DOI: 10.11134/btp.4.2026.1

УДК 579.873.21:004.42

РАЗРАБОТКА И АПРОБАЦИЯ АВТОНОМНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ РАБОТЫ С НОМЕНКЛАТУРОЙ MLVA MTBC159 ПО ДАННЫМ MIRU-VNTR-ТИПИРОВАНИЯ И ПОЛНОГЕНОМНОГО СЕКВЕНИРОВАНИЯ MYCOBACTERIUM TUBERCULOSIS

Найзабаева Д.А.[1,2], **Скиба Ю.А.**[1,2] *

[1]*РГП на ПХВ «Институт молекулярной биологии и биохимии им. М.А. Айтхожина», ул. Досмухамедова, 86, 050012, Алматы, Республика Казахстан*
[2]*Филиал ТОО «Национальный центр биотехнологии» в городе Алматы, ул. Жахангер, 14, Алматы, 050054, Казахстан*
**Автор-корреспондент: yuriy.skiba@gmail.com*

АННОТАЦИЯ

Молекулярная эпидемиология туберкулеза переходит от типирования по вариабельным тандемным повторам к полногеномному секвенированию (WGS), однако сведения о распространении генотипов *Mycobacterium tuberculosis* за прошедшие десятилетия записаны в номенклатуре MLVA MtbC159, а по коротким прочтениям профиль MIRU-VNTR определяется ненадежно. Номера этой номенклатуры присваивал единственный веб-сервис MIRU-VNTRplus (https://www.miru-vntrplus.org), интерактивные функции которого в 2026 г. стали недоступны. Цель работы состояла в том, чтобы сохранить связь новых данных с накопленными, перенеся функции сервиса на сторону лаборатории. Разработана автономная программа MTBC Lab Toolkit для Windows, которая присваивает номера по правилу точного совпадения панелей из 15 и 9 локусов, обрабатывает наборы профилей, указывает ближайшие зарегистрированные типы и сводит профили MIRU-VNTR с результатами полногеномного секвенирования. Справочные данные программы составляют 3 906 профилей с номерами из 152 публикаций и 23 560 типов MtbC15 официальной таблицы сервиса. Из 400 напечатанных в публикациях номеров программа воспроизвела все 392, согласующиеся с официальной таблицей, а остальные 8 оказались ошибками публикаций; неверного номера программа не выдала ни разу. Значение отсутствующего локуса не угадывается: программа ищет известные профили, совпадающие по остальным 23 локусам, и выдает ближайший номер, только если такой профиль один. Профиль с измененным локусом ни разу не получил прежний номер. Импорт результатов TB-Profiler и Mykrobe проверен на их выходных файлах, а пакет из 3 293 профилей обрабатывается за 7,3 с. Программа позволяет сопоставлять новые данные, в том числе профили, выведенные из длинных прочтений, с накопленными и опубликованными результатами MIRU-VNTR-типирования без обращения к внешнему серверу.

Ключевые слова: *Mycobacterium tuberculosis*; MIRU-VNTR; MLVA MtbC159; номенклатура генотипов; молекулярная эпидемиология; геномная эпидемиология; программное обеспечение; полногеномное секвенирование; Казахстан.

ВВЕДЕНИЕ

Типирование по 24 локусам, содержащим вариабельное число тандемных повторов микобактериальных диспергированных повторяющихся единиц (MIRU-VNTR), долгое время оставалось основным методом молекулярно-эпидемиологического надзора за туберкулезом. Набор из 24 локусов и входящая в него панель из 15 наиболее дискриминирующих локусов предложены в 2006 г. P. Supply с соавторами по итогам многоцентрового исследования: 15-локусная система – как стандарт рутинного эпидемиологического типирования, 24-локусная – как инструмент высокого разрешения для филогенетических исследований [1]; впоследствии метод стал международным стандартом [2]. Сопоставимость профилей между лабораториями обеспечивает номенклатура MLVA MtbC159 (MLVA – мультилокусный анализ вариабельного числа тандемных повторов): генотипу присваивается составной номер, первая часть которого отвечает типу по 15 локусам, вторая – типу по остальным 9 локусам. Номенклатура реализована как функция веб-сервиса MIRU-VNTRplus (https://www.miru-vntrplus.org), построенного на референсной коллекции из 186 штаммов основных линий комплекса *Mycobacterium tuberculosis*; номера присваиваются и хранятся централизованно [2, 3].

В Казахстане MIRU-VNTR-типирование применяется более десяти лет. На его основе дана общестрановая характеристика популяции возбудителя [4] и описан изолят генотипа RD-Rio с номером MtbC159 16113830 [5]. Позднее была оценена распространенность центрально-азиатского кластера 9432 генотипа Beijing среди изолятов с множественной лекарственной устойчивостью [6]. В последние годы популяцию возбудителя в Казахстане изучают и методом полногеномного секвенирования (WGS) [7], поэтому преемственность между обозначениями генотипов, накопленными по MIRU-VNTR, и данными секвенирования становится практической задачей.

Номера MtbC159 служат общепринятыми именами клонов возбудителя и в международных работах. Ими обозначены вспышка туберкулеза в учебном заведении Норвегии (кластер 10287189) [8] и клады генотипа Beijing, выявленные полногеномным секвенированием в Португалии и Гвинее-Бисау (кластеры 10032 и 9432) [9]; кластеры 107132 и 1471715 стали предметом отдельных филогеномных исследований [10, 11]. Поэтому сохранение единых обозначений важно не только для внутренних задач лаборатории.

В последнее десятилетие молекулярная эпидемиология туберкулеза переходит от типирования по тандемным повторам к полногеномному секвенированию: родство изолятов, установленное по однонуклеотидным полиморфизмам (SNP), точнее отражает передачу возбудителя, чем совпадение профилей MIRU-VNTR [12]. Од-

DOI: 10.11134/btp.4.2026.1

нако наиболее распространенные платформы WGS дают короткие прочтения, не перекрывающие протяженные повторы целиком, и число копий в части локусов по таким данным определяется ненадежно [13, 14]. Между тем сведения о распространении генотипов за прошедшие десятилетия записаны именно в номенклатуре MtbC159, а изоляты, исследованные 20–30 лет назад, зачастую утрачены и секвенированы быть не могут. Без сопоставления новых обозначений со старыми ретроспективный анализ распространения клонов, в том числе по литературным данным, становится невозможным [15].

Слабое место сложившейся практики – зависимость номенклатуры от единственного внешнего сервиса. По состоянию на октябрь 2026 г. интерактивные функции веб-сервиса MIRU-VNTRplus недоступны, и вместе с ними лаборатории лишаются сразу нескольких возможностей: присвоения номера по профилю, поиска ближайших штаммов референсной коллекции, обработки собственного набора изолятов и хранения своих данных. Эти возможности в разной степени воспроизводимы без централизованной базы.

Существующие программные средства решают смежные задачи. MIRU-profiler рассчитывает 24-локусный профиль по собранному геному [13], а MIREader – по длинным прочтениям секвенирования Oxford Nanopore и PacBio [14]; процедура на основе BioNumerics определяет аллели по результатам капиллярного электрофореза продуктов амплификации [16]. Все эти задачи предшествуют обозначению генотипа. Общедоступные средства, которые присваивали бы номера этой номенклатуры без обращения к серверу MIRU-VNTRplus, нам не известны, а опубликованные таблицы профилей с номерами рассеяны по отдельным работам и приводятся в разном порядке локусов.

Цель работы – сохранить возможность сопоставлять новые геномные данные с накопленными результатами MIRU-VNTR-типирования, перенеся на сторону лаборатории те функции веб-сервиса, которые воспроизводимы без централизованной базы. Они собраны в программе MTBC Lab Toolkit (далее – программа), работа которой оценена на опубликованных данных.

МАТЕРИАЛЫ И МЕТОДЫ

Правило номенклатуры. Номер MLVA MtbC159 состоит из двух частей, разделенных дефисом. Тип MtbC15 определяется пятнадцатью дискриминирующими локусами (424, 577, 580, 802, 960, 1644, 1955, 2163b, 2165, 2401, 2996, 3192, 3690, 4052, 4156), тип MtbC9 – остальными девятью (154, 2059, 2347, 2461, 2531, 2687, 3007, 3171, 4348); состав панелей принят по таблице 1 работы [1]. Тип каждой части присваивается только при полном совпадении всех локусов панели с зарегистрированным паттерном: отклонение хотя бы на один повтор означает новый паттерн, а не близкий тип. Это же правило служит для контроля данных: паттерн, которому отвечает более одного номера типа, и тип, которому отвечает более одного паттерна, считаются ошибкой и подлежат разбору.

Справочные данные. Источником данных служили публикации, в которых полные 24-локусные профили приведены вместе с номерами MtbC159 или в виде, пригодном для сверки с номенклатурой. Публикации отбирались поиском по полным текстам в Europe PMC (https://europepmc.org) и OpenAlex (https://openalex.org), по дополнительным материалам в Figshare (https://figshare.com) и по журналам России и Казахстана в КиберЛенинке (https://cyberleninka.ru); запросы включали разные написания названия номенклатуры и наиболее частые номера в сочетании со словами MIRU и VNTR. Профили извлекались из таблиц статей и дополнительных материалов, из записей в виде сплошной строки цифр в тексте и из рисунков; для одной работы (изоляты Иркутской области) таблица профилей с номерами получена от ее автора. Рисунки читались дважды независимо, по строкам и по столбцам; ячейки, в которых чтения расходились, проверялись третьим, независимым чтением, а результат сверялся с номерами, напечатанными на том же рисунке. Порядок локусов в каждом источнике устанавливался не по подписи к таблице, а сверкой с уже проверенными источниками и с официальной таблицей типов; выбранный порядок подтверждался на части строк, не участвовавших в подборе. Источники, данные которых не сходились ни с номенклатурой, ни между собой, исключались. В справочные данные программы входят профили, у которых номер напечатан полностью, профили референсных штаммов MIRU-VNTRplus [3] и профили из диссертационной работы по Кении [17]; строки, нарушающие правило «один паттерн – один тип», для определения типов не используются.

Официальная таблица типов MtbC15. Веб-сервис MIRU-VNTRplus содержал перечень всех зарегистрированных типов MtbC15 с их 15-локусными паттернами. Эта страница сохранилась в архиве Internet Archive (https://archive.org); использован снимок от 21 июня 2022 г. [18]. В справочные данные включены 23 560 типов, у которых значения всех 15 локусов заданы целыми числами. Таблицы типов MtbC9 в архиве нет, поэтому тип MtbC9 определяется только по паттернам, встреченным в публикациях. Напечатанный в публикации номер MtbC15 сверялся с официальной таблицей.

Алгоритм присвоения номера. Сначала программа проверяет точное совпадение всех 24 локусов профиля с профилем из справочных данных. При отсутствии такого совпадения 15-локусный и 9-локусный паттерны отыскиваются раздельно: если известны оба типа, из них составляется номер с указанием, встречается ли такое сочетание среди известных профилей; если известна лишь одна часть, выдается половинный номер с вопросительным знаком вместо неизвестной части – обозначение, принятое в лабораторной практике. Если значение какого-либо локуса отсутствует, программа не подставляет вместо него предполагаемую цифру: она отбирает в справочных данных все профили, совпадающие с запросом по всем имеющимся локусам. Если такой профиль один, выдается его номер с пометкой о том, что он получен по неполному профилю; если их несколько, номер составляется по частям, как описано выше, и часть, которую нельзя установить однозначно, заменяется вопросительным знаком. Дополнительно программа находит ближайшие зарегистрированные профили по числу несовпадающих локусов. Если соответствий нет, профилю

DOI: 10.11134/btp.4.2026.1

присваивается локальный идентификатор, вычисляемый из самого профиля и потому одинаковый при повторном запросе. Аллели нестандартного вида (например, запись «7(+6)» при смешанной культуре) сохраняются с отдельным статусом результата.

Реализация. Программа написана на языке Python 3.11 с использованием библиотек Tkinter, pandas, NumPy и openpyxl, собрана в автономный пакет средствами PyInstaller и распространяется установщиком для Windows 10. Сетевое соединение требуется только для поиска дополнительных материалов публикаций. Графический интерфейс объединяет в одном окне одиночный поиск, пакетную обработку, шаблоны входных таблиц, импорт результатов секвенирования, работу с дополнительными материалами публикаций, работу с лабораторной базой данных и журнал операций; те же действия доступны из командной строки. Входные таблицы CSV, TSV и XLSX принимаются как со столбцом строк профилей, так и с 24 столбцами локусов, в том числе таблицы MIRUReader; заголовки распознаются по номерам локусов и по опубликованным синонимам (MIRU 02, ETR A, QUB-11b, Mtub04 и др. [1]), кодировка и разделитель определяются автоматически. Результат сохраняется в книгу Excel с разбивкой по статусам присвоения. Данные лаборатории хранятся в локальной базе данных: изменения, вносимые при импорте, протоколируются с указанием источника и прежнего значения, а расхождение между имеющимся и поступившим значением не перезаписывается, а выносится на решение пользователя.

Модуль импорта читает результаты программ TB-Profiler [19] и Mykrobe [20]: версию программы и базы мутаций, линию и сублинию, семейство, категорию лекарственной устойчивости, результаты по отдельным препаратам и мутации, показатели качества картирования. Эти сведения записываются рядом с MIRU-профилем изолята. Отдельная функция находит по DOI ссылки на дополнительные материалы публикации и извлекает из таблицы приложения номера доступа к данным секвенирования.

Проверка работы программы. Присвоение номеров проверялось четырьмя прогонами пакетной обработки на профилях, у которых обе части номера напечатаны в публикациях (3 814 различных сочетаний профиля и номера); профили отбирались случайно, с фиксированным начальным значением генератора случайных чисел. В первом прогоне выданный номер сравнивался с напечатанным, а при расхождении напечатанный номер сверялся с официальной таблицей типов MtbC15. Во втором, как при непрочитанном локусе, из того же профиля удалялось значение одного случайно выбранного локуса панели MtbC9. Проверочные профили взяты из тех же публикаций, что и справочные данные, и прогон показывает, достаточно ли оставшихся 23 локусов, чтобы однозначно выделить исходный профиль среди известных: ожидался его номер, если им отвечает только он, и первая часть номера, если несколько; ошибкой считалась бы выдача полного номера, отличного от номера по полному профилю. В третьем у 200 профилей значение одного случайного локуса панели MtbC15 увеличивалось на единицу; ожидалось, что точное совпадение будет утрачено, а исходный тип окажется среди ближайших на расстоянии в один локус. Профили, которые после изменения сами совпадали с известным профилем, и профили, у которых выбранный локус не имел целочисленного значения, из этого прогона исключались. В четвертом подавались профили, отсутствующие в справочных данных, и проверялось, выдается ли локальный идентификатор и совпадает ли он при повторном прогоне. Импорт результатов секвенирования проверялся на выходных файлах программ TB-Profiler 4.0.3 и 6.7.0 и Mykrobe 0.13.0, установленных из Bioconda (https://bioconda.github.io) и запущенных с поставляемыми базами мутаций на тестовом наборе из 16 587 прочтений из репозитория TB-Profiler (https://github.com/jodyphelan/TBProfiler); набор содержит прочтения только генов-мишеней, поэтому проверялись разбор выходных файлов, а не качество предсказаний. Перенос номеров доступа проверялся на таблице дополнительных материалов работы по глобальной коллекции линии 3 [21], внесение опубликованных профилей с контролем повторов – на крупнейшей из разобранных работ [22]. Время обработки измерялось на ноутбуке (Intel Core i71165G7, 16 ГБ оперативной памяти) по трем прогонам пакета из 3 293 профилей. Работу отдельных функций программы дополнительно контролируют автоматические тесты (всего 141).

РЕЗУЛЬТАТЫ

Справочные данные. Из 292 отобранных публикаций данные удалось извлечь из 152. Они дали 47 246 наблюдений, из них 29 080 с полным набором 24 локусов (14 453 различных профилей); для 21 164 наблюдений 15-локусный паттерн отвечал типу официальной таблицы (таблица 1). Справочные данные программы составляют 3 906 профилей с номерами, 3 420 паттернов MtbC15 и 344 паттерна MtbC9. Все паттерны MtbC15 совпали с официальной таблицей и по профилю, и по номеру, противоречий между паттернами и типами не было. Доля напечатанных номеров MtbC15, не подтвержденных официальной таблицей, зависела от вида источника: в таблицах статей

Таблица 1 – Профили, извлеченные из публикаций, по виду источника

Вид источника	Публикаций	Наблюдений	Полных профилей	С номером	Номер не подтвержден, %
Таблицы статей и приложений	110	40 785	25 225	9 340	0,1
Профиль строкой в тексте или в одной ячейке	28	5 440	3 211	248	18,1
Рисунки статей (двойное чтение)	17	1 021	644	639	0,2
Итого	152	47 246	29 080	10 227	0,5

Примечание – одна публикация может содержать источники разного вида, поэтому сумма по строкам превышает число публикаций в строке «Итого». Неподтвержденным считался напечатанный номер MtbC15, расходящийся с официальной таблицей типов.

DOI: 10.11134/btp.4.2026.1

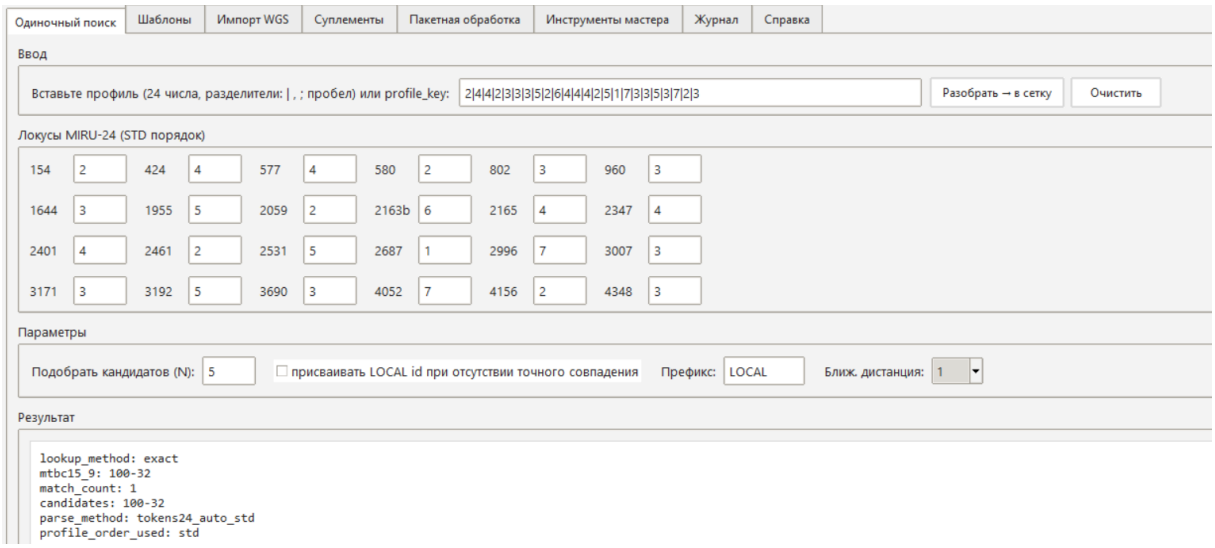


Рисунок 1 – Окно одиночного поиска: профилю по совпадению всех 24 локусов присвоен номер 10032

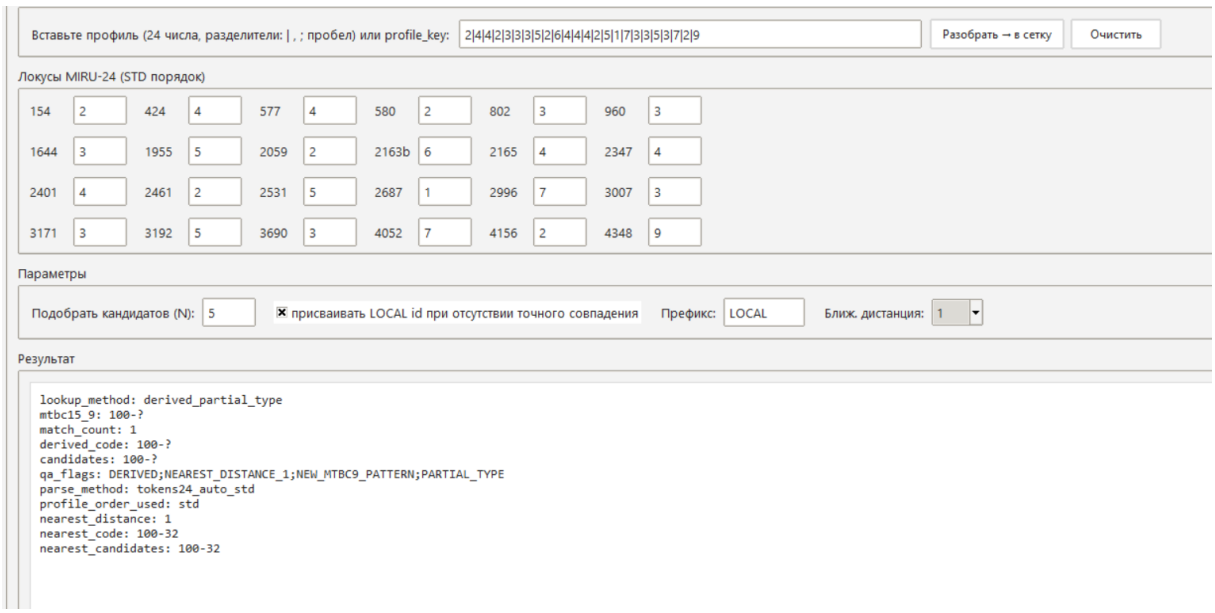


Рисунок 2 – Тот же профиль с измененным значением локуса 4348: точного совпадения нет, выдан половинный номер 100-? с пометкой о впервые встреченном паттерне MtbC9, ближайший зарегистрированный номер 10032 указан на расстоянии в один локус

и дополнительных материалов она составила 0,1 %, в распознанных рисунках – 0,2 %, а в профилях, напечатанных сплошной строкой цифр в тексте, – 18,1 %. Ни один неподтвержденный номер в справочные данные не вошел.

Замещение функций сервиса. Окно программы объединяет все этапы работы в отдельных вкладках. В окне одиночного поиска профиль вводится строкой или по отдельным локусам, а результат содержит статус присвоения, номер, список кандидатов и пометки контроля качества (рисунок 1). Изменение одного локуса панели MtbC9 лишает профиль точного совпадения, и вместо номера выдаются половинный номер и ближайший зарегистрированный тип (рисунок 2). Лаборатория получает не только обозначение, но и его основание.

Соответствие функций веб-сервиса и программы приведено в таблице 2. Воспроизведены присвоение номера, обработка набора профилей, указание ближайших зарегистрированных типов и хранение данных, причем хранение перенесено на сторону лаборатории. Регистрация нового номера не воспроизводится: профиль, отсутствующий в справочных данных, получает локальное обозначение и помечается как подлежащий регистрации. Построение деревьев и матриц расстояний в программу не входит, а сведения о линии и семействе, которые сервис выдавал вместе с номером, замещены импортом результатов секвенирования.

Проверка присвоения номеров. Результаты четырех прогонов приведены в таблице 3. Из 400 профилей с напечатанными номерами выданный номер совпал с напечатанным для 392, то есть для всех профилей, напечатанный номер которых согласуется с официальной таблицей типов MtbC15. В остальных 8 случаях официальной таблице

DOI: 10.11134/btp.4.2026.1

Таблица 2 – Функции веб-сервиса MIRU-VNTRplus и их воспроизведение в программе

Функция веб-сервиса	В программе
Присвоение номера MLVA MtbC159 по 24-локусному профилю	Есть: точное совпадение по панелям 15 и 9 локусов; справочные данные из 3 906 профилей и официальной таблицы типов MtbC15
Регистрация нового номера в единой номенклатуре	Нет и не может быть: номер присваивает держатель номенклатуры. Программа помечает профиль как подлежащий регистрации и дает ему устойчивый локальный идентификатор
Поиск по сходству с референсной коллекцией	Есть в части номенклатуры: ближайшие зарегистрированные типы по числу несовпадающих локусов, с перечнем кандидатов
Работа с набором изолятов пользователя, загрузка файлом	Есть: пакетная обработка таблиц CSV, TSV и XLSX со столбцом профилей или 24 столбцами локусов, результат книгой Excel с разбивкой по статусам
Хранение данных пользователя на сервере	Заменено локальной базой данных лаборатории с протоколом изменений; сеть не нужна
Деревья и матрицы расстояний	Нет: задача решается отдельными средствами филогенетического анализа
Сведения о линии, семействе и устойчивости	Заменено импортом результатов полногеномного секвенирования рядом с MIRU-профилем

Примечание – состояние веб-сервиса на октябрь 2026 г.

противоречит сам напечатанный номер, и программа выдала номер, согласующийся с таблицей, либо только ту часть номера, которую можно установить. Четыре записи принадлежали работе, в которой одно и то же значение первой части номера напечатано при разных 15-локусных паттернах, и программа выдала для них половинный номер. В двух случаях один и тот же профиль напечатан в публикациях под двумя разными номерами, и программа выдала тот из них, который подтверждается официальной таблицей. В одном случае напечатанный номер отличался от официального одной цифрой (14655 вместо 14665), а в последнем напечатанный профиль отличался от профиля номера 106632 в двух локусах, 3192 и 154; значение 5 в локусе 154 встречается в той же работе еще дважды и не встречается у изолятов генотипа Beijing в других источниках.

Без одного локуса оставшимся 23 локусам у 386 из 400 профилей отвечал единственный известный профиль, и программа выдала его номер, во всех случаях совпавший с номером по полному профилю. У 8 профилей оставшиеся локусы совпадали с двумя и более известными профилями, которые различаются как раз удаленным локусом: например, профилю 94387 без локуса 2347 отвечают также профили 9432, 9434 и 94332. В этих случаях программа выдала только первую часть номера, не выбирая между вариантами. Остальные 6 профилей – это описанные выше записи с ошибочными номерами; для них выдана часть номера или соответствие не найдено. Полного номера, отличного от номера по полному профилю, не выдано ни разу. После изменения одного локуса панели MtbC15 ни один из 194 профилей не получил прежний номер. Исходный тип указан среди ближайших, на расстоянии в один локус, для всех профилей с известным типом (191); у остальных 3, записей с ошибочными номерами, исходного типа в справочных данных нет. Все 100 профилей, отсутствующих в справочных данных, получили различные локальные идентификаторы, и повторный прогон дал те же значения.

Импорт результатов секвенирования и дополнительных материалов. На настоящих выходных файлах модуль импорта извлек все нужные поля (таблица 4). Обе версии TB-Profiler, различающиеся структурой выходного файла, определили линию 4, категорию устойчивости RR-TB (устойчивость к рифампицину) и мутацию rpoB p.Ser450Leu; различие в глубине определения сублинии (4.3.4.2 против 4.3.4) объясняется версиями базы му-

Таблица 3 – Проверочные прогоны на профилях из публикаций

Проверка	Профилей	Ожидаемое поведение	Получено
Профиль с напечатанным номером	400	Выдать тот же номер	Совпал у 392, то есть у всех, чей номер согласуется с официальной таблицей; у 8 номер ей противоречит
Тот же профиль без одного локуса MtbC9	400	Один подходящий известный профиль – его номер; несколько – первая часть номера	Один известный профиль у 386, номер тот же, что по полному профилю; несколько у 8, выдана первая часть; записи с ошибочными номерами 6; неверных номеров 0
Один локус MtbC15 изменен на единицу	194	Не выдавать прежний номер и указать исходный тип среди ближайших	Прежний номер не выдан ни разу; исходный тип среди ближайших у 191 из 191, у которых он есть в справочных данных
Профиль, отсутствующий в справочных данных	100	Выдать устойчивый локальный идентификатор	Выдан 100, различных 100, повтор дал те же 100

Примечание – профили отобраны случайно из 3 814 различных сочетаний профиля и номера, напечатанных в публикациях.

DOI: 10.11134/btp.4.2026.1

Таблица 4 – Проверка импорта результатов полногеномного секвенирования на настоящих выходных файлах программ

Программа, версия	Линия / сублиния	Категория устойчивости	Препараты и мутации	Картировано, %
TB-Profiler 4.0.3	lineage4 / 4.3.4.2	RR-TB	rpoB p.Ser450Leu; rifampicin R	98,75
TB-Profiler 6.7.0	lineage4 / 4.3.4	RR-TB	rpoB p.Ser450Leu; rifampicin R, rifapentine R	98,73
Mykrobe 0.13.0	lineage4 (4.10; 4.3.4.2)	Не сообщается	Rifampicin R, остальные препараты S	Нет поля

Примечание – все три программы запускались на одном и том же наборе из 16 587 прочтений.

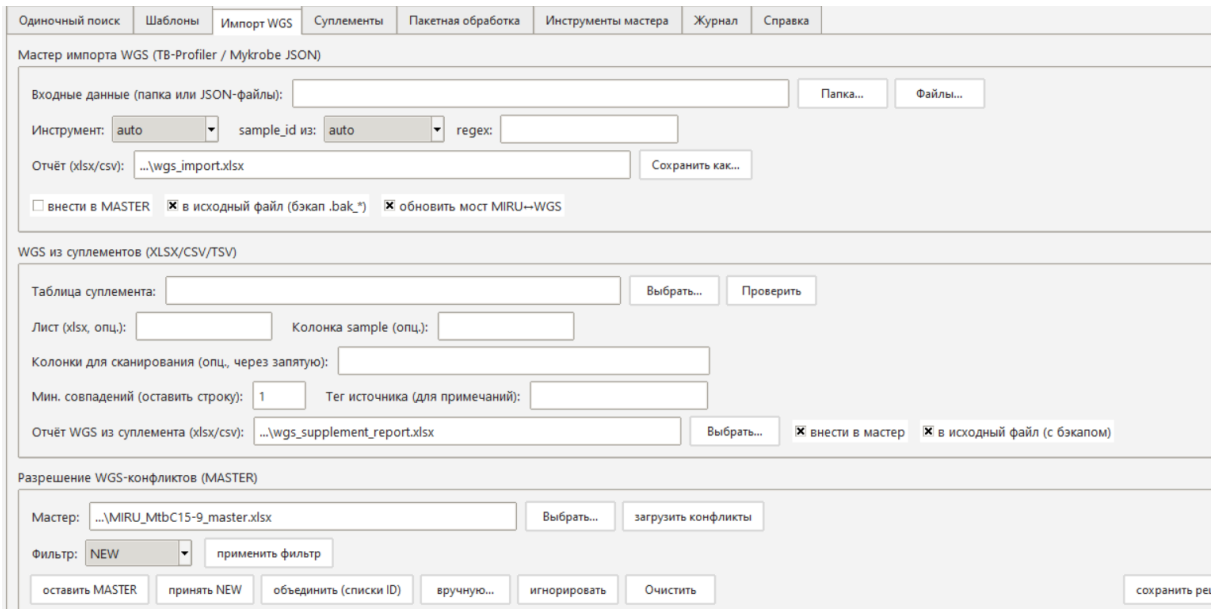


Рисунок 3 – Вкладка импорта результатов полногеномного секвенирования: чтение выходных файлов TB-Profiler и Mykrobe, перенос номеров доступа из дополнительных материалов публикаций и рассмотрение расхождений

таций. Mykrobe по тем же прочтениям определил устойчивость к рифампицину, чувствительность к остальным препаратам, линию 4 и видовую принадлежность.

Из таблицы дополнительных материалов работы по линии 3 [21] номера доступа извлечены для всех 373 изолятов и внесены в базу данных лаборатории с протоколом каждого изменения. Повторное внесение с измененными значениями не перезаписало имеющиеся данные: все 373 расхождения были вынесены на рассмотрение пользователя (рисунок 3). Из крупнейшей разобранной работы [22] в базу внесено 1 495 различных сочетаний профиля и номера; повторное внесение того же файла не добавило ни одной записи.

Время обработки. Пакет из 3 293 профилей обрабатывается за 7,3 с (медиана трех прогонов), из которых около 1 с занимает запуск программы; на один профиль приходится около 2 мс.

ОБСУЖДЕНИЕ

Когда централизованный сервис перестает работать, лаборатория теряет несколько разных возможностей, и они по-разному поддаются восстановлению. Присвоение уже существующего номера восстановимо полностью: правило номенклатуры опубликовано, а соответствие паттернов и типов можно собрать заново из открытых источников. Обработка набора изолятов, хранение данных и поиск ближайших обозначений требуют лишь локальных вычислений. Регистрация нового номера не восстановима в принципе: это действие держателя номенклатуры, и попытка присваивать новые номера локально разрушила бы единство обозначений. Автономное средство поэтому должно воспроизводить функции сервиса ровно в той части, в какой это не создает параллельной номенклатуры.

Главное методическое решение работы – воспроизведение правила точного совпадения панелей вместо поиска по сходству. При поиске по сходству профиль всегда получает какое-то обозначение, и расхождение между профилем и номером остается незамеченным; при требовании полного совпадения каждое такое расхождение сразу проявляется как противоречие. Проверочные прогоны подтвердили оба следствия этого правила. Изменение одного локуса не приводит к выдаче чужого номера как собственного, но сведения о сходстве не теряются: исходный тип указывается среди ближайших. То же правило действует при утрате локуса: номер выдается, только если оставшимся локусам отвечает единственный известный профиль, а при нескольких совпадениях выдается лишь часть номера, поэтому неполный профиль не получает номер наугад.

Надежность опубликованных номеров заметно зави-

DOI: 10.11134/btp.4.2026.1

села от того, как напечатан профиль. В таблице номер и профиль связаны структурой документа, а в сплошной строке цифр – только соседством в тексте, и утрата одной цифры номера внутри самого источника ничем не обнаруживается. Профиль из такой записи извлекать можно, но напечатанный рядом номер следует считать неподтвержденным, пока он не согласован с независимым источником; правило точного совпадения и официальная таблица типов дают такую возможность. Это стоит учитывать при пополнении подобных наборов данных.

По отношению к существующим средствам программа занимает свободную нишу. MIRU-profiler [13], MIRUReader [14] и процедура на основе BioNumerics [16] получают профиль из генома, прочтений или результатов электрофореза, то есть решают задачу, предшествующую обозначению генотипа. Веб-сервис MIRU-VNTRplus [2] присваивал номера, но требовал доступа к сети и централизованной базы; автономное средство переносит эту функцию на сторону лаборатории.

Смысл работы состоит не в поддержании устаревающего метода, а в связи двух периодов молекулярной эпидемиологии. Для изолята, секвенированного по длинным прочтениям (Oxford Nanopore, PacBio), связь устанавливается прямо: профиль, выведенный программой MIRUReader, расходился с лабораторным типированием лишь в 11 из 3024 локусов (0,36 %) [15], и такому профилю программа присваивает номер. По номеру тот же генотип отыскивается в опубликованных работах, и можно проследить, где и когда он встречался, даже если изоляты тех исследований давно недоступны. Это существенно: в разобранных публикациях год выделения указан для 16 947 наблюдений, и две трети из них приходятся на 1990–2000-е годы. По коротким прочтениям прямой путь ненадежен, и связь приходится устанавливать косвенно – через клоны, для которых известны и номер MtbC159, и положение на филогенетическом дереве по SNP. Такие соответствия уже установлены для ряда клонов генотипа Beijing [9–11], а конвейер simpiTB связывает данные секвенирования с международной базой генотипов [23]; чем больше их накоплено, тем увереннее изолят, известный только по SNP, соотносится с генотипом, описанным в литературе по MIRU-VNTR.

Структура выходных файлов TB-Profiler и Mykrobe [19, 20] различается не только между программами, но и между версиями одной программы, поэтому модуль импорта следует проверять на настоящих файлах.

Ограничения работы связаны прежде всего с полнотой справочных данных. Они включают только номера, для которых удалось найти полный профиль в открытых источниках. Для панели MtbC15 это ограничение смягчено официальной таблицей типов, и первая часть номера определяется для любого профиля, 15-локусный паттерн которого был зарегистрирован до июня 2022 г.; для панели MtbC9 подобной таблицы нет, и вторая часть определяется только для паттернов, встреченных в публикациях. Составной номер, обе части которого известны, но сочетание ранее не встречалось, следует рассматривать как предположение, а не как установленное обозначение. Это относится и к номеру, выданному по неполному профилю: он окажется неверным, если изолят несет еще не описанный вариант, отличающийся именно отсутствующим локусом, поэтому такой номер помечается особо. Проверочные прогоны на профилях из тех же публикаций характеризуют согласованность программы, а внешней проверкой служит официальная таблица типов. Естественное развитие работы – пополнение справочных данных из новых публикаций и обмен профилями между лабораториями региона.

ЗАКЛЮЧЕНИЕ

Функции веб-сервиса MIRU-VNTRplus, воспроизводимые без централизованной базы, были перенесены на сторону лаборатории. Разработанное программное обеспечение присваивает номера номенклатуры MLVA MtbC159 по правилу точного совпадения панелей, обрабатывает наборы профилей, указывает ближайшие зарегистрированные типы, хранит данные лаборатории с протоколом изменений и сводит профили MIRU-VNTR с результатами полногеномного секвенирования, не требуя доступа к сети. Регистрация новых номеров сознательно не воспроизводится: такой профиль получает воспроизводимое локальное обозначение.

Справочные данные, собранные из 152 публикаций и официальной таблицы типов MtbC15, позволяют воспроизводить опубликованные номера, а проверочные прогоны подтвердили заявленное поведение программы при полном, неполном, измененном и неизвестном профиле. Тем самым при переходе к полногеномному секвенированию сохраняется возможность соотносить новые данные с накопленными за десятилетия результатами MIRU-VNTR-типирования, в том числе опубликованными, и проводить ретроспективный анализ распространения генотипов. Программа может служить основой национального реестра генотипов возбудителя туберкулеза и средством обмена профилями между лабораториями региона.

ДОСТУПНОСТЬ ПРОГРАММЫ

Программа MTBC Lab Toolkit (установщик для Windows 10 вместе со справочными данными) предоставляется авторами по запросу.

БЛАГОДАРНОСТИ

Авторы благодарят создателей веб-сервиса MIRU-VNTRplus (https://www.miru-vntrplus.org) и номенклатуры MLVA MtbC159, на протяжении многих лет обеспечивавших единое обозначение генотипов *M. tuberculosis*, и авторов исследований, опубликовавших профили MIRU-VNTR вместе с номерами типов: на основе этих публикаций собраны справочные данные программы. Авторы признательны д-ру мед. наук С. Н. Ждановой (ФГБНУ «Научный центр проблем здоровья семьи и репродукции человека», Иркутск) за предоставленную таблицу профилей MIRU-VNTR изолятов Иркутской области с номерами MLVA MtbC159, а также Internet Archive (https://archive.org) за сохраненную копию официальной таблицы типов MtbC15.

ФИНАНСИРОВАНИЕ

DOI: 10.11134/btp.4.2026.1

Работа выполнена в рамках проекта BR27195585 «Создание новых отечественных тест-систем и поиск потенциальных биомаркеров для диагностики социально-значимых заболеваний», финансируемого Комитетом науки Министерства науки и высшего образования Республики Казахстан.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

ВКЛАД АВТОРОВ

Найзабаева Д.А. – поиск, сбор, систематизация и анализ литературных данных, тестирование и валидация программы, написание текста; Скиба Ю.А. – концепция и руководство работой, разработка программного обеспечения, реализация алгоритмов присвоения номенклатурных номеров, тестирование и валидация программы, написание текста.

СПИСОК ЛИТЕРАТУРЫ

1. Supply, P. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis / P. Supply, C. Allix, S. Lesjean [et al.] // J. Clin. Microbiol. – 2006. – Vol. 44, № 12. – P. 4498–4510. PMID: 17005759. DOI: 10.1128/JCM.01392-06

2. Weniger, T. MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria / T. Weniger, J. Krawczyk, P. Supply [et al.] // Nucleic Acids Res. – 2010. – Vol. 38, Web Server issue. – P. W326–W331. PMID: 20457747. DOI: 10.1093/nar/gkq351

3. Allix-Béguec, C. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of Mycobacterium tuberculosis complex isolates / C. Allix-Béguec, D. Harmsen, T. Weniger [et al.] // J. Clin. Microbiol. – 2008. – Vol. 46, № 8. – P. 2692–2699. PMID: 18550737. DOI: 10.1128/JCM.00540-08

4. Skiba, Y. Molecular snapshot of Mycobacterium tuberculosis population in Kazakhstan: a country-wide study / Y. Skiba, I. Mokrousov, G. Ismagulova [et al.] // Tuberculosis (Edinb.). – 2015. – Vol. 95, № 5. – P. 538–546. PMID: 26076582. DOI: 10.1016/j.tube.2015.04.012

5. Skiba, Y. Mycobacterium tuberculosis RD-Rio Strain in Kazakhstan / Y. Skiba, I. Mokrousov, D. Nabirova [et al.] // Emerg. Infect. Dis. – 2019. – Vol. 25, № 3. – P. 604–606. PMID: 30789328. DOI: 10.3201/eid2503.181179

6. Akhmetova, A. Prevalence of Beijing Central Asian/ Russian Cluster 94-32 among Multidrug-Resistant M. tuberculosis in Kazakhstan / A. Akhmetova, V. Bismilda, L. Chingissova [et al.] // Antibiotics (Basel). – 2024. – Vol. 13, № 1. – P. 9. PMID: 38275319. DOI: 10.3390/antibiotics13010009

7. Auganova, D. Genomic characterization and epidemiology of Mycobacterium tuberculosis lineage 2 isolates from Kazakhstan / D. Auganova, S. Atavliyeva, N. Gharbi [et al.] // Sci. Rep. – 2025. – Vol. 15, № 1. – P. 37715. PMID: 41152417. DOI: 10.1038/s41598-025-22485-2

8. Norheim, G. Tuberculosis Outbreak in an Educational Institution in Norway / G. Norheim, S. Seterelv, T.M. Arnesen [et al.] // J. Clin. Microbiol. – 2017. – Vol. 55, № 5. – P. 1327–1333. PMID: 28202795. DOI: 10.1128/JCM.01152-16

9. Perdigão, J. Emergence of multidrug-resistant Mycobacterium tuberculosis of the Beijing lineage in Portugal and Guinea-Bissau: a snapshot of moving clones by whole-genome sequencing / J. Perdigão, C. Silva, F. Maltez [et al.] // Emerg. Microbes Infect. – 2020. – Vol. 9, № 1. – P. 1342–1353. PMID: 32538300. DOI: 10.1080/22221751.2020.1774425

10. Mokrousov, I. Early ancient sublineages of Mycobacterium tuberculosis Beijing genotype: unexpected clues from phylogenomics of the pathogen and human history / I. Mokrousov, A. Vyazovaya, O. Pasechnik [et al.] // Clin. Microbiol. Infect. – 2019. – Vol. 25, № 8. – P. 1039.e1–1039.e6. PMID: 30528901. DOI: 10.1016/j.cmi.2018.11.024

11. Mokrousov, I. Insight into pathogenomics and phylogeography of hypervirulent and highly-lethal Mycobacterium tuberculosis strain cluster / I. Mokrousov, A. Vyazovaya, E. Shitikov [et al.] // BMC Infect. Dis. – 2023. – Vol. 23, № 1. – P. 426. PMID: 37353765. DOI: 10.1186/s12879-023-08413-7

12. Wyllie, D.H. A Quantitative Evaluation of MIRU-VNTR Typing Against Whole-Genome Sequencing for Identifying Mycobacterium tuberculosis Transmission: A Prospective Observational Cohort Study / D.H. Wyllie, J.A. Davidson, E. Grace Smith [et al.] // EBioMedicine. – 2018. – Vol. 34. – P. 122–130. PMID: 30077721. DOI: 10.1016/j.ebiom.2018.07.019

13. Rajwani, R. MIRU-profiler: a rapid tool for determination of 24-loci MIRU-VNTR profiles from assembled genomes of Mycobacterium tuberculosis / R. Rajwani, S. Shehzad, G.K.H. Siu // PeerJ. – 2018. – Vol. 6. – P. e5090. PMID: 30018852. DOI: 10.7717/peerj.5090

14. Tang, C.Y. MIRUReader: MIRU-VNTR typing directly from long sequencing reads / C.Y. Tang, R.T. Ong // Bioinformatics. – 2020. – Vol. 36, № 5. – P. 1625–1626. PMID: 31603462. DOI: 10.1093/bioinformatics/btz771

15. Buenestado-Serrano, S. Bridging the gap between molecular and genomic epidemiology in tuberculosis: inferring MIRU-VNTR patterns from genomic data / S. Buenestado-Serrano, M. Martínez-Lirola, A. Dippenaar [et al.] // J. Clin. Microbiol. – 2024. – Vol. 62, № 9. – P. e00741-24. PMID: 39136450. DOI: 10.1128/jcm.00741-24

16. Pedersen, M.K. Set-up and validation of mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR) analysis of Mycobacterium tuberculosis using BioNumerics software / M.K. Pedersen, A.B. Andersen, D.B. Folkvardsen [et al.] // PLoS One. – 2018. – Vol. 13, № 10. – P. e0205336. PMID: 30379832. DOI: 10.1371/journal.pone.0205336

17. Mangaa, G.K. Molecular epidemiology of Mycobacterium tuberculosis complex strains in Nairobi, Kenya [Электронный ресурс] : doctoral dissertation / G.K. Mangaa ; University of Lübeck, Research Center Borstel. – Lübeck, 2018. – 168 p. – Режим доступа: https://www.zhb.uni-luebeck.de/epubs/ediss1996.pdf (дата обращения: 24.09.2026). – Загл. с титул. экрана.

DOI: 10.11134/btp.4.2026.1

18. MLVA MtbC15 types [Электронный ресурс] // MIRU-VNTRplus : [сайт]. – Копия страницы от 21.06.2022 в веб-архиве Internet Archive. – Режим доступа: https://web.archive.org/web/20220621185438/https://www.miru-vntrplus.org/MIRU/types15.faces (дата обращения: 23.09.2026). – Загл. с экрана.

19. Phelan, J.E. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs / J.E. Phelan, D.M. O'Sullivan, D. Machado [et al.] // Genome Med. – 2019. – Vol. 11, № 1. – P. 41. PMID: 31234910. DOI: 10.1186/s13073-019-0650-x

20. Hunt, M. Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe / M. Hunt, P. Bradley, S.G. Lapierre [et al.] // Wellcome Open Res. – 2019. – Vol. 4. – P. 191. PMID: 32055708. DOI: 10.12688/wellcomeopenres.15603.1

21. Shuaib, Y.A. Origin and Global Expansion of Mycobacterium tuberculosis Complex Lineage 3 / Y.A. Shuaib, C. Utpatel, T.A. Kohl [et al.] // Genes (Basel). – 2022. – Vol. 13, № 6. – P. 990. PMID: 35741753. DOI: 10.3390/genes13060990

22. Merker, M. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage / M. Merker, C. Blin, S. Mona [et al.] // Nat. Genet. – 2015. – Vol. 47, № 3. – P. 242–249. PMID: 25599400. DOI: 10.1038/ng.3195

23. Couvin, D. simpiTB – a pipeline designed to extract meaningful information from whole genome sequencing data of Mycobacterium tuberculosis complex, allows to combine genomic, phylogenetic and clustering analyses in existing SITVIT databases / D. Couvin, E. Stattner, W. Segretier [et al.] // Infect. Genet. Evol. – 2023. – Vol. 113. – P. 105466. PMID: 37331497. DOI: 10.1016/j.meegid.2023.105466

DOI: 10.11134/btp.4.2026.1

DEVELOPMENT AND EVALUATION OF OFFLINE SOFTWARE FOR WORKING WITH THE MLVA MTBC159 NOMENCLATURE USING MIRU-VNTR TYPING AND WHOLE-GENOME SEQUENCING DATA OF MYCOBACTERIUM TUBERCULOSIS

Naizabayeva D.A.[1,2], **Skiba Y.A.** [1,2] *

[1]*M.A. Aitkhozhin's Institute of Molecular Biology and Biochemistry, 86 Dosmukhamedov St., Almaty 050012, Kazakhstan*
[2]*Almaty Branch of the National Center for Biotechnology, 14 Zhahanger St., Almaty 050054, Kazakhstan*
**Corresponding author: yuriy.skiba@gmail.com*

ABSTRACT

Molecular epidemiology of tuberculosis is moving from typing by variable-number tandem repeats to whole-genome sequencing (WGS), yet the record of how *Mycobacterium tuberculosis* genotypes spread over past decades is written in the MLVA MtbC159 nomenclature, and short sequencing reads do not resolve MIRU-VNTR profiles reliably. The codes were assigned by a single web service, MIRU-VNTRplus (https://www.miru-vntrplus.org), whose interactive functions became unavailable in 2026. This work moved the service functions to the laboratory to keep new data linked to that record. Offline software for Windows, MTBC Lab Toolkit, was developed that assigns codes by the exact-match rule of the 15-locus and 9-locus panels, processes sets of profiles, reports the nearest registered types and links MIRU-VNTR profiles with whole-genome sequencing results. Its reference data comprise 3 906 coded profiles from 152 publications and 23 560 MtbC15 types from the service's official table. Of 400 published codes, the software reproduced all 392 agreeing with the official table; the other 8 were publication errors, and no wrong code was returned. A missing locus is not guessed: the software searches known profiles matching the other 23 loci and returns the nearest code only if a single such profile exists. A profile with one locus changed never received its previous code. TB-Profiler and Mykrobe import was verified on their output files; a batch of 3 293 profiles is processed in 7.3 s. The software compares new data, including profiles inferred from long reads, with accumulated and published MIRU-VNTR typing results without an external server.

Keywords: *Mycobacterium tuberculosis*; MIRU-VNTR; MLVA MtbC159; genotype nomenclature; molecular epidemiology; genomic epidemiology; software; whole-genome sequencing; Kazakhstan.

MYCOBACTERIUM TUBERCULOSIS-ТІҢ MIRU-VNTR ТИПТЕУ ЖӘНЕ ТОЛЫҚ ГЕНОМДЫҚ СЕКВЕНИРЛЕУ ДЕРЕКТЕРІ БОЙЫНША MLVA MTBC159 НОМЕНКЛАТУРАСЫМЕН ЖҰМЫС ІСТЕУГЕ АРНАЛҒАН АВТОНОМДЫ БАҒДАРЛАМАЛЫҚ ҚАМТАМАСЫЗ ЕТУДІ ӘЗІРЛЕУ ЖӘНЕ СЫНАҚТАН ӨТКІЗУ

Найзабаева Д.А. [1,2], **Скиба Ю.А.** [1,2] *

[1]*«М.А. Айтхожин атындағы молекулярлық биология және биохимия институты» ШЖҚ РМК, Досмұхамедов көшесі, 86, 050012, Алматы, Қазақстан Республикасы*
[2]*«Ұлттық биотехнология орталығы» ЖШС-нің Алматы қаласындағы филиалы, Жахангер көшесі, 14, Алматы, 050054, Қазақстан*
**Хат алмасуға жауапты автор: yuriy.skiba@gmail.com*

ТҮЙІН

Туберкулездің молекулалық эпидемиологиясы тандемдік қайталанымдар санының өзгергіштігіне негізделген типтеуден толық геномдық секвенирлеуге (WGS) көшуде. Алайда соңғы онжылдықтардағы *Mycobacterium tuberculosis* генотиптерінің таралуы туралы деректер MLVA MtbC159 номенклатурасы бойынша тіркелген, ал қысқа оқылымдар негізінде MIRU-VNTR профилі сенімді түрде анықталмайды. Осы номенклатура бойынша нөмірлерді тек MIRU-VNTRplus (https://www.miru-vntrplus.org) веб-сервисі беріп келді, бірақ оның интерактивті функциялары 2026 жылы қолжетімсіз болып қалды. Жұмыстың мақсаты – сервис функцияларын зертхананың өзінде іске асыру арқылы жаңа деректердің бұрын жинақталған деректермен байланысын сақтау.

Windows-қа арналған автономды MTBC Lab Toolkit бағдарламасы әзірленді. Ол 15 және 9 локустан тұратын панельдер бойынша дәл сәйкестік ережесіне сәйкес нөмір береді, профильдерді топтап өңдейді, тіркелген ең жақын типтерді көрсетеді және MIRU-VNTR профильдерін толық геномдық секвенирлеу нәтижелерімен байланыстырады. Бағдарламаның анықтамалық базасын 152 жарияланымнан алынған, нөмірлері көрсетілген 3 906 профиль және сервистің ресми кестесіндегі 23 560 MtbC15 типі құрайды.

Жарияланымдарда келтірілген 400 нөмірдің ішінен бағдарлама ресми кестеге сәйкес келетін 392 нөмірдің барлығын дұрыс берді, ал қалған 8 нөмірдің жарияланымдарда қате көрсетілгені анықталды. Бағдарлама бірде-бір қате нөмір берген жоқ. Дерегі жоқ локустың мәні болжанбайды: бағдарлама қалған 23 локус бойынша сәйкес келетін белгілі профильдерді іздейді және мұндай профиль біреу болған жағдайда ғана ең жақын нөмірді береді. Бір локусы өзгертілген профильге ешбір жағдайда бұрынғы нөмір берілген жоқ. TB-Profiler және Mykrobe нәтижелерін импорттау

DOI: 10.11134/btp.4.2026.1

нақты шығыс файлдарында тексерілді. 3 293 профильден тұратын жиынтық 7,3 с ішінде өңделеді.

Бағдарлама жаңа деректерді, соның ішінде ұзын оқылымдардан алынған профильдерді, бұрын жинақталған және жарияланған MIRU-VNTR типтеу нәтижелерімен сыртқы серверге жүгінбей салыстыруға мүмкіндік береді.

Түйін сөздер: *Mycobacterium tuberculosis*; MIRU-VNTR; MLVA MtbC159; генотиптер номенклатурасы; молекулалық эпидемиология; геномдық эпидемиология; бағдарламалық қамтамасыз ету; толық геномдық секвенирлеу; Қазақстан.