

Чаплин А.В., Коржанова М., Коростин Д.О.

ВЫЯВЛЕНИЕ ГЕНОВ АНТИБИОТИКОРЕЗИСТЕНТНОСТИ БАКТЕРИЙ В ДАННЫХ ПОЛНОГЕНОМНОГО СЕКВЕНИРОВАНИЯ (ОБЗОР ЛИТЕРАТУРЫ)

ФГАОУ ВО Российский национальный исследовательский медицинский университет им. Н. И. Пирогова Минздрава РФ, 119571, Москва, Россия

Одной из важнейших проблем современной медицины является распространение резистентности к антибиотикам среди актуальных бактериальных возбудителей патологии человека. Диагностические технологии, направленные на определение спектра устойчивости каждого бактериального изолята, лежат в основе оптимизации лечебного процесса и мероприятий по борьбе с распространением резистентности. Совершенствование методов оценки резистентности становится приоритетной задачей лабораторной медицины.

Обзор посвящён важнейшей проблеме современной медицины, связанной с распространением антибиотикорезистентности – анализу возможностей биоинформатических инструментов обработки данных полногеномного секвенирования бактериальных возбудителей. Поиск источников для анализа проблемы выполнялся по базам данных PubMed, Российской научной электронной библиотеке eLIBRARY, поисковым системам Всемирной организации здравоохранения, Европейского общества микробиологии и инфекционных болезней (ESCMID).

Описаны существующие технологические платформы полногеномного секвенирования для определения генетических детерминант резистентности у бактерий. В основе систем обнаружения детерминант антибиотикорезистентности лежат базы данных известных белковых или нуклеотидных последовательностей. К числу наиболее часто используемых баз данных принадлежат Resfinder, CARD, Bacterial Antimicrobial Resistance Reference Gene Database.

Классическим этапом поиска генов резистентности является сверка новых последовательностей с базой данных с учётом разных порогов сходства. Предварительная сборка бактериального генома изучаемого изолята делает результаты поиска генов резистентности более корректными (по сравнению с поиском генов в контигах, не собранных в геном). Рассматриваются новые биоинформатические подходы к поиску генов резистентности, включая использование нейросетей и машинного обучения. Критический обзор достоинств и недостатков существующих баз данных антибиотикорезистентности позволил предложить протокол предсказания антибиотикорезистентности на основе массивов полногеномного секвенирования. Предлагаемый протокол может быть использован в качестве основы для создания стандартизированного алгоритма качественной и количественной оценки чувствительности микроорганизмов к АМП на основе данных полногеномного секвенирования.

Ключевые слова: антибиотики; резистентность; бактерии; секвенирование; биоинформатика; обзор.

Для цитирования: Чаплин А.В., Коржанова М., Коростин Д.О. Выявление генов антибиотикорезистентности бактерий в данных полногеномного секвенирования (обзор литературы). Клиническая лабораторная диагностика. 2021; 66 (11): 684-688. DOI: <https://dx.doi.org/10.51620/0869-2084-2021-66-11-684-688>

Для корреспонденции: Коржанова Маргарита, лаборант центра высокоточного редактирования и генетических технологий для биомедицины; e-mail: marory.kmm@gmail.com

Благодарности. Выражаем благодарность Центру высокоточного редактирования и генетических технологий для биомедицины за поддержку в методической части исследования/работы.

Финансирование. Работа выполнена при поддержке гранта Российского научного фонда (проект № 20-15-00235).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила 04.09.2020

Принята к печати 23.03.2021

Опубликовано 29.11.2021

Chaplin A.V., Korzhanova M., Korostin D.O.

IDENTIFICATION OF BACTERIAL ANTIBIOTIC RESISTANCE GENES IN NEXT-GENERATION SEQUENCING DATA (REVIEW OF LITERATURE)

Pirogov Russian National Research Medical University, 119571, Moscow, Russia

The spread of antibiotic-resistant human bacterial pathogens is a serious threat to modern medicine. Antibiotic susceptibility testing is essential for treatment regimens optimization and preventing dissemination of antibiotic resistance. Therefore, development of antibiotic susceptibility testing methods is a priority challenge of laboratory medicine. The aim of this review is to analyze the capabilities of the bioinformatics tools for bacterial whole genome sequence data processing. The PubMed database, Russian scientific electronic library eLIBRARY, information networks of World health organization and European Society of Clinical Microbiology and Infectious Diseases (ESCMID) were used during the analysis. In this review, the platforms for whole genome sequencing, which are suitable for detection of bacterial genetic resistance determinants, are described. The classic step of genetic resistance determinants searching is an alignment between the query nucleotide/protein sequence and the subject (database) nucleotide/protein sequence, which is performed using the nucleotide and protein sequence databases. The most commonly used databases are Resfinder, CARD, Bacterial Antimicrobial Resistance Reference Gene Database. The results of the resistance determinants searching in genome assemblies is more correct in comparison to results of the searching in contigs. The new resistance genes searching bioinformatics tools, such as neural networks and machine learning, are discussed in the review. After critical appraisal of the current antibiotic resistance databases we designed a protocol for predicting antibiotic resistance using whole genome sequence data. The designed protocol can be used as a basis of the algorithm for qualitative and quantitative antimicrobial susceptibility testing based on whole genome sequence data.

Key words: antibiotics; resistance, bacteria; sequencing; bioinformatics; review.

For citation: Chaplin A.V., Korzhanova M., Korostin D.O. Identification of bacterial antibiotic resistance genes in next-generation sequencing data (review of literature). *Klinicheskaya Laboratornaya Diagnostika (Russian Clinical Laboratory Diagnostics)*. 2021; 66 (11): 684-688 (in Russ.). DOI: <https://dx.doi.org/10.51620/0869-2084-2021-66-11-684-688>

For correspondence: Korzhanova M., laboratory assistant Center for Precision Genome Editing and Genetic Technologies for Biomedicine; e-mail: marory.kmm@gmail.com

Information about authors:

Chaplin A., <https://orcid.org/0000-0003-1377-7153>;
Korzhanova M., <https://orcid.org/0000-0003-2512-2517>;
Korostin D., <https://orcid.org/0000-0003-1343-2550>.

Acknowledgment. *The study was supported by the Russian Science Foundation (Project ID 20-15-00235). We thank the Center for Precision Genome Editing and Genetic Technologies for Biomedicine (Moscow) for the genetic research methods.*

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

Received 04.09.2020
Accepted 23.03.2021
Published 29.11.2021

Борьба с заболеваниями, в основе которых лежит инфекционный процесс, осложняется распространением резистентности возбудителей к антибиотикам. Антибиотикорезистентность получила глобальное распространение и расценивается специалистами в качестве одной из угроз для человечества [1]. Только в США от проблем, связанных с резистентными микробами, ежегодно погибает около 35000 пациентов [2]. Важной задачей медицины является совершенствование технологий оценки антибиотикорезистентности каждого возбудителя. Фенотипические методы выявления резистентности, широко применяемые в микробиологической практике, часто не позволяют расшифровать механизмы её возникновения. Развитие и удешевление молекулярно-генетических методов и, прежде всего, полногеномного секвенирования (ПГС) в начале XXI века позволяет рассматривать ПГС в качестве альтернативы фенотипической диагностике. Европейский центр профилактики и контроля заболеваний рассматривает ПГС в качестве основной технологии будущего в изучении молекулярной эпидемиологии инфекционных заболеваний и антибиотикорезистентности бактерий [3]. Внедрение ПГС задерживается из-за отсутствия доступных и универсальных инструментов биоинформатической обработки полученных данных. Стандартные пакеты биоинформатического обеспечения, находящиеся на серверах компаний, предоставляющих услуги по обработке данных ПГС (Illumina BaseSpace Sequence Hub, Thermo Fisher Cloud Connect), не позволяют находить генетические детерминанты антибиотикорезистентности.

Цель обзора – проанализировать возможности биоинформатических инструментов анализа данных бактериальных сиквенсов и предложить протокол выявления генов резистентности в массивах данных ПГС.

Платформы для полногеномного секвенирования. К технологиям коротких прочтений относят методы второго поколения, основанные на детекции сигнала от множества случайно фрагментированных молекул ДНК длиной от 80 до 800 пар оснований и генерировании больших объёмов выходных данных [4].

Большую часть рынка ПГС-платформ коротких прочтений занимают системы Illumina. В основе их работы лежит технология SBS (от *англ.* – sequencing by synthesis), при которой считываются сигналы флуоресцентно меченных дезоксирибонуклеотидов с обратимыми терминаторами на 3'-конце, используемых для постройки комплементарной молекулы ДНК, при их присоединении ДНК-полимеразой. Секвенаторы MiniSeq и MiSeq отно-

сятся к бюджетной линейке и обладают низкой производительностью и понятным рабочим процессом, подходят для небольших лабораторий. Более дорогие HiSeq and NovaSeq обеспечивают высокую производительность при снижении себестоимости образца, но требуют дополнительной автоматизации процесса подготовки библиотек для регулярной загрузки приборов. Платформа IonTorrent от Thermo Fisher Scientific основана на полупроводниковом секвенировании, где детектируется изменение pH в ходе синтеза комплементарной цепи ДНК на прочитываемой матрице [5]. Модель этой платформы S5 сравнима с MiSeq по производительности, но превосходит его по длительности секвенирования в 2-3 раза. Пробоподготовка запуска может автоматизироваться при помощи системы IonChef. Недостатком является высокая частота ошибок в гомополимерных регионах.

Технологии длинных прочтений относятся к третьему поколению и основаны на детекции сигнала от единичной молекулы ДНК. При этом отсутствие этапа клональной амплификации библиотеки снижает количество ошибок в процессе подготовки ДНК, однако повышает требования к технологической части секвенаторов из-за высокого шума при детекции. Несмотря на преимущества для *de novo* сборки, III поколение все ещё используется реже, чем технологии коротких прочтений [6, 7]. Вероятно, это происходит из-за того, что технологии длинных прочтений нуждаются в собственных алгоритмах первичного и вторичного анализа, исправляющих ошибки, которых на 11-15% больше, чем у платформ коротких прочтений [8, 9]. Платформы PacBio от Pacific Biosciences используют секвенирование синтезом одиночных молекул ДНК и прочитывают фрагменты ДНК до 175 тыс. пар оснований в течение нескольких часов. Приборы Sequel и Sequel II характеризуются средней производительностью, но высокой скоростью секвенирования [10]. Платформа Oxford Nanopore Technologies MinION использует метод «протаскивания» ДНК через нанопоры, в ходе которого каждый тип азотистых оснований специфически изменяет ионный ток на мембране, длина прочтения достигает даже миллионов пар оснований [11]. Базовая модель MinION рассчитана на прямое подключение к компьютеру, обеспечивает высокую производительность при низкой скорости секвенирования: 30 Гб данных за 48 час. Из недостатков стоит отметить: трудности с прочтением GC-богатых регионов и невысокую точность секвенирования, которая составляет от 65% до 88% [12], ошибки при секвенировании гомополимерных участков [13].

Главным недостатком ПГС являются необходимость в квалифицированном персонале на каждом этапе выполнения методики, включая подготовку библиотеки, обогащение целевыми фрагментами, секвенирование, последующую обработку и интерпретацию данных. Другой недостаток связан с тем, что ПГС в России находится в «серой» правовой зоне: для большинства реактивов, приборов, технологий и биоинформатических решений нет регистрационных удостоверений, допускающих к использованию в клинической практике.

Биоинформатическая обработка данных полногеномного секвенирования с целью выявления генов антибиотикорезистентности. В основе систем обнаружения детерминант антибиотикорезистентности лежат базы данных известных белковых или нуклеотидных последовательностей. Среди самых популярных баз данных, которые курируются вручную и регулярно обновляются следует отметить базы Resfinder [14], CARD [15], Bacterial Antimicrobial Resistance Reference Gene Database [16].

Стандартным этапом поиска детерминант резистентности является сверка просеквенированных последовательностей изучаемого бактериального изолята с последовательностями детерминант резистентности, представленными в используемой базе данных. Эта задача может быть сравнительно простой для случая активно изучаемых условно-патогенных или патогенных микроорганизмов – зачастую детерминанты резистентности будут на 100% соответствовать представленным в базах данных. Бывают и другие случаи. В настоящее время активно изучаются резистомы – совокупности всех генов антибиотикорезистентности, представленных в бактериальных сообществах, населяющих определённые экологические ниши (например, кишечник человека) [17]. Поиск детерминант резистентности в метагеномах для оценки состава резистомов может представлять значительные сложности – такие детерминанты мало представлены в существующих базах данных, преимущественно сформированных на основании изучения легко культивируемых микроорганизмов, вследствие чего является задача поиска далёких гомологов [18].

Изначально для сверки новых последовательностей с базой данных использовался просто поиск гомологов, обладающих определённым процентом сходства последовательности с помощью алгоритмов BLAST или DIAMOND. Данный вариант используется в программах Resfinder [14], использующей одноимённую базу данных, и ABRicate (<https://github.com/tseemann/abricate>), способной осуществлять поиск во множестве баз данных. Развитием данного подхода можно считать установление разных порогов сходства для разных детерминант резистентности, как сделано в программном обеспечении RGI, использующем базу данных CARD [15]. Недавно предложен более сложный вариант, воплощённый в алгоритме DeerpARG, при котором результаты поиска гомологов по сходству передаются в глубокую нейросеть, что может обеспечить более точное распознавание генов резистентности, далёких от референсных [19].

Перспективная альтернатива перечисленных способов связана с подходом на основе использования скрытых моделей Маркова (СММ), которые позволяют по-разному оценивать несовпадения и инсерции/делеции в разных позициях белковой последовательности, что даёт возможность различать консерватив-

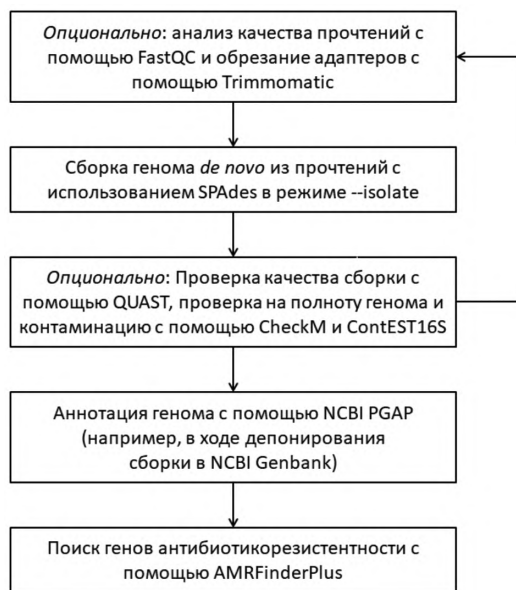
ные и вариабельные участки белка [20]. С помощью СММ возможно выявление генов резистентности, которые являются эволюционно отдалёнными от ранее известных генетических детерминант [16, 20]. За счёт этого подход с использованием СММ оказался более чувствительным в выявлении генов резистентности в метагеномах почв и кишечника человека, чем поиск с помощью BLAST по базе данных CARD [20]. Комбинированная схема, в которой применяются как BLAST, так и СММ, используется в программном обеспечении AMRFinderPlus, осуществляющем поиск с использованием Bacterial Antimicrobial Resistance Reference Gene Database [16].

Для поиска отдалённых гомологов генов резистентности разработан и ряд других методов, пока не воплощённый в виде универсального и широко используемого программного обеспечения. Для этой цели применяется анализ таких признаков, как аминокислотный состав, вторичная структура, физико-химические свойства белка [21, 22]. Наиболее ресурсоёмкий и вычислительно сложный из применённых подходов – сравнительное моделирование трёхмерных структур белков, которые являются более консервативными, чем аминокислотная последовательность [18]. Возможно, такие методы впоследствии будут применяться для расширения существующих баз данных, чтобы потом поиск по ним можно было осуществлять более простыми средствами.

Классические подходы к предсказанию антибиотикорезистентности по последовательности генома используют собранный геном (или метагеном). Используется сборка генома *de novo*, при которой прочтения объединяются в длинные фрагменты непрерывной геномной последовательности (контиги). Фрагментированность такой сборки может привести к потере некоторых генов антибиотикорезистентности, если они будут разорваны границами контигов [23]. Сложность могут представлять многокопийные гены (такие, как рибосомальные РНК), так как небольшие вариации между копиями могут быть потеряны в результате сборки. Если главной задачей исследования является обнаружение аллелей генов «домашнего хозяйства», вызывающих антибиотикорезистентность, альтернативой сборке *de novo* может стать сборка по референсному геному, однако приобретённые гены резистентности (располагающиеся, например, в мобильных генетических элементах) могут быть потеряны при таком подходе.

Использование собранных геномов позволяет тщательно проверить чистоту образца, качество секвенирования и качество предсказаний, однако, сборка является медленным процессом, требовательным к вычислительным ресурсам. Для упрощения разработан ряд подходов, которые осуществляют поиск непосредственно в прочтениях, получаемых в результате полногеномного секвенирования. Программное обеспечение ARIBA [24] позволяет осуществлять картирование парно-концевых прочтений на кластеры, составленные из генов антибиотикорезистентности, с последующей локальной сборкой прочтений и выравниванием этой сборки на гены антибиотикорезистентности.

Ещё одним важным аспектом является простота использования. Большинство вариантов программного обеспечения рассчитано на запуск через командную строку в системе Linux, что требует наличия квалифицированного сотрудника, осуществляющего обработку данных. Обнадеживающим исключением является



Возможный протокол предсказания антибиотикорезистентности по последовательности генома чистой культуры бактерий.

система Mykrobe с графическим интерфейсом, рассчитанная на анализ прочтений, полученных в результате полногеномного секвенирования *Mycobacterium tuberculosis* [25].

Серьёзная проблема выявления генетических детерминант антибиотикорезистентности проявляется в отсутствии общепринятого подхода к валидации результатов с использованием различных программ [26, 27]. Непонятно, должны ли быть представлены в тестовых выборках «пограничные» случаи антибиотикорезистентности, насколько должны тестовые выборки отражать реальное генетическое разнообразие микробных сообществ, какое соотношение различных механизмов антибиотикорезистентности должно быть в них отображено [27]. Среди наиболее крупных валидационных работ на реальных данных необходимо отметить сравнительное исследование предсказательной способности AMRFinder и Resfinder, в котором первая программа показала более высокую чувствительность [16]. Данное исследование проведено на 6242 изолятах, однако, они относились к всего четырём видам возбудителей острых кишечных инфекций человека.

Отсутствие общепринятой методики валидации не позволяет выявить, какой именно из современных подходов к обнаружению детерминант резистентности окажется оптимальным для решения конкретной задачи. К тому же, базы данных генов антибиотикорезистентности постоянно расширяются, а в уже выпущенное программное обеспечение добавляются новые функции, такие как корректное предсказание инактивирующих мутаций в генах антибиотикорезистентности.

На основе анализа возможностей современных программных инструментов мы синтезировали протокол предсказания антибиотикорезистентности на основе данных полногеномного секвенирования генома чистой культуры бактерий (см. рисунок). Он может быть использован для данных, полученных с любой платформы высокопроизводительного секвенирования или некото-

рых их комбинаций за счёт универсальности программы для сборки *de novo* SPAdes [28]. Применение полной сборки генома позволяет провести полноценную проверку качества и отсутствия контаминации в образце на основании уровня различий между копиями консервативных белок-кодирующих генов (что осуществляется программой CheckM) и 16S рРНК (что можно проверить с помощью веб-сервиса ContEst16S) [29, 30]. Использование AMRFinderPlus позволяет проводить не только корректную аннотацию генов антибиотикорезистентности, присутствующих в базе данных, но и предсказывать функции у сравнительно далёких гомологов [16]. Недостатком данного варианта анализа можно считать значительные требования к вычислительной мощности и объёму оперативной памяти для осуществления сборки *de novo*, многоэтапность, не позволяющая масштабировать подход для одновременной работы со множеством штаммов. В случае необходимости такого масштабирования можно обратиться к программам для предсказания генов антибиотикорезистентности на уровне исходных прочтений, например, ARIBA [24], которые позволяют добиться желаемой цели за один шаг.

Часто предсказание антибиотикорезистентности по полному геному обычно осуществляется на основе правила «наличие гена – наличие резистентности». Данный подход несёт в себе допущения: (а) за резистентность отвечает только один локус, а не несколько; (б) локус резистентности имеет высокую пенетрантность и его эффект не может быть замаскирован другими участками генома или зависеть от условий внешней среды [23]. В некоторых случаях эти допущения нарушаются [31]. Это может приводить к значительному числу ошибок предсказания, что показано при валидации системы AMRFinder в случае предсказания резистентности *Salmonella enterica* к хинолонам и аминогликозидам [16].

Проведён ряд работ, в которых данная задача предсказания фенотипа по генотипу решалась с помощью подходов машинного обучения. Для этого необходимо создать большую базу результатов полногеномного секвенирования штаммов с известными профилями резистентности, из этой базы сформировать обучающую выборку для классификатора. При этом в некоторых работах использовались ранее полученные данные о генах антибиотикорезистентности [32], в других работах обучение классификатора осуществлялось с нулевого уровня [33]. При этом последние работы в этой области используют не алгоритмы обучения по типу «чёрного ящика» (например, нейросети), а легко интерпретируемые алгоритмы, такие как деревья решений [32,33]. Это позволяет не просто сформировать систему предсказания, но и проследить логику её работы, что откроет возможность более глубокого понимания механизмов приобретения резистентности.

Интересным направлением является предсказание резистентности не как бинарного признака «чувствительный/резистентный», а как численного показателя минимальной подавляющей концентрации препарата (МПК), что может помочь в исследовании ситуаций, когда данная концентрация близка к пограничному значению. Этот подход успешно применён как для случая предварительно заданного массива детерминант резистентности, так и для обучения «с нуля» [34, 35].

Такие системы машинного обучения зачастую видоспецифичны, поэтому развитие данного подхода ограничивается количеством геномных данных для конкрет-

ных видов возбудителей заболеваний человека. Можно ожидать, что в будущем подобные системы предсказания будут внедрены в программное обеспечение, используемое для рядовых задач предсказания антибиотикорезистентности.

ЛИТЕРАТУРА / REFERENCES

1. Banin E., Hughes D., Kuipers O.P. Bacterial pathogens, antibiotics and antibiotic resistance. *FEMS Microbiol. Rev.* 2017; 41(3): 450-2. DOI: 10.1093/femsre/fux016.
2. CDC's Antibiotic Resistance Threats in the United States, 2019 (2019 AR Threats Report). Available at: <https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf> (accessed 26 August 2020).
3. Revez J., Espinosa L., Albiger B., Leitmeyer K. C. Struelens M.J., Tóth Á. et al. Survey on the use of whole-genome sequencing for infectious diseases surveillance: rapid expansion of European national capacities, 2015-2016. *Front. Public Health.* 2017; 5: 347. DOI: 10.3389/fpubh.2017.00347.
4. Schadt E.E., Turner S., Kasarskis A. A window into third-generation sequencing. *Hum. Mol. Genet.* 2010; 19(R2): R227-R240. DOI: 10.1093/hmg/ddq416.
5. Rothberg J., Myers J. Semiconductor sequencing for life. *J. Biomol. Tech.* JBT. 2011; 22(Suppl): S41-S2.
6. Loman N.J., Quick J., Simpson J.T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods.* 2015; 12(8): 733-5. DOI: 10.1038/nmeth.3444.
7. Schmid M., Frei D., Patrignani A., Schlapbach R., Frey J. E., Remus-Emsermann, M. N., Ahrens, C. H. Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. *Nucleic Acids Res.* 2018; 46(17): 8953-65. DOI: 10.1093/nar/gky726.
8. Chin C.S., Alexander D.H., Marks P., Klammer A.A., Drake J., Heiner C. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. methods.* 2013; 10(6): 563-9. DOI: 10.1038/nmeth.2474.
9. Nagarajan N., Pop M. Sequence assembly demystified. *Nat. Rev. Genet.* 2013. 2013; 14(3): 157-67. DOI: 10.1038/nrg3367.
10. Rhoads A., Au K.F. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics.* 2015; 13(5): 278-89. DOI: 10.1016/j.gpb.2015.08.002.
11. Stoddart D., Heron A.J., Mikhailova E., Maglia G., Bayley H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl. Acad. Sci. USA.* 2009; 106(19): 7702-7. DOI: 10.1073/pnas.0901054106.
12. Lu H., Giordano F., Ning Z. Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics.* 2016; 14(5): 265-79. DOI: 10.1016/j.gpb.2016.05.004.
13. Jain M., Koren S., Miga K.H., Quick J., Rand A.C., Sasani T.A. et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. biotechnol.* 2018; 36(4): 338-45. DOI: 10.1038/nbt.4060.
14. Zankari E., Hasman H., Cosentino S., Vestergaard M., Rasmussen S., Lund O. et al. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 2012; 67(11): 2640-4. DOI: 10.1093/jac/dks261.
15. Alcock B.P., Raphenya A.R., Lau T.T., Tsang K.K., Bouchard M., Edalatmand, A. et al. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 2020; 48(D1): D517-D25. DOI: 10.1093/nar/gkz935.
16. Feldgarden M., Brover V., Haft D.H., Prasad A.B., Slotta D.J., Tolstoy I. et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* 2019; 63(11): e00483-19. DOI: 10.1128/AAC.00483-19.
17. Lim M.Y., Cho Y., Rho M. Diverse distribution of resistomes in the human and environmental microbiomes. *Curr. Genomics.* 2018; 19(8): 701-11. DOI: 10.2174/1389202919666180911130845.
18. Ruppé E., Ghoulane A., Tap J., Pons N., Alvarez A.S., Maziers N. et al. Prediction of the intestinal resistome by a three-dimensional structure-based method. *Nat. Microbiol.* 2019; 4(1): 112-23. DOI: 10.1038/s41564-018-0292-6.
19. Arango-Argoty G., Garner E., Pruden A., Heath L.S., Vikesland P., Zhang L. DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome.* 2018; 6(1): 1-15. DOI: 10.1186/s40168-018-0401-z.
20. Gibson M.K., Forsberg K.J., Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 2015; 9(1): 207-16. DOI: 10.1038/ismej.2014.106.
21. Chowdhury A.S., Khaledian E., Broschat S.L. Capreomycin resistance prediction in two species of Mycobacterium using a stacked ensemble method. *J. Appl. Microbiol.* 2019; 127(6): 1656-64. DOI: 10.1111/jam.14413.
22. Chowdhury A.S., Call D.R., Broschat S.L. Antimicrobial resistance prediction for gram-negative bacteria via game theory-based feature evaluation. *Sci. Rep.* 2019; 9: 14487. DOI: 10.1038/s41598-019-50686-z.
23. Su M., Satola S.W., Read T.D. Genome-based prediction of bacterial antibiotic resistance. *J. Clin. Microbiol.* 2019; 57(3): e01405-18. DOI: 10.1128/JCM.01405-18.
24. Hunt M., Mather A.E., Sánchez-Busó L., Page A.J., Parkhill J., Keane J. A., Harris S.R. ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genomics.* 2017; 3(10): e000131. DOI: 10.1099/mgen.0.000131.
25. Hunt M., Bradley P., Lapierre S.G., Heys S., Thomsit M., Hall M.B. et al. Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. *Wellcome Open Res.* 2019; 4: 191. DOI: 10.12688/wellcomeopenres.15603.1.
26. Yao H., Yiu S.-M. Deep analysis and optimization of CARD antibiotic resistance gene discovery models. *BMC Genomics.* 2019; 20: 914. DOI: 10.1186/s12864-019-6318-5.
27. Angers-Loustau A., Petrillo M., Bengtsson-Palme J., Berendonk T., Blais B., Chan K.G. et al. The challenges of designing a benchmark strategy for bioinformatics pipelines in the identification of antimicrobial resistance determinants using next generation sequencing technologies. *Fl1000Res.* 2018; 7: 459. DOI: 10.12688/fl1000research.14509.2.
28. Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 2012; 19(5): 455-77. DOI: 10.1089/cmb.2012.0021.
29. Lee I., Chalita M., Ha S.M., Na S.I., Yoon S.H., Chun J. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 2017; 67(6): 2053-7. DOI: 10.1099/ijsem.0.001872.
30. Parks D.H., Imelfort M., Skennerton C.T., Hugenholtz P., Tyson G.W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015; 25: 1043-55. DOI: 10.1101/gr.186072.114.
31. Hughes D., Andersson D.I. Environmental and genetic modulation of the phenotypic expression of antibiotic resistance. *FEMS Microbiol. Rev.* 2017; 41(3): 374-91. DOI: 10.1093/femsre/fux004.
32. Van Camp P.J., Haslam D.B., Porollo A. Prediction of antimicrobial resistance in gram-negative bacteria from whole-genome sequencing data. *Front. Microbiol.* 2020; 11: 1013. DOI: 10.3389/fmicb.2020.01013.
33. Drouin A., Letarte G., Raymond F., Marchand M., Corbeil J., Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. *Scientific Reports.* 2019; 9: 4071. DOI: 10.1038/s41598-019-40561-2.
34. Nguyen M., Long S.W., McDermott P.F., Olsen R.J., Olson R., Stevens R.L. et al. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal *Salmonella*. *J. Clin. Microbiol.* 2019; 57(2): e01260-18. DOI: 10.1128/JCM.01260-18.
35. Nguyen M., Brettin T., Long S.W., Musser J.M., Olsen R.J., Olson R., et al. Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Sci. Rep.* 2018; 8: 428. DOI: 10.1038/s41598-017-18972-w.