

# WHOLE-GENOME SEQUENCING OF CLINICAL *Mycobacterium tuberculosis* ISOLATES WITH DIFFERENT DRUG SENSITIVITY PROFILES

**Kozhamkulov U., Kairov U., Yerezhepov D.,  
Akhmetova A., Molkenov A., Akilzhanova A.**

*Nazarbayev University, National Laboratory Astana  
53, Kabanbai Batyr Avenue, 010000, Astana, Kazakhstan  
ulan.kozhamkulov@nu.edu.kz*

## ABSTRACT

Tuberculosis (TB) remains a major cause of morbidity and mortality in many countries and thus, represents a major public health problem in Kazakhstan and globally. An alternative and more promising option in the fight against this disease is the application of genetic diagnostic methods, for example, the detection of point mutations or other genetic determinants of TB drug resistance and genotyping of *M. tuberculosis*. Whole-genome sequencing and next-generation sequencing (NGS) technologies have developed rapidly in recent decades and are now applied in many branches of biomedical research. New NGS technologies allow us to study whole-genome sequences of pathogens associated with various infectious diseases, to study their molecular and genetic features, and thus, to understand these pathogens on a fundamental and practical basis in terms of microbiology, virology, and epidemiology.

This paper describes the application of high-performance “shotgun” next-generation sequencing, using a Roche 454 GS FLX + Titanium platform to determine the complete genome sequences of twenty clinical isolates of *M. tuberculosis* with different drug sensitivity profiles. The paper furthermore defines the quality indicators of genome sequencing, describes the de novo assembly of the complete genomes of *M. tuberculosis* isolates, and details the mapping and alignment to the reference genome of *M. tuberculosis* (H37Rv strain; GenBank database). The assembled genomes of two *M. tuberculosis* isolates were deposited in the NCBI GenBank database and are available for public access under the numbers AZBA00000000 and AZAZ00000000. The other genomes are available on request.

**Keywords:** whole-genome sequencing, tuberculosis, *Mycobacterium tuberculosis*, DNA, pyrosequencing, genome

УДК577.21

## ПОЛНОГЕНОМНОЕ СЕКВЕНИРОВАНИЕ КЛИНИЧЕСКИХ ИЗОЛЯТОВ *M. TUBERCULOSIS* РАЗЛИЧНЫМ ПРОФИЛЕМ ЛЕКАРСТВЕННОЙ ЧУВСТВИТЕЛЬНОСТИ

**Кожамкулов У.А., Каиров У.Е., Ережепов Д.А.  
Ахметова А.Ж., Молкенова А.Б., Акильжанова А.Р.**

*Назарбаев Университет, Частное учреждение «National Laboratory Astana»  
пр.Кабанбай батыра, 53, Астана, 010000, Казахстан  
ulan.kozhamkulov@nu.edu.kz*

## АБСТРАКТ

Туберкулез остается одной из основных причин заболеваемости и смертности во многих странах и серьезной проблемой общественного здравоохранения в Казахстане и в мире. Альтернативным и более перспективным вариантом в борьбе с этим заболеванием является использование генетических методов диагностики, например, выявление точечных мутаций или других генетических детерминант, определяющих резистентность к противотуберкулезным препаратам и генотипирование *M. tuberculosis*.

Полногеномное секвенирование или секвенирование следующего поколения «next-generation sequencing (NGS)» получило бурное развитие в последние десятилетия и используется во многих отраслях биомедицинских исследований.

Использование новых технологий NGS позволяют изучать полногеномные последовательности возбудителей различных инфекционных заболеваний, изучать их молекулярно-генетические особенности, развивать фундаментальные и практические основы микробиологии, вирусологии, эпидемиологии.

В данной работе описаны этапы проведения высокопроизводительного «shotgun» секвенирования нового поколения на платформе Roche 454 GS FLX+Titanium для определения полной

последовательности генома двадцати клинических изолятов *M.tuberculosis* с различным профилем лекарственной устойчивости. Определены показатели качества полногеномного секвенирования, проведена *de novo* сборка полных геномов *M.tuberculosis*, а также картирование и выравнивание на референсный геном штамма *M.tuberculosis*H37Rv из международной базы данных GenBank. Результаты сборки двух геномов *M.tuberculosis* были загружены в базу данных NCBI GenBank и доступны для публичного доступа под номерами AZBA00000000, AZAZ00000000, остальные геномы доступны по запросу.

**Ключевые слова:** полногеномное секвенирование, туберкулез, микобактерии туберкулеза, ДНК, пиросеквенирование, геном.

## ВВЕДЕНИЕ

Туберкулез остается одной из основных причин заболеваемости и смертности во многих странах и серьезной проблемой общественного здравоохранения в Казахстане и в мире, поражая одну треть населения земного шара и унося жизни более полутора миллионов человек в год [1]. Значительные усилия исследователей ориентированы на выявление биомаркеров с потенциальным применением в качестве антигенов (иммуногенных факторов вирулентности), специфических средств диагностики латентной или активной формы болезни, прогностических инструментов оценки тяжести заболевания, определения лекарственной устойчивости к основным противотуберкулезным препаратам [2,3]. Альтернативным и более перспективным вариантом в борьбе с этим заболеванием является использование генетических методов анализа, например, выявление точечных мутаций или других генетических детерминант, определяющих резистентность к противотуберкулезным препаратам первого и второго ряда и генотипирование *M. tuberculosis*. Этот подход стал возможным благодаря определенному успеху в исследовании молекулярно-генетических характеристик микобактерий и молекулярных основ резистентности *Mycobacterium tuberculosis* к противотуберкулезным препаратам [4-6].

Так называемое полногеномное секвенирование или секвенирование следующего поколения «next-generation sequencing (NGS)», получило бурное развитие в последние 10-15 лет. Классическим методом для секвенирования нуклеотидных последовательностей с высокой точностью, но небольшой производительностью, является метод терминации растущей цепи, разработанный Фредериком Сэнгером [7].

Разработанные новые технологии NGS позволяют одновременно определять нуклеотидные последовательности множества нитей ДНК, что на выходе дает массив данных и чтение огромного количества нуклеотидов в день [8,9]. Использование новых технологий NGS позволяет изучать полногеномные последовательности возбудителей различных инфекционных заболеваний, изучать их молекулярно-генетические особенности, развивать фундаментальные и практические основы микробиологии, вирусологии, эпидемиологии [10,11].

В настоящее время изучаются и обсуждаются возможности использования различных платформ NGS как в научных исследованиях, так и в повседневной практике лечебных учреждений для решения проблем диагностики инфекционных заболеваний, детекции и обнаружения новых мутаций, обуславливающих лекарственную устойчивость вирусов и бактерий [10, 12,13].

С быстрым развитием технологии секвенирования нового поколения появилась возможность для изучения и выявления особенностей генетического разнообразия микобактерий туберкулеза [14].

Полная последовательность генома микобактерии туберкулеза H37Rv (вирулентный штамм, выделенный в 1905 году, а затем размноженный в лабораторных условиях) была опубликована в 1998 году [15]. Кольцевой геном состоит из 4411532 пар оснований и имеет среднее содержание гуанина и цитозина 65,6%. Оригинальная аннотация определила 3974 гена, кодирующих 3924 белка и 50 стабильных РНК. С тех пор были добавлены первоначально упущенные, дополнительные 82 белок-кодирующих гена. Гены были определены с помощью определения последовательности гомологии с известными белками других микроорганизмов, экспериментов с использованием двумерного электрофореза и масс-спектрометрии, а также биоинформатических методов, которые проверяют роль кодонов *M. tuberculosis*. Около 4000 генов классифицируются на 11 широких функциональных групп. Из них 52% относятся к группам с точной или предполагаемой функцией, а остальные 48% приходятся на консервативные гипотетические или функционально неизвестные гены.

За публикацией полного генома H37Rv последовало завершение секвенирования генома клинического штамма микобактерии туберкулеза CDC1551 (выделенный в 1995 году и ответственный за вспышку ТБ в США) [16], а также частичное секвенирование генома микобактерий туберкулеза штамма 210 (представитель W/Beijing штаммов, ответственных за большинство случаев ТБ в странах Азии и бывшего Советского Союза, а также вспышки в США) [17,18]. Анализ этих последовательностей показывает наличие одиночных нуклеотидных полиморфизмов (SNP, включая синонимичные и несинонимичные замены), полиморфизмов большой последовательности (LSP; генетические делеции/вставки), а также изменение количества и типов мобильных элементов (например, транспозонов и профагов) между изолятами микобактерий туберкулеза.

В Казахстане ранее проводились исследования по изучению мутаций в генах, определяющих устойчивость к основным и отдельным противотуберкулезным препаратам, генотипированию штаммов методом MIRU-VNTR анализа, по полногеномному секвенированию клинических изолятов *M. tuberculosis* [19-24].

Целью данного исследования является изучение методологических подходов к проведению секвенирования полного генома *M. tuberculosis* и определение полной последовательности генома клинических изолятов *M. tuberculosis* различной лекарственной чувствительностью.

## МАТЕРИАЛЫ И МЕТОДЫ

### *Клинические изоляты M. tuberculosis*

Клинические изоляты *M. tuberculosis* были получены от пациентов, проходивших лечение с 2013 по 2014гг. в Национальном центре проблем туберкулеза Республики Казахстан. Идентификация и выделение чистой культуры *M. tuberculosis* проводились в референс-лаборатории Национального центра проблем туберкулеза Республики Казахстан, г. Алматы.

Клинические изоляты микобактерии туберкулеза, ДНК микобактерий туберкулеза: 1) *M. tuberculosis* множественной лекарственной устойчивостью (MDRTB); 2) *M. tuberculosis* широкой лекарственной устойчивостью (XDRTB); 3) *M. tuberculosis* полирезистентные – различные сочетания лекарственной устойчивости, кроме сочетания H (изониазид)+R (рифампицин); 4) *M. tuberculosis* монорезистентный, устойчивые к одному из противотуберкулезных препаратов (например, устойчивость только к рифампицину, или только к изониазиду, этамбутолу, стрептомицину); 5) лекарственно-чувствительные *M. tuberculosis* ко всем противотуберкулезным препаратам.

### *Выделение ДНК микобактерий*

Выделение тотальной геномной ДНК образцов коллекции было проведено с использованием метода, описанного Van Soolingen D. [25].

### *Подготовка ДНК библиотеки M. tuberculosis и полногеномное секвенирование*

Подготовка библиотек для полногеномного секвенирования проводилась по протоколам производителя RapidLibraryPreparationMethodManualGSFLX+Series-XL+. Подготовка библиотеки ДНК состоит из 6 этапов: ДНК фрагментация с помощью небулизации, затупление концов фрагментов ДНК, подготовка AMPurebeads, лигирование адаптеров, удаление маленьких фрагментов, количественная оценка библиотеки, качественная оценка библиотеки (опционально), подготовка рабочих аликвот.

В самом начале для получения библиотеки проводится фрагментация исходной ДНК до размеров фрагментов длиной 400-800 нуклеотидов, которые в дальнейшем будут амплифицированы и секвенированы на геномном анализаторе GS FLX+. Данные фрагментированные участки ДНК получают путем пропускания раствора через узкое отверстие под высоким давлением в процессе «небулизации». Для этого использовали 0,5 мкг изучаемого ДНК в 100 мкл ТЕ буфера, к ДНК добавляли 500 мкл небулизационного буфера. В небулизаторе (GSRapidLibraryPrepNebulizer, Roche) исследуемый ДНК в ТЕ буфере пропускали азот под давлением в 2,0 атм. в течение 1 минуты.

ДНК очищали с помощью набора MinElute (Qiagen) с добавлением к образцу ДНК 2,5 мл буфера PVI (Qiagen) и очисткой ДНК на колонке MinElutePCR Purification Kit (Qiagen), в соответствии с протоколом производителя. Готовый исследуемый ДНК вымывали с колонки с помощью 16 мкл ТЕ буфера и сохраняли для дальнейших исследований и получали двухнитевые фрагменты исследуемой ДНК.

Далее проводили обработку концов нити ДНК и лигировали специфические «адапторы» – 4-нуклеотидную уникальную последовательность для распознавания библиотеки и краситель FAM (рисунок 1).

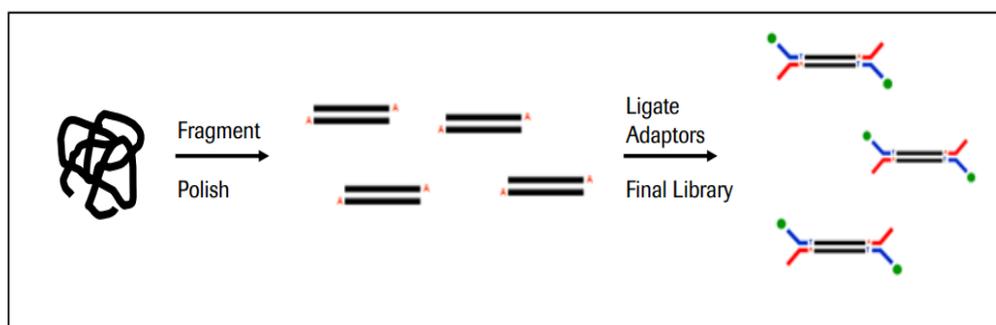


Рис.1. Схема лигирования адаптеров и получение конечной библиотеки ДНК

Fig. 1. Ligation of adaptors and preparation final DNA library

В итоге в результате небулизации были получены фрагменты ДНК как оптимальной для

пиросеквенирования длины 400-800 нуклеотидов, так и фрагменты, которые содержали фрагменты меньшей длины и несвязавшиеся адапторы. Секвенирование этих фрагментов нецелесообразно и приведет к снижению общего объема прочтенных последовательностей. Поэтому для очистки от фрагментов меньшей необходимой длины и ненужных адапторов проводили очистку препарата от коротких фрагментов с помощью набора AmpurebeadsXL(Agencourtampurebeads, BeckmanCoulterGenomics). Набор Ampurebeadsпозволят быстро и эффективно очистить препараты ДНК от фрагментов меньше 300 нуклеотидов.

Качественную оценку подготовленной библиотеки ДНК проводили с помощью электрофоретического метода на приборе AgilentBioanalyzer 2100 с использованием HighSensitivityDNAchip, согласно инструкции производителя. Для дальнейшего секвенирования ожидаемый диапазон 1400bp-1800bp.

Для количественной оценки библиотеки ДНК использовали флюорометрический метод с помощью прибора FluorometerHoefel DQ300.

Следующий этап – титрование, осуществлялся с помощью метода эмульсионного титрования, где проводились несколько эмульсионных ПЦР и использовались различные соотношения ДНК фрагментов к магнитным бусинкам (микрочастицы). Далее проводится отбор бусинок, содержащих амплифицированную ДНК, от не содержащих ДНК и высчитывается процент обогащения ДНК-содержащих бусинок от общего первоначального количества. Соответственно, если к большему количеству бусинок будет присоединена ДНК, то эмульсионнаяПЦР приведет к высокому проценту обогащения; если к меньшему количеству магнитных бусинок присоединяется ДНК, в этом случае процент будет низким.

Для осуществления эмульсионной ПЦР и обогащения магнитных бусинок с фрагментами ДНК использовали руководство завода-изготовителя RocheemPCRMethodManual – Lib-L SV, где использовался набор emPCRKitLib-LSV. Ниже представлены этапы проведения титрования библиотеки случайных фрагментов:

- 1) подготовка растворов для титрования и очистки;
- 2) связывание ДНК фрагментов библиотеки с магнитными бусинками;
- 3) формирование водно-масляных сфер;
- 4) эмульсионная ПЦР (амплификация);
- 5) этап очистки магнитных бусинок;
- 6) расчет процента обогащения ДНК-содержащих бусинок.

После подсчета и анализа процента обогащения проводится секвенирование библиотеки фрагментов искомой ДНК на генетическом анализаторе GS FLX+, который включает два этапа: проведение эмульсионной ПЦР (амплификация библиотеки) и пиросеквенирование фрагментов ДНК, связанных с магнитными бусинками на геномном анализаторе.

Полногеномное «shotgun» секвенирование было выполнено с использованием платформы Roche 454 GS FLX+ Titanium (Roche diagnostics) по стандартному протоколу GS-FLXPlus\_Sequencing\_Method-Manual\_XLPlusKit\_May2011.

#### **Биоинформатические методы**

Сборка полных геномов проводилась с использованием *NEWBLERde novo* assembler (454 Life Sciences, Branford, CT). Параметры сборки: ожидаемая глубина – 0; минимальная длина рида – 20; минимальная длина перекрытия – 40; минимальная идентичность перекрытия – 90%; параметр идентичности выравнивания – 2; параметр различности выравнивания – «-3»; пограничное значение для всех контигов – 100; пограничное значение для больших контигов – 500; пограничное значение для скафолдов – 2000.

Выравнивание и картирование сиквенсовых ридов проводилось на референсный штамм *M.tuberculosis* H37Rv (NC\_000962.3, GCF\_000195955.2) с использованием программного обеспечения GS Reference Mapping (454 Life Sciences, Branford, CT).

## **РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ**

Подготовлены разведения изучаемых образцов в альтернативной концентрации 1E+09 molecule/ul для каждого образца ранее подготовленной ДНК библиотеки.

#### **Начальная подготовка образцов**

ДНК библиотеки (Rapid DNA libraries) были подготовлены для 20 клинических изолятов *M. tuberculosis*: из них 8 MDR, 3 XDR, 1 монорезистентный, 1 полирезистентный и 7 чувствительных образцов *M. tuberculosis*.

Были использованы MID адапторы для подготовки библиотек. Список MID адапторов с соответствующими образцами показан в таблице 1.

**Таблица 1.** MID адапторы, использованные для подготовки ДНК библиотеки

**Table 1.** MID adaptors used to DNA library preparation

Образец	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
MID	RL001	RL002	RL003	RL004	RL005	RL006	RL007	RL008	RL009	RL010
Образец	N11	N12	N13	N14	N15	N16	N17	N18	N19	N20
MID	RL001	RL002	RL003	RL004	RL005	RL006	RL007	RL008	RL009	RL010

Для качественной оценки библиотеки ДНК подготовлены два пула по 10 образцов *M.tuberculosis* в каждой (первый пул P1=N1-N10), и таким же образом были объединены образцы второго пула (второй пулP2=N11-N20).

Правильное соотношение молекул ДНК и числа магнитных бусинок зависит от многих факторов: качества библиотеки, размера ДНК и других непредсказуемых факторов. Для получения правильных результатов ДНК-секвенирования необходимо провести титрование.

По рекомендациям протоколов компании-производителя и на основании практики для проведения пиросеквенирования оптимальным является процент обогащения от 5 до 20%, в этом случае большая часть микрочастиц содержит ПЦР-ампликоны единичной молекулы ДНК. Протокол титрования основан на руководстве emPCRAMplificationMethodManual-LibLSVGSFLX+ Series-XL+(May 2011, Roche).

Для подсчета числа обогащенных ДНК-содержащих микрочастиц и определения процента обогащения подсчитывали число частиц с помощью счетчика CASY модель TT, согласно инструкции производителя. Процент обогащения для каждой реакции ПЦР определяли по формуле: % обогащения = (число обогащенных микрочастиц с ДНК)/(общее число собранных микрочастиц).

В нашем исследовании оптимальным для эмульсионной ПЦР и секвенирования для P1 было соотношение 6 молекул на магнитную бусинку, и для P2 является соотношение 5 молекул на магнитную бусинку.

#### **Эмульсионный ПЦР фрагментов ДНК *M.tuberculosis* для пиросеквенирования**

Для успешной амплификации фрагментов ДНК библиотеки необходимо, чтобы одна магнитная бусинка связывалась с одним фрагментом ДНК, что обеспечивается за счет комплементарного взаимодействия ранее описанных адапторов олигонуклеотидами, пришитыми к бусинкам. Для проведения эмульсионной ПЦР все необходимые компоненты смеси добавляются в водно-масляные эмульсии. В результате в эмульсионных каплях (диаметр 50-100 мкм) фрагменты ДНК остаются связанными с магнитными бусинами. Изопропанолом удалялось эмульсионное масло и ДНК-содержащие бусинки выделялись и были готовы для последующего пиросеквенирования на GS FLX+.

Эмульсионный ПЦР и очистку микрочастиц проводили в соответствии с методикой (emPCRMethodManual -Lib-LLV) фирмы Roche, схожей для титрования, но с использованием набора для большого объема (GSTitaniumLVemPCRKit (Lib-L) v2), где содержание магнитных бусинок намного выше и доходит до 60-80млн.

После проведения эмульсионного ПЦР, отмывки бусинок, процент обогащения составил более 8%. Отсюда следует, что ДНК-содержащих бусинок достаточно для проведения 2 циклов пиросеквенирования на пикотитровальной пластине размером 70\*75 мм (рекомендуемое количество – 4 млн. частиц на цикл).

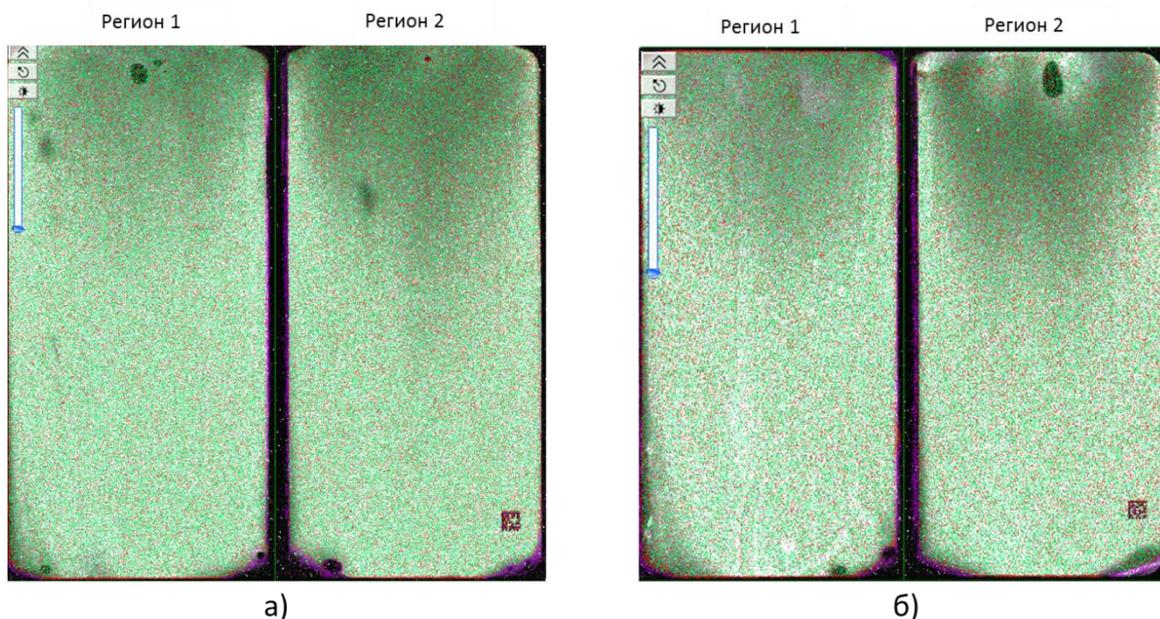
Припиросеквенировании бусинок со связанными фрагментами ДНК они помещаются в отдельные ячейки пикотитровальной пластины. Во время пиросеквенирования в емкость с пикотитровальной пластиной периодически поступает раствор, содержащий один из четырех нуклеотидов; включение используемого на данном цикле нуклеотида в комплементарную цепь ДНК-полимеразой сопровождается излучением света, что регистрируется высокочувствительной цифровой камерой в каждой из ячеек. Результатом является определенная в каждой из ячеек нуклеотидная последовательность соответствующего фрагмента ДНК средней длиной 400-600 нуклеотидов.

В данном исследовании было проведено два эксперимента по пиросеквенированию на GSFLX+ с использованием набора GSTitaniumSequencingKitXL+ и пикотитровальной пластины GSTitaniumPicoTiterPlateKit 70x75. Пластины разбивали на два «больших» региона. В первом и втором эксперименте использовали оба региона. Процедуру проводили в соответствии с протоколом, приведенным

в SequencingMethodManualGSFLX+ (июнь 2013). Полученные результаты секвенирования в каждом эксперименте сохраняли в формате SFF, содержащем информацию по каждой ячейке – флуограмма и данные о качестве прочтения каждого нуклеотида. Данные файлы в этом формате используются в качестве исходных программами по «сборке» геномных последовательностей, в т.ч. с использованием GSDenovoassembler.

**Показатели качества полногеномного секвенирования**

Полногеномное секвенирование проводилось в два этапа: а) пулирование первых 10 изолятов на 2-х регионах сиквенсового чипа; б) пулирование остальных 10 изолятов также на 2-х регионах. На рисунке 2 изображены поверхности сиквенсового чипа после проведения 2-этапного процесса полногеномного секвенирования двадцати изолятов *M.tuberculosis*.



а – первый этап; б– второй этап

**Рис.2.**Изображение поверхностей чипа после проведения 2- этапного процесса полногеномного секвенирования двадцати изолятов *M.tuberculosis*

a– first stage;b– second stage

**Fig. 2.**The image of the chip surface after the 2-step process genome-sequencing of twenty *M.tuberculosis* isolates

Красные точки – это области на чипе, которые не прошли один из трех фильтров оценки качества сиквенса ридов и представляют собой ячейки сбусинками, на которых, как правило, не было обнаружено достаточного количества нуклеотидных участков. Тогда как зеленые точки – это ячейки, успешно прошедшие три этапа фильтров. Данная картина свидетельствует об успешности двухэтапного полногеномного секвенирования. Детекция отдельных изолятов проводилась на этапе анализа сиквенсовых ридов по индексам из двух регионов на чипе. На каждом этапе применялось по 10 индексов группы RL.

Кодировка изолятов, согласно определенной лекарственной устойчивости, представлена в таблице 2. В анализе полных геномов использовали 8 изолятов MDR типа, 3 изолята XDR типа, 1 изолят полирезистентного типа, 1 изолят монорезистентного типа и 7 изолятов чувствительного типа.

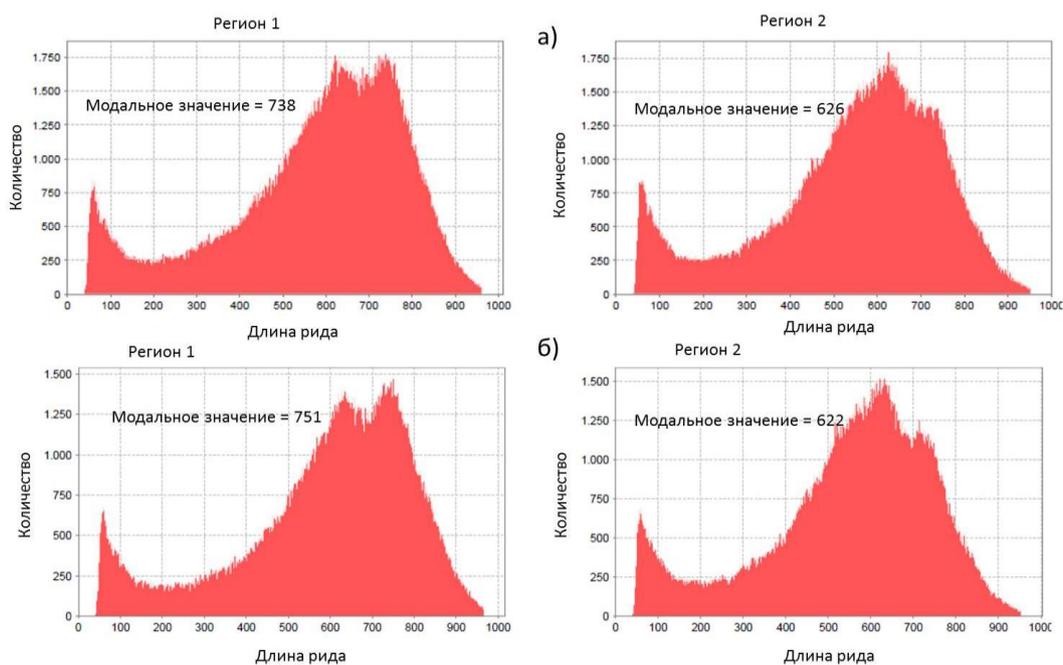
**Таблица 2.** Кодировка клинических изолятов *M.tuberculosis*

**Table 2.** Encoding clinical isolates of *M.tuberculosis*

№	Идентификатор	Лекарственная устойчивость	№	Идентификатор	Лекарственная устойчивость
1	MTB-06-001	MDR	11	MTB-07-006	XDR
2	MTB-06-002	MDR	12	MTB-07-008	монорезистентный
3	MTB-06-003	MDR	13	MTB-07-009	полирезистентный
4	MTB-06-009	MDR	14	MTB-06-004	чувствительный

5	MTB-07-004	MDR	15	MTB-06-006	чувствительный
6	MTB-07-005	MDR	16	MTB-06-007	чувствительный
7	MTB-07-007	MDR	17	MTB-06-008	чувствительный
8	MTB-07-010	MDR	18	MTB-07-001	чувствительный
9	MTB-06-005	XDR	19	MTB-07-002	чувствительный
10	MTB-06-010	XDR	20	MTB-07-003	чувствительный

На рисунке 3 графически представлено общее распределение количества ридов относительно их длины, с модальным значением. Для двух этапов двух регионов модальные значения сиквенсовых ридов составили от 622 до 751 нуклеотидов. В целом, характерно преобладание большого числа длинных ридов (около 600 нуклеотидов).



а – первый этап; б – второй этап

**Рис.3.** Распределение сиквенсовых ридов для двадцати полных геномов *M.tuberculosis*

а – firststage; б – second stage

**Fig.3.** The distribution of sequences reads for twenty complete genomes of *M.tuberculosis*

Длинные риды – отличительная особенность платформы секвенирования нового поколения Roche 454, что является большим преимуществом, в отличие от платформы секвенирования нового поколения Illumina (длина рида не более 100 нуклеотидов). Более длинные риды облегчают и улучшают проведение дальнейшего биоинформатического анализа, повышают качество *de novo* сборки геномов и аннотацию полных геномов.

#### ***De novo* сборка полных геномов *M.tuberculosis***

*De novo* сборка полных геномов проводилась с использованием *NEWBLERdenovoassembler* (454 Life Sciences, Branford, CT) со сходными параметрами для каждого изолята.

В таблице 3 представлены основные статистические показатели сиквенсовых ридов и сборки полных геномов для двадцати изолятов. Как видно из таблицы 4, количество секвенированных нуклеотидов варьирует от 31 млн. до 121,5 млн. с относительным покрытием на геном от 7X до 27,6X и средним показателем 15,8X.

**Таблица 3.** Основные показатели качества сборки полных геномов *M.tuberculosis***Table 3.** The main indicators of quality of DNA sequence assembly of complete genomes *M.tuberculosis*

Изолят	Общее количество нуклеотидных оснований	Общее количество сиквенсовых ридов	Относительное покрытие на геном	Показатель Q40, %
МТВ-06-001	76320315	138102	17,3	99,64
МТВ-06-002	121537822	217003	27,6	99,66
МТВ-06-003	51063921	94214	11,6	99,1
МТВ-06-004	70459742	129815	16,0	99,6
МТВ-06-005	78654146	137386	17,8	99,64
МТВ-06-006	45570743	89254	10,3	98,8
МТВ-06-007	56439724	102358	12,8	99,34
МТВ-06-008	65752951	123020	14,9	99,35
МТВ-06-009	107732099	199130	24,4	99,71
МТВ-06-010	90432039	160624	20,5	99,71
МТВ-07-001	87837690	153839	19,9	99,65
МТВ-07-002	115787490	202140	26,2	99,75
МТВ-07-003	50410022	89480	11,4	98,99
МТВ-07-004	82683221	148752	18,7	99,66
МТВ-07-005	31072505	57404	7,0	96,26
МТВ-07-006	44530906	86279	10,1	98,73
МТВ-07-007	44315584	79028	10,0	98,68
МТВ-07-008	35035785	67232	7,9	97,32
МТВ-07-009	86068944	153515	19,5	99,61
МТВ-07-010	53676819	96850	12,2	99,22
Средняя величина	69769123,4	126271,25	15,81516883	99,121

В зависимости от конкретных целей исследований, чем больше величина относительного покрытия, тем лучше качество секвенирования и выше вероятность обнаружить определенный полиморфизм.

В целом величина покрытия зависит от конечной производительности сиквенсовой платформы. Большая производительность платформы позволяет секвенировать большее количество образцов с максимальным покрытием за один запуск. В нашем случае, относительные показатели покрытия являются достаточно высокими и достаточными для дальнейшей загрузки геномов в международные базы данных и научной публикации.

Параметр Q40 рассчитывался для оценки точности детекции нуклеотида и связан с увеличением числа ошибок на концевых участках сиквенсовых ридов. Q40 эквивалентен вероятности неточной детекции нуклеотида один раз на 10 000, что соответствует 99,99% точности. Исходя из полученных нами данных, средняя величина составила 99,1% сиквенсовых ридов, соответствуют показателю точности Q40, что свидетельствует о высоком качестве проведенного секвенирования.

#### **Картирование и выравнивание на референсный геном *M.tuberculosis***

Для проведения сравнительного анализа полных геномов 20 изолятов был использован референсный штамм *M. tuberculosis*H37Rv из международной базы данных GenBank, NationalCenterforBiotechnologyInformation ([http://www.ncbi.nlm.nih.gov/assembly/GCF\\_000195955.2](http://www.ncbi.nlm.nih.gov/assembly/GCF_000195955.2)). Референсный штамм *M.tuberculosis*H37Rv является наиболее изученным, аннотированным и охарактеризованным, считается «золотым стандартом» для сравнения.

Выравнивание сиквенсовых ридов проводилось отдельно для каждого из двадцати изолятов. В таблице 4 представлены основные статистические показатели выравнивания сиквенсовых ридов 20 изолятов на референсный геном *M.tuberculosis*H37Rv.

**Таблица 4.** Основные статистические показатели полногеномного выравнивания на референсный геном H37Rv*M.tuberculosis*

**Table 4.** Basic statistical indicators of genome alignment to the reference genome H37Rv*M.tuberculosis*

Изолят	Общее количество контигов	Количество оснований в контигах	Максимальная длина контига	Количество картированных ридов в %
МТВ-06-001	132	4333752	281952	98,05
МТВ-06-002	130	4344004	303877	98,5
МТВ-06-003	148	4333561	222565	97,56
МТВ-06-004	135	4337566	303556	97,43
МТВ-06-005	135	4321916	303356	98,99
МТВ-06-006	157	4347096	200344	95,68
МТВ-06-007	153	4329021	253875	98,31
МТВ-06-008	156	4357278	215250	96,74
МТВ-06-009	133	4337388	276710	98,09
МТВ-06-010	131	4331203	303334	98,64
МТВ-07-001	123	4329325	303273	98,83
МТВ-07-002	121	4336934	285682	98,77
МТВ-07-003	173	4332878	250629	98,51
МТВ-07-004	122	4334359	214743	97,88
МТВ-07-005	337	4318793	141748	97,36
МТВ-07-006	142	4340700	250625	96,07
МТВ-07-007	167	4321411	218004	98,69
МТВ-07-008	237	4325477	108480	96,15
МТВ-07-009	127	4335715	303961	98,23
МТВ-07-010	141	4339554	289314	98,35
Ср.величина	155	4334396,55	251563,9	97,8415

В среднем, по всем изолятам, 97,8% сиквенсовыхридов было картировано на референсный штамм, что в нуклеотидном выражении составило 4334396 оснований. Также в таблице 5 приведены статистические показатели для контигов. Контиги представляют собой участки нуклеотидной последовательности, собранные на основе коротких ридов в процессе *denovo* сборки полных геномов.

## ВЫВОДЫ

Основным направлением данной работы является обоснование и разработка методологических подходов к проведению секвенирования полного генома *M.tuberculosis* с использованием современных подходов и инструментов секвенирования нового поколения.

В данной работе была описана методология проведения полногеномного секвенирования и биоинформатического анализа двадцати изолятов *M.tuberculosis* на высокопроизводительной платформе секвенирования нового поколения Roche 454 GS FLX+ методом «shotgun» секвенирования. Результаты сборки двух геномов *M.tuberculosis* были загружены в базу данных NCBI GenBank и доступны для публичного доступа под номерами AZBA00000000, AZAZ00000000, остальные геномы доступны по запросу.

Следующим этапом исследования будет изучение генетических маркеров лекарственной устойчивости *M.tuberculosis*, а также проведение сравнительного биоинформатического анализа между основными группами штаммов *M.tuberculosis* с различным профилем лекарственной чувствительности.

## Финансирование

Исследование проведено в рамках гранта «Создание предпосылок персонализированного подхода в диагностике и лечении туберкулеза на основе полногеномного секвенирования *M.tuberculosis*» по бюджетной программе МОН РК «Грантовое финансирование научных исследований» на 2013-2015 гг.

## ЛИТЕРАТУРА

1. World Health Organisation. Global tuberculosis report (2015) 20<sup>th</sup> edition. Available at: [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/) (accessed 10 June 2016).
2. Walzl G., Ronacher K., DjobaSiawaya J.F., Dockrell H.M. Biomarkers for TB treatment response: challenges and future strategies // J Infect.– 2008.–Vol.57. – P.103-109.
3. Wallis R.S., Pai M., Menzies D. et al. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice // Lancet.– 2010. – Vol.375. – P.1920-1937.
4. Telenti A., Imboden P., Marchesi F., Schmidheini T., Bodmer T. Direct, automated detection of rifampin-resistant *Mycobacterium tuberculosis* by polymerase chain reaction and single-strand conformation polymorphism analysis // Antimicrob. Agents Chemother.– 1993. – Vol.37.–P.2054-2058.
5. Zhang Y., Heym B., Allen B., Young D., Cole S. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis* // Nature.– 1992. – Vol.358.– P.591-593.
6. Takiff H.E., Salazar L., Guerrero C. et al. Cloning and nucleotide sequence of *M.tuberculosis* *gyrA* and *gyrB* genes and detection of quinolone resistance mutations // Antimicrob Agents Chemother.– 1994. – Vol.38. – P.773-780.
7. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors // PNAS.– 1977.– Vol.74 (12).– P. 5463-5467.
8. Voelkerding K.V., Dames S.A., Durtschi J.D. Next-generation sequencing: from basic research to diagnostics // Clin. Chem.– 2009. – Vol.55 (4).– P.641-658.
9. Moorthie S., Mattocks C.J., Wright C.F. Review of massively parallel DNA sequencing technologies // Hugo J.– 2011.– Vol.5.– P.1-12.
10. Koser C.U., Ellington M.J., Cartwright E.J.P. et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology // PLoS Pathogen.– 2012.– Vol.8.
11. Harris S.R., Cartwright E.J.P., Török M.E. et al. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study // Lancet Infect Dis.– 2013. – Vol.13.– P. 130-136.
12. Nakamura S., Maeda N., Miron I.M. et al. Metagenomic diagnosis of bacterial infections // Emerg. Infect. Dis.– 2008.– Vol.14.– P.1784-1786.
13. Köser C.U., Holden M.T., Ellington M.J. et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak // N. Engl. J. Med.– 2012.– Vol.366.– P.2267-2275.
14. Filliol I., Motiwala A.S., Cavatore M. et al. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set // J Bacteriol.– 2006.– Vol.188 (2).– P.759-772.
15. Cole S.T., Brosch R., Parkhill J. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence // Nature.– 1998.– Vol.393.– P. 537-544.
16. Fleischmann R.D., Alland D., Eisen J.A. et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains // J. Bacteriol.– 2002. – Vol.184.– P. 5479-5490.
17. Glynn J.R., Whiteley J., Bifani P.J., Kremer K. and van Soolingen D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review // Emerg Infect. Dis.– 2002.– Vol.8.– P.843-849.
18. Van Soolingen D., Qian L., de Haas P.E. et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia // J. Clin. Microbiol.– 1995.– Vol. 33.–P.3234-3238.
19. Kozhamkulov U., Akhmetova A., Rakhimova S. et al. Molecular characterization of rifampicin- and isoniazid-resistant *Mycobacterium tuberculosis* strains isolated in Kazakhstan // Japanese Journal of Infectious Disease.– 2011.– Vol.64.– P.253-255.
20. Akhmetova A., Kozhamkulov U., Bismilda V. et al. Mutations in the *pncA* and *rpsA* genes among 77 *Mycobacterium tuberculosis* isolates in Kazakhstan // International Journal of Tuberculosis and Lung Diseases.– 2015. – T.19 (2).– P.179-184.
21. Ibrayeva A., Kozhamkulov U., Raiymbek D. et al. Molecular epidemiology of *Mycobacterium tuberculosis* strains circulating in the penitentiary system of Kazakhstan // International Journal Tuberculosis and Lung Diseases.– 2014. – Vol.18(3).– P.298-301.
22. Skiba Y., Mokrousov I., Ismagulova G. et al. Molecular snapshot of *Mycobacterium tuberculosis* population in Kazakhstan: A country-wide study // Tuberculosis.– 2015.– Vol.95 (5).– P.538-546.
23. Kairov U., Kozhamkulov U., Molkenov A., Rakhimova S. et al. Draft Genome Sequences of Two Clinical Isolates of *Mycobacterium tuberculosis* from Sputum of Kazakh Patients // Genome announcements.– 2015.– Vol.3 (3).
24. Ilin A.I., Kulmanov M.E., Korotetskiy I.S. et al. Complete Genome Sequence of Multidrug-Resistant Clinical Isolate *Mycobacterium tuberculosis* 187.0, Used To Study the Effect of Drug Susceptibility Reversion by the New Medicinal Drug FS-1 // Genome announcements.– 2015.– Vol.3(6).
25. Van Soolingen D., de Haas P.E., Hermans P.W., van Embden J.D. DNA fingerprinting of *Mycobacterium tuberculosis* // Methods Enzymol.– 1994.– Vol.235.– P. 196-205.

## REFERENCES

1. World Health Organisation. Global tuberculosis report (2015) 20<sup>th</sup> edition. Available at: [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/) (accessed 10 June 2016).
2. Walzl G., Ronacher K., DjobaSiawaya J.F., Dockrell H.M. Biomarkers for TB treatment response: challenges and future strategies. *J Infect.*, 2008, vol.57, pp.103-109. PMID:18649943.
3. Wallis R.S., Pai M., Menzies D.et al. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice.*Lancet.*,2010,vol.375, pp. 1920-1937.PMID: 20488517.
4. Telenti A., Imboden P., Marchesi F., Schmidheini T., Bodmer T. Direct, automated detection of rifampin-resistant *Mycobacterium tuberculosis* by polymerase chain reaction and single-strand conformation polymorphism analysis.*Antimicrob. Agents Chemother.*, 1993,vol.37, pp. 2054-2058.PMID: 8257122.
5. Zhang Y., Heym B., Allen B., Young D., Cole S. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*.*Nature*, 1992,vol.358, pp. 591-593.PMID:1501713.
6. TakiffH.E., Salazar L., Guerrero C.et al. Cloning and nucleotide sequence of *M.tuberculosisgyrA* and *gyrB* genes and detection of quinolone resistance mutations. *Antimicrob Agents Chemother.*, 1994,vol.38, pp. 773-780.PMID:8031045.
7. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors.*PNAS*, 1977,vol.74 (12),pp. 5463-5467.PMID:271968.
8. Voelkerding K.V., Dames S.A., Durtschi J.D. Next-generation sequencing: from basic research to diagnostics.*Clin. Chem.*, 2009,vol.55 (4),pp.641-658.PMID:19246620.
9. Moorthie S., Mattocks C.J., Wright C.F. Review of massively parallel DNA sequencing technologies. *Hugo J.*, 2011, vol.5, pp. 1-12. PMID:23205160.
10. Koser C.U., Ellington M.J., Cartwright E.J.P. et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathogen*, 2012, vol.8, e1002824. PMID:22876174.
11. Harris S.R., Cartwright E.J.P., Török M.E. et al. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis.*, 2013,vol.13, pp. 130-136. PMID:23158674.
12. Nakamura S., Maeda N., Miron I.M. et al. Metagenomic diagnosis of bacterial infections. *Emerg. Infect. Dis.*, 2008, vol.14,pp.1784-1786. PMID:18976571.
13. Köser C.U., Holden M.T., Ellington M.J. et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.*, 2012, vol. 366,pp. 2267-2275. PMID:22693998.
14. Filliol I., Motiwala A.S., Cavatore M. et al. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol.*, 2006, vol.188 (2),pp. 759-772.PMID:16385065.
15. Cole S.T., Brosch R., Parkhill J. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998,vol.393, pp. 537-544. PMID:9634230.
16. Fleischmann R.D., Alland D., Eisen J.A. et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.*, 2002, vol.184, pp. 5479-5490. PMID:12218036.
17. Glynn J.R., Whiteley J., Bifani P.J., Kremer K. and van Soolingen D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *EmergInfect. Dis.*, 2002,vol.8, pp. 843-849. PMID:12141971.
18. Van Soolingen D., Qian L., de Haas P.E. et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. *J. Clin. Microbiol.*, 1995, vol.33,pp. 3234-3238. PMID:8586708.
19. Kozhamkulov U., Akhmetova A., Rakhimova S. et al. Molecular characterization of rifampicin- and isoniazid-resistant *Mycobacterium tuberculosis* strains isolated in Kazakhstan. *Japanese Journal of Infectious Disease.*,2011,vol.64, pp. 253-255. PMID:21617314.
20. Akhmetova A., Kozhamkulov U., Bismilda V. et al. Mutations in the *pncA* and *rpsA* genes among 77 *Mycobacterium tuberculosis* isolates in Kazakhstan. *International Journal of Tuberculosis and Lung Diseases*, 2015, vol.19 (2), pp. 179-184. PMID: 25574916.
21. IbrayevaA., Kozhamkulov U., Raiymbek D.et al. Molecular epidemiology of *Mycobacterium tuberculosis* strains circulating in the penitentiary system of Kazakhstan. *International Journal Tuberculosis and Lung Diseases*, 2014, vol.18(3), pp. 298-301. PMID: 24670565.
22. Skiba Y., Mokrousov I., Ismagulova G. et al. Molecular snapshot of *Mycobacterium tuberculosis* population in Kazakhstan: A country-wide study. *Tuberculosis*, 2015, vol. 95 (5), pp. 538-546. PMID: 26076582.
23. Kairov U.,Kozhamkulov U., Molkenov A., RakhimovaS. et al. Draft Genome Sequences of Two Clinical Isolates of *Mycobacterium tuberculosis* from Sputum of Kazakh Patients.*Genome announcements*, 2015, vol.3 (3), e00466-15. doi:10.1128/genomeA.00466-15. PMID: 25977436.
24. Ilin A.I.,Kulmanov M.E.,Korotetskiyet I.S. et al. Complete Genome Sequence of Multidrug-Resistant Clinical Isolate *Mycobacterium tuberculosis* 187.0, Used To Study the Effect of Drug Susceptibility Reversion by the New Medicinal Drug FS-1.*Genome announcements*, 2015, vol.3(6), e01272-15.doi:10.1128/genomeA.01272-15. PMID: 26543112.

## ДӘРІЛІК ТӨЗІМДІЛІКТІҢ ӘРТҮРЛІ ПРОФИЛЬДЕРІ БАР *M. TUBERCULOSIS* КЛИНИКАЛЫҚ ИЗОЛЯТТАРЫНЫҢ ТОЛЫҚГЕНОМДЫҚ СЕКВЕНИРЛЕНУІ

Қожамқұлов Ұ.Ә., Қайров Ұ.Е., Ережепов Д.Ә.,  
Ахметова А.Ж., Молкенов А.Б., Ақылжанова А.Р.

Назарбаев Университеті, «National Laboratory Astana» Жеке Мекемесі  
Қабанбай батыр даңғ., 53, Астана, 010000, Қазақстан  
ulan.kozhamkulov@nu.edu.kz

### ТҮЙІН

Туберкулез көптеген елдерде ауру-сырқаудың және өлімнің негізгі себепкері және Қазақстан мен дүниежүзінің денсаулық сақтаудың маңызды мәселесі болып қалуда. Осы аурумек күресудің альтернативті әрі үмітті жолы ретінде диагностиканың генетикалық әдістерін қолдану, мысалы, туберкулезге қарсы қолданылатын препараттарға төзімділікті набықтайтын нүктелік мутацияларды немесе басқа генетикалық детерминанттарды анықтау мен *M. tuberculosis* бактерияларының генотипирленуін қарастыруға болады.

Толықгеномды секвенирлеу немесе жаңа буынды секвенирлеу (next generation sequencing – NGS) соңғы онжылдықта өте қарқынды түрде даму ықпалын алды және биомедицинаның көптеген салаларында қолданылады. NGS-тің жаңа әдістері әртүрлі инфекциялық және жұқпалы аурулардың қоздырғыштарының толықгеномдық тізбектерін, олардың молекулалық және генетикалық ерекшеліктерін зерттеуге, микробиология, вирусология мен эпидемиологияның фундаменталдық және практикалық негіздерін дамытуға мүмкіншілік береді.

Осы жұмыста дәрілік төзімділіктің әртүрлі профильдері бар *M. tuberculosis*-тің 20 клиникалық изоляттарының толықгеномдық тізбегін Roche454GS FLX+Titanium платформасының негізінде жоғары өнімді «shotgun» секвенирленуінің қадамдары сипатталған. Толықгеномдық секвенирлеу сапасының көрсеткіштері анықталды, *M.tuberculosis*-тің толық геномдарының *de novo* құрастырылуы, сонымен қатар карталау мен халықаралық GenBank дерекқорындағы *M.tuberculosis*-тің H37Rv референсті геномына теңестірілу іске асырылды. *M.tuberculosis*-тің екі геномның құрастыру нәтижелері халықаралық NCBI GenBank дерекқорына AZBA00000000 және AZAZ00000000 нөмірлері астында енгізілді және көпшілік үшін қолжетімді. Қалғандары сұраныс бойынша алына алады.

Негізгі сөздер: толықгеномдық секвенирлеу, туберкулез, туберкулез микобактериялары, ДНҚ, пиросеквенирлеу, геном.