

Ю.И.Яшечкин, Е.В.Найденова, Т.В.Бугоркова, С.А.Щербакова

СОЗДАНИЕ БАЗЫ ДАННЫХ ПО ХАРАКТЕРИСТИКАМ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНОМОВ ШТАММОВ ВОЗБУДИТЕЛЕЙ БАКТЕРИАЛЬНЫХ И ВИРУСНЫХ ИНФЕКЦИЙ I-II ГРУПП ПАТОГЕННОСТИ

ФКУЗ «Российский научно-исследовательский противочумный институт «Микроб», Саратов

Разработана база данных (БД), содержащая характеристики переменных тандемных повторов, олигонуклеотидных праймеров и зондов геномов штаммов возбудителей бактериальных и вирусных инфекций I-II групп патогенности. БД позволяет накапливать информацию об участках геномов, а с использованием встроенных VB-приложений быстро проводить анализ данных и осуществлять информативно обоснованный расчет генодиагностических препаратов, основанных на различных форматах полимеразной цепной реакции.

Ключевые слова: база данных, ПЦР, VNTR, генодиагностические препараты.

Yu.I.Yashechkin, E.V.Naydenova, T.V.Bugorkova, S.A.Shcherbakova

Setting-up of the Database on the Nucleotide Sequences of the Genomes of the Strains of Bacterial and Viral Infections Agents of the I-II Pathogenicity Groups

Russian Research Anti-Plague Institute "Microbe", Saratov

Set-up is database containing characteristics of variable tandem repeats, oligo-nucleotide primers and probes of the genomes of the strains bacterial and viral infections agents of the I-II pathogenicity groups. It allows for accumulation of information on genome regions. Using the incorporated VBA one can rapidly analyze the data and carry out substantiated calculation of gene-diagnostic preparations based on various PCR frameworks.

Key words: database, PCR, VNTR, gene-diagnostic preparations.

К настоящему времени разработано большое количество методов генотипирования, с помощью которых получены разнообразные данные по геномной идентификации микроорганизмов, в том числе и возбудителей особо опасных инфекционных заболеваний [5]. Мультилокусный анализ числа переменных тандемных повторов (MLVA), определяющий количество переменных тандемных повторов на определенных участках генома возбудителя, позволяет составить генетический портрет штамма. Выполнение этого анализа основано на выборе достаточно большого количества локусов, содержащих переменные тандемные повторы (VNTR). Полученные MLVA- и VNTR-профили выделенных культур возбудителей особо опасных инфекционных болезней позволяют определить происхождение штаммов, источник и пути распространения инфекции, что является важной задачей при проведении эпидемиологических исследований.

В связи с этим разработка и внедрение в практику молекулярно-генетических методов диагностики возбудителей бактериальных и вирусных инфекций I-II групп патогенности, обеспечивающих быстрое обнаружение и идентификацию возбудителя, является актуальной задачей. Совершенствование аналитического расчета праймеров и зондов при разработке ПЦР-тест-систем значительно повысит качество конструируемых препаратов и сократит время их разработки.

Цель настоящей работы – создание пополняемой базы данных, позволяющей анализировать и си-

стематизировать переменные тандемные повторы, олигонуклеотидные праймеры и зонды, рассчитанные на нуклеотидные последовательности геномов штаммов микроорганизмов и РНК-содержащих вирусов I-II групп патогенности.

В Интернет-ресурсах существуют различные базы данных, решающие отдельные задачи, например, по подбору праймеров (<http://pga.mgh.harvard.edu/primerbank>), VNTR-локусам (<http://minisatellites.upsud.fr>) [6], зондам (<http://www.ncbi.nlm.nih.gov/sites/entrez>) и другие, однако они разработаны в разных форматах, и интегрирование полученных из них данных связано с определенными проблемами.

Проведенный анализ структуры этих информационных продуктов выявил необходимость создания единой интегрированной базы данных для решения всех перечисленных выше частных задач. Основываясь на этом, была сформирована пополняемая база данных, которая позволяет анализировать и систематизировать переменные тандемные повторы, олигонуклеотидные праймеры и зонды, рассчитанные на нуклеотидные последовательности геномов штаммов микроорганизмов и РНК-содержащих вирусов I-II групп патогенности.

База данных формата Access 8.0 создана в интерактивной среде разработки программного обеспечения Visual Basic (VB) 6.0 [7] с использованием запросов VB и (или) SQL (Structured Query Language – «язык структурированных запросов»). В базе данных накоплено около 1000 нуклеотидных последовательностей высокопатогенных микроорганизмов, относя-

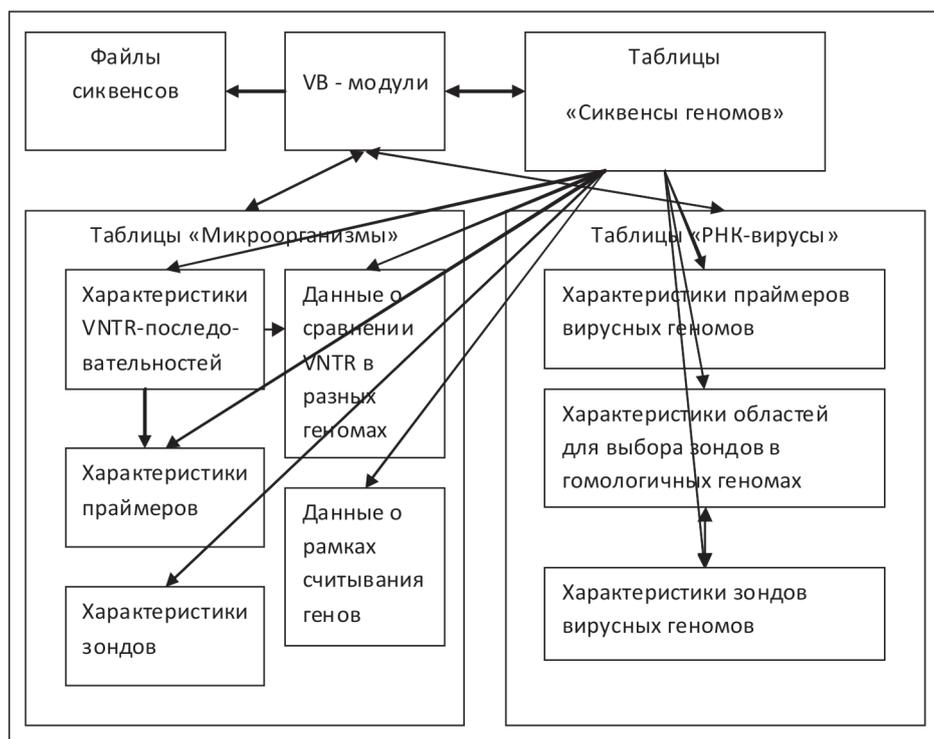


Рис. 1. Структурно-логическая схема базы данных и схема связи данных таблиц БД Access. Рамками выделены структурные блоки таблиц

щихся к родам *Francisella*, *Yersinia*, *Bacillus*, *Vibrio*, и около 8000 нуклеотидных последовательностей РНК-содержащих вирусов I–II групп патогенности, опасных для человека, которые относятся к семействам *Filoviridae* (Эбола и Марбург), *Arenaviridae* (Ласса, Хунин, Мачупо), *Flaviviridae* (желтой лихорадки, клещевого энцефалита, Денге, Западного Нила), *Bunyaviridae* (ККГЛ, ГЛПС).

Таблицы в базе данных организованы таким образом, что можно подбирать наборы праймеров для использования их в мультиплексной ПЦР, в результате чего значительно повышается диагностический уровень этих препаратов. Это является важным положительным критерием, который принципиально отличает разработанную базу данных от других информационных и аналитических ресурсов. Расчет таких систем осуществляется с использованием интегрированного с ней приложения: «Программы расчета согласованных по термодинамическим параметрам олигонуклеотидных праймеров для мультиплексной ПЦР и формирования базы данных» (Свидетельство о Госрегистрации № 2010612350 от 9.02.2010 г.).

С использованием запросов SQL и программных модулей VB в базе данных накоплены полученные нами сведения о переменных тандемных повторах геномов микроорганизмов, олигонуклеотидных праймерах и зондах геномов микроорганизмов и вирусов. Запросы позволяют выбирать праймеры и зонды на определенный ген, область генома или VNTR в зависимости от решаемой генодиагностической задачи.

Структурно база данных состоит из трех блоков таблиц: первый – «Сиквенсы геномов», второй – «Микроорганизмы», третий – «РНК-вирусы» (рис. 1).

Таблицы соединены между собой посредством

уникальных кодов («ID») и имеют подчиненную организацию. Главной по иерархии является таблица сиквенсов, удаление записи из нее приводит к удалению соответствующих записей из всех остальных таблиц. Изменение данных в подчиненных таблицах не влияет на другие данные.

Наличие ключевых полей в таблицах исключает дублирование записей.

Для расширения информационных возможностей базы данных применена файловая структура сиквенсов, т.е. каждая нуклеотидная последовательность содержится в отдельном файле, а связь с базой данных Access осуществляется через программные модули VB. Таким образом, общий объем базы данных ограничен лишь дисковым пространством.

При заполнении всех таблиц, просмотре и отборе сиквенсов, выборе праймеров, зондов и т.д. используют оконные VB-формы. В программе реализован удобный и быстрый доступ к таблицам базы данных, а также полям и записям, который осуществляется с помощью запросов, написанных на языке SQL и VB.

Пополнение базы данных, загрузка сиквенсов происходит с использованием программного кода, вызываемого из VB-формы командной кнопкой. При загрузке файла сиквенса выполняется предварительная проверка имени и содержимого файла на отсутствие ранее загруженного в базу данных аналогичного файла, что исключает дублирование данных в базе.

Из VB-формы «Праймеры» (рис. 2) блока «Микроорганизмы» можно с помощью запроса SQL получить данные из таблиц: «Сиквенсы», «VNTR» и «Праймеры». Аналогично из VB-формы, связанной с блоком таблиц «РНК-вирусы», можно с помощью запроса SQL получить данные из таблиц «Сиквенсы» и «Гомологичные праймеры». Из форм можно про-

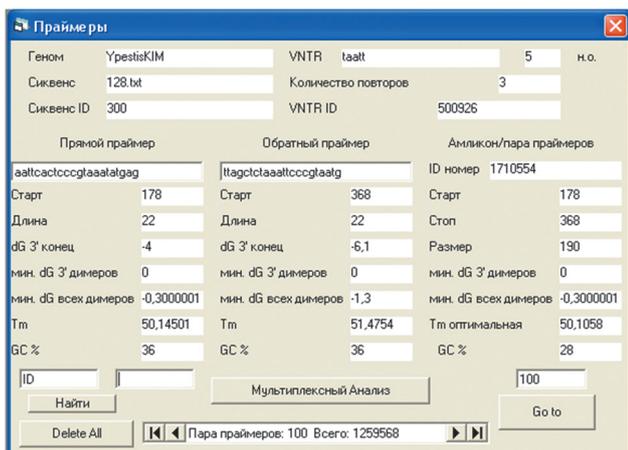


Рис. 2. Форма доступа к таблицам, содержащим характеристики олигонуклеотидных праймеров, сиквенсов и VNTR

водить запросы SQL по поиску, выбору и удалению записей.

При загрузке форм реализуется запрос SQL на объединение записей из разных таблиц. Альтернативно доступ и работа с таблицами возможна при использовании систем управления базами данных MS Access [4] и VisData (Visual Basic). Экспорт запросов SQL в MS Excel позволяет проводить статистическую обработку данных.

Структура базы данных представлена так, чтобы пользователь, в соответствии с поставленной задачей (независимо от сочетания бактерия/вирус), мог сконфигурировать запрос на выбор, расчет или выбор и расчет праймеров и зондов. Это значительно облегчает, ускоряет и информационно обосновывает разработку препаратов для генной диагностики возбудителей бактериальной и вирусной природы.

Использование базы данных при разработке генодиагностических препаратов иллюстрируется следующим примером. При разработке ПЦР-тест-системы для выявления РНК вируса Западного Нила были проанализированы 73 геномные нуклеотидные последовательности изолятов вируса I–V генетических вариантов. При помощи запроса SQL и VB-приложения значения гомологии выровненных

нуклеотидных последовательностей, выраженные в процентах, из базы данных были экспортированы в программу MS Excel, где эти данные были проанализированы и графически отображены с использованием встроенных функций (рис. 3).

На основании полученных результатов определены наиболее консервативные участки, перспективные для аналитического расчета праймеров, которые локализованы в участках РНК, кодирующих GPE, NS2B, NS5 гены и 5'-область генома. Из них для дальнейшей работы была выбрана нуклеотидная последовательность области NS5 гена. С применением программных модулей VB рассчитаны праймеры на этот участок, их характеристики пополнили базу данных. Экспорт этих характеристик в программу MS Excel и использование «фильтра данных» позволило выбрать оптимальные праймеры и разработать структуру ПЦР-тест-системы.

Специфичность реакции с выбранными праймерами была подтверждена экспериментально, что позволило сконструировать «Набор реагентов для выявления РНК вируса Западного Нила методом обратной транскрипции и полимеразной цепной реакции с электрофоретическим учетом результатов (ГенНил-РЭФ)» [1].

С использованием разработанной базы данных также были созданы: «Набор реагентов для выявления РНК вируса Крымской-Конго геморрагической лихорадки (ККГЛ) методом обратной транскрипции и полимеразной цепной реакции с электрофоретическим учетом результатов (ГенКонго-РЭФ)» [2], «Набор реагентов для выявления РНК вирусов Денге (I–IV типов) методом обратной транскрипции и полимеразной цепной реакции с электрофоретическим учетом результатов (ГенДенге-РЭФ)» [3].

Таким образом, разработанная база данных «Нуклеотидные последовательности и характеристики геномов, вариабельных tandemных повторов, олигонуклеотидных праймеров и зондов штаммов возбудителей бактериальных и вирусных инфекций I–II групп патогенности», зарегистрированная в Реестре баз данных (№ 2011620585 19.08.2011 г.),

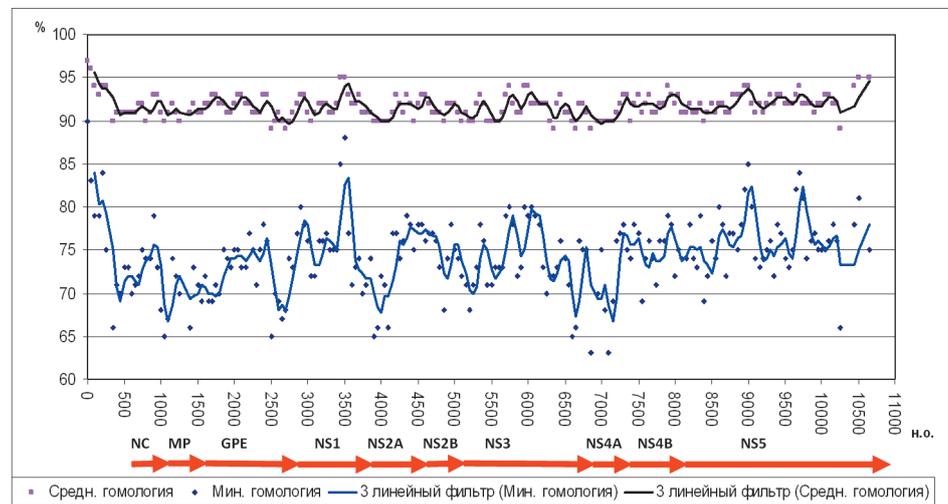


Рис. 3. Сравнительный анализ гомологии полных нуклеотидных последовательностей штаммов вируса Западного Нила

позволяет не только накапливать информацию об участках (структурах) геномов бактерий и вирусов, но и проводить анализ данных, что абсолютно необходимо при конструировании генодиагностических тест-систем.

СПИСОК ЛИТЕРАТУРЫ

1. Красовская Т.Ю., Шарова И.Н., Щербакова С.А., Яшечкин Ю.И., Хуторецкая Н.В., Ларичев В.Ф., Бутенко А.М., Куличенко А.Н. Разработка и внедрение тест-системы для лабораторной диагностики лихорадки Западного Нила методом ПЦР. *ЗНИСО*. 2007; 6(171):42–6.
2. Найденова Е.В., Шарова И.Н., Щербакова С.А., Яшечкин Ю.И., Хуторецкая Н.В., Ларичев В.Ф., Саркисян К.А., Воробьева М.С., Бутенко А.М., Куличенко А.Н., Львов Д.К., Кутырев В.В. Тест-система для выявления РНК вируса Крымской-Конго геморрагической лихорадки методом ОТ-ПЦР. *Пробл. особо опасных инф.* 2006; 2(92):54–6.
3. Найденова Е.В., Яшечкин Ю.И., Щербакова С.А. Набор реагентов для выявления РНК вирусов Денге (I–IV типов) методом обратной транскрипции и полимеразной цепной реакции. *Национальные приоритеты России*. 2011; 2(5):53.
4. Нортон П., Андерсен В. Разработка приложений в Access 97 в подлиннике. Пер. с англ. СПб: БНВ-Санкт-Петербург; 2000. 656 с.
5. Попов Ю.А., Ерошенко Г.А., Булгакова Е.Г., Смирнова Н.И. Разработка комплексного алгоритма генотипирования и методов оценки генетического разнообразия природных штаммов возбудителей чумы и холеры. *Пробл. особо опасных инф.* 2009; 4(102):5–11.
6. Chang C.H., Chang Y.C., Underwood A., Chiou C.S., Kao C.Y. VNTRDB: a bacterial variable number tandem repeat locus database. *Nucl. Acids Res.* 2007; 35:416–21.
7. Visual Basic 6.0. Пер. с англ. СПб: БНВ-Санкт-Петербург; 1998. 992 с.

References

1. Krasovskaya T.Yu., Sharova I.N., Shcherbakova S.A., Yashechkin Yu.I., Khutoretskaya N.V., Larichev V.F., Butenko A.M., Kulichenko A.N. [Development and implementation of the test-system for laboratory diagnostics of West-Nile fever using PCR]. *Zdor. Nasel. Sreda Obit.* 2007; 6(171): 42–6.
2. Naidenova E.V., Sharova I.N., Scherbakova S.A., Yashechkin Yu.I., Khutoretskaya N.V., Larichev V.F., Sarkisyan K.A., Vorobyova M.S., Butenko A.M., Kulichenko A.N., L'vov D.K., Kutuyev V.V. [A test system to reveal RNA of the Congo-Crimean hemorrhagic fever virus using RT-PCR procedure]. *Probl. Osobo Opasn. Infek.* 2006; (92):54–6.
3. Naidenova E.V., Yashechkin Yu.I., Scherbakova S.A. [A panel for I-IV type Dengue virus RNA detection using reverse transcription and PCR]. *Natsional. Prior. Rossii.* 2011; 2(5):53.
4. Norton Yu.A., Andersen V. [Design of Applications on the base of Original Version of Access 97 Software]. Translated from the English. SPb.; 2000. 656 p.
5. Popov Yu.A., Eroshenko G.A., Bulgakova E.G., Smirnova N.I. [Development of complex genotyping algorithm and methods of evaluation of the genetic diversity of plague and cholera agent natural strains]. *Probl. Osobo Opasn. Infek.* 2009; (102):5–11.
6. Chang C.H., Chang Y.C., Underwood A., Chiou C.S., Kao C.Y. VNTRDB: a bacterial variable number tandem repeat locus database. *Nucl. Acids Res.* 2007; 35:416–21.
7. Visual Basic 6.0. Translated from the English. SPb.; 1998. 992 p.

Authors:

Yashechkin Yu.I., Naidenova E.V., Bugorkova T.V., Shcherbakova S.A. Russian Research Anti-Plague Institute "Microbe". 46, Universitetskaya St., Saratov, 410005, Russia. E-mail: rusrapi@microbe.ru

Об авторах:

Яшечкин Ю.И., Найденова Е.В., Бугоркова Т.В., Щербакова С.А. Российский научно-исследовательский противочумный институт «Микроб». 410005, Саратов, ул. Университетская, 46. E-mail: rusrapi@microbe.ru

Поступила 30.01.12.