

УДК 004.9

МЕТОД ИСПРАВЛЕНИЯ ОШИБОК ВСТАВКИ И УДАЛЕНИЯ В НАБОРЕ ЧТЕНИЙ НУКЛЕОТИДНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

А.В. Александров^а, А.А. Шальто^а

^а Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

Адрес для переписки: alantbox@gmail.com

Информация о статье

Поступила в редакцию 19.10.15, принята к печати 07.12.15

doi:10.17586/2226-1494-2016-16-1-108-114

Язык статьи – русский

Ссылка для цитирования: Александров А.В., Шальто А.А. Метод исправления ошибок вставки и удаления в наборе чтений нуклеотидной последовательности // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 1. С. 108–114.

Аннотация

Предмет исследования. Разработан метод исправления ошибок вставки и удаления в наборе геномных чтений гаплоидного организма. Приведены результаты тестирования на двух библиотеках – искусственно сгенерированном наборе чтений из генома бактерии *Escherichia coli* и реальной библиотеке чтений бактерии *Pseudomonas stutzeri*.

Метод. Работа метода основана на использовании *k*-меров. В отличие от большинства распространенных методов исправления ошибок, *k*-меры используются только для поиска похожих друг на друга чтений. Для похожих чтений вычисляется строка-консенсус, которая затем используется для исправления ошибок в самих чтениях. **Основные результаты.** Алгоритм реализован в виде самостоятельного программного модуля. Программный модуль протестирован на реальных и синтезированных данных. Качество исправления ошибок разработанного метода выше, чем у известных современных аналогов. Для сравнения использовалась метрика N50, а также суммарная и максимальная длина контига. **Практическая значимость.** Разработанный метод может быть использован в связке с распространенными методами сборки генома, не приспособленными для использования с чтениями, содержащими ошибки вставки и удаления.

Ключевые слова

сборка генома, исправление ошибок, ошибки вставки и удаления

ERROR CORRECTION METHOD FOR SEQUENCING DATA WITH INSERTIONS AND DELETIONS

A.V. Alexandrov^а, A.A. Shalyto^а

^а ITMO University, Saint Petersburg, 197101, Russian Federation

Corresponding author: alantbox@gmail.com

Article info

Received 19.10.15, accepted 07.12.15

doi:10.17586/2226-1494-2016-16-1-108-114

Article in Russian

For citation: Alexandrov A.V., Shalyto A.A. Error correction method for sequencing data with insertions and deletions. Part I. Research of detonation engines. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 1, pp. 108–114.

Abstract

Subject of Research. A method for error correction for sequencing reads of a haploid organism with insertions and deletions was developed. It was tested on two libraries: a synthesized dataset for *Escherichia coli* bacterium and a real dataset of reads for *Pseudomonas stutzeri*. **Method.** The method is based on using *k*-mers but only for finding reads that are close to each other. For the close reads a consensus string is created which is then used for correcting errors in the initial reads. **Main Results.** The algorithm is implemented as a separated program. The program has been tested on both real and synthesized data. The method performance is higher than that of the other known methods (N50 metric was used as well as total contig length and maximal contig length as metrics for comparison). **Practical Relevance.** The method can be used together with known genome assembly methods not suitable for application with the reads containing insertion and deletion errors.

Keywords

genome assembly, error correction, insertions and deletions errors

Введение

Существует несколько компаний, выпускающих устройства для получения коротких чтений. Самыми распространенными на рынке этих устройств являются продукты компании Illumina [1], однако в последнее время секвенаторы Ion Torrent [2] стремительно набирают популярность благодаря своей дешевизне. Сравнительные характеристики секвенаторов этих двух компаний представлены в табл. 1.

Параметр	Ion Torrent Proton	Illumina HiSeq 2500
Цена системы, \$	243000	740000
Стоимость годового обслуживания, \$	19400	59200
Стоимость на 1 миллиард пар нуклеотидов, \$	16,67	46,00
Длительность одного запуска, ч	8	27

Таблица 1. Сравнительные характеристики секвенаторов Illumina и Ion Torrent

В последнее время популярным стало использование так называемых парных чтений [3]. При получении парных чтений секвенатором выделяется расположенный в случайном месте последовательности ДНК-фрагмент, из которого затем считываются префикс и суффикс. Важно отметить, что эти префикс и суффикс считываются с разных нитей ДНК, причем неизвестно, какой был считан с прямой нити, а какой – с обратной. Исходя из этого, удобно рассматривать не исходные геном и набор чтений, а дополненные своими обратными-комплементарными копиями.

Результатом работы секвенатора в случае использования парных чтений являются пары последовательностей, про которые примерно известно, на каком расстоянии они располагались в исходной последовательности ДНК.

В процессе чтения секвенаторами допускаются технические ошибки. Ошибки бывают трех типов:

1. ошибки вставки – в основном проявляются в прочтении более длинных, чем в исходном геноме, последовательностей одинаковых нуклеотидов (например, вместо «AA» было прочитано «AAA»);
2. ошибки удаления – в этом случае в прочитанной нуклеотидной последовательности может не хватать одного нуклеотида (например, вместо «ACGT» было прочитано «AGT»);
3. ошибки замены – в таких случаях некоторые нуклеотиды были прочитаны неверно (например, вместо нуклеотида A был прочитан нуклеотид G).

Секвенаторы компании Ion Torrent совершают ошибки вставки и удаления значительно чаще [4], чем ошибки замены, в связи с чем возникает необходимость разработки методов их исправления.

Поскольку задача сборки генома не является новой, существует набор методов, осуществляющих исправление ошибок в чтениях. Большинство из них не приспособлено для ошибок вставки и удаления, поэтому на таких данных они работают очень плохо. Распространенные методы исправления ошибок можно разделить на две группы: одни основаны на исправлении k -меров (сборщики ITMO-assembler [5], ABySS [6], Quake [7], Hammer [8], ALLPATHS [9]), а другие используют граф де Брейна [10] (сборщики EULER [11], Velvet [12]).

Методы, основанные на исправлении k -меров, работают не с чтениями, а с их подстроками длины k – k -мерами. Если в чтениях присутствуют только ошибки замены, то можно для каждого k -мера перебрать все возможные варианты ошибок, которые могли в нем произойти (их $3 \times k$ варианта). Такой подход работает для ошибок вставки и замены, так как при таких ошибках k -меры превращаются соответственно в $(k+1)$ -меры и $(k-1)$ -меры.

Методы, основанные на графе де Брейна, в целом менее эффективны, чем основанные на k -мерах, так как требуют хранения в памяти большого графа, из-за чего затрудняется их масштабируемость, а также из-за сложности графа, получаемого в случае больших геномов.

Таким образом, необходимо разработать метод исправления ошибок, который будет оптимизирован для работы с чтениями, содержащими как «классические» ошибки замены, так и «современные» ошибки вставки и удаления.

Предлагаемый метод

Предлагаемый в настоящей работе метод основан на определении перекрывающихся чтений и не использует графа де Брейна. Для эффективного исправления ошибок необходимо, чтобы каждая позиция генома была прочитана несколько раз, так как это единственный способ отличить правильно прочитанный нуклеотид от прочитанного неверно. Это, ввиду небольшой вероятности ошибки, дает право считать, что наибольшее число раз нуклеотид на каждой позиции был прочитан верно. На практике используются наборы чтений, покрывающие геном несколько десятков раз. Важно отметить, что не только отдельные позиции всего генома были прочитаны несколько десятков раз, но и небольшие его подстроки (не длиннее самих чтений) встречаются в чтениях несколько раз, причем, чем длиннее подстрока, тем меньше шансов, что несколько различных чтений ее содержат. Рассмотрим чтения, полученные из одного фрагмента генома (рис. 1).

Поскольку доля ошибок невелика (значительно меньше половины), они должны быть похожи. По этим чтениям по принципу консенсуса можно восстановить фрагмент генома, из которого они были прочитаны. После этого по строке можно восстановить сами чтения.

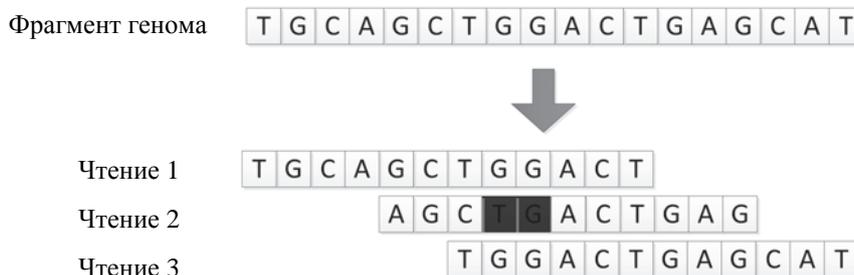


Рис. 1. Фрагмент генома и чтения, произведенные из этого фрагмента. Темно-серым выделено место, содержащее ошибки

Поскольку сама геномная последовательность в процессе сборки недоступна, для выделения группы чтений, полученных из одного фрагмента генома, нельзя просто найти подстроку генома, больше всего похожую на набор чтений. Вместо этого рассмотрим k -меры – подстроки чтений длины k .

Если бы ошибок не было, каждый k -мер задавал бы фрагмент генома (или несколько, если в геноме есть повторы длины хотя бы k). В случае наличия в чтениях ошибок некоторые k -меры не являются подстроками чтений.

Поскольку ошибки происходят с небольшой вероятностью, вероятность того, что один и тот же k -мер будет прочитан несколько раз с одинаковым набором ошибок, очень мала. Из этого вытекает, что те k -меры, которые встречаются в наборе чтений мало раз, являются ошибочными, остальные же являются реальными подстроками генома (рис. 2). Будем называть редко встречающиеся k -меры «плохими», а часто встречающиеся – «хорошими».

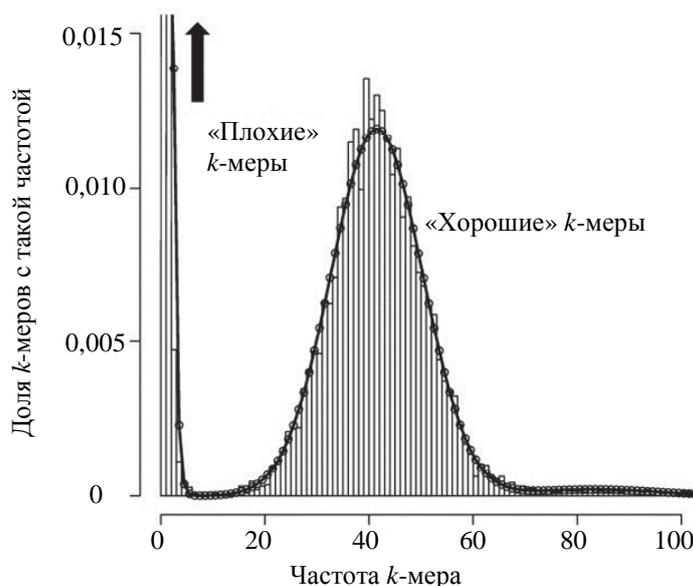


Рис. 2. Распределение частот k -меров в чтениях

В качестве порогового значения выбирается значение, соответствующее первому максимуму числа k -меров в чтениях.

После нахождения всех «хороших» k -меров для каждого из них вычисляется набор чтений, которые содержат рассматриваемый k -мер. Эти чтения составляют одну группу. Группы обрабатываются независимо друг от друга.

Кластеризация. Поскольку в геноме обычно есть повторы небольшой длины, некоторые k -меры встречаются в нем в нескольких местах. Это означает, что чтения, содержащие какой-то определенный k -мер, могут сильно отличаться друг от друга. В этом случае их можно разделить на группы, внутри которых они будут различаться гораздо меньше, после чего группы можно будет обрабатывать отдельно друг от друга (рис. 3).

Понятно, что повторы, длина которых приближается снизу к длине чтений, так разрешить невозможно. Однако повторы, длина которых не превышает длины k -мера, таким образом учесть можно.

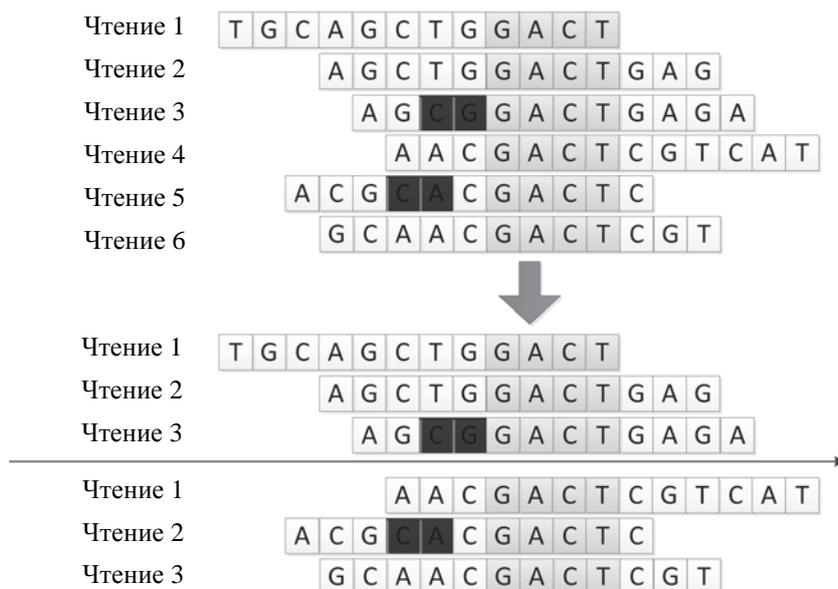


Рис. 3. Кластеризация чтений, полученных из разных мест генома

Разрезание чтений. Поскольку все чтения в одном кластере имеют общий k -мер, удобно каждое чтение разделить на три части – левую (слева от общего k -мера), среднюю (общий k -мер) и правую (справа от общего k -мера) (рис. 4). Затем левые части разворачиваются, после чего группы левых и правых частей обрабатываются отдельно друг от друга. Это делается для того, чтобы все обрабатываемые строки в группе начинались в одной позиции.

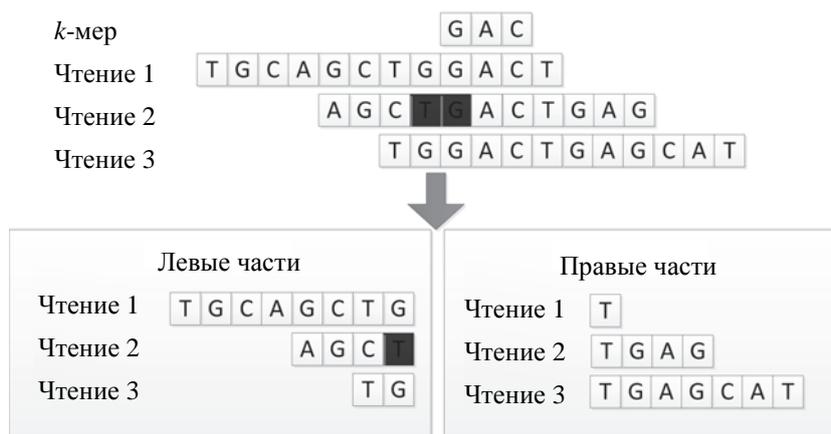


Рис. 4. Выделение из чтений левой и правой частей

Консенсус. Строка-консенсус восстанавливается инкрементально. При этом на шаге с номером i поддерживаются текущий префикс искомой строки, имеющий длину i , и множество индексов для каждой из данных строк. Индекс i для строки x означает, что подозревается, что префикс строки x длины i соответствует текущему префиксу строки-консенсуса. В начале работы алгоритма искомая строка пуста, а индексы каждой из строк содержат лишь ноль. Для увеличения длины текущего префикса на единицу перебирается все 4 варианта продолжения, для каждого из них для каждого индекса вычисляется, как изменяется расстояние Левенштейна между префиксом строки-консенсуса и соответствующей индексу подстрокой. Если увеличение длины префиксов строки-консенсуса и данной строки происходит без увеличения расстояния, то данная строка «поддерживает» добавляемый к строке-консенсусу символ. В итоге к строке-консенсусу добавляется тот символ, который поддерживает больше всего чтений.

Таким образом, алгоритм состоит из следующих шагов:

1. сбор статистики по k -мерам;
2. определение порога и построение множества «хороших» k -меров;
3. построение индекса чтений для всех «хороших» k -меров;
4. для каждого k -мера:
 1. разделение каждого чтения на 3 части;
 2. нахождение строки-консенсуса для каждой части чтений;
 3. исправление ошибок в чтениях.

Важно отметить, что алгоритм поиска ошибок в k -мерах легко распараллеливается, так как соответствующие различным k -мерам чтения могут обрабатываться независимо друг от друга.

Апробация метода

Основная работа алгоритма происходит при построении строки-консенсуса для группы строк. Пусть имеется N строк, для которых необходимо найти консенсус. Пусть самая длинная из них имеет длину L . Тогда всего алгоритм сделает не более L шагов. На каждом шаге необходимо пройтись по всем N строкам и для каждой из них обновить расстояние Левенштейна до текущего префикса. Это делается за $O(K)$, где K – длина текущего префикса. Таким образом, суммарное время обработки группы чтений не превосходит

$$N \times (O(1) + O(2) + \dots + O(L)) = O(N \times L^2),$$

где значения L в данной оценке различаются для каждой группы, однако не превосходит максимальной длины чтений R .

Если средняя частота k -мера в чтениях равна C , то время работы алгоритма не превосходит $O(L \times C \times R^2)$,

где L – число «хороших» k -меров (длина генома). Таким образом, число элементарных операций, необходимых для сборки бактериального генома (длина – несколько миллионов пар нуклеотидов) со средним покрытием k -мера, равным 30, и средней длиной чтения, равной 100, примерно равно 1×10^{13} . Вычислительному узлу с 24 процессорами, совершающими по 10^8 элементарных операций в секунду, на это требуется около часа, что, принимая во внимание, что остальные стадии сборки аналогичного по размеру генома могут требовать несколько часов работы, является неплохим результатом.

Экспериментальные исследования. Для оценки эффективности предложенного метода было проведено несколько экспериментов. Тестирование проводилось на двух библиотеках – искусственно сгенерированном наборе чтений из генома бактерии *Escherichia coli* (K-12 MG1655) [13] и реальной библиотеке чтений бактерии *Pseudomonas stutzeri* (CGMCC 1.1803) [14].

Средства тестирования. Для оценки качества тестирования одни и те же библиотеки собирались разными сборщиками – сборщиком ABySS и ITMO-denovo-assembler (<http://genome.ifmo.ru>). Для каждого сборщика проводилось два эксперимента – сборка осуществлялась как из исходных чтений, так и из исправленных. ABySS не предназначен для работы с чтениями, содержащими ошибками вставки и удаления, поэтому результаты работы ABySS на исправленных чтениях отражают качество работы этапа исправления ошибок. Также использовался сборщик MIRA-assembler [15], поддерживающий чтения секвенаторов Ion Torrent и умеющий работать с ошибками вставки и удаления. В качестве сравниваемой величины была выбрана метрика N50 [16]. Эта величина показывает такую максимальную длину контига, что хотя бы половина генома покрыта контигами такой длины или длиннее.

Библиотека E.coli. Бактерия *Escherichia coli* имеет геном длиной около 4,5 миллионов нуклеотидов. Сгенерированная библиотека обеспечивала 20-кратное покрытие генома чтениями длиной около 100 нуклеотидов.

Библиотека была сгенерирована в два этапа. На первом этапе были сгенерированы безошибочные подстроки генома (равномерно распределенные по линейному геному, длина чтения 100 нуклеотидов). На втором этапе в эти подстроки были внесены ошибки. Такой способ генерации библиотеки позволил сравнить результаты сборки безошибочных чтений и сборки чтений с ошибками с применением исправления ошибок.

Результаты сборки представлены в табл. 2.

Библиотека	Суммарная длина контигов	N50	Максимальная длина контига
Безошибочные чтения, ABySS	4582767	43154	162361
Неисправленные чтения с ошибками, ABySS	3003406	163	981
Исправленные чтения с ошибками, ABySS	4593871	28492	92287
Неисправленные чтения с ошибками, MIRA-assembler	4593118	21162	83750

Таблица 2. Результаты сборки искусственных чтений E.coli

Первая строка показывает, насколько хорошо вообще ABySS может обработать данные чтения.

Вторая строка таблицы доказывает, что ABySS не приспособлен для обработки чтений с ошибками вставки и удаления. Это выражается в маленькой суммарной длине контигов (примерно 2/3 длины всего генома), а также в маленьком значении N50 (меньше удвоенной длины чтения).

Третья строка таблицы показывает, что исправление ошибок значительно улучшает ситуацию, потому что результаты сборки ABySS на исправленных чтениях всего в полтора раза хуже сборки из безошибочных чтений.

Четвертая строка таблицы показывает, что приспособленный для чтений Ion Torrent сборщик MIRA-assembler справляется с задачей хуже, чем разработанный метод.

Библиотека P.stutzeri. Бактерия *Pseudomonas stutzeri* имеет геном длиной около 4,5 миллионов нуклеотидов. Библиотека состояла из 4 файлов: I3_1_in.iontor.fastq, I3_2_in.iontor.fastq, 100_in.iontor.fastq и 200_in.iontor.fastq. Файлы I3_1 и I3_2 содержали парные чтения и обеспечивали суммарное 35-кратное покрытие. Средняя длина чтений в файлах I3_1 и I3_2 – 83 нуклеотида. Библиотека 100 содержала 4,6 миллионов чтений со средней длиной 117 нуклеотидов, обеспечивая 119-кратное покрытие генома. Наконец, библиотека 200 содержала 5,5 миллионов чтений со средней длиной 231 нуклеотид, обеспечивая 282-кратное покрытие. Результаты сборки представлены в табл. 3.

Библиотека	Суммарная длина контигов	N50	Максимальная длина контига
Неисправленные чтения I3_1 и I3_2, ABySS	5134815	3227	15646
Исправленные чтения I3_1 и I3_2, ABySS	4863666	9926	68660
Неисправленные чтения I3_1 и I3_2, ITMO-denovo-assembler	5321832	7118	49527
Исправленные чтения I3_1 и I3_2, ITMO-denovo-assembler	4672051	11619	66366
Неисправленные чтения I3_1 и I3_2, MIRA-assembler	5205159	5285	47004
Неисправленные чтения I3_1, I3_2 и 100, ABySS	4495333	17881	69269
Исправленные чтения I3_1, I3_2 и 100, ABySS	4558926	18878	76335
Неисправленные чтения I3_1, I3_2 и 100, ITMO-denovo-assembler	5076230	13267	67933
Исправленные чтения I3_1, I3_2 и 100, ITMO-denovo-assembler	4878434	20290	98357
Неисправленные чтения I3_1, I3_2 и 100, MIRA-assembler	5250041	15252	95547
Неисправленные чтения I3_1, I3_2, 100 и 200, ABySS	4500845	18034	61569
Исправленные чтения I3_1, I3_2, 100 и 200, ABySS	4501487	22017	63725
Неисправленные чтения I3_1, I3_2, 100 и 200, ITMO-denovo-assembler	5126675	14755	51078
Исправленные чтения I3_1, I3_2, 100 и 200, ITMO-denovo-assembler	4799664	22265	74866

Таблица 3. Результаты сборки реальных чтений P.stutzeri

Как и на искусственных данных, на настоящих чтениях разработанный алгоритм значительно улучшает качество сборки сборщиками ABySS и ITMO-denovo-assembler. Также заметно, что качество сборки исправленных чтений любым сборщиком выше, чем качество сборки при помощи MIRA-assembler.

Заключение

Разработан метод исправления ошибок, основанный на поиске перекрытий между чтениями. Проведено экспериментальное исследование разработанного метода, показавшее работоспособность метода как на искусственных, так и на реальных данных. Результаты, полученные с использованием разработанного метода, превосходят по основным характеристикам (N50, суммарная и максимальная длина контигов) результаты, достижимые при помощи распространенных аналогов предложенного метода. Разработанный метод может быть распараллелен на большое число процессоров, что делает возможным его использование для больших объемов данных.

В настоящий момент исследуются пути увеличения качества исправления ошибок. Так, планируется добавить возможность загрузки чтений не по одному k -меру, а по нескольким соседним. Это позволит загружать за один раз больше чтений и значительно уменьшит число итераций алгоритма. Также планируется использовать информацию о качестве прочтения нуклеотидов, предоставляемую секвенатором, для улучшения работы процедуры построения консенсуса.

Литература

1. Rothberg J.M., Hinz W., Rearick T.M. et al. An integrated semiconductor device enabling non-optical genome sequencing // Nature. 2011. V. 475. N 7356. P. 348–352. doi: 10.1038/nature10242
2. Bentley D.R., Balasubramanian S., Swerdlow H. et al. Accurate whole human genome sequencing using reversible terminator chemistry // Nature. 2008. V. 456. N 7218. P. 53–59. doi: 10.1038/nature07517

3. Roach J., Boysen C., Wang K., Hood L. Pairwise end sequencing: a unified approach to genomic mapping and sequencing // *Genomics*. 1995. V. 26. N 2. P. 345–353. doi: 10.1016/0888-7543(95)80219-C
4. Bragg L.M., Stone G., Butler M.K., Hugenholtz P., Tyson G.W. Shining a light on dark sequencing: characterizing errors in ion torrent PGM data // *PLOS Computational Biology*. 2013. V. 9. N 4. Art. e1003031. doi: 10.1371/journal.pcbi.1003031
5. Александров А.В., Казаков С.В., Мельников С.В., Сергушичев А.А., Царев Ф.Н., Шалыто А.А. Метод исправления ошибок в наборе чтений нуклеотидной последовательности // *Научно-технический вестник СПбГУ ИТМО*. 2011. № 5 (75). С. 81–84.
6. Simpson J.T., Wong K., Jackman S.D., Schein J.E., Jones S.J.M., Birol I. ABySS: a parallel assembler for short read sequence data // *Genome Research*. 2009. V. 19. N 6. P. 1117–1123. doi: 10.1101/gr.089532.108
7. Kelley D.R., Schatz M.C., Salzberg S.L. Quake: quality-aware detection and correction of sequencing errors // *Genome Biology*. 2010. V. 11. N 11. Art. R116. doi: 10.1186/gb-2010-11-11-r116
8. Medvedev P., Scott E., Kakaradov B., Pevzner P. Error correction of high-throughput sequencing datasets with non-uniform coverage // *Bioinformatics*. 2011. V. 27. N 13. P. i137–i141. doi: 10.1093/bioinformatics/btr208
9. Butler J., MacCallum I., Kleber M., Shlyakhter I.A., Belmonte M.K., Lander E.S., Nusbaum C., Jaffe D.B. ALLPATHS: de novo assembly of whole-genome shotgun microreads // *Genome Research*. 2008. V. 18. N 5. P. 810–820. doi: 10.1101/gr.7337908
10. de Bruijn N.G. A combinatorial problem // *Koninklijke Nederlandse Akademie v. Wetenschappen*. 1946. V. 49. P. 758–764.
11. Pevzner P.A., Tang H., Waterman M.S. An Eulerian path approach to DNA fragment assembly // *Proceedings of the National Academy of Sciences of the USA*. 2001. V. 98. N 17. P. 9748–9753. doi: 10.1073/pnas.171285098
12. Zerbino D.R., Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs // *Genome Research*. 2008. V. 18. N 5. P. 821–829. doi: 10.1101/gr.074492.107
13. Riley M., Abe T., Arnaud M.B., Berlyn M.K., Blattner F.R., Chaudhuri R.R., Glasner J.D., Horiuchi T., Keseler I.M., Kosuge T., Mori H., Perna N.T., Plunkett III G., Rudd K.E., Serres M.H., Thomas G.H., Thomson N.R., Wishart D., Wanner B.L. *Escherichia coli* K-12: a cooperatively developed annotation snapshot-2005 // *Nucleic Acids Research*. 2006. V. 34. N 1. P. 1–9. doi: 10.1093/nar/gkj405
14. Chen M., Yan Y., Zhang W., Lu W., Wang J., Ping S., Lin M. Complete genome sequence of the type strain *Pseudomonas stutzeri* CGMCC 1.1803 // *Journal of Bacteriology*. 2011. V. 193. N 21. P. 6095. doi: 10.1128/JB.06061-11
15. Chevreaux B., Wetter T., Suhai S. Genome sequence assembly using trace signals and additional sequence information // *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*. 1999. V. 99. P. 45–56.
16. Miller J.R., Koren S., Sutton G. Assembly algorithms for next-generation sequencing data // *Genomics*. 2010. V. 95. N 6. P. 315–327. doi: 10.1016/j.ygeno.2010.03.001

- Александров Антон Вячеславович* – аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, alantbox@gmail.com
- Шалыто Анатолий Абрамович* – доктор технических наук, профессор, заведующий кафедрой, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, shalyto@mail.ifmo.ru
- Anton V. Alexandrov* – postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, alantbox@gmail.com
- Anatoly A. Shalyto* – D.Sc., Professor, Head of Chair, ITMO University, Saint Petersburg, 197101, Russian Federation, shalyto@mail.ifmo.ru