

**КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ
УНИВЕРСИТЕТ**
Институт фундаментальной медицины и биологии

**И. И. ЗАДОРИНА, М. И. МАРКЕЛОВА, И. Г. СТАРОСТИНА,
З. И. ИСХАКОВА, Е. А. БУЛЫГИНА**

ДЕПОНИРОВАНИЕ ЧЕРНОВЫХ СБОРОК ГЕНОМОВ В NSVI

Учебно-методическое пособие

КАЗАНЬ

2024

УДК 004.9

ББК 28.0

315

*Печатается по рекомендации учебно-методической комиссии
Института фундаментальной медицины и биологии КФУ
(протокол № 1 от 29.08.2024 г.)*

Рецензенты:

Кандидат биологических наук, старший преподаватель кафедры генетики ИФМиБ КФУ **Д. Э. Журавлева**, кандидат биологических наук, старший преподаватель кафедры микробиологии ИФМиБ КФУ **Д. С. Бульмакова**.

Задорина И. И. Депонирование черновых сборок геномов в NCBI. Учебно-методическое пособие / И. И. Задорина, М. И. Маркелова, И. Г. Старостина, З. И. Исхакова, Е. А. Булыгина // Казань, Казанский федеральный университет, 2024. – 40 с.

В учебно-методическом пособии приведена пошаговая схема депонирования сборки в NCBI. Рекомендовано для научно-исследовательской работы и выпускной квалификационной работы медицинских и биологических направлений.

© Задорина И. И., Маркелова М. И., Старостина И. Г.,
Исхакова З. И., Булыгина Е. А.
© КФУ, 2024

Содержание

Введение	4
1.1. Требования к геномной сборке	8
2. Регистрация и загрузка данных в NCBI	10
2.1. Заполнение вкладки Submitter	15
2.2. Заполнение вкладки General information	16
2.3. Заполнение вкладки BioProject General Info.....	22
2.4. Заполнение вкладки Sample Type	25
2.5. Заполнение вкладки BioSample Attributes	31
2.6. Заполнение вкладки Source	36
2.7. Заполнение вкладки Files	36
2.8. Заполнение вкладки Assignment	38
2.9. Заполнение вкладки References	39
2.10. Заполнение вкладки Review & Submit	40
Заключение.....	42
Список использованных источников	43
Список электронных ресурсов	43

Введение

National Center for Biotechnology Information – Национальная медицинская библиотека, является крупнейшей в мире биомедицинской библиотекой и лидером в исследованиях в области вычислительной медицинской информатики. Играет ключевую роль во внедрении биомедицинских исследований в практику. Исследовательские и информационные службы поддерживают научные открытия, здравоохранение и охрану общественного здоровья. Внедряют новые способы повышения доступности биомедицинских данных и информации, создают инструменты для улучшения управления данными и личного здоровья. Позволяют исследователям бесплатно использовать огромное количество биомедицинских данных для валидации и сравнения собственных результатов. В структуре имеется много подразделений, интересующее подразделение: Национальный центр биотехнологической информации (NCBI).

Конгресс учредил Национальный центр биотехнологической информации в 1988 году как подразделение NLM и как национальный ресурс информации по молекулярной биологии. С тех пор NCBI является ведущим источником общедоступных биомедицинских баз данных, программных средств для анализа молекулярных и геномных данных и исследований в области вычислительной биологии

NCBI создает и поддерживает более 40 баз данных для медицинского и научного сообщества, к ним относятся литературные, молекулярные и геномные базы данных. Основной базой данных литературы NCBI является PubMed, которая предоставляет аннотации

и цитаты для миллионов статей из тысяч биомедицинских журналов. Записи PubMed содержат ссылки на полнотекстовые версии статей из центрального электронного архива PubMed NCBI (PMC) и с веб-сайтов журналов, а также ссылки на соответствующую информацию с других сайтов NCBI, такую как геномные и молекулярные базы данных.

Некоторые из основных геномных и генетических ресурсов NCBI:

- GenBank – аннотированная коллекция всех общедоступных последовательностей ДНК;
- RefSeq – коллекция последовательностей ДНК, РНК и белка;
- SRA – архив сырых прочтений, в котором хранятся данные высокопроизводительного секвенирования;
- dbSNP – базу данных однонуклеотидных полиморфизмов;
- dbGaP – базу данных, которая объединяет данные о генотипах и фенотипах из общегеномных источников;
- GWAS – полногеномные исследования ассоциаций;
- ClinVar – коллекция зарегистрированных взаимосвязей между генетическими вариациями и наблюдаемым состоянием здоровья;
- реестр генетического тестирования,
- база данных генетических тестов
- ресурс по обнаружению патогенов, который объединяет данные о последовательности бактериальных патогенов, полученных из пищевых продуктов, окружающей среды и пациентов-людей, для облегчения активного наблюдения за патогенами и болезнями пищевого происхождения в режиме реального времени;

- ClinicalTrials.gov – база данных клинических исследований, проводимых по всему миру, финансируемых из частных и государственных источников.

Дополнительные базы данных включают информацию о химических структурах и их биологической активности, белковых последовательностях, структуре белка и экспрессии генов, таксономии и многом другом.

Развитие технологий секвенирования следующего поколения (Next-generation sequencing, высокопроизводительное секвенирование) позволило снизить стоимость прочтения геномов, а также повлекло за собой развитие науки-биоинформатики. И в 00-е – 10-е годы XXI века начался экспоненциальный рост количества расшифрованных геномов, в том числе геномов прокариот (Рисунок 1). В настоящее время база данных NCBI содержит информацию о 2,5 миллионах геномах различных организмов. Наибольшее количество геномов принадлежат бактериям (2,17 миллионов геномов) и эукариотам (42,5 тыс. геномов).

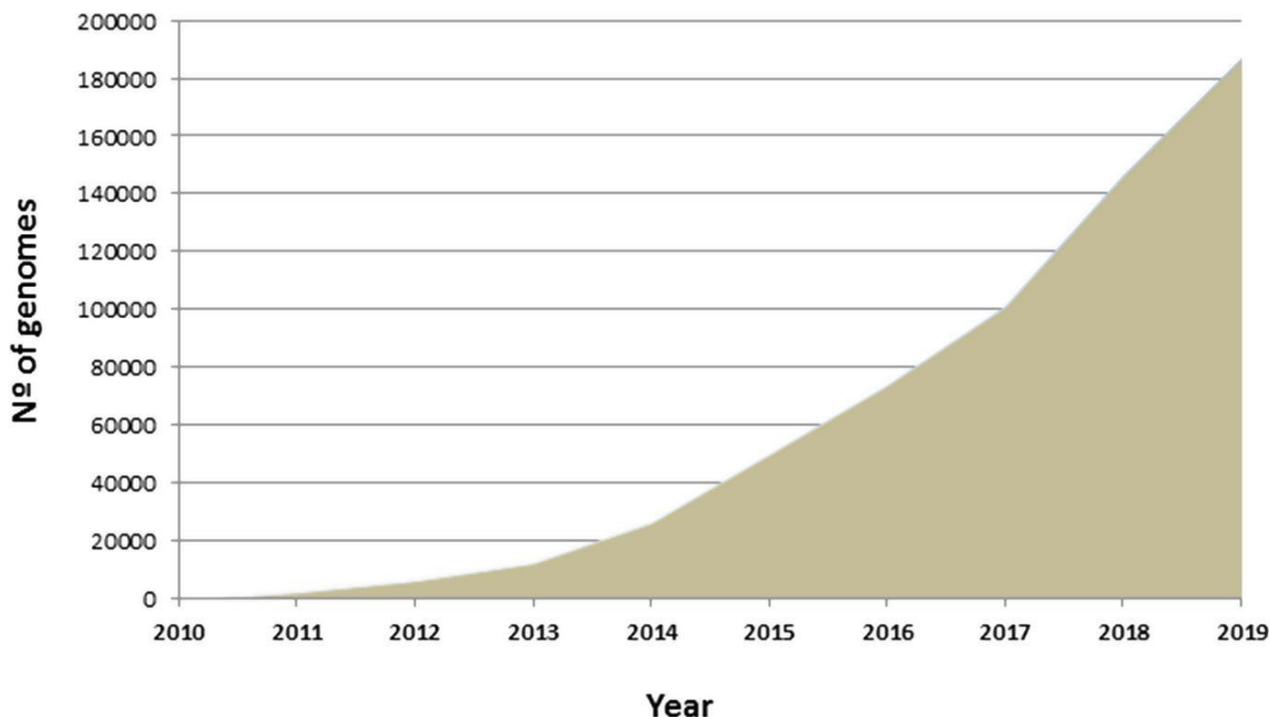


Рисунок 1 – Число прокариотических геномов, опубликованных с 2010 по 2019 гг. [Carro *et al.*, 2021]

Одним из этапов исследования, включающего секвенирование генома бактерии, является депонирование геномной сборки в общедоступные мировые базы данных. Депонирование данных обеспечивает большую прозрачность, большее доверие к вашему исследованию и лучшую воспроизводимость результатов. Практически все высокорейтинговые журналы (как зарубежные, так и российские) требуют предварительной публикации геномныхборок и/или сырых прочтений. Базы данных NCBI не только позволяют хранить информацию о собранном геноме бактерии или археи, но и проводить автоматическую аннотацию генома с помощью программы NCBI Prokaryotic Genome Annotation Pipeline (PGAP), что позволяет адекватно сравнивать разные депонированные геномы между собой, избегая ошибок, которые могут возникнуть из-за использования

разных методов аннотации. Кроме того, сама программа PGAP регулярно обновляется, и все геномы, депонированные в RefSeq и Genbank, переаннотируются с использованием последней версии PGAP (несколько раз в год), что позволяет всегда иметь доступ к актуальной информации.

1.1. Требования к геномной сборке

Перед подготовкой к депонированию в любые общедоступные базы, убедитесь, что полученная вами сборка полная и не контаминирована последовательностями других организмов. Оценку качества сборки можно произвести с помощью программы CheckM [Parks *et al.*, 2015].

Типичная команда для запуска CheckM:

```
>checkm lineage_wf -t 8 -x fasta /dir_with_fasta /output_checkm
```

где:

- `-t` – количество используемых ядер процессора, зависит от конкретного компьютера или сервера
- `-x fasta` – обозначение типа файла, в котором записана геномная сборка (может быть `fasta`, `fa`, `fna`, `faa`)
- `/dir_with_fasta` – путь до папки, где лежит файл с геномной сборкой, будут проанализированы все файлы формата `fasta`, находящиеся в данной папке
- `/output_checkm` – папка с результатами анализа CheckM

Для депонирования максимальная контаминация сборки не должна превышать 5%, для полноты геномной сборки нет четких критериев, но не менее 80% для геномных сборок из метагеномных

прочтений (Metagenome-assembled genome, MAG), и около 90% для классических геномных сборок. При невозможности достижения указанных параметров стоит воздержаться от депонирования генома и повторить предыдущие этапы анализа – выделение чистой культуры, выделение ДНК, подготовку библиотек и секвенирование с достаточным покрытием и геномную сборку.

Также перед депонированием необходимо самостоятельно верифицировать таксономическую принадлежность исследуемой бактерии. Для этого необходимо выявить наиболее похожий геном микроорганизма:

- Выровнять последовательность маркерных генов вашего микроорганизма на базу данных нуклеотидных последовательностей NCBI с помощью программы BlastN (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Такими генами обычно являются 16S рРНК, *gyrA*, 23S рРНК, *rpoB*, *gyrB*, *dnaK* и другие. Данный способ не всегда позволяет точно определить вид, т.к. в силу настоящего состояния систематики, часто несколько видов или даже родов бактерий имеют схожие последовательности маркерных генов. Например, *Escherichia* и *Shigella* не отличаются по последовательности 16S рРНК.

- После предыдущего этапа скачать полные геномы бактерий, которые имеют максимальное сходство последовательности маркерных генов с маркерным геном вашей исследуемой бактерии. Провести анализ Average nucleotide identity (ANI) для сравнения полной последовательности геномов. Обычно, ANI между геномами одного вида превышает 95%.

- При наличии значительных вычислительных мощностей для определения таксономической принадлежности можно использовать методы, применяемые в метагеномных исследованиях, например MetaPhlAN. Это позволяет выравнивать одновременно около 200 маркерных генов исследуемого генома на референсную базу данных.

- Кроме того, для определения таксономической принадлежности можно картировать всю геномную сборку на референсную базу прокариотических геномов NCBI с помощью BlastN – только на персональном компьютере или сервере, в веб-версии BlastN невозможно.

После определения видовой или родовой (при невозможности определения вида допускается вид назвать как sp.) принадлежности выберите имя для штамма. Важно, чтобы оно не повторялось с уже депонированными штаммами во избежание путаницы. Дублирование имени штамма допустимо в случае повторного секвенирования генома конкретного лабораторного штамма (например, *E. coli* Nissle 1917).

2. Регистрация и загрузка данных в NCBI

Заходите на сайт NCBI (<https://www.ncbi.nlm.nih.gov/>), как видно из рисунка 2 на текущей странице представлено несколько вкладок, нажимаете на вкладку *Submit*.

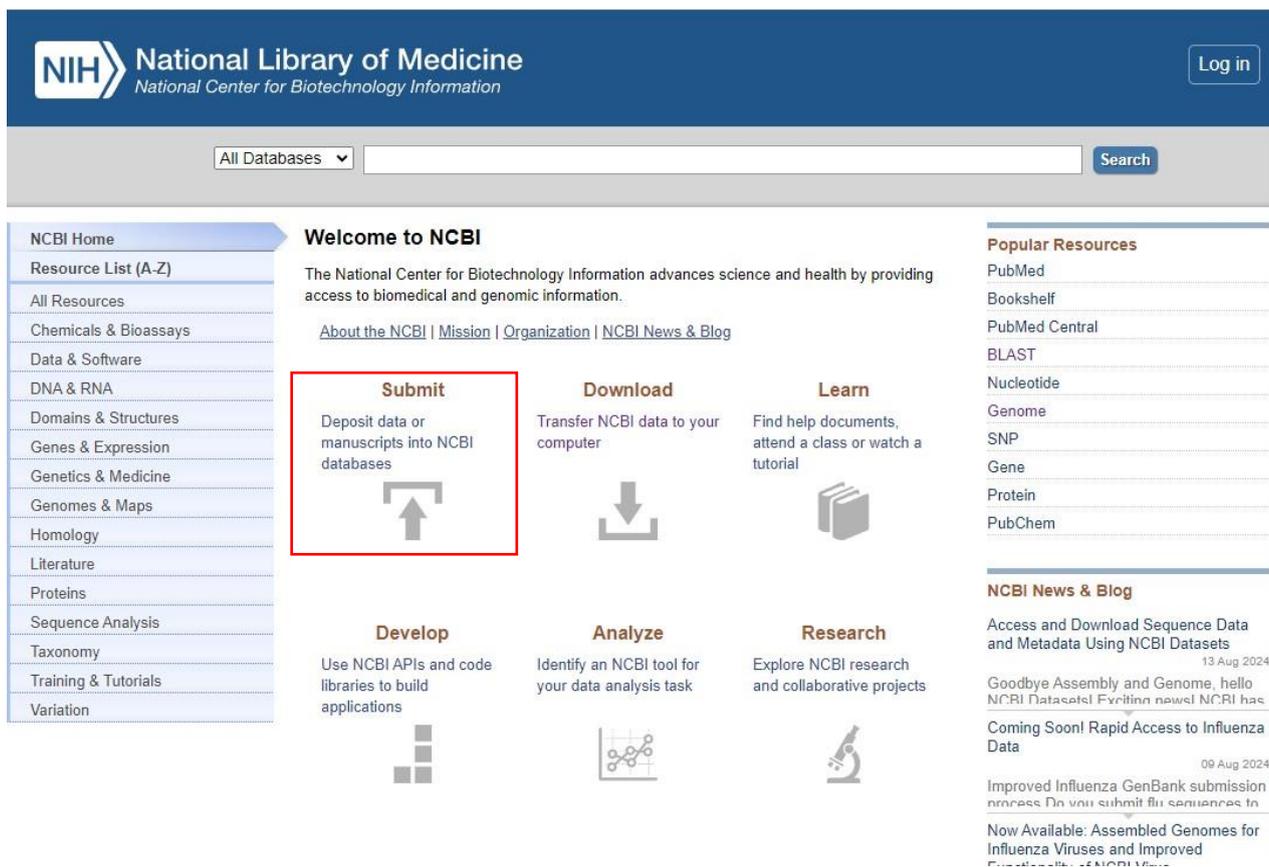


Рисунок 2 – Главная страница NCBI

Переходите на GenBank, выбираете вкладку собранные (Рисунок 3) геномы эукариот и прокариот Submit assembled eukaryotic and prokaryotic genomes (WGS or Complete), нажимаете **Submit**.

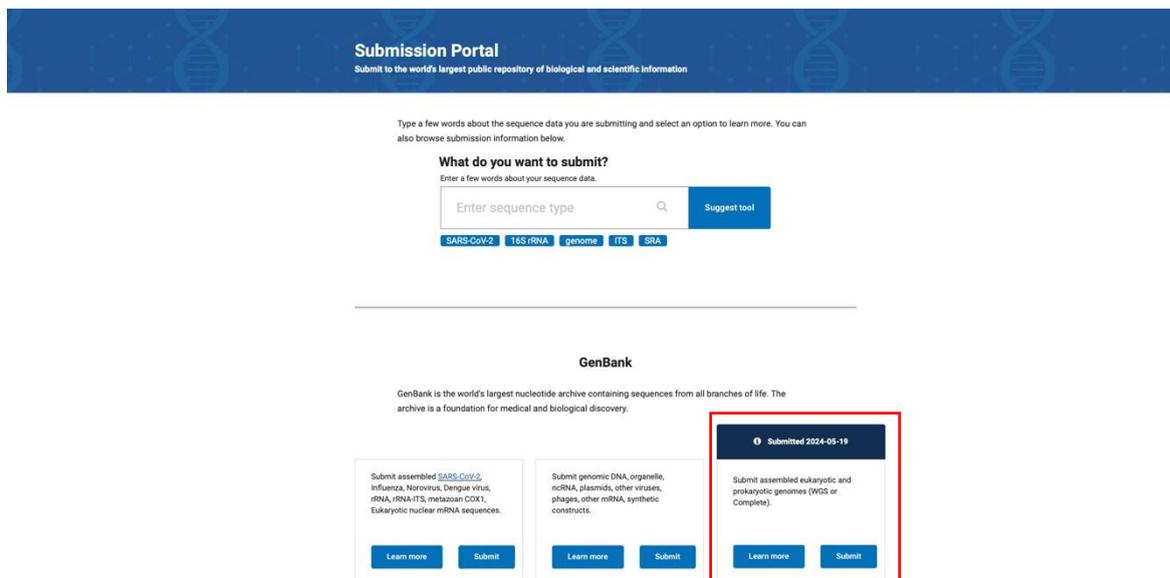


Рисунок 3 – Вкладки отправки последовательностей

Для работы с NCBI необходима регистрация учётной записи. Нажимаете на **Log in** (Рисунок 4), после чего выбираете варианты входа в систему (Рисунок 5).

Удобным способом является вход через аккаунт Google, при его наличии, что не требует создания и запоминания дополнительных логинов и паролей.

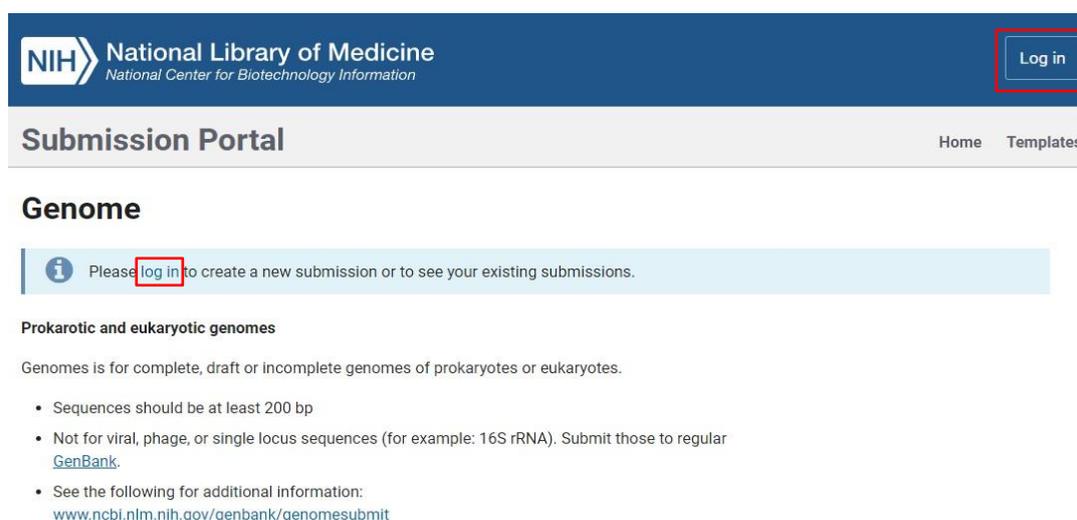


Рисунок 4 – Вход в систему для подачи заявок

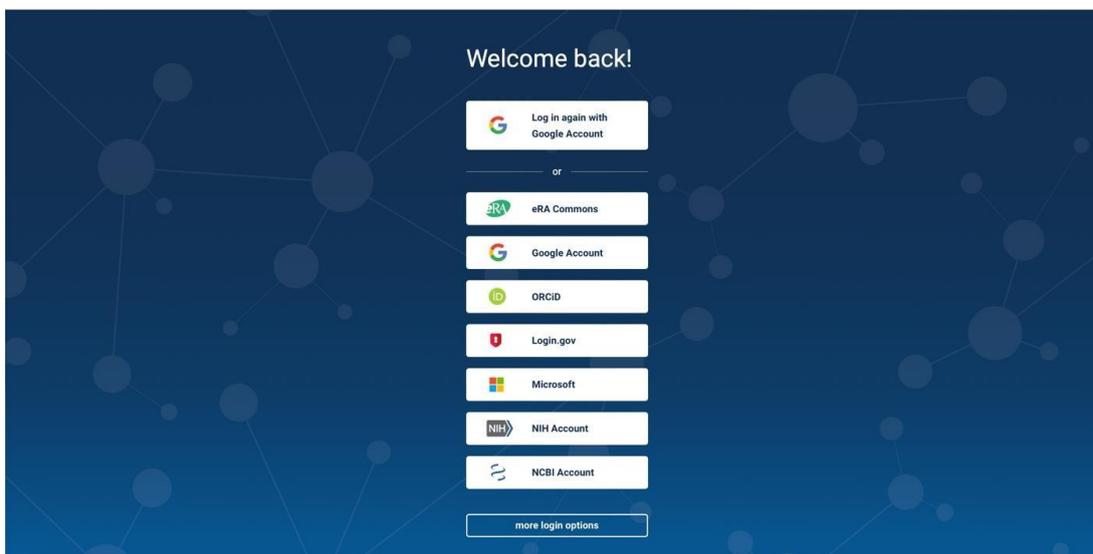


Рисунок 5 – Дополнительные варианты входа в систему NCBI

После входа в систему, можете ознакомиться с предложенными вариантами загрузки генома на портал. Наиболее простым способом является загрузка через веб-браузер, которая не требует установки каких-либо расширений или навыков работы с командной строкой, но позволяет выгрузить файлы общим объемом не более 2Гб, чего вполне достаточно для депонирования бактериальных геномов. Если необходимо будет выгружать сырые риды в Short read archive (SRA) или партией несколько сотен бактериальных геномов (до 10Гб), то необходимо установить плагин Aspera Connect plugin. Для этого в веб-браузере Chrome перейдите по ссылке (https://chromewebstore.google.com/category/extensions?hl=ru&utm_source=ext_sidebar), в строке поиска вводите «IBM Aspera Connect», кликните на появившийся плагин, нажимаете на «Установить». После выбора способа загрузки данных, нажимаете на *New Submission*, как показано на Рисунке 6.

Submission Portal Home **My submissions** Manage data Templates My profile

Genome **New submission**

i **New:** NCBI has a Foreign Contamination Screen (FCS) tool suite available in [GitHub](#) and now running on new submissions to help improve the quality of your genome submissions. See [NCBI Insights](#) and our preprint for more details.

Note: This Genomes wizard is not for viral, phage, or single locus sequences (for example: 16S rRNA). Submit those to regular [GenBank](#).

Prokaryotic and eukaryotic genomes

Genomes is for complete, draft or incomplete genomes of prokaryotes or eukaryotes.

- Sequences should be at least 200 bp
- Not for viral, phage, or single locus sequences (for example: 16S rRNA). Submit those to regular [GenBank](#).
- See the following for additional information: www.ncbi.nlm.nih.gov/genbank/genomesubmit

Options to upload data:

Upload via Web browser or Aspera browser plugin UPDATED FEBRUARY 2024 +

Upload via Aspera command line or FTP UPDATED FEBRUARY 2024 +

i **Info:** You have not submitted anything yet.

Рисунок 6 – Загрузка нового проекта

Далее выбираете, как отправить свои данные: если один геном, ставите *Single genome*, несколько — *Batch/multiple genomes*, гаплотипы — *Pseudohaplotypes (also called haplotypes) of one or more diploid assemblies*. После нажимаете на *Continue* (Рисунок 7).

Submission Portal Home **My submissions** Manage data Templates My profile

Genome submission: SUB14664884 Delete submission

New

Submission Type Required fields are marked with * asterisk

*** How do you want to submit your data?**

Single genome

Manually complete a web form to describe one genome assembly and to upload its sequences.

Batch/multiple genomes (maximum 400 per submission)

Use this to submit at most 400 genomes that have some common information. Provide or fill in the "Genome Info" file, a tab-delimited text file that describes each of your genome assemblies and their attributes/metadata, plus the genome sequences. Use one file per genome.

Information that must be common to all genomes in the batch are:

- BioProject
- (initial) release date
- assembly type (either WGS or non-wgs, not a mix of both types)
- file type (FASTA or SQN)
- gap/Ns details
- publication information (for FASTA submissions only)
- PGAP request status (Yes/No; for prokaryotic genomes only)
- [See more details here](#)

Pseudohaplotypes (also called haplotypes) of one or more diploid assemblies

NEW More haplotype options are available and the columns of the embedded table are reordered to be more intuitive

The haplotypes of a diploid or polyploid assembly have the same restrictions as Batch/Multiple (see above) and must also have:

- the BioSample of the individual sequenced
- separate BioProjects, which will be connected by an Umbrella BioProject
- the haplotype pairs identified as principal and alternate when one is much better than the other; haplotype 1 and haplotype 2 when they are of similar quality; or maternal and paternal when that information is known.

For more details see [Submitting Multiple Haplotype Assemblies](#)

Continue

Рисунок 7 – Тип отправки

2.1. Заполнение вкладки **Submitter**

Во вкладке **Submitter** заполняете все обязательные поля, они обозначены звездочкой. Вводите информацию о вас (имя, фамилия, электронная почта) и о вашей организации (организация, отделение, улица, город, индекс, страна), также можете заполнить и

дополнительные поля (дополнительную электронную почту, адрес сайта организации, телефон, факс), после чего нажимаете на **Continue** (Рисунок 8). Данные необходимо заполнять аккуратно, т. к. они будут опубликованы на странице с депонированным геномом.

Submission Portal Home My submissions Manage data Templates My profile

Genome submission: SUB14664884 Delete submission
new generic WGS submission

1 SUBMITTER 2 GENERAL INFO 3 FILES 4 ASSIGNMENT 5 REFERENCES 6 REVIEW & SUBMIT

Submitter Required fields are marked with * asterisk

* First (given) name Middle name * Last (family) name
Ivan Ivanov

* Email (primary) Email (secondary)
ivan.ivanov@gmail.com

At least one email should be from the organization's domain.

Group for this submission
No group (affiliation from my personal profile)
You can create a group for shared submissions in your profile

* Submitting organization Submitting organization URL * Department
Kazan Federal University Institute of Fundamental Medicine and B

Phone Fax
Street * City State/Province * Postal code
Kremlevskaya Kazan 420

* Country
Russia

Continue

Рисунок 8 – Заполненная вкладка *Submitter*

2.2. Заполнение вкладки **General information**

Заполнение *General information*:

1. Do you already a BioProject accession number for this research?

Если ранее не создавали в GenBank проект для этого генома, группы геномов или сырых прочтений в SRA, вставляете в окно *No*. Будет

создан новый уникальный идентификатор BioProject.

2. Do you already have BioSample accession number for there samples? Если ранее не создавали BioSample для этого генома, вставляете в окно *No*. Будет создан новый уникальный идентификатор BioSample.

3. When should this submission be released to the public? Выберите *Release following processing*, чтобы геном стал доступен сразу после аннотации. Кроме того, есть возможность сделать геном доступным в определенную дату, но это нужно сделать как максимум в день подачи публикации в рецензируемый журнал, чтобы редактор и рецензенты смогли иметь доступ к депонированным геномам. *Release on specified date or upon publication, whichever is first* (Рисунок 9) и вводите в поле *Projected release date* дату в формате «год-месяц-день» (Рисунок 10). Если вы хотите дополнить существующий проект, в поле *Do you already have a BioProject accession number for this research?* вставляете *Yes* и вписываете номер существующего проекта (*PRJNAXXXXXX*), если нужно дополнить *BioSample*, в поле *Do you already have BioSample accession numbers for these samples?* вставляете *Yes* и вписываете номер BioSample в формате (*SAMNXXXXXXXXX*) (Рисунок 9).

Genome submission: SUB14664884

[Delete submission](#)

New

1 SUBMITTER 2 GENERAL INFO 3 FILES 4 ASSIGNMENT 5 REFERENCES 6 REVIEW & SUBMIT

General Information

Required fields are marked with * asterisk

BioProject

The BioProject bundles the data for this research project.

* Do you already have a BioProject accession number for this research?

 Yes No (a BioProject will be created within this submission)

BioSample

The BioSample stores the detailed metadata of the sample that was sequenced.

* Do you already have BioSample accession numbers for these samples?

 Yes No (BioSamples will be created within this submission)

Release date

Note: Release of BioProject or BioSample is also triggered by the release of linked data.

* When should this submission be released to the public?

 Release following processing
 Release on specified date or upon publication, whichever is firstРисунок 9 – Заполнение *General info*

1 SUBMITTER 2 GENERAL INFO 3 FILES 4 ASSIGNMENT 5 REFERENCES 6 REVIEW & SUBMIT

General Information

Required fields are marked with * asterisk

BioProject

The BioProject bundles the data for this research project.

* Do you already have a BioProject accession number for this research?

 Yes No (a BioProject will be created within this submission)

* Existing BioProject

PRJNAXXXXXX

BioSample

The BioSample stores the detailed metadata of the sample that was sequenced.

* Do you already have BioSample accession numbers for these samples?

 Yes No (BioSamples will be created within this submission)

* Existing BioSample

SAMNXXXXXXXX

Release date

Note: Release of BioProject or BioSample is also triggered by the release of linked data.

* When should this submission be released to the public?

 Release following processing
 Release on specified date or upon publication, whichever is first

* Projected release date

YYYY-MM-DD

Рисунок 10 – Заполнение *General information* для дополнения проекта

Обязательны для заполнения поля, обозначенные звездочкой, у каждого поля сверху имеется вопросительный знак, если на него кликнуть, открывается подсказка по заполнению (Рисунок 11).

Genome assembly metadata

Genome Assembly structured comment is in the contig .sqn file(s)

Assembly date Name of algorithm, eg Newbler OR SOAPdenovo. Select the assembly program from the pulldown list, or enter the name of the program in the box.

* Assembly method ? Version or date program was run ? Delete

[Add another assembly method](#)

Assembly name ? If you have a meaningful assembly name like UCLA_Agam_2.1 (see naming recommendations), please provide it here, otherwise we will auto-generate it.

* Genome coverage ?

* Sequencing technology ? Delete

[Add another sequencing technology](#)

Рисунок 11 – Продолжение заполнения вкладки *General information*

Заполнение *Genome info*:

1. *Assembly date* – дата сборки в формате «год-месяц-день».
2. *Assembly method* – вписываете название программы, которая использовалась для сборки генома: например, **SPAdes**, **Unicycler** и т.п.
3. *Version or date program was run* – вписываете версию используемой программы, например, **4.0.0**, либо дату запуска, если сборка была произведена в онлайн-версии программы.
4. *Assembly name* – можете вписать название геномной сборки, иначе оно будет сформировано автоматически.
5. *Genome coverage* – покрытие генома при секвенировании,

например, **100x**. Считается самостоятельно по формуле:

$$\text{Покрытие} = \frac{\text{Количество прочтений} * \text{длину прочтения}}{\text{Размер геномной сборки}}$$

При расчете количества прочтений необходимо учесть, в каком режиме было осуществлено секвенирование – парноконцевом (paired-end) или одноконцевом (single-end). При парноконцевом секвенировании не забудьте умножить количество ридов на 2.

Длина прочтения также зависит от режима секвенирования – 50 – 300 п.о. при секвенировании на платформе Illumina в зависимости от модели секвенатора. Размер геномной сборки – общая длина собранного генома. Ее можно оценить с помощью программы QUAST [Gurevich *et al.*, 2013].

6. Sequencing Technology – прибор, на котором было произведено секвенирование генома, например, **Illumina MiSeq**

7. Did your sample include the full genome? – Ваша сборка содержит полный геном? Если полнота сборки более 90%, выберите *Yes*, если меньше, то не загружайте такой бактериальный геном, эта функция полезна для сборок отдельных хромосом эукариот.

8. Is this the final version? Если это окончательная версия сборки, то выберите *Yes*. Даже если у вас в дальнейшем появится возможность улучшить сборку путем секвенирования длинных прочтений, вы сможете обновить текущую версию сборки, сделанный на этом этапе выбор не является окончательным.

9. Is it a de novo assembly? Если это *de novo* сборка, выберите *Yes*. Если сборка производилась с использованием референсного

генома, то выберите *No* и укажите идентификатор геномной сборки, которая была использована в качестве референса.

10. *Is it an update of existing submission?* Если вы впервые загружаете данную сборку, выберите *No*. Если это обновление уже депонированной сборки генома, то выберите *Yes* и заполните идентификатор старой версии сборки.

11. *Do not automatically trim or remove sequences identified as contamination* – НЕ ставьте галочку, чтобы при выявлении последовательностей-артефактов (индексы, адаптеры) на концах контигов они были автоматически удалены, иначе этим придется заниматься самостоятельно. При обнаружении артефактов внутри контига, вам придет письмо от команды NCBI с просьбой самостоятельно удалить артефакты и разделить контиги, где они обнаружились на несколько штук. Даже при аккуратной обрезке технических последовательностей, выполненной до сборки генома, на этапе депонирования они могут быть обнаружены, это не является критерием плохой сборки.

12. *Select a category for your submission.* Если депонируемый геном был секвенирован вашей лабораторией (даже при заказе секвенирования у сторонних организаций), выберете *Original*. Если вы собрали геном из сырых данных, ранее депонированных кем-то, кто не имеет отношения к вашей рабочей группе, выберете *Third Party Data*.

13. После чего нажимаете *Continue* (Рисунок 12).

Genome info

Genome assembly metadata

Genome Assembly structured comment is in the contig .sqn file(s)

Assembly date ⓘ

*** Assembly method** ⓘ *** Version or date program was run** ⓘ **Delete**

[Add another assembly method](#)

Assembly name ⓘ
 ⓘ If you have a meaningful assembly name like UCLA_Agam_2.1 ([see naming recommendations](#)), please provide it here, otherwise we will auto-generate it.

*** Genome coverage** ⓘ

*** Sequencing technology** ⓘ **Delete**

[Add another sequencing technology](#)

*** Did your sample include the full genome?**
 Yes (even for draft genomes or if a prokaryotic genome assembly may not include plasmids)
 No, I deliberately selected a subset of the genome (e.g. only one chromosome of a eukaryote or only the non-repetitive regions of the genome)

*** Is this the final version?** ⓘ
 Yes No

*** Is it a *de novo* assembly?**
 Yes No

*** Is it an update of existing submission?**
 Yes No

Do not automatically trim or remove sequences identified as contamination

ⓘ GenBank staff will automatically remove contaminants that are found to be the entire sequence or at the end of a sequence, and will post the reports and edited fasta file to the submission portal. Any Ns at the end of sequences will be removed, and sequences shorter than 200bp after trimming will be removed. Note that internal contamination will not be automatically removed since the sequence may be misassembled and therefore should be split at the contamination and resubmitted as separate sequences.

Submission Category

*** Select a category for your submission:** ⓘ
 Original (directly sequenced by submitter)
 Third Party Data (derived from other primary sequence data)

[Read about Third Party data \(TPA\) submission requirements](#)

Submission title ⓘ

Private comments to NCBI staff ⓘ

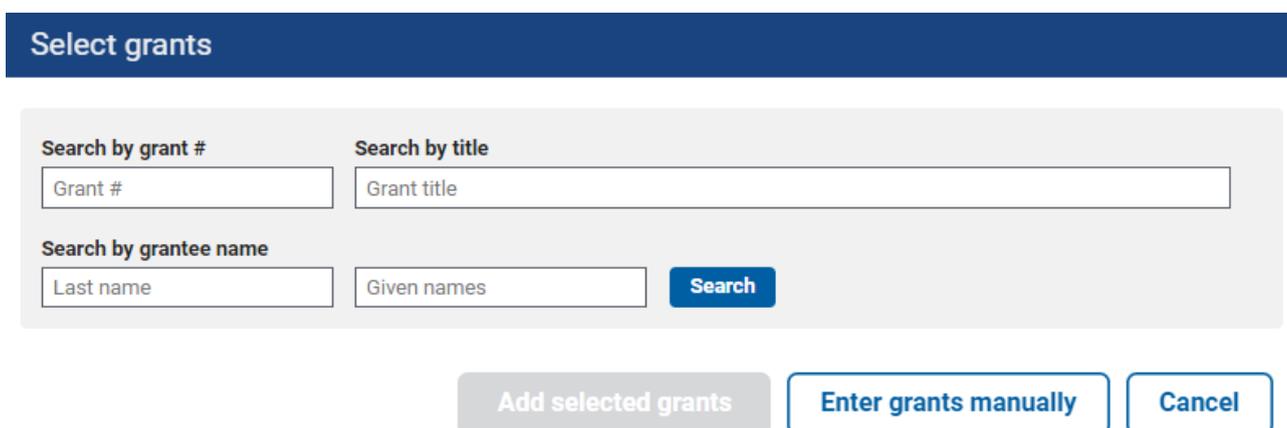
Continue

Рисунок 12 – Заполненная вкладка *General information*

2.3. Заполнение вкладки BioProject General Info

В обязательном поле *Public description* описываем цель исследования, актуальность (допустимый объем - 4000 символов), в дополнительном поле *Relevance* можете отметить область, более

подходящую вашему исследованию, например, *Agricultural, Medical, Evolution, Model organism* и т.д. В **Relevance description** описываете его, также можно ввести сайт, имеющий отношение к данному исследованию (при необходимости можем добавить сайт, для этого нажимаете на **Add another link**). Кроме того, можете добавить название и идентификатор гранта (но депонирование данных в базу NCBI не считается рецензируемой публикацией), для этого нажимаете на **Add grants** и заполняете всплывающее окно (Рисунок 13), после чего нажимаете на **Continue** (Рисунок 14).



Select grants

Search by grant #
Grant #

Search by title
Grant title

Search by grantee name
Last name Given names Search

Add selected grants Enter grants manually Cancel

Рисунок 13 – Окно *Select grants*

BioProject General Info

★ Public description

Lactiplantibacillus plantarum p999 isolated from quail cecum to characterize the probiotic potential for chicken

3888 characters left

Relevance

Agricultural 

External links

Description

URL

Delete



 [Add another link](#)

Select your grants

-  Use this tool to look up grants from many subscribed governmental funding agencies (eg NIH, CDC, FDA and VA) and some non-governmental funding sources (eg HHMI). You can search by grant number, title or grantee name. If your grant is not included, you can select the "Add grants manually" option within this tool to add your grant.

 [Add grants](#)

Consortium name

Consortium URL

Continue

Рисунок 14 – Вкладка *BioProject General Info*

2.4. Заполнение вкладки *Sample Type*

Во вкладке *Sample Type* имеются три пакета, выбираете подходящий пакет фильтров, который лучше всего описывает ваш геном: *All packages* (рис. 15), *Packages for MAG submitters* (Рисунок 16), *Packages for metagenome submitters* (Рисунок 17). Пакет *All packages* разделяется на *NCBI packages* (используется, когда нецелесообразно использовать минимальную информацию) и *GSC MxS packages for genomes, metagenomes, and marker sequences* (для формального описания и стандартизации метаданных образца), в поле ***(Optional) Filter packages by organism name*** для автоматической фильтрации можете ввести полное название микроорганизма, например *Escherichia coli* и нажать на ***Reset and show all packages***, после чего автоматически подберутся предположительно подходящие пакеты. Выбираете в правой колонке *GSC MxS packages for genomes, metagenomes, and marker sequences* -> ***MIGS Cultured Bacterial/Archaeal***, если вы депонируете классическую сборку генома культивируемой бактерии, после чего выбираете подходящую среду обитания - (Рисунок 18), после чего нажимаете на ***Continue***. От выбора настоящего пакета зависит набор полей, необходимых для заполнения.

Наличие на портале загрузки разных наборов пакетов, характеризующих геномы, является попыткой унификации информации, заполняемой для каждого генома организма, выделенного из соответствующего источника. Не пытайтесь найти пакет с минимальным количеством полей. Отсутствие необходимой информации об организме, чей геном вы депонируете, не позволит

другим исследователям использовать ваши данные для сравнения. Дополнительно набор полей, необходимых для заполнения представлены по пакетам на рисунках 19 - 23.

Sample Type

★ Select the package that best describes your samples.

All packages Packages for MAG submitters Packages for metagenome submitters

(Optional) Filter packages by organism name

Enter the full scientific name of your samples, e.g., *Escherichia coli*

Reset and show all packages

- To filter for relevant BioSample packages, enter the **full scientific name** of the organism of your samples.
- If your BioSamples are derived from a species **not represented in NCBI's Taxonomy database**, enter the genus-level name, e.g., *Escherichia*
- If your BioSamples are derived from **more than one organism**, enter the common species, genus, or family, e.g., *Enterobacteriaceae*
- If your BioSamples are **metagenomic/environmental**, or **metagenome-assembled genomes (MAG)**, select the appropriate tab above
- For more information about organism names, see [Organism information](#).

NCBI packages [More...](#)

- Pathogen**
Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.
- One Health Enteric**
Use for microbial isolates that are collected for genomic surveillance of enteric pathogens. Sample spaces include the following: 1. human/animal hosts; 2. food samples; 3. food facilities; 4. environmental samples (farm, water, and the environment).
US public health agencies have created customized versions of this package that include more specific guidance, controlled vocabulary picklists, and sub-packages for each of the 4 sample types.
 - [GitHub repository](#)
 - [Validation for the OHE package](#)
- Microbe**
Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use [MixS](#), Pathogen or Virus packages.
- Model organism or animal**
Use for multicellular samples or cell lines derived from common laboratory model organisms, e.g., mouse, rat, *Drosophila*, worm, fish, frog, or large mammals including zoo and farm animals.
- Metagenome or environmental**
Use for metagenomic and environmental samples when it is not appropriate or advantageous to use [MixS](#) packages.
- Invertebrate**
Use for any invertebrate sample.
- Human**
WARNING: Only use for human samples or cell lines that have no privacy concerns. For all studies involving human subjects, it is the submitter's responsibility to ensure that the information supplied protects participant privacy in accordance with all applicable laws, regulations and institutional policies. Make sure to remove any direct personal identifiers from your submission. If there are patient privacy concerns regarding making data fully public, please submit samples and data to NCBI's [dbGaP](#) database. [dbGaP](#) has controlled access mechanisms and is an appropriate resource for hosting sensitive patient data. For samples isolated from humans use the Pathogen, Microbe or appropriate [MixS](#) package.
- Plant**
Use for any plant sample or cell line.

GSC [MixS](#) packages for genomes, metagenomes, and marker sequences [More...](#)

- MIGS Cultured Bacterial/Archaeal**
Use for cultured bacterial or archaeal genomic sequences. Organism must have lineage [Bacteria](#) or [Archaea](#).
- MIGS Eukaryotic**
Use for eukaryotic genomic sequences. Organism must have lineage [Eukaryota](#).
- MIMAG Metagenome-assembled Genome**
Use for metagenome-assembled genome sequences produced using computational binning tools that group sequences into individual organism genome assemblies starting from metagenomic data sets. Organism cannot contain the term 'metagenome'. Use the MIUVIG package for virus genomes. Before creating BioSamples for prokaryotic and eukaryotic MAGs, please read and follow the [MAG submission instructions](#).
- MIMS Environmental/Metagenome**
Use for environmental and metagenome sequences. Organism must be a metagenome, where lineage starts with [unclassified sequences](#) and scientific name ends with 'metagenome'.
- MISAG Single Amplified Genome**
Use for single amplified genome sequences produced by isolating individual cells, amplifying the genome of each cell using whole genome amplification, and then sequencing the amplified DNA. Organism cannot contain the term 'metagenome'.

Continue

Рисунок 15 – Пакет *All packages*

Sample Type

★ Select the package that best describes your samples.

[All packages](#) [Packages for MAG submitters](#) [Packages for metagenome submitters](#)

NCBI packages [More...](#)

- Pathogen**
Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.

GSC [MixS](#) packages for genomes, metagenomes, and marker sequences [More...](#)

- MIMAG Metagenome-assembled Genome**
Use for metagenome-assembled genome sequences produced using computational binning tools that group sequences into individual organism genome assemblies starting from metagenomic data sets. Organism cannot contain the term 'metagenome'. Use the MIUVIG package for virus genomes. Before creating BioSamples for prokaryotic and eukaryotic MAGs, please read and follow the [MAG submission instructions](#).

Continue

Рисунок 16 – Пакет *Packages for MAG submitters*

Sample Type

★ Select the package that best describes your samples.

[All packages](#) [Packages for MAG submitters](#) [Packages for metagenome submitters](#)

NCBI packages [More...](#)

- Metagenome or environmental**
Use for metagenomic and environmental samples when it is not appropriate or advantageous to use [MixS](#) packages.
- Pathogen**
Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.

GSC [MixS](#) packages for genomes, metagenomes, and marker sequences [More...](#)

- MIMS Environmental/Metagenome**
Use for environmental and metagenome sequences. Organism must be a metagenome, where lineage starts with [unclassified sequences](#) and scientific name ends with 'metagenome'.

Continue

Рисунок 17 – Пакет *Packages for metagenome submitters*

- MIGS Cultured Bacterial/Archaeal**
Use for cultured bacterial or archaeal genomic sequences.
Organism must have lineage [Bacteria](#) or [Archaea](#).

- No environmental package
- agriculture
- air
- built
- food-animal and animal feed
- food-farm environment
- food-food production facility
- food-human foods
- host-associated
- human-associated
- human-gut
- human-oral
- human-skin
- human-vaginal
- hydrocarbon resources-cores
- hydrocarbon resources-fluids/swabs
- microbial mat/biofilm
- miscellaneous or artificial
- plant-associated
- sediment
- soil
- symbiont-associated
- wastewater/sludge
- water

Рисунок 18 – *MIGS Cultured Bacterial/Archaeal*

NCBI packages

Use when it is not advantageous to use the Genome Standards Consortium (GSC) minimum information ([MixS](#)) packages.

[Less](#)

- Pathogen**
Use for pathogen samples that are relevant to public health.
Required attributes include those considered useful for the rapid analysis and trace back of pathogens.

- Pathogen: clinical or host-associated
- Pathogen: environmental/food/other
- Combined pathogen submission

Рисунок 19 – *Pathogen*

MIGS Eukaryotic
Use for eukaryotic genomic sequences. Organism must have lineage [Eukaryota](#).

- No environmental package
- agriculture
- air
- built
- food-animal and animal feed
- food-farm environment
- food-food production facility
- food-human foods
- host-associated
- human-associated
- human-gut
- human-oral
- human-skin
- human-vaginal
- hydrocarbon resources-cores
- hydrocarbon resources-fluids/swabs
- microbial mat/biofilm
- miscellaneous or artificial
- plant-associated
- sediment
- soil
- symbiont-associated
- wastewater/sludge
- water

Рисунок 20 – *MIGS Eukaryotic*

MIMAG Metagenome-assembled Genome
Use for metagenome-assembled genome sequences produced using computational binning tools that group sequences into individual organism genome assemblies starting from metagenomic data sets. Organism cannot contain the term 'metagenome'. Use the MIUVIG package for virus genomes. Before creating BioSamples for prokaryotic and eukaryotic MAGs, please read and follow the [MAG submission instructions](#).

- No environmental package
- agriculture
- air
- built
- food-animal and animal feed
- food-farm environment
- food-food production facility
- food-human foods
- host-associated
- human-associated
- human-gut
- human-oral
- human-skin
- human-vaginal
- hydrocarbon resources-cores
- hydrocarbon resources-fluids/swabs
- microbial mat/biofilm
- miscellaneous or artificial
- plant-associated
- sediment
- soil
- symbiont-associated
- wastewater
- water

Рисунок 21 – *MIMAG Metagenome-assembled Genome*

- MIMS Environmental/Metagenome**
Use for environmental and metagenome sequences. Organism must be a metagenome, where lineage starts with [unclassified sequences](#) and scientific name ends with 'metagenome'.

- agriculture
- air
- built
- food-animal and animal feed
- food-farm environment
- food-food production facility
- food-human foods
- host-associated
- human-associated
- human-gut
- human-oral
- human-skin
- human-vaginal
- hydrocarbon resources-cores
- hydrocarbon resources-fluids/swabs
- microbial mat/biofilm
- miscellaneous or artificial
- plant-associated
- sediment
- soil
- symbiont-associated
- wastewater
- water

Рисунок 22 – *MIMS Environmental/Metagenome*

- MISAG Single Amplified Genome**
Use for single amplified genome sequences produced by isolating individual cells, amplifying the genome of each cell using whole genome amplification, and then sequencing the amplified DNA. Organism cannot contain the term 'metagenome'.

- No environmental package
- agriculture
- air
- built
- food-animal and animal feed
- food-farm environment
- food-food production facility
- food-human foods
- host-associated
- human-associated
- human-gut
- human-oral
- human-skin
- human-vaginal
- hydrocarbon resources-cores
- hydrocarbon resources-fluids/swabs
- microbial mat/biofilm
- miscellaneous or artificial
- plant-associated
- sediment
- soil
- symbiont-associated
- wastewater
- water

Рисунок 23 – *MISAG Single Amplified Genome*

2.5. Заполнение вкладки BioSample Attributes

В разделе *Attributes* (Рисунок 24). В *Sample name*: прописываете название образца (оно должно быть уникальным, кратким информативным), например, *AG10*, в поле *Organism*: (к какому организму относится), например, *Lactiplantibacillus plantarum*, в *Strain*: штамм, например, *AG10* (Рисунок 25). О выборе имени штамма см. пункт 1.1. «Требования к геномной сборке» настоящего пособия.

Genome submission: SUB14664884 Delete submission

New

1 SUBMITTER 2 GENERAL INFO 3 BIOPROJECT GENERAL INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SOURCE 7 FILES 8 ASSIGNMENT 9 REFERENCES 10 REVIEW & SUBMIT

Attributes Required fields are marked with * asterisk

Package MIGS: cultured bacteria/archaea, food-animal and animal feed; version 6.0

* Sample Name

* Organism

* strain

Рисунок 24 – Раздел *Attributes*

Attributes Required fields are marked with * asterisk

Package MIGS: cultured bacteria/archaea, food-animal and animal feed; version 6.0

* Sample Name

* Organism

* strain

Рисунок 25 – Заполненный раздел *Attributes*

В разделе *Environment* (Рисунок 26):

1. **Collection date:** указать дату сбора образца, дату вносите в формате: «день-месяц-год», «месяц-год», «год».

2. **Broad-scale environmental context:** выбираете наиболее подходящий тип окружающей среды, из которой был собран ваш образец, в формате ENVO (можете воспользоваться сайтом <https://bioportal.bioontology.org/>), (например, desert biome, ENVO:01000179 - https://bioportal.bioontology.org/ontologies/ENVO/?p=classes&conceptid=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FENVO_01000179).

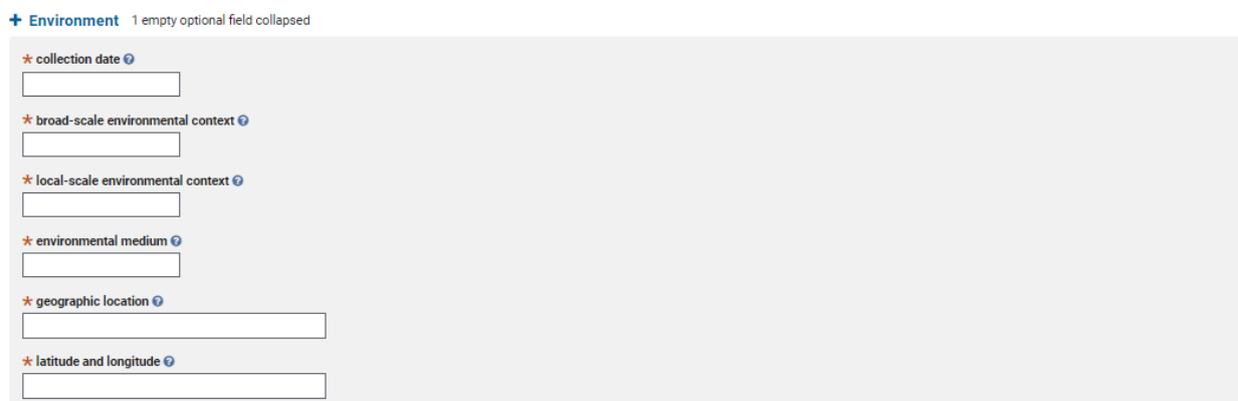
3. **Local-scale environmental context:** выбираете факторы окружающей среды, оказывающие влияние на объект во время отбора проб (например, shoreline, ENVO:00000486 - https://bioportal.bioontology.org/ontologies/ENVO/?p=classes&conceptid=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FENVO_00000486).

4. **Environmental medium:** выбираете факторы, которые идентифицируют материал во время отбора (например, soil, ENVO:00001998 - https://bioportal.bioontology.org/ontologies/ENVO/?p=classes&conceptid=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FENVO_00001998).

5. **Geographic location:** указываете географическое происхождение образца, например, *Russia*.

6. **Latitude and longitude:** вписываете географические координаты места, где был взят образец, например, 55.79 N 49.12 E для г. Казань (Рисунок 27).

В разделе *Environment* есть дополнительное поле для заполнения (*1 empty optional field collapsed*), для того чтобы его открыть можете кликнуть на плюс (Рисунок 28).

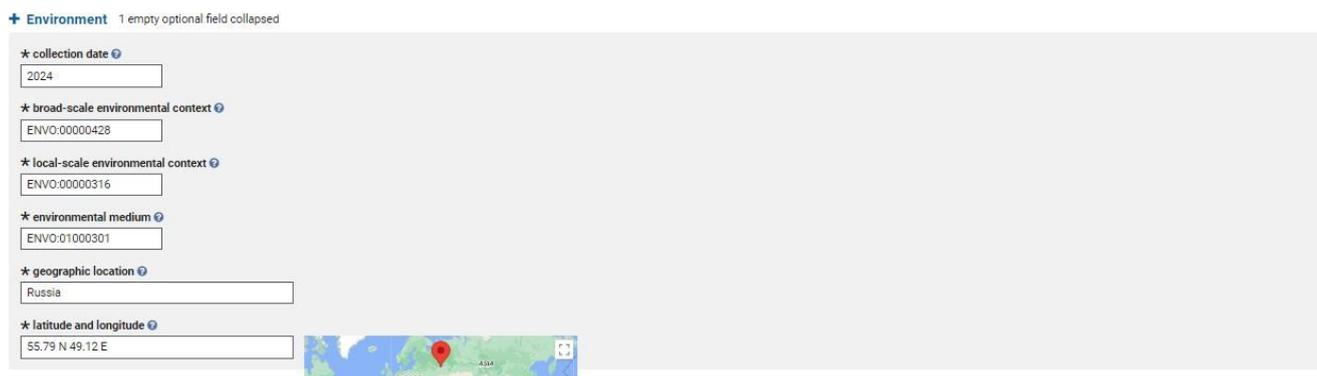


The screenshot shows the 'Environment' section of a form. At the top, there is a header '+ Environment 1 empty optional field collapsed'. Below this, there are seven input fields, each with a red asterisk and a help icon:

- * collection date
- * broad-scale environmental context
- * local-scale environmental context
- * environmental medium
- * geographic location
- * latitude and longitude

All fields are currently empty.

Рисунок 26 – Раздел *Environment*

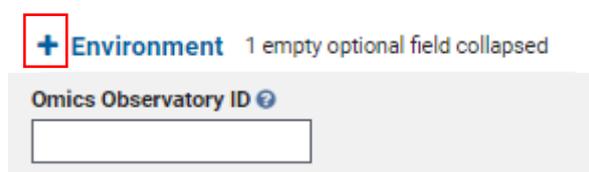


The screenshot shows the 'Environment' section of a form, now filled with data. The header is '+ Environment 1 empty optional field collapsed'. The fields contain the following information:

- * collection date: 2024
- * broad-scale environmental context: ENVO:0000428
- * local-scale environmental context: ENVO:0000316
- * environmental medium: ENVO:0100301
- * geographic location: Russia
- * latitude and longitude: 55.79 N 49.12 E

Below the 'latitude and longitude' field, there is a small map showing a red location pin over Russia.

Рисунок 27 – Заполненный раздел *Environment*



This is a close-up screenshot of the 'Environment' section. The header '+ Environment 1 empty optional field collapsed' is visible. Below it, there is a field labeled 'Omics Observatory ID' with a help icon. The field is currently empty. A red square highlights the plus sign icon in the header.

Рисунок 28 – Дополнительное поле в разделе *Environment*

В разделе *Nucleic Acid Sequence Source* (Рисунок 29):

1. ***Isolation and growth condition:*** указываете PubMed ID или ссылку на статью с описанием метода выделения.

2. ***Number of replicons:*** 1 – для бактерии, если геномная сборка не содержит плазмиды, или вы в этом не уверены. Если точно знаете, сколько плазмид в депонируемой вами сборке, посчитайте их вместе с хромосомой. Например, геномная сборка содержит хромосому и две плазмиды – укажите значение 3.

3. ***Reference for biomaterial:*** публикация, в которой впервые упоминается организм.

Также в *Nucleic Acid Sequence Source* есть дополнительные поля для заполнения, для того чтобы их открыть нужно кликнуть на плюс (Рисунок 30). Остальные обязательные разделы заполняете, которые выделены звездочками, в зависимости от выбранной среды обитания (Рисунок 31), еще по желанию можете заполнить дополнительный раздел *Custom Attributes* (Рисунок 32), после чего нажимаете на ***Continue.***

The screenshot shows a form titled '+ Nucleic Acid Sequence Source' with a sub-header '17 empty optional fields collapsed'. The form is divided into two main sections. The first section contains three required fields, each marked with a red asterisk and a blue question mark icon: 'isolation and growth condition', 'number of replicons', and 'reference for biomaterial'. The second section contains three optional fields: 'isolation source', 'negative control type', and 'positive control type'. Each field is represented by a text input box.

Рисунок 29 – Раздел *Nucleic Acid Sequence Source*

+ Nucleic Acid Sequence Source 17 empty optional fields collapsed

* isolation and growth condition ⓘ

* number of replicons ⓘ

* reference for biomaterial ⓘ

isolation source ⓘ

negative control type ⓘ

positive control type ⓘ

Рисунок 30 – Заполненный раздел *Nucleic Acid Sequence Source*

+ Food-animal and animal feed 81 empty optional fields collapsed

* collection site geographic feature ⓘ

* food product origin geographic location ⓘ

* food production characteristics ⓘ

* food product type ⓘ

* Interagency Food Safety Analytics Collaboration (IFSAC) category ⓘ

* intended consumer ⓘ

* purpose of sampling ⓘ

Рисунок 31 – Раздел *Food-animal and animal feed*

Custom Attributes

Name

Value

Delete

Рисунок 32 – Раздел *Custom Attributes*

2.6. Заполнение вкладки *Source*

В разделе *Prokaryote source* в поле *Annotate this prokaryotic genome in the NCBI Prokaryotic Annotation Pipeline (PGAP) before its release* (для того, чтобы сделать аннотацию с помощью PGAP) - *Yes*, в *Bacteria and/or source DNA is available from* (указываете название и адрес лаборатории или частного лица, а также идентификатор коллекции культур), после чего нажимаете на *Continue* (Рисунок 33).

The screenshot shows the 'Submission Portal' interface. At the top right, there are navigation links: Home, My submissions (active), Manage data, Templates, and My profile. Below this, the submission ID 'SUB14664884' is displayed with a 'Delete submission' button. A progress bar indicates the current step is '6 SOURCE'. The 'Source' section is titled 'Prokaryote source' and contains a form with the following elements:

- A text input field labeled 'Bacteria and/or source DNA is available from'.
- A required field question: '★ Annotate this prokaryotic genome in the NCBI Prokaryotic Annotation Pipeline (PGAP) before release?' with radio buttons for 'Yes' (selected) and 'No'.
- An informational note: 'Note that annotation is not required, but you may request to have your prokaryotic genome annotated by PGAP before it is released. The pre-release annotated genome will be posted back to this submission for you to view. If no Release date has been requested in this submission, the PGAP-annotated genome will be released upon completion of processing. In addition, PGAP is now available as a stand-alone software package to run on your own machine, local cluster or the Cloud.'
- A blue 'Continue' button at the bottom left, which is highlighted with a red rectangular box.

Рисунок 33 – Вкладка *Source*

2.7. Заполнение вкладки *Files*

В разделе *Files for Submission* в поле *Which of these 3 options describes this genome submission?* выбираете из 3 вариантов:

1. *Each chromosome is in a single sequence and there are no extra sequences* (если каждая хромосома состоит из одной последовательности, и никаких дополнительных последовательностей нет).

2. *One or more chromosomes are still in multiple pieces and/or some*

sequences are not assembled into chromosomes (одна или несколько хромосом все еще состоят из нескольких частей или некоторые последовательности не собраны в хромосомы).

3. *We are submitting just the AGP file(s) for a genome assembly; the components of the AGP file are already in GenBank* (отправляем только AGP-файл для сборки генома).

В данном примере, сборка представляет собой несколько контигов – *One or more chromosomes are still in multiple pieces and/or some sequences are not assembled into chromosomes*, после чего в поле *How do you want to provide files for this submission?* проставляете - *Web browser upload via HTTP or Aspera Connect plugin* (если вы грузите до пары сотен бактериальных геномов общим объемом до 2 Гб), в поле *Files* загружаете файл со сборкой (в самом файле удалить контиги короче 200 п.о., т.к. они все равно будут исключены из сборки при депонировании по правилам NCBI), после чего нажимаете на *Continue* (Рисунок 34).

Submission Portal Home **My submissions** Manage data Templates My profile

Genome submission: SUB14664884 Delete submission

New

1 SUBMITTER 2 GENERAL INFO 3 BIOPROJECT GENERAL INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SOURCE **7 FILES** 8 ASSIGNMENT 9 REFERENCES 10 REVIEW & SUBMIT

Files for Submission Required fields are marked with * asterisk

*** Which of these 3 options describes this genome submission?**

1. Each chromosome is in a single sequence and there are no extra sequences

- There can still be gaps within the sequences.
- We will prompt you to provide the information for any Ns that represent gaps.
- Internal sequences must be arranged in the correct order and orientation. Sequences concatenated in unknown order are not allowed.
- Plasmids and organelles can still be in multiple pieces.
- If the sequences are assembled using an AGP file, choose the next option.

2. One or more chromosomes are still in multiple pieces and/or some sequences are not assembled into chromosomes

- This will be processed as a WGS genome and may include AGP files in the submission
- There can still be gaps within the sequences.
- We will prompt you to provide the information for any Ns that represent gaps.
- Internal sequences must be arranged in the correct order and orientation. Sequences concatenated in unknown order are not allowed.

3. We are submitting just the AGP file(s) for a genome assembly; the components of the AGP file are already in GenBank

*** How do you want to provide files for this submission?**

FTP or Aspera Command Line file preload

All files for a submission must be uploaded into a single folder.

Web browser upload via HTTP or Aspera Connect plugin

Do not use web browser HTTP upload if you are uploading files over 10 GB or more than 500 files.

To upload large files (larger than 2 GB), please use [Aspera Connect plugin](#).

*** Files**

or drag and drop them here

Name	Size	Created	Delete
contigs_ncbi.fasta	11.6 MB	2024-08-14 20:32	<input type="button" value="X"/>

Рисунок 34 – Заполненная вкладка *Files*

2.8. Заполнение вкладки *Assignment*

Assignment:

1. *Is any sequence a complete chromosome?* (сборка проведена до хромосом) - *No*.

2. *Does any sequence belong to a plasmid?* (если в контигах нет плазмид) – *No*, после чего нажимаете на *Continue* (Рисунок 35).

Submission Portal Home My submissions Manage data Templates My profile

Genome submission: SUB14664884 Delete submission

New

1 SUBMITTER 2 GENERAL INFO 3 BIOPROJECT GENERAL INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SOURCE 7 FILES 8 ASSIGNMENT 9 REFERENCES 10 REVIEW & SUBMIT

Assignment Required fields are marked with * asterisk

★ Is any sequence a complete chromosome?
 Yes No

★ Does any sequence belong to a plasmid?
 Yes No

Continue

Рисунок 35 – Заполненная вкладка *Assignment*

2.9. Заполнение вкладки *References*

References:

1. *Sequence authors* перечисляете всех соавторов, причастных к выделению, культивированию, секвенированию и биоинформатическому анализу.
2. *Publication status* ставите – *Unpublished*.
3. *Reference title* примерное название будущей статьи.
4. *Reference authors* выбираете *Same as sequence authors*, после чего нажимаете на *Continue* (Рисунок 36).

Submission Portal Home My submissions Manage data Templates My profile

Genome submission: SUB14664884 Delete submission

contigs_ncbl.fasta genome submission

1 SUBMITTER 2 GENERAL INFO 3 BIOPROJECT GENERAL INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SOURCE 7 FILES 8 ASSIGNMENT 9 REFERENCES 10 REVIEW & SUBMIT

References Required fields are marked with * asterisk

Sequence authors
Who should be publicly credited as the submitter of this sequence data? Enter authors below. Drag and drop to reorder authors.

* First (given) name MI * Last (family) name Delete

[Add another sequence author](#)

Reference

* Publication status
 Unpublished In-press Published

Reference title

Reference authors
 Same as sequence authors Specify authors

Continue

Рисунок 36 – Вкладка *References*

2.10. Заполнение вкладки *Review & Submit*

Проверяете правильность внесенных данных, если все введено верно, нажимаете на *Submit* (Рисунок 37). Если допущена ошибка переходите во вкладку, в которой была допущена ошибка, корректируете и возвращаетесь на вкладку *Review & Submit*. Статус депонирования геномной сборки можно отследить в личном кабинете, а также на привязанную электронную почту будут приходить письма с оповещением об успешном/неуспешном завершении каждого этапа. Конечным результатом являются уникальные идентификаторы геномной сборки, а также ее аннотация с помощью PGAP.

Genome submission: SUB14664884

contigs_ncbi.fasta genome submission

[Delete submission](#)**1** SUBMITTER **2** GENERAL INFO **3** BIOPROJECT GENERAL INFO **4** BIOSAMPLE TYPE **5** BIOSAMPLE ATTRIBUTES **6** SOURCE **7** FILES **8** ASSIGNMENT **9** REFERENCES **10** REVIEW & SUBMIT**Review & Submit**

This WGS submission will be released following processing.

Submitter

Submitter
igorsimankov93@gmail.com

Submitting organization

Department

Street

City

State/Province

Postal code

Country

General Information

BioProject ID

BioSample ID

Genome assembly structured comment is in the contig_sqn file

Assembly methods

Genome coverage

Sequencing technologies

Did your sample include the full genome? yes

Is this the final version? yes

Is it *de novo* assembly? yes

Is it an update of existing submission? no

GenBank will remove detected contamination, if possible Yes

BioSample general information

Organism name

Organism type

BioSample type

Package

BioProject general information

Public description

Prokaryote source

Annotate this prokaryotic genome in the NCBI Prokaryotic Annotation Pipeline before being released? yes

Biosample attributes

Sample name

Files

FASTA contigs

Sequence authors**References**

Publication status

Authors

Your submission is not yet complete. Finish your submission to get accession(s) sooner. You may need to upload your data again if your submission remains unfinished.

To proceed, please review your submission, make necessary changes on any tab, then click the 'Submit' button.

[Submit](#)[Ask for help](#)Рисунок 37 – Вкладка *Review & Submit*

Заключение

В данном учебно-методическом пособии приведена пошаговая схема депонирования геномной сборки в NCBI.

Данная схема включает в себя заполнение вкладок Submitter, General information, BioProject General Info, Sample Type, BioSample Attributes, Source, Files, Assignment, References, Review & Submit.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Carro, L. The Taxonomy of Bacteria in the Genomic Era / L. Carro, Á. Peix, E. Velázquez // *Developmental Biology in Prokaryotes and Lower Eukaryotes*. Springer, Cham. – 2021. – P. 289–309.
2. Gurevich, A. QUAST: quality assessment tool for genome assemblies / A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler // *Bioinformatics*. – 2013. – V. 29 (8). – P. 1072–1075.
3. Parks, D. H. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes / D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson // *Genome Res.* – 2015. – V. 25 (7). – P. 1043–1055.

Список электронных ресурсов

- BlastN. – URL: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (дата обращения: 20.07.2024).
- Aspera Connect plugin. – URL: https://chromewebstore.google.com/category/extensions?hl=ru&utm_source=ext_sidebar (дата обращения: 20.07.2024).
- National Center for Biotechnology Information. – URL: <https://www.ncbi.nlm.nih.gov> (дата обращения: 21.07.2024).
- BioPortal. – URL: <https://bioportal.bioontology.org> (дата обращения: 21.07.2024).