

разработаны оригинальные способы генотипирования, повышающие точность ДНК-анализа, с дополнительным обоснованием достоверности тестов секвенированием ПЦР-продуктов и депонированием расшифрованных нуклеотидных последовательностей в GenBank NCBI (GenBank A/N: JX649155-JX649158).

Таким образом, из общего числа проанализированных образцов растений, наиболее перспективными генотипами, рассматриваемыми в качестве исходного материала для дальнейшей селекционной работы по созданию сортов яровой пшеницы с крахмалом амилопектинового типа (с низким содержанием амилозы) являются две линии Кк-8/06-6 и О-192/03-5, несущие в своем геноме нулевой *Wx-B1b*-аллель, с последующим введением *Wx-A1b*- и (или) *Wx-D1b*-аллеля путем скрещивания с донорами нуль-аллелей по локусам *Waxy*-генов.

СОВМЕСТНОЕ ПРИМЕНЕНИЕ ГРАФОВ ДЕ БРЁЙНА, ГРАФОВ ПЕРЕКРЫТИЙ И МИКРОСБОРКИ ДЛЯ *DE NOVO* СБОРКИ ГЕНОМА

Александров А.В., Казаков С.В., Мельников С.В., Сергушичев А.А., Царев Ф.Н., Шалыто А.А.

НИУ ИТМО, Санкт-Петербург

В работе предлагается метод сборки генома, состоящий из трех этапов: сборка квазиконтигов, сборка контигов из квазиконтигов и микросборка. Результаты эксперимента по сборке генома бактерии *E. coli* показали, что качество сборки сравнимо с другими сборщиками.

В качестве входных данных выступают парные чтения геномной последовательности, полученные из фрагментов длиной от 100 до 600 нуклеотидов. На первом этапе строится граф де Брёйна, вершинами которого являются все k -меры, встречающиеся более одного раза. Для каждой пары чтений из входных данных выполняется поиск всех путей, соединяющих начальные k -меры этих чтений. При этом рассматриваются только пути, длины которых ограничены снизу и сверху исходя из априорных знаний о распределении длин фрагментов. Для этого используются два одновременных обхода в ширину от начальных k -меров. Если все найденные пути имеют одинаковую длину и не сильно отличаются друг от друга, то можно говорить о том, что найдена последовательность, являющаяся подстрокой генома. Такие последовательности мы называем квазиконтигами.

На втором этапе из квазиконтигов с помощью метода *overlap-layout-consensus* собираются контиги. Если все квазиконтиги нельзя обработать в оперативной памяти, то часть их отбрасывается.

На третьем этапе производится упорядочивание контигов и заполнение промежутков между ними. Для этого все чтения картируются на контиги с помощью программного средства Bowtie. Находятся все парные чтения, картированные на разные контиги. Для каждой пары соединенных таким образом контигов определяется, в каком порядке они идут в геноме, и выбираются все пары чтений, хотя бы одно из которых картируется на эти контиги. Из них строится граф де Брёйна. Его размер существенно меньше графа, используемого на первом этапе (поэтому этап называется микросборкой). Путем поиска путей в этом графе промежутки между контигами заполняются, тем самым они объединяются.

Для эксперимента использовалась библиотека *SRR001665* (<http://www.ncbi.nlm.nih.gov/sra?term=SRR001665>) парных чтений генома бактерии по 36 нуклеотидов с длиной фрагмента около 200, дающая 160-кратное покрытие. В результате было получено 247 контигов с $N50=53720$, покрывающих 98% референсного генома.

SRAGEN: ВЕБ-СЕРВИС ДЛЯ СРАВНИТЕЛЬНОГО ГЕНОМНОГО АНАЛИЗА ПРОКАРИОТ

^{1,2}Алтухов И.А., ^{1,2}Ищенко Д.С., ^{1,2}Коган В.И., ¹Базалеев Н.А.,
^{1,2}Кулемин Н.А., Тяхт А.В.¹, ¹Шитиков Е.А., ^{1,2}Алексеев Д.Г.

¹ФГБУН НИИ ФХМ ФМБА России, Москва

²МФТИ (ГУ), Долгопрудный

Одной из особенностей текущего положения в области геномики является стихийное накопление геномных данных. В текущей версии GenBank от 15 октября 2012 года содержится 145431 миллион оснований из 158 миллионов генетических последовательностей [1]. При этом, более половины этих данных были получены за последние 5 лет.

Большинство современных исследований проводятся с целью сравнения очень схожих близкородственных организмов имеющих разные фенотипические свойства (лабораторные, патогенные, резистентные). Сейчас это принимает все большую социальную значимость ввиду роста возникновения мутаций влияющих на резистентность у таких бактерий, как *M.tuberculosis*, *N.gonorrhoeae* и прочие.

В настоящее время существует много программных решений, позволяющих проводить сравнительный геномный анализ. Но часто, биологи не имеют специальных навыков для работы с такими подходами и сталкиваются с проблемой получения определенных результатов.

SRAGEN пользовательский интернет портал, позволяющий проводить