

Оценка потенциала библиотеки парноконцевых ридов, полученных на платформе Illumina, для полногеномной и локальной сборки

Научный руководитель – Кулакова Милана Анатольевна

Зорина Наталья Александровна

Студент (бакалавр)

Санкт-Петербургский государственный университет, Биологический факультет,
Санкт-Петербург, Россия

E-mail: natalyack@yandex.ru

В данной работе положено начало большому исследованию, конечной целью которого является сборка и аннотирование генома нереидной полихеты *Alitta virens*.

Сборка генома - сложная задача, которая требует тщательного предварительного анализа качества библиотеки, состоящей из коротких прочтений, полученных при помощи технологии NGS (Next Generation Sequencing).

В этом исследовании была использована библиотека парноконцевых ридов, полученных на платформе Illumina HiSeq 2500. Оценка качества проводилась с помощью программы FastQC. Далее был построен график распределения k - меров с помощью программы Jellyfish. Фильтрация библиотеки и очистка от адаптерных последовательностей осуществлялась программами Cookiecutter и Trimmomatic.

Оценка потенциала библиотеки для полногеномной сборки проводилась при помощи ассемблера SOAPdenovo и программы Quast.

После этого было сделано выравнивание ридов с помощью программы Bowtie и построена гистограмма распределения частот встречаемости вставок разной длины.

Оценка потенциала библиотеки для локальной сборки Нох- кластера осуществлялась с помощью программы Geneious.

В результате было установлено, что данная библиотека не может быть использована для полногеномной сборки ($N50 = 2833$) без дополнительных библиотек с большей длиной вставки и большей длиной прочтений, но может быть использована для локальной сборки отдельных локусов интереса, в частности Нох- кластера.