

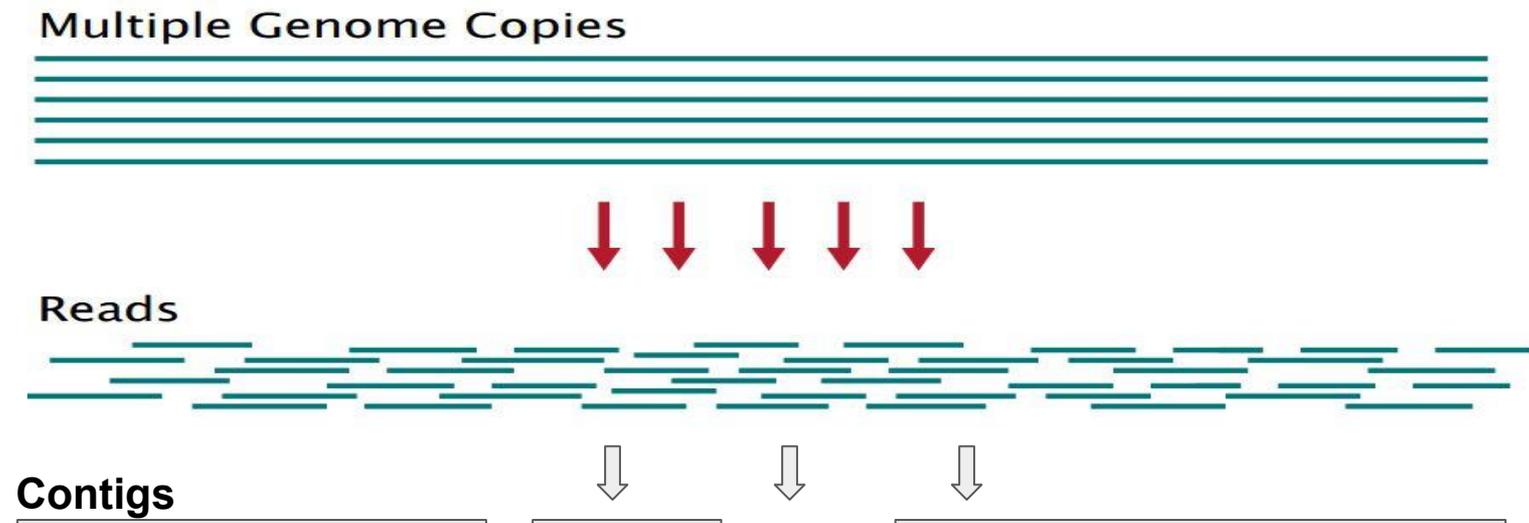
Анализ ошибок SPAdes

Даниил Плющенко

Руководитель: Юрий Горшков, лаборатория "ЦАБ"

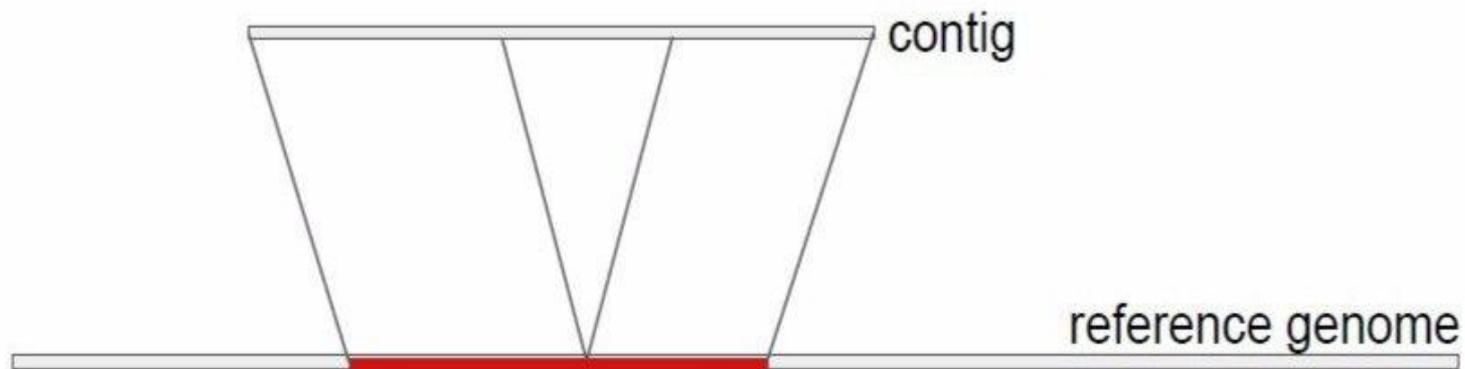
Суть задачи

Научиться сообщать в удобном формате об ошибках сборки (misassemblies).
Полученный результат встроить в диагностику SPAdes.



Пример ошибки

Фрагмент приложен к неправильной позиции.



Чем пользовался? С чем приходилось разбираться?

SPAdes - непосредственно сборка генома. Для экспериментов использовалась сборка кишечной палочки, дававшая 11 ошибок.

QUAST - оценка качества сборки путём приложений контигов (результата сборки) и референса (“идеального” генома)

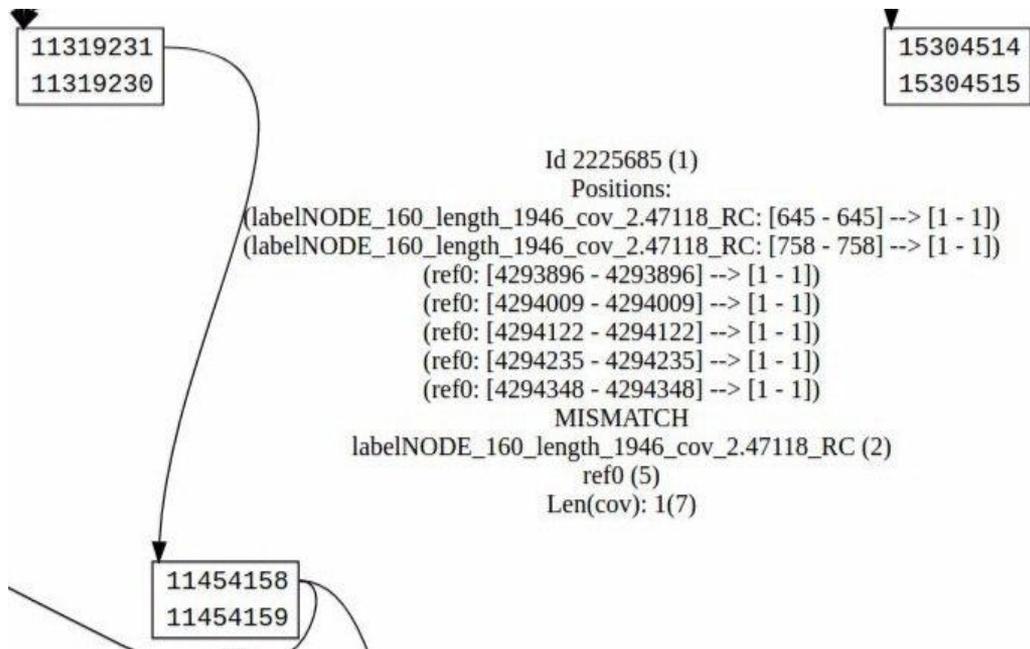
OnlineVis - диагностическая отрисовка фрагментов графа сборки на определённой стадии работы SPAdes с информацией на рёбрах (кол-во приложений и их позиции).

Код на C++.

Понятный вид ошибки

Ошибки проявляются различным кол-вом приложений к ребру.

Добавление меток “MISMATCH”
на такие рёбра. Заметить проще,
но кардинально ситуацию не
улучшает, т.к. подобные места
всё ещё надо искать “руками”.



Улучшение

Хотелось реализовать команду (в OnlineVis) для отрисовки всех подобных участков графа и их окрестностей. За основу была взята ранее имеющаяся, но неработающая команда `draw_misassemblies`.

Были исправлены некоторые ошибки в коде, в некоторых случаях командой стало можно пользоваться.

Добавлена новая версия алгоритма.

Что не получилось?

Довести исправления/алгоритм до полноценно работающей единой программы.

На данный момент есть несколько разрозненных кусков кода, которые частично выполняют задачу.