

ПРИЛОЖЕНИЕ ДЛЯ АВТОМАТИЗАЦИИ СБОРКИ, СТРУКТУРНОЙ И ФУНКЦИОНАЛЬНОЙ АННОТАЦИИ ГЕНОМОВ ПРОКАРИОТ

Зилов Д.С. (Университет ИТМО)

Научный руководитель – в.н.с, к.б.н Комиссаров А.С. (Университет ИТМО)

Появление секвенирования нового поколения впервые позволило значительно ускорить и удешевить определение полной последовательности геномов. Для автоматизации и ускорения обработки данных секвенирования мы создали Rappori - масштабируемый, простой в установке и использовании пайплайн для сборки и аннотации геномов прокариот. Программа выполняет крупномасштабный анализ последовательности нуклеиновых кислот бактерий от подготовки данных до функциональной аннотации.

Введение. Появление секвенирования нового поколения впервые позволило значительно ускорить и удешевить определение полной последовательности геномов миллионов организмов, от бактерий до человека. Со стороны биоинформатики очевидно, что по мере снижения стоимости секвенирования, количество обрабатываемых данных будет увеличиваться. Необходимо определить области анализа, которые являются рутинными, и автоматизировать их.

Основная часть. Мы создали Rappori - масштабируемый, простой в установке и использовании пайплайн для сборки и аннотации геномов прокариот. Программа выполняет крупномасштабный анализ последовательности нуклеиновых кислот бактерий от подготовки данных до функциональной аннотации.

Rappori включает в себя самые передовые и эффективные инструменты для геномной аннотации и позволяет гибко настраивать их использование. Процесс протекает от подготовки прочитываний последовательности до сборки генома, через очистку генома от контаминаций к структурной и функциональной аннотации. Rappori позволяет пользователям выбирать таксоны для построения пан-геномных графов. На протяжении всего процесса осуществляется контроль качества проведенных преобразований. Также, инструмент может использоваться как в виде программы командной строки, так и через веб-интерфейс.

Для тестирования качества работы пайплайна был создан датасет, содержащий в себе 16 геномов: 8 для тестирования аннотации, 8 для полноценной работы программы, включая как сборку генома, так и его аннотацию. Для получения отчета в среднем требуется от 90 до 180 минут при запуске программы на 32 потоках и 128 Гб оперативной памяти.

Выводы. Мы разработали быстрое и удобное приложение для анализа данных секвенирования геномов прокариот. На данный момент похожие приложения не поддерживаются либо устарели. В отличие от уже существующих, Rappori, основан на современных приложениях и выдает точный результат за короткое время. Приложение будет находиться в свободном доступе для биоинформатиков и в виде веб-интерфейса для биологов и медиков без опыта в биоинформатике.

Зилов Д.С. (автор)

Подпись

Комиссаров А.С. (научный руководитель)

Подпись