

© М. Ф. Вечерковская, В. В. Тец, 2015 г.
УДК 616-093/-098-004.651

М. Ф. Вечерковская, В. В. Тец

СОЗДАНИЕ БАЗЫ ДАННЫХ ДЛЯ РЕШЕНИЯ ЗАДАЧ ПО СИ- СТЕМАТИЗАЦИИ РЕЗУЛЬТА- ТОВ ОПЫТОВ, ПРОВОДИМЫХ В ХОДЕ НАУЧНЫХ ИССЛЕДОВА- НИЙ В МИКРОБИОЛОГИИ

Кафедра микробиологии, вирусологии и иммунологии имени академика Д. К. Заболотного Первого Санкт-Петербургского государственного медицинского университета имени академика И. П. Павлова

ВВЕДЕНИЕ

После реализации проекта «Геном человека» в 2008 г. стартовал проект «Микробиом человека», направленный на изучение состава микробиоты человека [3, 4, 9]. Целью нового международного проекта является идентификация и изучение микроорганизмов, живущих в ассоциации с организмом человека. Традиционно микробиологические исследования базируются на выделении чистой культуры — получении популяции клеток, возникающих при делении одной исходной бактерии и обладающих одинаковыми культуральными, морфологическими и биохимическими признаками. Получение чистой культуры является ключевым этапом в систематике бактерий, изучении их ультраструктуры, биологических свойств, изменчивости и способности вызывать заболевания человека [5]. Одним из лидирующих направлений в изучении представителей микробиоты человека стало выявление бактерий, относящихся к «пока не культивируемым». Такие бактерии на данном этапе не удается получить в лабораторных условиях в виде чистых культур. Во многом это обосновано сложностью изоляции из стойких микробных сообществ — биопленок. «Биопленка» — термин, обобщающий формы существования микроорганизмов в сообществах. Биопленки, включающие много видов микроорганизмов, составляют микрореконсоциумы, или смешанные биопленки, — сложно устроенные комплексы, с высоко организованной архитектурой и ультраструктурой, с высоким уровнем межклеточной коммуникации и горизонтальной передачей генов [2, 6, 7].

Изучение состава бактериальных биопленок, и в особенности смешанных, включающих неизвестные микроорганизмы, требует оперативной работы с большими объемами информации. Современная микробиология использует множество методов исследования — от традиционных, основанных на культивировании микроорганизмов, до молекулярно-

биологических, позволяющих предсказывать свойства микроба путем изучения генома единственной бактериальной клетки, полученной непосредственно из исследуемого материала. На каждом этапе такой работы исследователю необходимо оперировать данными, полученными с помощью различных методов (и поэтому в различных форматах): тексты с описанием биопленок и отдельных колоний, морфологии и ультраструктуры микроорганизмов по результатам световой и электронной микроскопии, описание условий культивирования и модификации среды; таблицы с учетом результатов проводимых биохимических и иных тестов, направленных на выявление свойств бактерий; фотографии, полученные при световой и электронной микроскопии; последовательности нуклеотидов, полученных при секвенировании отдельных генов и геномов микроорганизмов; аннотации геномов; результаты метагеномного и филогенетического анализов. Отслеживание во времени и систематизация таких данных с использованием одной только файловой системы компьютера становятся все более трудоемкими. Еще сложнее проводить поиск и сравнение данных по одному или нескольким параметрам одновременно. Невозможность синхронизации данных между удаленно расположенными компьютерами и их пользователями приводит к некоррелированному учету результатов опытов и потере единообразия учета экспериментов. В результате замедляется процесс обработки информации, данные перемешиваются и даже могут быть утрачены в связи с отсутствием четко отлаженной системы резервного копирования. Для систематизации и обработки информации, получаемой в ходе экспериментов, а также для увеличения доступа к этим данным создаются базы данных, как с открытым, так и с закрытым доступом. База данных — это интегрированная компьютерная структура (хранится и обрабатывается в вычислительной системе), содержащая совокупность систематизированных данных и метаданных (характеристики данных), предоставляющая конечному пользователю возможность поиска и обработки информации, содержащейся в ней [1]. Биологические базы данных отличаются сложной структурой, поскольку содержат в себе информацию в разнообразных форматах (сиквенсы, изображения, описания, статьи и др.). Построены они, как правило, на основе реляционной модели [8]. Создание и администрирование биологических баз данных — одно из актуальных направлений в биоинформатике. Несмотря на то, что на рынке информационных услуг и в открытом доступе существует немало баз данных с готовыми шаблонами (в основном направленных на систематизацию клиентских баз), среди них нет ни одной, удовлетворяющей потребности такого исследования. Большинство биологических баз данных, рас-

положенных в Интернете, имеют узконаправленный характер и предназначены для предоставления доступа к готовым результатам, полученным различными исследовательскими группами.

В связи с этим на кафедре микробиологии, вирусологии и иммунологии им. акад. Д. К. Заболотного ПСПбГМУ им. акад. И.П. Павлова была создана база данных «Microbes».

МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ

Программа «FileMakerPro» (*FileMakerInc.*, США) представляет широкий набор инструментов для работы с реляционными базами данных. Основным преимуществом этой системы управления базами данных является возможность формирования взаимосвязей между записями, что делает ее универсальным инструментом для гибкого и быстрого анализа информации.

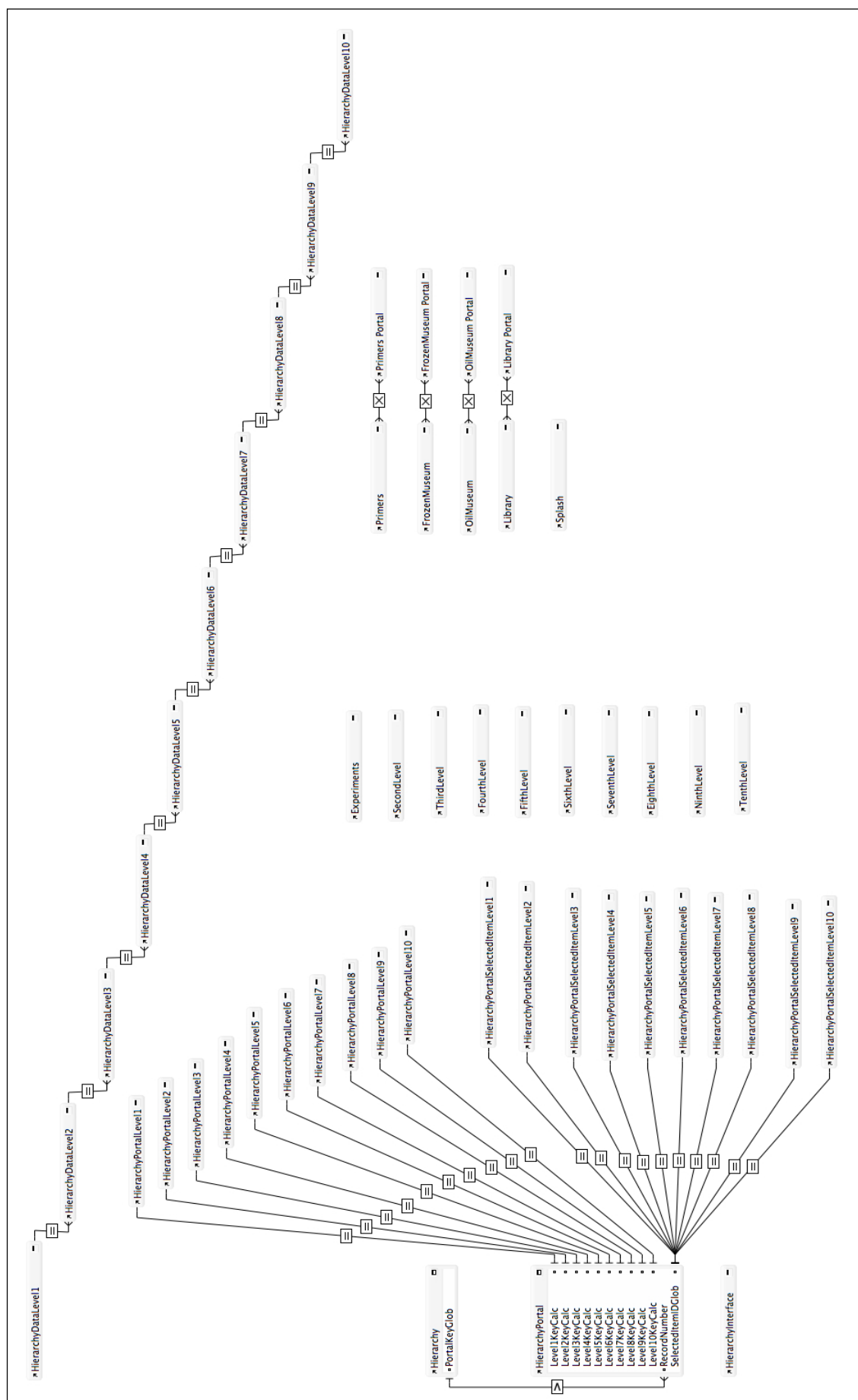
Для создания базы данных «Microbes» использовалась программа «FileMakerProAdvanced 12.0». Основа иерархической системы — модифицированная «SeedCodeHierarchyPro». Поскольку базу планировалось использовать как на OSX, так и на Windows, для написания скриптов, обращающихся к операционной системе, использовались AppleScript и VBScript соответственно. Для создания удаленного доступа использовали «FileMaker-Server 12». Сервер имеет статический IP-адрес. Подключение к серверу осуществляется в защищенном режиме с использованием VPN (virtual private network — виртуальная частная сеть). Структура базы представлена таблицами, предназначенными для хранения атрибутивной и графической информации, а также таблицами для графического представления информации пользователю.

В базе используются одновременно две различные модели систематизации материала. Основной является иерархическая модель, в которой каждый этап опыта связан с предыдущим. Таким образом, легко отслеживать динамику результатов этапов, относящихся к разным уровням (этапы выделения бактерий, входящих в микробные биопленки), и собирать в одной вкладке результаты, относящиеся к одному уровню, в том числе получаемые в разные дни (биохимическая идентификация, анализ гена 16S rRNA и др.). Вторая модель используется для учета хранящихся штаммов бактерий, праймеров для молекулярно-биологических исследований и публикаций. Хранение информации происходит на созданном нами удаленном сервере. Каждый сотрудник, работающий с базой, имеет индивидуальные имя пользователя и пароль. Подключение к базе может производиться с любого авторизованного автором базы компьютера через защищенный Интернет-канал. Таким образом достигается возможность работы с базой одновременно нескольких сотрудников, при этом информация автоматически об-

новляется, производится резервное копирование данных. Приложение «FileMakerGo 12» позволяет авторизованным пользователям работать с мобильных устройств.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

База данных имеет многоуровневую структуру. Все таблицы базы данных, хранящие информацию, взаимосвязаны (рисунок). Открывая базу, пользователь попадает на основную страницу, на которой в виде раскрывающегося списка представлены все опыты. Для каждого указана основная информация. На этой странице можно создать, удалить опыт целиком или отдельные его этапы, внести изменения в основные параметры, а также проводить поиск по нескольким критериям. С этой страницы можно перейти к поиску и просмотру имеющихся штаммов и праймеров. Отсюда же происходит переход к вкладкам, содержащим полную информацию о любом этапе опыта. При создании нового опыта на его вкладке можно внести разнообразную информацию об источнике исследуемого материала, исходных условиях культивирования и о помещении части исследуемого материала на длительное хранение. В течение эксперимента каждый этап имеет свой «уровень». В зависимости от структуры опыта таких уровней может быть любое количество. При этом с использованием базы легко проследить их связь между собой. Так, например, при расщеплении смешанной микробной биопленки можно получить несколько различающихся колоний. Каждая из этих колоний будет иметь индивидуальную вкладку, при этом все они будут на одном уровне, поскольку имеют один источник. Каждая такая вкладка будет содержать несколько дополнительных сведений — описание, фотографии, биохимическая идентификация и чувствительность к антибиотикам, результаты молекулярно-биологических методов исследования и анализа генов, а также итог, — предназначенных для внесения результатов, получаемых при использовании различных методов для изучения данной колонии. В описании указываются номер культуры, дата, среда, на которой получена данная колония, маркировка пробирок для длительного хранения, вносится описание колонии и ее фотография. Фотографии вносятся в каждый день культивирования. Макрофотографии колоний и микрофотографии мазков сопровождаются подробным описанием. На вкладке «биохимическая идентификация» указываются результаты определения вида (или рода) с помощью традиционных методов, полуавтоматических и автоматических систем. К результатам прикладывается файл с отчетом об анализе, создаваемый системами, или отсканированный бланк. При описании чувствительности исследуемого микроорганизма к антианти-



Структура базы данных «Микроб»

кам результаты можно внести как с помощью выпадающего меню, так и прикрепив файл с таблицей или отчетом полуавтоматической системы определения. На вкладке молекулярно-биологических методов исследования прикрепляются файлы, содержащие последовательность сиквенированного гена или генома, указываются результаты сравнения этих последовательностей с различными, наиболее признанными научным сообществом, базами, приводятся результаты филогенетического анализа. Здесь же хранится аннотация генома. На итоговой вкладке содержится резюме в свободной форме, отражающее наиболее важные, по мнению исследователя, результаты этапа. На разных этапах программа может обращаться к операционной системе для создания на удаленном сервере дополнительных папок.

Разработанная база данных предназначена, в первую очередь, для исследований, направленных на многопрофильное научное изучение большого числа микроорганизмов. База позволяет не только контролировать и анализировать большие объемы разнородной информации, но и оперативно сравнивать данные, полученные разными специалистами, работающими в одном проекте, а также демонстрировать результаты в процессе обучения.

База зарегистрирована в ФИПС 8 августа 2013 г. под № 2013620895.

ЛИТЕРАТУРА

1. Coronel C., Morris S. A., Rob P. Database Systems: Design, Implementation, and Management. — 10th ed. — Cengage Learning, 2012.
2. Flemming H.-C., Wingender J. The biofilm matrix // Nature Publishing Group. Nature Publishing Group. — 2010. — Vol. 8. — № 9. — P. 623–633.
3. Gevers D. et al. The Human Microbiome Project: a community resource for the healthy human microbiome // Plos. Biol. — 2012. — Vol. 10. — № 8. — P. e10013771.
4. McGuire A. L. et al. Ethical, legal, and social considerations in conducting the Human Microbiome Project // Genome Research. — 2008. — Vol. 18. — № 12. — P. 1861–1864.

5. Rossello-Mora R., Amann R. The species concept for prokaryotes // FEMS Microbiology Reviews. — 2001. — Vol. 25. — № 1. — P. 39–67.

6. Tetz V. V. Formation and structure of mixed bacterial communities // APMIS. Blackwell Publishing Ltd. — 1999. — Vol. 107. — № 7–12. — P. 645–654.

7. Tetz V. V., Rybalchenko O. V. Ultrastructure of colony-like communities of bacteria // APMIS. Blackwell Publishing Ltd. — 1997. — Vol. 105. — № 1–6. — P. 99–107.

8. Tisdall J. Beginning Perl for bioinformatics. — O'Reilly Media, Inc., 2009.

9. Turnbaugh P. J. et al. The human microbiome project // Nature. — 2007. — Vol. 449. — № 7164. — P. 804–810.

РЕЗЮМЕ

М. Ф. Вечерковская, В. В. Тец

Создание базы данных для решения задач по систематизации результатов опытов проводимых в ходе научных исследований в микробиологии

Работа посвящена созданию базы данных для решения задач систематизации результатов научных исследований, проводимых в микробиологии. Описан опыт написания и внедрения базы данных «Microbes» на основе системы управления базами данных «FileMakerPro 12». Разработанная база позволяет не только контролировать и анализировать большие объемы разнородной информации, но и оперативно сравнивать данные, полученные разными специалистами, работающими в одном проекте.

Ключевые слова: база данных, систематизация результатов экспериментов, микробиология.

SUMMARY

M. F. Vecherkovskaya, V. V. Tets

Design and implementation of a new database for microbiology research test result classification

The article describes design and implementation of a new personalized database to solve the problems of classification of test results in microbiology. The practice of design and implementation of the database «Microbes» based on the database management system FileMakerPro has been described. The resulting database allows not only for control and for analysis of large amounts of heterogeneous data, but also for dynamic comparison of the data generated through collaboration of multidisciplinary research teams.

Keywords: database, test result classification, microbiology.