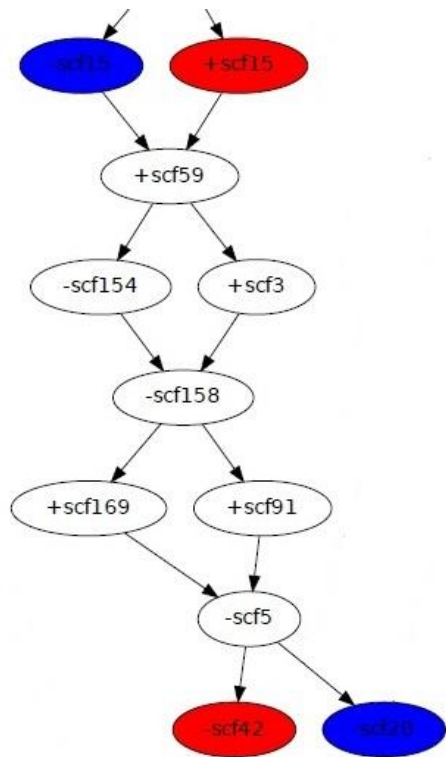
- На какое значение перекрывать контиги?
- Перекрывать ровно на это значение, на больше или равно, на интервал значений?
- Состав контигов для построения графа (контиги из скаффолда, комплиментарные контиги из скаффолда, короткие контиги)?

Вставка коротких контигов

Для каждой пары соседних контигов (вершин u , v) из скаффолда:

Ищем уникальный кратчайший путь, проходящий только через короткие контиги, причем длина пути должна не превосходить наперед заданной константы.

Вставка коротких контигов



Вставка коротких контигов

Давайте искать точки сочленения!

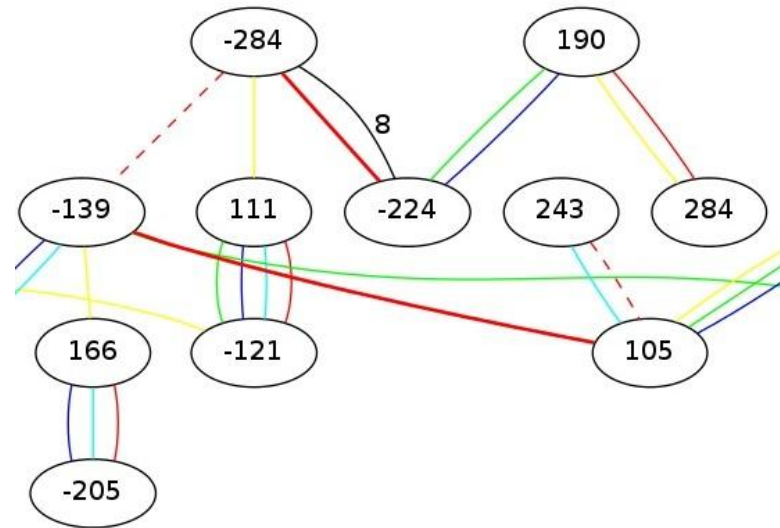
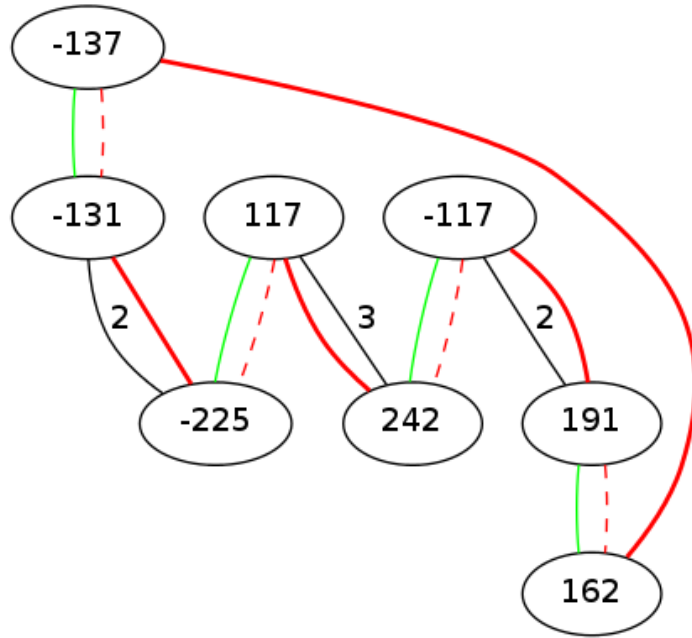
Для каждой пары соседних контигов (вершин u , v) из скаффолда:

1. В графе перекрытий, если нет ребра (u, v) , добавляем ребро (v, u)
2. Находим сильно связный подграф, содержащий вершины u , v .
3. Удалив вставленное ребро, строим граф блоков-точек сочленения.
4. Добавляем в scaffold, все точки сочленения лежащие на пути в графе блоков точек сочленения между вершинами u , v

Результаты

| | Old result | | New result | |
|------------|----------------|---------------|----------------|---------------|
| | Miss - ordered | Total contigs | Miss - ordered | Total contigs |
| E.Coli | 0 | 133 | 0 | 148 |
| H.Pylori | 0 | 71 | 0 | 142 |
| S.Aureus | 2 | 161 | 0 | 199 |
| V.Cholerae | 2 | 298 | 16 | 850 |
| Drosophila | - | - | 33 | 1793 |

Перестройки и исправление не правильно вставленных КОНТИГОВ



Итого

Задачи, которые были поставлены

1. Улучшить построение графа перекрытий
2. Улучшить вставку коротких контигов в скаффолд
3. Посмотреть как можно находить и исправлять перестройки

выполнены

Спасибо за
внимание!