

УДК 575.113.12:579.238]:57.088.7

<https://doi.org/10.51523/2708-6011.2023-20-1-19>

Опыт проведения секвенирования генома *Klebsiella pneumoniae* с использованием метода коротких прочтений на платформе Illumina

А. С. Шафорост¹, А. А. Зятков¹, Е. В. Воропаев¹, О. В. Осипкина¹,
А. В. Воропаева², Н. А. Бонда³, И. О. Стома¹

¹Гомельский государственный медицинский университет, г. Гомель, Беларусь

²Республиканский научно-практический центр радиационной медицины и экологии человека, г. Гомель, Беларусь

³Гомельский областной центр гигиены, эпидемиологии и общественного здоровья, г. Гомель, Беларусь

Резюме

Цель исследования. Рассмотреть основные этапы секвенирования генома *Klebsiella pneumoniae* с использованием метода коротких прочтений на платформе Illumina и описать особенности процесса пробоподготовки библиотек и анализа полученных данных.

Материалы и методы. Дезоксирибонуклеиновую кислоту (ДНК) для высокопроизводительного секвенирования выделяли из культур *Klebsiella pneumoniae*. Пробоподготовку выполняли согласно инструкции производителя к набору Nextera XT DNA Library Prep. Секвенирование проводили на платформе Illumina MiSeq с использованием картриджа 2x151. Сборку генома до уровня контигов производили с помощью приложения SPAdes Genome Assembler на сервисе Illumina BaseSpace Sequence Hub и набора программ в среде Linux. Оценку качества сборки генома проводили с помощью сервиса QUAST.

Результаты. Проведено секвенирование генома образцов культур *K. pneumoniae* с последующей оценкой качества запуска, сборки генома и определением его основных параметров.

Заключение. Рассмотрены основные этапы секвенирования генома *K. pneumoniae* с использованием метода коротких прочтений на платформе Illumina. Отмечены основные параметры оценки качества пробоподготовки, запуска и сборки генома.

Ключевые слова: высокопроизводительное секвенирование, Illumina, сборка генома, прокариоты

Вклад авторов. Шафорост А.С., Зятков А.А.: проведение исследования, анализ и обработка результатов; Воропаев Е.В.: аналитическая оценка содержания исследования; Осипкина О.В., Воропаева А.В.: подготовка статьи к публикации; Бонда Н.А.: сбор и подготовка материала; Стома И.О.: концепция и организация исследования.

Благодарность. Авторы выражают благодарность заведующему лабораторией «Центр аналитических и генно-инженерных исследований» ГНУ «Институт микробиологии Национальной академии наук Беларуси», к.м.н., доценту Валентовичу Л.Н. за помощь и консультирование в процессе сборки генома и при проведении анализа его результатов.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Источники финансирования. При проведении исследования стороннее финансирование и спонсорская поддержка не осуществлялись.

Для цитирования: Шафорост АС, Зятков АА, Воропаев ЕВ, Осипкина ОВ, Воропаева АВ, Бонда НА, Стома ИО. Опыт проведения секвенирования генома *Klebsiella pneumoniae* с использованием метода коротких прочтений на платформе Illumina. Проблемы здоровья и экологии. 2023;20(1):152–159. DOI: <https://doi.org/10.51523/2708-6011.2023-20-1-19>

Experience of *Klebsiella pneumoniae* genome sequencing using the short read method on the Illumina platform

Alexander S. Shaforost¹, Alexey A. Ziatskov¹, Evgenii V. Voropaev¹,
Olga V. Osipkina¹, Alla V. Voropaeva², Nadezhda A. Bonda³, Igor O. Stoma¹

¹Gomel State Medical University, Gomel, Belarus

²Republican scientific and practical center for radiation medicine and human ecology, Gomel, Belarus

³Gomel Regional Center for Hygiene, Epidemiology and Public Health, Gomel, Belarus

Abstract

Objective. To review the main stages of *Klebsiella pneumoniae* genome sequencing using the Illumina short-read method and describe the peculiarities of sample library preparation and analysis of the obtained data.

© А. С. Шафорост, А. А. Зятков, Е. В. Воропаев, О. В. Осипкина, А. В. Воропаева, Н. А. Бонда, И. О. Стома, 2023

Materials and methods. Deoxyribonucleic acid (DNA) for high-throughput sequencing was isolated from *Klebsiella pneumoniae* cultures. Sample preparation was performed according to the manufacturer's instructions for the Nextera XT DNA Library Prep kit. Sequencing was performed on an Illumina MiSeq platform using a 2x151 cartridge. Genome assembly to the contigs was performed using the SPAdes Genome Assembler application on the Illumina BaseSpace Sequence Hub service and a set of programs in a Linux environment. The quality of genome assembly was assessed using the QUAST service.

Results. Genome sequencing of *K. pneumoniae* culture samples was performed, followed by an evaluation of the quality of the launch, assembly of the genome, and determination of its main parameters.

Conclusion. The main steps of *K. pneumoniae* genome sequencing have been considered using the short-read method on the Illumina platform. The main parameters for assessing the quality of sample preparation, launch and genome assembly are described.

Keywords: high throughput sequencing, Illumina, genome assembly, prokaryotes

Author contributions. Shaforost A.S., Ziatskov A.A.: research, analysis and processing of results; Voropaev E.V.: analytical evaluation of the content of the study; Osipkina O.V., Voropaeva A.V.: preparation of the article for publication; Bonda N.A.: collection and preparation of material; Stoma I.O.: concept and organization of the study.

Conflict of interest. The authors declare no conflict of interest.

Funding. The study did not receive any external funding or sponsorship.

For citation: Shaforost AS, Ziatskov AA, Voropaev EV, Osipkina OV, Voropaeva AV, Bonda NA, Stoma IO. Experience of *Klebsiella pneumoniae* genome sequencing using the short reads method on the Illumina platform. *Health and Ecology Issues*. 2023;20(1):152–159. DOI: <https://doi.org/10.51523/2708-6011.2023-20-1-19>

Введение

На современном этапе развития медицины появился ряд технических решений, которые значительно повышают качество диагностики различных, в том числе и инфекционных, заболеваний и способствуют разработке эффективных методов их лечения [1]. Было предложено множество инноваций как в области диагностики, так и в терапевтической сфере [1]. Для решения сложных и трудоемких задач молекулярной биологии микроорганизмов у современного исследователя в настоящее время имеются различные диагностические инструменты, из которых наиболее известны и часто используются различные модификации секвенирования (определения последовательности нуклеотидов) ДНК [1].

Классическое, или секвенирование ДНК первого поколения по методу Сенгера, получило мировое признание научного сообщества, выразившееся в получении его автором — Фредериком Сенгером — Нобелевской премии по химии в 1980 г. [1].

Технологии секвенирования следующего поколения являются наиболее эффективным и современным методом исследования генетических особенностей различных организмов. Наиболее распространенными являются решения, предлагаемые компаниями Illumina (MiniSeq, MiSeq, HiSeq и др.) и Thermo Fisher Scientific (Ion GeneStudio S5) [2]. Получаемые данные применяются для улучшения диагностики социально значимых заболеваний и разработки персонализированной терапии благодаря применению

широкого спектра панелей, позволяющих определять наличие точечных мутаций и предрасположенности к развитию отдельных патологий (онкологических, неврологических, сердечно-сосудистых, респираторных и др.), а также для идентификации и классификации онкологических заболеваний, определения антибиотикорезистентности микроорганизмов и др. [3].

Для исследования микробных сообществ, секвенирования геномов вирусов и прокариот *de novo* наиболее предпочтительным является технология компании Illumina (секвенатор MiSeq), сочетающая в себе достаточную для секвенирования таксономически значимых локусов метагеномной ДНК длину прочтения (до 2×301 нуклеотидов), что имеет решающее значение для сборки участков генома с повторяющимися последовательностями (гомополимерами), высокую пропускную способность и относительно низкую стоимость секвенирования [4].

Цель исследования

Рассмотреть основные этапы секвенирования генома *Klebsiella pneumoniae* с использованием метода коротких прочтений на платформе Illumina и описать особенности процесса пробоподготовки библиотек и анализа полученных данных.

Материалы и методы

Пробоподготовку образцов ДНК, выделенной из культур *Klebsiella pneumoniae*, для высокопроизводительного секвенирования на платформе

Illumina MiSeq проводили с использованием набора Nextera XT DNA Library Prep. Сборку генома до уровня контигов выполняли с помощью приложения SPAdes Genome Assembler на сервисе Illumina BaseSpace Sequence Hub и набора программ в среде Linux с последующей оценкой качества сборки генома с помощью приложения QUAST.

Результаты и обсуждение

Получение качественных выходных данных высокопроизводительного секвенирования напрямую зависит от корректности выполнения отдельных этапов (экстракции ДНК, получении геномных библиотек и их пулирование, настройки запуска) и выбора метода обработки получаемой информации.

Для получения максимального количества ДНК из культур *K. pneumoniae* использовали набор PureLink™ PCR Purification Kit (Thermo Fisher Scientific, США), возможно применение аналогичных наборов, способных разрушать клеточную стенку грамотрицательных микроорганизмов с последующим выделением нуклеиновой кислоты.

Концентрацию ДНК измеряли на спектрофотометре NanoDrop 1000. Для дальнейшего

анализа использовали образцы с соотношением экстинкции A260/280 $\geq 1,8$, если образцы имели загрязнение, требовалась дополнительная их очистка этанолом с дальнейшим измерением количества и качества ДНК.

С помощью набора Nextera XT DNA Library Prep, проводили пробоподготовку для полногеномного секвенирования, которая включала в себя следующие стадии: ферментативную фрагментацию геномной ДНК, штрихкодирование образцов и их очистку с использованием магнитных частиц.

Для ферментативной фрагментации проводили разведение образцов ДНК до 0,2 нг/мкл, измерение концентрации геномной ДНК выполняли на флуориметре Qubit4 (Thermo Fisher Scientific, Германия) с помощью набора Qubit™ 1x dsDNA High Sensitivity Assay Kit (1x dsDNA HS).

Для оценки эффективности фрагментации, качества полученных библиотек можно использовать электрофоретическое разделение в агарозном геле или капиллярный электрофорез (рисунки 1) и измерение концентрации библиотек на флуориметре Qubit4 (набор 1x dsDNA HS).

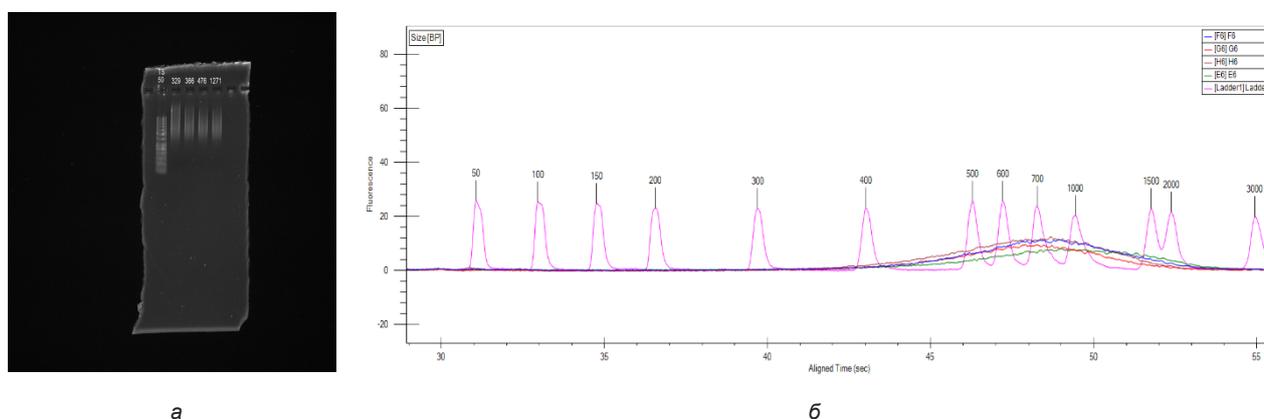


Рисунок 1. Электрофореграмма геномных библиотек *K. pneumoniae* после индексной ПЦР:
 а — 1,7 % агарозный гель (маркер — GeneRuler 50 bp DNA Ladder SM0371 (ThermoFisher Scientific, США));
 б — результат капиллярного электрофореза (набор Perkin Elmer The LabChip® NGS 3K)
 Figure 1. Electrophoregram of *K.pneumoniae* genomic libraries after index PCR:
 а — 1.7 % agarose gel (marker — GeneRuler 50 bp DNA Ladder SM0371 (ThermoFisher Scientific, USA));
 б — results of capillary electrophoresis (Perkin Elmer The LabChip® NGS 3k set)

Размер полученных библиотек для геномного секвенирования находился в интервале от 250 до 1500 п. н., что соответствует инструкции к набору Nextera XT DNA Library Prep.

Измеряли концентрацию (нг/мкл) очищенных библиотек с помощью флуориметра Qubit 4 (набор 1x dsDNA HS) или амплификатора «в реальном времени» методом измерения флу-

оресценции по конечной точке с применением интеркалирующего красителя 2x Zuber Green-1 («ПраймТех», Республика Беларусь) (таблица 1). При использовании последнего метода необходимо предварительное приготовление стандартов с последующим построением калибровочной кривой.

Таблица 1. Концентрация геномных библиотек ДНК при измерении с помощью флуориметра Qubit 4 и методом измерения флуоресценции по конечной точке

Table 1. Concentration of genomic DNA libraries when measured with the Qubit 4 fluorometer and the endpoint fluorescence measurement method

Номер образца	Концентрация «4 нМ» библиотек (Qubit 4), нг/мкл	Концентрация «4 нМ» библиотек (по конечной точке), нг/мкл
329	1,62	1,14
366	1,69	1,20
476	1,48	1,06
1271	1,62	1,15

Отмеченные отличия в величине концентрации библиотек, полученные с помощью различных методов, можно интерпретировать различной химической структурой используемых интеркалирующих красителей или погрешностями измерений. Определение концентрации библиотек является критически важным показателем, влияющим на количество генерируемых в процессе запуска кластеров и в конечном итоге — на объем получаемой по результатам секвенирования информации.

Завышенная концентрация объединенной геномной библиотеки может привести к перекластеризации и стать причиной снижения производительности выполнения анализа и показателя Q30 (доля нуклеотидов, идентифицированных с точностью 99,9 %), появлению артефактов секвенирования, снижению общего объема выходных данных и даже к досрочному прерыванию запуска. Уменьшение количества получаемых данных может происходить и при недокластеризации (недозагрузке) проточной ячейки [5, 6].

Для полногеномного секвенирования с использованием реагентов версии v2 оптимальным считается количество «сырых кластеров» (до прохождения фильтра качества) в диапазоне 1000–1200 кластеров на 1 мм² [6]. Согласно рекомендациям компании Illumina для секвенатора MiSeq загрузочная концентрация объединенной геномной библиотеки составляет 6–20 пМ [7]. Контролем качества для генерации кластеров, секвенирования и выравнивания прочтений, а также генерации матрицы перекрестных помех является примесь ДНК вируса ФХ174 (PhiX). Минимальное содержание PhiX в объединенной геномной библиотеке для секвенатора MiSeq составляет 5 % [8]. Это значение во многом определяется разнообразием последовательностей нуклеотидов в образце: чем оно ниже, тем доля PhiX должна быть выше.

Первоначальная обработка данных, полученных в результате анализа коротких фрагментов, осуществляется в автоматическом режиме с

помощью встроенного программного обеспечения Illumina MiSeq с предоставлением выходных данных в виде 2 файлов, содержащих последовательности для прямого и обратного прочтения в формате FASTQ.

Перед сборкой генома необходимо провести оценку качества полученных данных. Для этого использовали программу FastQC [9]. Одной из значимых метрик является «Per base sequence quality», которая отражает изменение качества прочтения нуклеотидов в процессе секвенирования. Как правило, отмечается его снижение в начале и в конце прочтения. На основании анализа этого параметра принимается решение об удалении начальных или конечных нуклеотидов с качеством менее Q30. Метрика «Sequence Length Distribution» отображает распределение прочтений по длине и позволяет получить информацию о доле прочтений требуемого размера (в нашем случае — 151 нуклеотид) и необходимости фильтрации прочтений.

Полученные данные были использованы для сборки генома до уровня контигов с применением приложения SPAdes Genome Assembler (инструмент с открытым исходным кодом для секвенирования *de novo*) на сервисе Illumina BaseSpace Sequence Hub [10] или набора программ в среде Linux. Сборку генома можно проводить, используя аналоги данных приложений [11].

При выполнении сборки генома в среде Linux для предварительной обработки (удаление прочтений с длиной, меньше заданной, последовательностей адаптеров и нуклеотидов с низким качеством) исходных данных секвенирования, полученных на платформе Illumina, использовали java программу Trimmomatic 0.39 в режиме PE (парноконцевых прочтений) [12].

Пример кода в терминале выглядит следующим образом:

```
java -jar trimmomatic-0.39.jar PE 329_R1_sequence.fastq 329_R2_sequence.fastq 329_forward_paired.fastq 329_forward_unpaired.fastq 329_reverse_paired.fastq 329_reverse_unpaired.fastq
```

```
ILLUMINACLIP:/usr/share/trimmomatic/
NexteraPE-PE.fa:2:40:15 LEADING:3 TRAILING:3
SLIDINGWINDOW:4:15 MINLEN:36
```

Первая часть позволяет запустить пакет Trimmomatic в режиме работы с парноконцевыми прочтениями и из исходных данных, полученных во время прямого (R1) и обратного (R2) чтения библиотек, получить 4 файла в формате FASTQ: 2 — для «парных» выходных данных и 2 — для соответствующих «непарных» выходных данных. Опция ILLUMINACLIP позволяет удалить последовательности адаптеров из прочтений, выбрать соответствующий файл с адаптерными последовательностями (например, NexteraPE-PE.fa), удалить начальные (LEADING) и конечные (TRAILING) нуклеотиды с качеством ниже Q3 (для удаления нуклеотидов, идентифицированных с точностью менее 50,12 %) при сканировании последовательности с помощью скользящего окна (SLIDINGWINDOW) шириной в 4 нуклеотида, обрезаая, когда среднее качество падает ниже Q15 (для удаления нуклеотидов, идентифицированных с точностью менее 96,84 %), и удаляя прочтения короче 36 нуклеотидов (MINLEN).

Собственно сборка генома до набора контигов производится с помощью пакета SPAdes-3.15.5. Согласно литературным данным, он является одним из лучших решений для сборки малых геномов [13, 14].

Пример кода в терминале выглядит следующим образом:

```
./spades.py --threads 8 --careful \
--pe1-1 /home/user/329_forward_paired.fastq \
--pe1-2 /home/user/329_reverse_paired.fastq \
--pe1-s /home/user/329_forward_unpaired.
fastq \
--pe1-s /home/user/329_reverse_unpaired.
fastq \
-o /home/user/329spades
```

Программа выполняет сборку генома в «аккуратном» режиме (--careful), используя одновременно 8 потоков, и формирует набор контигов и скэффолдов на базе k-меров различного размера. В итоге программа выбирает наилучший результат и сохраняет его в выходной директории (-o /home/user/329spades).

Для обнаружения смежных контигов и расчета величины покрытия и ожидаемого размера генома при заданном размере k-мера (-k 127) используется пакет combinator-FQ [15].

Пример кода в терминале выглядит следующим образом:

```
./combinator-FQ.py contigs.fasta -k 127 (-o /
home/user/combinator-FQ_report
```

В качестве входных данных используется набор контигов, полученный на предыдущем этапе, а на выходе получаем таблицу в формате tsv с указанием совместимых концов контигов и отчет следующего вида:

```
=== Summary ===
```

```
290 contigs were processed.
```

```
Sum of contig lengths: 5699908 bp
```

```
Expected length of the genome: 6059787 bp
```

```
Min coverage: 0.727723
```

```
Max coverage: 725.0
```

```
Mean coverage: 66.04
```

```
Median coverage: 34.43
```

```
LQ-coefficient: 48.3
```

Оценку качества сборки генома можно проводить с помощью сервиса QUASt (Quality Assessment Tool for Genome Assemblies) посредством анализа рассчитываемых им показателей (рисунок 2) [16].

QUASt работает как с эталонными геномами, так и без них. Инструмент принимает несколько сборок, поэтому подходит для их сравнения.

Для предварительной оценки качества сборки генома использовали следующие параметры: общее количество контигов, N50 и L50 (в некоторых случаях данные параметры могут быть неинформативными).

Общее количество контигов (рисунок 2а) и их распределение по длине (рисунок 2а и 2б) позволяют получить информацию об уровне фрагментации собранного генома. Чем меньше значение данного показателя, тем выше качество сборки. Параметр N50 характеризует качество сборки с точки зрения ее непрерывности и равен по значению длине самого короткого контига при суммировании длин минимального числа контигов (L50) для получения суммы, большей или равной половине длины сборки генома.

Использованные настройки Trimmomatic и SPAdes на платформе Linux позволили получить сборку генома с большей величиной N50 и меньшим количеством контигов для покрытия 50 % генома (L50) по сравнению с приложением SPAdes Genome Assembler (рисунок 2а и 2б). Однако оба варианта показывают сходные результаты при сравнении полученных с их помощью сборок генома на веб-ресурсах, позволяющих проводить анализ антибиотикорезистентности, что дает возможность выбора методики сборки генома на усмотрение исследователя.

Полученные файлы в формате *.fasta могут быть использованы для дальнейшего анализа генетических детерминант антибиотикорезистентности на веб-ресурсах ResFinder 4.1 [17, 18], CARD RGI [19, 20] или др.

При необходимости определения сиквенс-типа микроорганизма можно использовать ресурс Multi Locus Sequence Typing (MLST) [21].

All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

Worst Median Best Show heatmap

	contigs_329spades_ubuntu	contigs329BS
# contigs	123	153
# contigs (≥ 0 bp)	290	277
# contigs (≥ 1000 bp)	110	135
# contigs (≥ 5000 bp)	71	87
# contigs (≥ 10000 bp)	61	73
# contigs (≥ 25000 bp)	48	49
# contigs (≥ 50000 bp)	34	36
Largest contig	395 721	477 769
Total length	5 671 778	5 664 312
Total length (≥ 0 bp)	5 699 908	5 691 676
Total length (≥ 1000 bp)	5 662 404	5 651 343
Total length (≥ 5000 bp)	5 566 665	5 535 226
Total length (≥ 10000 bp)	5 500 805	5 443 147
Total length (≥ 25000 bp)	5 253 479	4 997 049
Total length (≥ 50000 bp)	4 735 024	4 533 502
N50	132 893	115 489
N75	82 539	70 139
L50	14	16
L75	27	31
GC (%)	56.74	56.75
Mismatches		
# N's	0	0
# N's per 100 kbp	0	0

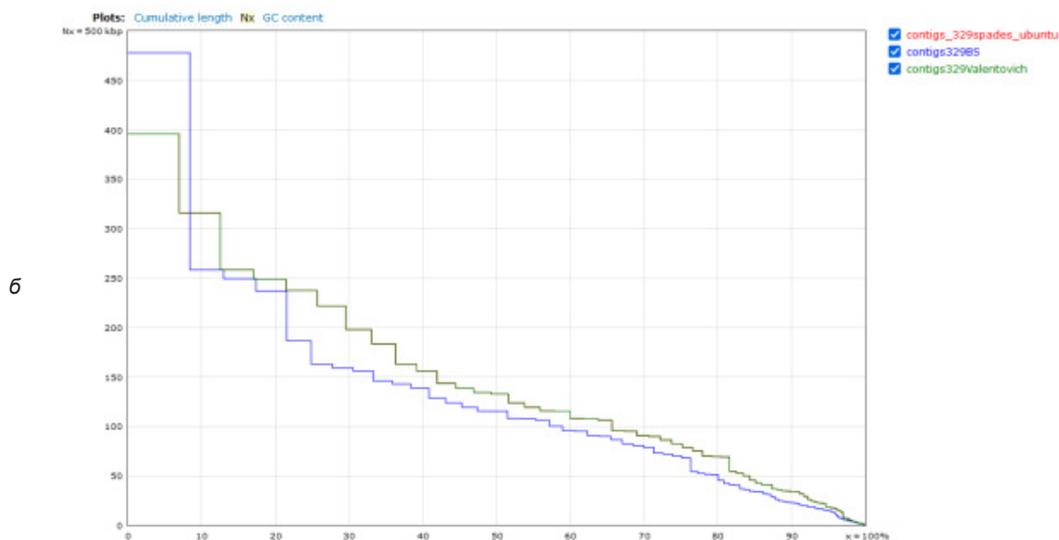


Рисунок 2. Результаты оценки трех сборок геномов с помощью сервиса QUASt: а — количественные показатели анализа сборки генома; б — распределение контигов по длине
 Figure 2. The results of the evaluation of three genome assemblies using the QUASt service: a — quantitative indicators of the genome assembly analysis; b — contigs distribution by length

Заключение

Платформа Illumina MiSeq позволяет выполнять широкий спектр задач, направленных на целевое секвенирование малых геномов, а также проведение исследований в области метагеномики и клинической молекулярной диагностики.

Рассмотрены основные этапы секвенирования генома *K. pneumoniae* с использованием метода коротких прочтений на платформе Illumina и

освещены ключевые особенности процесса пробоподготовки библиотек и анализа полученных данных. Показана значимость контроля качества на всех этапах высокопроизводительного секвенирования.

Данная работа будет полезна для начинающих исследователей в молекулярно-генетической биологии.

Список литературы

- Gupta N, Verma VK. Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. In: Arora PK, editor. *Microbial Technology for the Welfare of Society*, Singapore: Springer; 2019. p. 313-341. DOI: https://doi.org/10.1007/978-981-13-8844-6_15
- Pervez MT, Hasnain MJU, Abbas SH, Moustafa MF, Aslam N, Shah SSM. A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *Biomed Res Int*. 2022 Sep 2022;29:345780. DOI: <https://doi.org/10.1155/2022/3457806>
- AmpliSeq for Illumina Community Panels. [дата обращения 2023 январь 24]. Режим доступа: <https://emea.illumina.com/products/by-brand/ampliseq/community-panels.html>
- Goodwin, S., McPherson, J. & McCombie, W. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17:333-335. DOI: <https://doi.org/10.1038/nrg.2016.49>
- Quality Scores for Next-Generation Sequencing. [дата обращения 2023 январь 24]. Режим доступа: https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf
- Cluster density guidelines for Illumina sequencing platforms using non-patterned flow cells. [дата обращения 2023 январь 23]. Режим доступа: <https://emea.support.illumina.com/bulletins/2016/10/cluster-density-guidelines-for-illumina-sequencing-platforms-.html>
- MiSeq System Denature and Dilute Libraries Guide (15039740). [дата обращения 2022 декабрь 15]. Режим доступа: https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-10.pdf
- How much PhiX spike-in is recommended when sequencing low diversity libraries on Illumina platforms? [дата обращения 2022 сентябрь 21]. Режим доступа: <https://emea.support.illumina.com/bulletins/2017/02/how-much-phix-spike-in-is-recommended-when-sequencing-low-divers.html>
- Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. [дата обращения 2022 октябрь 5]. Режим доступа: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- SPAdes-Genome-Assembler Details - BaseSpace Sequence Hub. [дата обращения 2023 январь 16]. Режим доступа: <https://basespace.illumina.com/apps/3047044/SPAdes-Genome-Assembler?preferredversion>
- Bioinformatics portal Galaxy | Europe. [дата обращения 2023 январь 24]. Режим доступа: https://usegalaxy.eu/root?tool_id=toolshed.g2.bx.psu.edu%2Frepos%2Fnm1%2Fspades%2Fspades%2F3.15.4%2Bgalaxy1
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120. DOI: <https://doi.org/10.1093/bioinformatics/btu170>
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*. 2012;19:455-477. DOI: <https://doi.org/10.1089/cmb.2012.0021>
- Releases ablab/spades. GitHub [дата обращения 2023 январь 16]. Режим доступа: <https://github.com/ablab/spades/releases>
- Сиколенко МА, Сергеев РС, Валентович ЛН. Метод оценки полноты нуклеотидных данных для сборки геномных последовательностей на основе расчёта доли смежных контигов. В: Материалы II междунар. науч.-практ. конф. «Компьютерные технологии и анализ данных (СТДА'2020)» (Минск, 23–24 апреля 2020 г). Минск: БГУ; 2019. с. 162-166. [дата обращения 2023 январь 18]. Режим доступа: <https://elib.bsu.by/handle/123456789/248660>
- QUAST - Quality Assessment Tool for Genome Assemblies [дата обращения 2023 январь 23]. Режим доступа: <https://cab.cc.spbu.ru/quast/>
- Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*. 2020;75:3491-3500. DOI: <https://doi.org/10.1093/jac/dkaa345>
- ResFinder 4.1 - Identification of acquired antibiotic resistance genes [дата обращения 2023 январь 10]. Режим доступа: <https://cge.food.dtu.dk/services/ResFinder/>
- McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 2013;57:3348-3357. DOI: <https://doi.org/10.1128/AAC.00419-13>
- The Comprehensive Antibiotic Resistance Database. [дата обращения 2023 январь 09]. Режим доступа: <https://card.mcmaster.ca/analyze/rqi>
- MLST 2.0 - Multi Locus Sequence Typing [дата обращения 2023 январь 09]. Режим доступа: <https://cge.food.dtu.dk/services/MLST/>

References

- Gupta N, Verma VK. Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. In: Arora PK, editor. *Microbial Technology for the Welfare of Society*, Singapore: Springer; 2019. p. 313-341. DOI: https://doi.org/10.1007/978-981-13-8844-6_15
- Pervez MT, Hasnain MJU, Abbas SH, Moustafa MF, Aslam N, Shah SSM. A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *Biomed Res Int*. 2022 Sep 2022;29:345780. DOI: <https://doi.org/10.1155/2022/3457806>
- AmpliSeq for Illumina Community Panels. [дата обращения 2023 январь 24]. Режим доступа: <https://emea.illumina.com/products/by-brand/ampliseq/community-panels.html>
- Goodwin, S., McPherson, J. & McCombie, W. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17:333-335. DOI: <https://doi.org/10.1038/nrg.2016.49>
- Quality Scores for Next-Generation Sequencing. [дата обращения 2023 январь 24]. Режим доступа: https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf
- Cluster density guidelines for Illumina sequencing platforms using non-patterned flow cells. [дата обращения 2023 январь 23]. Режим доступа: <https://emea.support.illumina.com/bulletins/2016/10/cluster-density-guidelines-for-illumina-sequencing-platforms-.html>
- MiSeq System Denature and Dilute Libraries Guide (15039740). [дата обращения 2022 декабрь 15]. Режим доступа: https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-10.pdf
- How much PhiX spike-in is recommended when sequencing low diversity libraries on Illumina platforms? [дата обращения 2022 сентябрь 21]. Режим доступа: <https://emea.support.illumina.com/bulletins/2017/02/how-much-phix-spike-in-is-recommended-when-sequencing-low-divers.html>
- Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. [дата обращения 2022 октябрь 5]. Режим доступа: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- SPAdes-Genome-Assembler Details - BaseSpace Sequence Hub. [дата обращения 2023 январь 16]. Режим доступа: <https://basespace.illumina.com/apps/3047044/SPAdes-Genome-Assembler?preferredversion>
- Bioinformatics portal Galaxy | Europe. [дата обращения 2023 январь 24]. Режим доступа: https://usegalaxy.eu/root?tool_id=toolshed.g2.bx.psu.edu%2Frepos%2Fnm1%2Fspades%2Fspades%2F3.15.4%2Bgalaxy1

12. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120. DOI: <https://doi.org/10.1093/bioinformatics/btu170>
13. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*. 2012;19:455-477. DOI: <https://doi.org/10.1089/cmb.2012.0021>
14. Releases ablab/spades. GitHub [дата обращения 2023 январь 16]. Режим доступа: <https://github.com/ablab/spades/releases>
15. Sikolenko MA, Sergeev RS, Valentovich LN. A method for assessing the completeness of nucleotide data for genomic sequences based on the calculation of the proportions of contiguous contig assemblies. In: Proceedings of the II Intern. scientific-practical. conf. «Computer technologies and data analysis (CTDA'2020)» (Minsk, April 23-24, 2020) Minsk: BSU; 2019. p. 162-166. [Date of access 2023 January 18]. Mode of access: <https://elib.bsu.by/handle/123456789/248660> (in Russ.).
16. QUASt - Quality Assessment Tool for Genome Assemblies [Date of access 2023 January 23]. Mode of access: <https://cab.cc.spbu.ru/quast/>
17. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;75:3491–5000 [Date of access 2023 January 09]. Mode of access: <https://doi.org/10.1093/jac/dkaa345>
18. ResFinder 4.1 - Identification of acquired antibiotic resistance genes [Date of access 2023 January 10]. Mode of access: <https://cge.food.dtu.dk/services/ResFinder/>
19. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 2013;57:3348-3357. DOI: <https://doi.org/10.1128/AAC.00419-13>
20. The Comprehensive Antibiotic Resistance Database. [Date of access 2023 January 09]. Mode of access: <https://card.mcmaster.ca/analyze/rqi>
21. MLST 2.0 - Multi Locus Sequence Typing [Date of access 2023 January 09]. Mode of access: <https://cge.food.dtu.dk/services/MLST/>

Информация об авторах / Information about the authors

Шафорост Александр Сергеевич, старший научный сотрудник научно-исследовательской лаборатории, УО «Гомельский государственный медицинский университет», Гомель, Беларусь

ORCID: <https://orcid.org/0000-0002-6725-5353>

e-mail: asofocl@mail.ru

Зяцьков Алексей Александрович, старший научный сотрудник научно-исследовательской лаборатории, УО «Гомельский государственный медицинский университет», Гомель, Беларусь

ORCID: <https://orcid.org/0000-0001-9542-3791>

e-mail: ziatskovaa@gmail.com

Воропаев Евгений Викторович, к.м.н., доцент, проректор по научной работе, УО «Гомельский государственный медицинский университет», Гомель, Беларусь

ORCID: <https://orcid.org/0000-0002-9435-6109>

e-mail: voropaev.evgenii@gmail.com

Осипкина Ольга Викторовна, заведующий научно-исследовательской лабораторией, УО «Гомельский государственный медицинский университет», Гомель, Беларусь

ORCID: <https://orcid.org/0000-0002-1931-4224>

e-mail: olga.osipkina@mail.ru

Воропаева Алла Викторовна, к.б.н., доцент, врач клинической лабораторной диагностики, ГУ «Республиканский научно-практический центр радиационной медицины и экологии человека», Гомель, Беларусь

ORCID: <https://orcid.org/0000-0003-0213-2421>

e-mail: allo4ka3665@mail.ru

Бонда Надежда Александровна, заведующий микробиологической лабораторией, ГУ «Гомельский областной центр гигиены, эпидемиологии и общественного здоровья», Гомель, Беларусь

ORCID: <https://orcid.org/0000-0003-2104-585X>

e-mail: bondana8448@gmail.com

Стома Игорь Олегович, д.м.н., доцент, ректор, УО «Гомельский государственный медицинский университет», Гомель, Беларусь

ORCID: <https://orcid.org/0000-0003-0483-7329>

e-mail: rektor@gsmu.by

Alexander S. Shaforst, Senior Researcher at the Research Laboratory of Gomel State Medical University

ORCID: <https://orcid.org/0000-0002-6725-5353>

e-mail: asofocl@mail.ru

Alexey A. Ziatskov, Senior Researcher at the Research Laboratory at Gomel State Medical University

ORCID: <https://orcid.org/0000-0001-9542-3791>

e-mail: ziatskovaa@gmail.com

Evgenii V. Voropaev, Candidate of Medical Sciences, Associate Professor, Vice-Rector for Scientific Work, Gomel State Medical University

ORCID: <https://orcid.org/0000-0002-9435-6109>

e-mail: voropaev.evgenii@gmail.com

Olga V. Osipkina, Head of the Research Laboratory, Gomel State Medical University

ORCID: <https://orcid.org/0000-0002-1931-4224>

e-mail: olga.osipkina@mail.ru

Alla V. Voropaeva, Candidate of Biological Sciences, Associate Professor, Physician at Clinical Laboratory Diagnostics, Republican Scientific and Practical Center for Radiation Medicine and Human Ecology

ORCID: <https://orcid.org/0000-0003-0213-2421>

e-mail: allo4ka3665@mail.ru

Nadezhda A. Bonda, Head of the Microbiology Laboratory, Gomel Regional Center for Hygiene, Epidemiology and Public Health

ORCID: <https://orcid.org/0000-0003-2104-585X>

e-mail: bondana8448@gmail.com

Igor O. Stoma, Doctor of Medical Sciences, Associate Professor, Rector of Gomel State Medical University

ORCID: <https://orcid.org/0000-0003-0483-7329>

e-mail: rektor@gsmu.by

Автор, ответственный за переписку / Corresponding author

Воропаев Евгений Викторович
e-mail: voropaev.evgenii@gmail.com

Evgenii V. Voropaev
e-mail: voropaev.evgenii@gmail.com

Поступила в редакцию / Received 02.02.2023

Поступила после рецензирования / Accepted 08.02.2023

Принята к публикации / Revised 28.02.2023