

Глубокий анализ ошибок сборки

Студент: Татьяна Кривошеева
Руководитель проекта: Сергей Нурк ¹

¹Лаборатория алгоритмической биологии

Защита проектов, весна 2015

Ошибки геномного ассемблера



Read1



Read2



Assembly

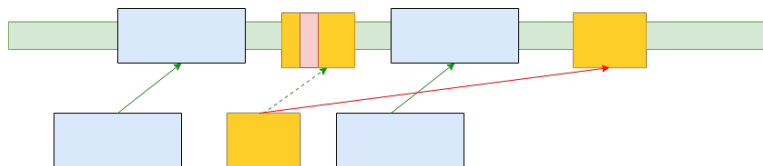


Задачи

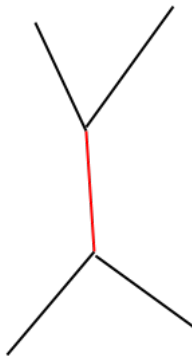
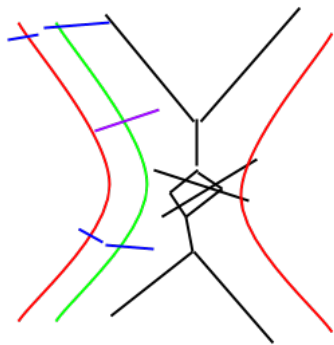
- ▶ Автоматизация процесса поиска ошибок (extensive misassemblies)
- ▶ Автоматизация анализа причин их возникновения

Поиск ошибок ошибок

1. SPAdes
2. QUAST
3. Выявление false positives



Поиск ошибок ошибок



Планы



4. True positives
5. Автоматизация процесса анализа ошибок
6. Визуализацию выделенных областей графа сборки

Несоответствие референсу

1. Часть референса прикладывается к геному, но совершенно в другом месте
2. В графе нет ребра соответствующего заданной части референса

Снова Quast и настоящие ошибки сборки

1. Повторы заставляют Quast выбирать неправильный путь
2. Несоответствие путей из-за циклов или повторов

Алгоритм анализа ошибок

1. Загрузка extensive misassemblies из результатов обработки QUAST
2. Проверка соответствия стрендов
3. Проверка: путь контига — подпоследовательность пути референса
4. Проверка: путь контига в районе ошибки — подпуть референса
5. Выравнивание путей контига и референса
6. Нахождение циклов
7. Вывод выравнивания

Итоги

1. Фильтрация ложных ошибок
2. Базовая классификация настоящих ошибок
3. Вывод результатов выравнивания