



Протокол работы с информационной системой по биоресурсным коллекциям институтов ФАНО России на примере коллекции микроорганизмов

Ф.В. Казанцев^{1, 2}, А.А. Смирнова¹, А.С. Розанов¹, Ю.Е. Уварова¹, Д.А. Афонников^{1, 2}, С.Е. Пельтек¹, С.А. Лашин^{1, 2}

¹ Федеральное исследовательское учреждение Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Во многих научных организациях России имеются коллекции микроорганизмов, по которым накоплены большие объемы информации. Эти данные представляют описание объектов разной природы (бактерии, археи, грибы, протисты) и их свойств, которые были собраны и каталогизированы поколениями исследователей. Не каждая организация, располагающая такими коллекциями, имеет электронный каталог с открытым доступом, что осложняет работу с этими уникальными материалами для самих держателей коллекций, приводит к фактическому отсутствию доступа к биологическим образцам для широкого круга сторонних исследователей. В целях облегчения обмена информацией между держателями коллекций из разных организаций и сообществом исследователей требуется разработка обобщенного электронного каталога коллекций микроорганизмов, обеспечивающего единообразный свободный доступ научного сообщества к информации о большом списке коллекций. Для объединения информации по коллекциям микроорганизмов в рамках проекта по созданию информационной системы для биоресурсных коллекций институтов ФАНО России (<http://www.biores.cytogen.ru>) создан портал биоресурсных коллекций микроорганизмов (<http://www.biores.cytogen.ru/microbes/>), который является площадкой, на которой организации-держатели коллекций могут разместить информацию о единицах хранения своих коллекций, а также другие данные по коллекциям, включая ссылки на собственные каталоги. В настоящей статье мы описываем принципы работы с порталом в рамках направления коллекций микроорганизмов. Графический интерфейс портала позволяет получать пользователям, как зарегистрированным, так и незарегистрированным, следующую информацию о коллекциях микроорганизмов: список коллекций, представленных в базе данных, контактные данные организации и сведения о кураторе коллекции, сводную статистику по каждой из них, а также информацию о единицах хранения. Зарегистрированные пользователи-держатели коллекций имеют возможность создавать и модифицировать записи о единицах хранения своих коллекций, а также актуализировать их описание. Для автоматизации работы с порталом реализован также доступ к базе данных посредством программного протокола (REST API, <http://api.biores.cytogen.ru/microbes/>). В настоящее время происходит наполнение портала, который уже содержит описание более чем 13 тыс. единиц хранения (из них 3.5 тыс. приходится на биоресурсные коллекции микроорганизмов) 65 биоресурсных коллекций организаций ФАНО России (из них 12 коллекций микроорганизмов с суммарным разнообразием фондов порядка 50000 штаммов).

Ключевые слова: биоресурсная коллекция; база данных; интернет-портал; микробиологическая коллекция.

Information system on microbial collections as a part of bioresource collections portal for Russia's FASO organizations: a working protocol

F.V. Kazantsev^{1, 2}, A.A. Smirnova¹, A.S. Rozanov¹, Yu.E. Uvarova¹, D.A. Afonnikov^{1, 2}, S.E. Peltek¹, S.A. Lashin^{1, 2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
² Novosibirsk State University, Novosibirsk, Russia

Nowadays, many scientific organizations of Russia own collections of microorganisms on which large volumes of information have been generated. These data represent the descriptions of objects of diverse nature (bacteria, archaea, fungi, protists) and their properties, which have been carefully collected and cataloged by generations of researchers. Not every organization that has such collections has an open access electronic catalog, which not only complicates work with these unique materials, but also even hides the fact of the existence of such collections. This state of affairs requires the development of electronic resources for presenting these materials to the scientific community. To put together the information on microorganism collections, we have developed an internet portal (<http://www.biores.cytogen.ru/microbes/>) of microbial bioresource collections of FASO organizations in the Russian Federation. The portal was created under the project developing the information system for bioresource collections of FASO institutes. It is a platform where collection organizations can place information about the storage units of their collections, as well as other information on collections, including links to their own catalogs. In this paper, we describe the principles of working with the portal. The portal's graphical interface allows users, both registered and unregistered, to receive the following information about collections of microorganisms: a list of collections represented in the database, contact details of the organization and information about the curator of the collection, summary statistics for each collection, as well as information on storage units. Registered users – owners of collections – have the opportunity to create and modify records about the storage units of their collections, and to update their description. To automate work with the portal, software access to the database through the REST API has been implemented (<http://api.biores.cytogen.ru/microbes/>). At present, the portal is still being filled, but it already contains a description of more than 13,000 items of storage (of which 3500 are in the microorganisms' part)

of 65 bioresource collections in Russia's FASO organizations. Of these collections, 12 with microorganisms have a total diversity of funds of about 50,000 strains).

Key words: bioresource collection; database; internet portal; microbial collection.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Казанцев Ф.В., Смирнова А.А., Розанов А.С., Уварова Ю.Е., Афонников Д.А., Пельтек С.Е., Лашин С.А. Протокол работы с информационной системой по биоресурсным коллекциям институтов ФАНО России на примере коллекции микроорганизмов. Вавиловский журнал генетики и селекции. 2018;22(2):279-284. DOI 10.18699/VJ18.361

HOW TO CITE THIS ARTICLE:

Kazantsev F.V., Smirnova A.A., Rozanov A.S., Uvarova Yu.E., Afonnikov D.A., Peltek S.E., Lashin S.A. Information system on microbial collections as a part of bioresource collections portal for Russia's FASO organizations: a working protocol. Vavilovskii Zhurnal Genetikii i Selekcii = Vavilov Journal of Genetics and Breeding. 2018;22(2):279-284. DOI 10.18699/VJ18.361 (in Russian)

Белки, катализаторы, биологически активные вещества, биотопливо и биополимеры, полученные на основе синтеза в клетках микроорганизмов, широко применяются во многих отраслях экономики, таких как пищевая промышленность, сельское хозяйство, текстильная промышленность, медицина, химическая промышленность, энергетика и др. Основой производства этих веществ служат штаммы-суперпродуценты, выделенные из природных источников и адаптированные для производственных целей путем селекции (Wang et al., 2001) или метаболической инженерии (Larkum et al., 2012; Lee, Kim, 2015). Современная генетическая инженерия, основанная на применении методов системной биологии, позволяет успешно оптимизировать выход биологического продукта у микроорганизмов (Розанов и др., 2014). Однако этап первичного скрининга штаммов на способность к продукции того или иного метаболита остается одним из важнейших и во многом определяет успех дальнейшей работы по созданию суперпродуцентов (Schallmey et al., 2014; Xiao, Lu, 2014). Источником первичного материала для создания штаммов, применяемых в практике и научных исследованиях, служат коллекции микроорганизмов (биоресурсные коллекции), полученных из различных местообитаний. Такие коллекции содержат штаммы вирусов, бактерий, грибов или протистов, которые являются таксономически определенными, хорошо охарактеризованными физиологически, а также генетически стабильными, аутентичными, контролируемыми и воспроизводимыми на больших отрезках времени, при этом их свойства хорошо задокументированы (Overmann, 2015). Повышенный интерес к коллекциям микроорганизмов не случаен и обусловлен активным поиском новых метаболических путей и особых механизмов адаптации к условиям, значительно отличающимся от обычных. Например, богатейшим источником новых ферментов и метаболитов для промышленности, науки и медицины являются экстремофилы (Chen, Jiang, 2018).

Коллекции микроорганизмов представлены в различных странах и научных центрах, большинство из них описаны на сайте Всемирной федерации коллекций культур (WFCC, <http://www.wfcc.info/home/>) (Sun et al., 2015). В настоящее время во многих научных организациях России имеются также коллекции микроорганизмов, по которым накоплены большие объемы информации. Эти данные содержат описание объектов разной природы и их свойств, которые были бережно собраны и каталогизированы поколениями исследователей. Не каждая организация,

обладающая такими коллекциями, имеет электронный каталог с открытым доступом, что осложняет работу с этими уникальными материалами для самих держателей коллекций и приводит к фактическому отсутствию доступа к биологическим образцам для широкого круга сторонних исследователей. Такое положение дел требует разработки и создания электронных ресурсов для их предоставления научному сообществу.

В области исследования микроорганизмов существуют каталоги с описанием коллекций, но, как правило, используемые в них форматы описания данных часто содержат недостаточно информации о единицах хранения. В России довольно много организаций, поддерживающих подобные коллекции. Среди них:

- Всероссийская коллекция микроорганизмов (ВКМ), ФГБУН Институт биохимии и физиологии микроорганизмов РАН, Пущино (<http://www.vkm.ru/rus/index.htm>);
- Коллекция микроорганизмов UNIQEM, ФИЦ Биотехнологии РАН, Москва (<http://www.fbras.ru/katalog-kollekcii-mikroorganizmov-uniqem>);
- Коллекция микроорганизмов для виноделия «Магарач», ФГБУН Всероссийский научно-исследовательский институт виноградарства и виноделия «Магарач» РАН, Ялта (<http://magarach-institut.ru/institut/kollekcziya-mikroorganizmov-vinodeliya-lmagarachr.html>);
- Региональная профилированная коллекция алканотрофных микроорганизмов, Институт экологии и генетики микроорганизмов Уральского отделения РАН, Пермь (<http://iegmc.ru/>) и др.

Как правило, эти ресурсы нацелены на хранение коллекции одной отдельной организации и имеют достаточно подробные форматы описания единиц хранения. Несмотря на то что принцип работы с такими коллекциями достаточно стандартен, работа с десятком ресурсов затруднительна, если необходимо обращаться к ним для оперативного проведения биологических исследований или написания обзорных статей. Таким образом, актуальной задачей стало объединение биоресурсных каталогов с унифицированным интерфейсом обработки данных, предоставляющих средства приведения данных к унифицированному формату, и площадки по каталогизации и хранению данных.

Для объединения данных по коллекциям микроорганизмов в рамках проекта по созданию информационной системы для биоресурсных коллекций институтов ФАНО России создан портал биоресурсных коллекций микроорганизмов, который является площадкой, на которой орга-

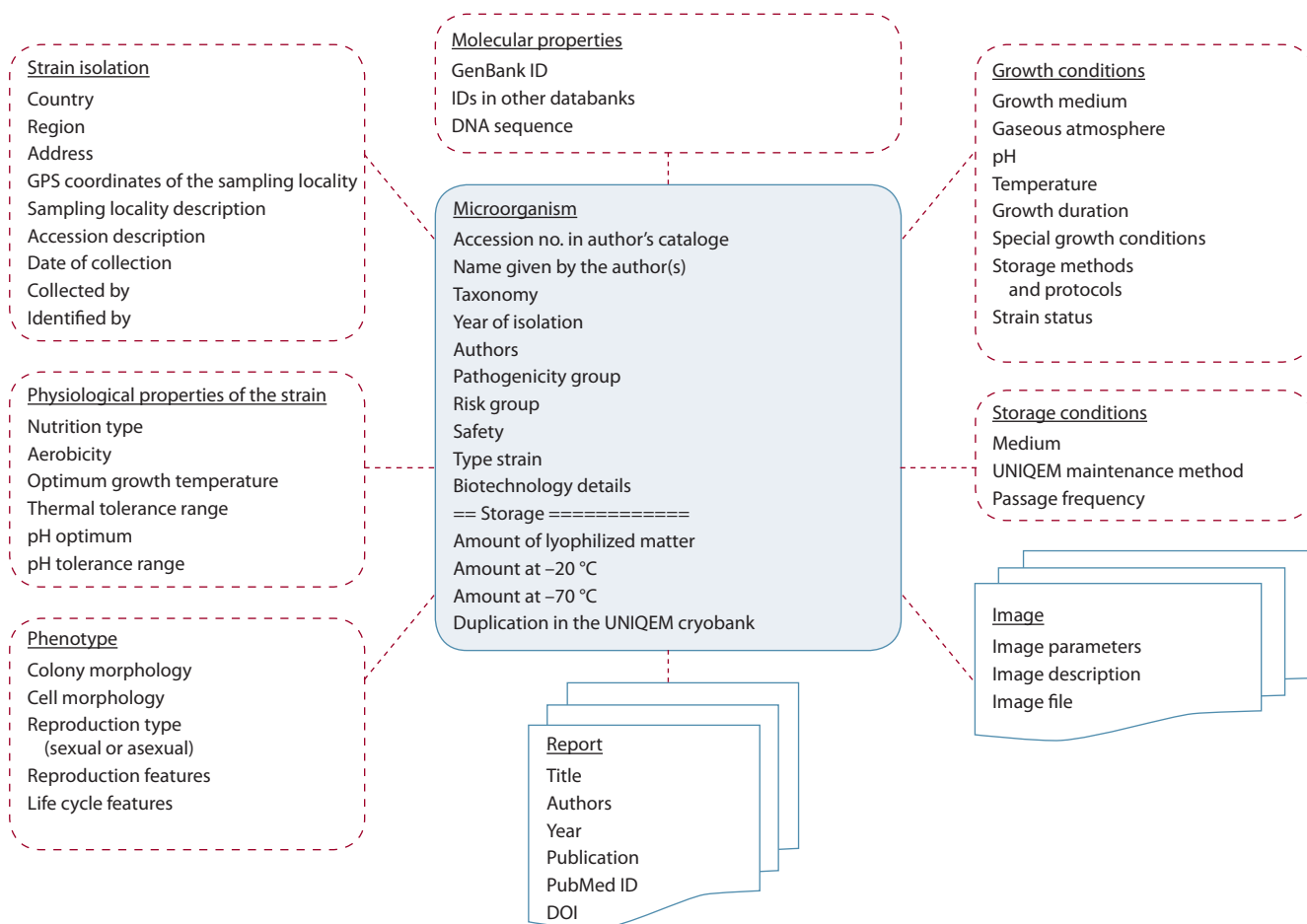


Fig. 1. Description of an accession in the microbial collection.

низации-держатели коллекций могут разместить данные о единицах хранения своих коллекций, а также другую информацию по коллекциям, включая ссылки на собственные каталоги. В этой статье мы описываем принципы работы с порталом в рамках направления коллекций микроорганизмов.

Материалы и методы

Описание единицы хранения базы данных микроорганизмов

Описание единицы хранения коллекции в базе данных включает несколько информационных блоков (рис. 1).

- **Паспортные данные.** Содержит информацию о номере образца в реестре, его таксономическое положение, характеристику патогенности микроорганизма и его биотехнологических свойств. Указываются количество сохраняемого образца (при различных режимах хранения) и информация о наличии дубликатов.
- **Данные о сборе образца** включают расположение и характеристику места сбора, а также список лиц, выделивших и описавших образец.
- **Условия культивирования.** Служит для описания среды и физиологических параметров культивирования, методов и режимов консервации, особых требований к условиям культивирования.

- **Условия сохранения.** Содержит параметры хранения культуры.
- **Физиологические свойства образца.** Содержит информацию о типе питания, оптимальных параметрах для поддержания жизнедеятельности штамма.
- **Молекулярно-генетическая информация.** Содержит информацию об имеющихся геномных последовательностях образца со ссылками на соответствующие базы данных.
- **Фенотипические характеристики.** Дает описание свойств колоний, морфологии клеток, особенностей жизненного цикла.
- **Ссылки на литературные источники.** Содержит данные о публикациях, связанных с образцом (выходные данные журнала, список авторов, ссылка на идентификатор базы PubMed и DOI).
- **Изображения.** Для образца может быть представлено несколько изображений различного характера: внешний вид образца, фотографии геля после электрофореза ДНК, изображения пиков масс-спектрограмм и др.

Реализация базы данных и интерфейса пользователя

База данных каталога коллекций реализована при помощи системы управления базами данных (СУБД) PostgreSQL (<https://www.postgresql.org>). Графический интерфейс поль-

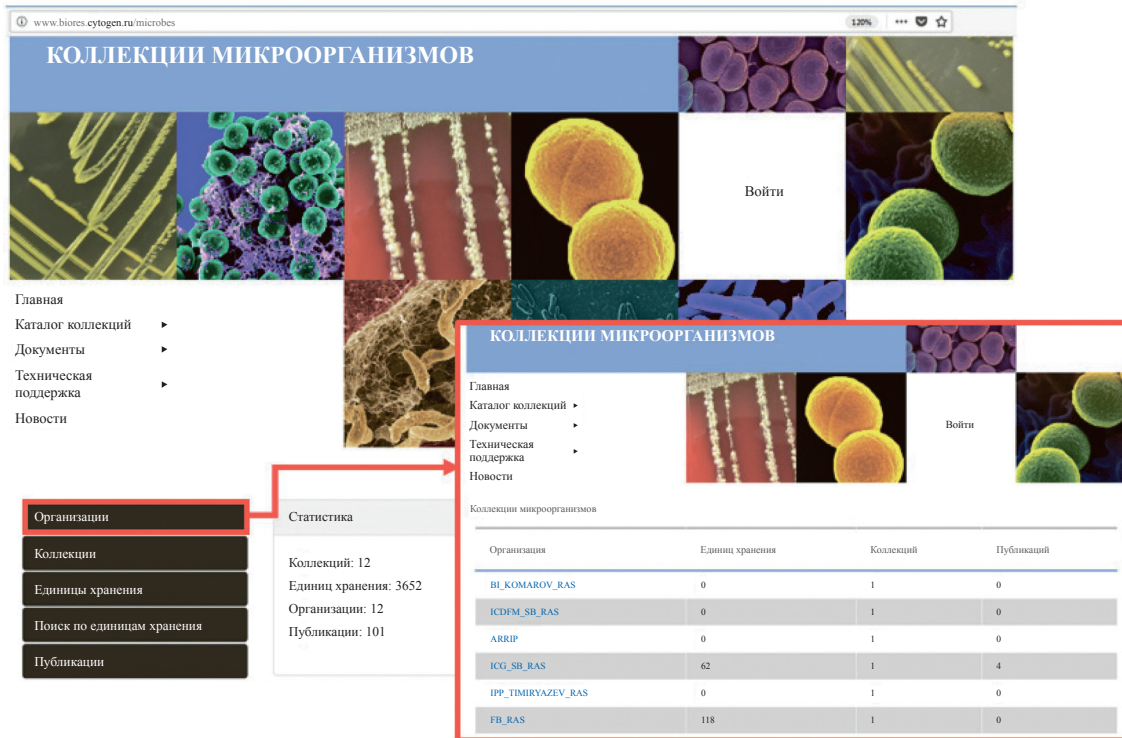


Fig. 2. Graphical user interface of the microbial collections portal.

зователя разработан по схеме Модель–Контроллер–Представление (МКП). Доступ к данным осуществляется через модуль программного интерфейса на основе технологии Representational State Transfer (REST) в специальном формате удаленных запросов к базам данных через Интернет. Программный модуль, обеспечивающий этот интерфейс, реализован на языке Java с применением библиотек Spring (<http://spring.io/>). Графический Web-интерфейс пользователя реализован с использованием инструментария и библиотек DRUPAL 7.0 (<https://www.drupal.org>).

Результаты

Доступ к базе данных. База данных коллекций микроорганизмов доступна по адресу: http://www.biores.cytogen.ru/portal_microbes_collections (короткая ссылка – <http://tiny.cc/wfx8qy>). Незарегистрированный пользователь может получить следующую информацию о коллекциях микроорганизмов: список коллекций, представленных в базе данных, контактные данные организации и сведения о кураторе коллекции, сводную статистику по каждой из них, а также информацию о единицах хранения.

Чтобы информация о новой коллекции была внесена в базу данных и отображалась на портале, необходима регистрация на сайте. Для регистрации необходимо послать заявку администрации портала с указанием лиц, ответственных за поддержание коллекции и свои контактные данные. Зарегистрированный пользователь получает права редактирования информационной карточки своей коллекции (включая контактную информацию и краткое описание коллекции), права на создание и редактирование записей, соответствующих единицам хранения. Для редактирования информации о коллекции можно подключить

несколько пользователей с разграничением ролей: оператор – пользователь(-ли), который вводит/редактирует данные коллекции; куратор коллекции – пользователь, который следит за актуальностью и корректностью введенной операторами информации.

Работа с графическим интерфейсом пользователя.

При входе на портал (<http://www.biores.cytogen.ru/>) пользователь попадает на главную страницу, которая содержит информацию о портале, новости и т. д. (рис. 2). На этой же странице пользователь может авторизоваться (если уже зарегистрирован на портале).

Для того чтобы перейти к интересующей коллекции, пользователь должен выбрать ее из раздела «Каталог коллекций» в меню портала. В результате пользователь попадает на страницу описания коллекции, содержащую контактную информацию о держателе коллекции (которая была введена держателями этой коллекции) и сводную информацию о ее размере.

Чтобы получить информацию о единицах хранения коллекции, нужно выбрать ссылку «Перейти к коллекции», находящуюся под названием коллекции. Данная ссылка может отсутствовать в случае, если организация не разместила информацию о единицах хранения на портале. Если организация имеет собственный электронный каталог, в разделе описания коллекции возможно добавление ссылок на собственный электронный ресурс.

Пользователь попадает на страницу со списком единиц хранения коллекции, сводной статистикой, ссылкой на список публикаций, ассоциированных с этой коллекцией. В списке единиц хранения отображается краткая информация: название штамма, описание образца и место, откуда он был взят.

REST requests to the microbial collection database

URL request	Description
http://api.biores.cytogen.ru/microbes/	Shows a list of API methods and examples of usage.
http://api.biores.cytogen.ru/microbes/stat	3Requests currents statistics of the database.
http://api.biores.cytogen.ru/microbes/collections	Requests a list of registered collections on the portal. Returns both collections identifiers and information on the number of depository units and publications.
http://api.biores.cytogen.ru/microbes/collections/id_col	Requests information about collection with id_col identifier. Returns information on the number of depository units and publications.
http://api.biores.cytogen.ru/microbes/collections/id_col/entities	Requests a list of all depository items stored in a certain collection with id_col identifier. Returns a list of short descriptions.
http://api.biores.cytogen.ru/microbes/collections/id_col/entities/id_entity	Requests the full description of a certain depository item from a particular collection (collection identifier id_col, depository item identifier id_entity).
http://api.biores.cytogen.ru/microbes/publications	Requests a list of all publications registered on the portal. For each publication, associated depository items can be requested.
http://api.biores.cytogen.ru/microbes/images	Requests a list of all images registered on the portal. For each image, associated depository items can be requested.

Выбрав элемент из списка, пользователь переходит на страницу с полным описанием карточки единицы хранения, содержащей следующие разделы: основные характеристики штамма, информация о выделении штамма, условия культивирования, молекулярно-генетические свойства, фенотип, физиология, условия хранения, описание таксономии, изображение (важные характеристики штамма, представленные в графическом виде).

Регистрация и пополнение информации о коллекциях. Чтобы ввести информацию коллекции на портале, необходимо зарегистрироваться. Для этого необходимо связаться с администрацией портала по адресу: biores-support@bionet.nsc.ru. После подтверждения регистрации пользователь может вносить правки.

Для авторизации на портале (<http://www.biores.cytogen.ru/>) пользователю нужно:

- 1) перейти по ссылке «Войти», расположенной в правой верхней части страницы;
- 2) ввести в форму полученные при регистрации логин и пароль, после чего пользователь попадает на страницу своей учетной записи, где ему доступна возможность ее редактирования;
- 3) редактирование информации о коллекции доступно пользователям с правом внесения первичной информации в конкретную базу данных биоресурсной коллекции. Для этого пользователь должен перейти к коллекции через раздел «Каталог коллекций» главного меню портала, где по ссылке «Редактировать» пользователю открывается форма для внесения и редактирования описания коллекции;
- 4) при переходе к списку единиц хранения коллекции пользователю с правом внесения первичной информации в конкретную базу данных биоресурсной коллекции

доступна ссылка для внесения новой карточки единицы хранения и редактирования уже существующих.

Программный доступ к данным коллекций. Для решения задач компьютерного поиска и сбора статистики по данным портала биоресурсных коллекций реализован программный интерфейс доступа (Application Programming Interface, API) по технологии REST. В случае базы данных коллекций микроорганизмов интерфейс расположен по адресу (<http://api.biores.cytogen.ru/microbes/>). REST API позволяет получать необходимую информацию через серию URL запросов. Данные из базы возвращаются в виде структурированного текста в формате JSON (<http://json.org/>). JSON-объект может быть открыт любым текстовым редактором или обработан одним из множества программных инструментов. Большинство универсальных сред моделирования (Matlab, Scilab и др.) и сред программирования на языках высокого уровня, таких как Python, R, C++, Java, имеют инструментарию по работе с JSON текстами. Пример работы с базой данных через REST API на языке Python можно найти в работе (Kazantsev et al., 2017).

REST API портала имеет набор унифицированных запросов для выводов информации различного типа. В таблице описан формат формирования таких запросов. Запросы могут быть выполнены в строке Интернет-браузера или приложения, написанного на каком-либо языке программирования.

Содержание портала. В настоящее время база данных коллекций микроорганизмов включает описание 3.5 тыс. единиц хранения из 12 микробиологических коллекций ФАНО России с суммарным разнообразием фондов порядка 50000 штаммов. Ведется его активное наполнение.

Заклучение

Разработку и наполнение информационных ресурсов, подобных порталу, представленному в нашей статье, можно рассматривать в нескольких аспектах. С одной стороны, хранение информации о микробиологических ресурсах организаций России и их фондах в единой базе данных с унифицированным описанием единиц хранения и с унифицированным же доступом к данным имеет большое утилитарное значение для различных отраслей отечественной микробиологии. С другой стороны, анализ всего содержимого базы данных с помощью современных методов работы с большими данными, несомненно, будет иметь и фундаментальное значение.

Acknowledgments

This work was supported by the Federal Agency for Scientific Organizations, program for the support of bioresource collections, project 0324-2017-0050.

Conflict of interest

The authors declare no conflict of interest.

References

Chen G.Q., Jiang X.R. Next generation industrial biotechnology based on extremophilic bacteria. *Curr. Opin. Biotechnol.* 2018;50:94-100. DOI 10.1016/j.copbio.2017.11.016.

- Kazantsev F., Akberdin I., Lashin S., Ree N., Timonov V., Ratushny A., Khlebodarova T., Likhoshvai V. MAMMOTH: a new database for curated MAtheMatical Models of bioMOlecular sysTems. *J. Bioinform. Comput. Biol.* 2017. DOI 10.1142/S0219720017400108.
- Larkum A.W., Ross I.L., Kruse O., Hankamer B. Selection, breeding and engineering of microalgae for bioenergy and biofuel production. *Trends Biotechnol.* 2012;30(4):198-205. DOI 10.1016/j.tibtech.2011.11.003.
- Lee S.Y., Kim H.U. Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* 2015;33(10):1061-1072. DOI 10.1038/nbt.3365.
- Overmann J. Significance and future role of microbial resource centers. *Syst. Appl. Microbiol.* 2015;38(4):258-265. DOI 10.1016/j.syapm.2015.02.008.
- Rozanov A.S., Kotenko A.V., Akberdin I.R., Peltek S.E. Recombinant strains of *Saccharomyces cerevisiae* for ethanol production from plant biomass. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding.* 2014;18(4/2):989-998. (in Russian)
- Schallmeyer M., Frunzke J., Eggeling L., Marienhagen J. Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. *Curr. Opin. Biotechnol.* 2014;26:148-154.
- Sun Q., Liu L., Wu L., Li W., Liu Q., Zhang J., Ma J. Web resources for microbial data. *Genomics Proteomics Bioinformatics.* 2015;13(1):69-72. DOI 10.1016/j.gpb.2015.01.008.
- Xiao Z., Lu J.R. Strategies for enhancing fermentative production of acetoin: a review. *Biotechnol. Adv.* 2014;32(2):492-503. DOI 10.1016/j.biotechadv.2014.01.002.
- Wang Z., Zhuge J., Fang H., Prior B.A. Glycerol production by microbial fermentation: a review. *Biotechnol. Adv.* 2001;19(3):201-223. DOI 10.1016/S0734-9750(01)00060-X.