

Apr 25, 2024

## Загрузка сиквенсов в базу данных коллекции (Specify 7) и экспорт данных в GenBank и GBIF

DOI

[dx.doi.org/10.17504/protocols.io.3byl4975jgo5/v1](https://dx.doi.org/10.17504/protocols.io.3byl4975jgo5/v1)



Nina Filippova<sup>1</sup>, Elena Zvyagina<sup>1</sup>

<sup>1</sup>Yugra State University



Nina Filippova

Yugra State University

### Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.3byl4975jgo5/v1>

External link: <https://fungariumysu.org/>

**Protocol Citation:** Nina Filippova, Elena Zvyagina 2024. Загрузка сиквенсов в базу данных коллекции (Specify 7) и экспорт данных в GenBank и GBIF. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.3byl4975jgo5/v1>

**License:** This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited



**Protocol status:** Working

**We use this protocol and it's working**

**Created:** April 14, 2024

**Last Modified:** April 25, 2024

**Protocol Integer ID:** 98165

**Keywords:** dnaDerivedData, Specify\_7, GBIF, DwC, openData, sequence data to gbif, exporting sequence data, actual uploading of sequence, data to genbank, export of sequence, uploading sequence, sequences to genbank, term storage location for the sequence, preparing sequence, numbers in genbank, subsequent export of data, sequences open for publication, database, metadata, subsequent export, files into the specify, exporting data, edited sequence, metadata into the specify, term storage location, sequence, file, gbif, genbank, export, initial processing, processing, table into the integrated publishing toolkit portal, data, storage, загрузка

**Funders Acknowledgements:**

**Grant from the federal budget for the implementation of the state assignment "Molecular-genetic methods in the study and assessment of biodiversity in the Northern regions**

**Grant ID:** FENG-2024-0003

## Abstract

Protocol for uploading sequences and their metadata into the Specify 7 database and subsequent export of data to GenBank and GBIF.

1. The first stage of preparing sequences involves their initial processing (trimming ends, assembling), saving the edited sequence, and adding comments to the laboratory journal.
2. The second stage involves the actual uploading of sequences, their metadata, and trace files into the Specify 7 database. The database serves as the primary long-term storage location for the sequences. From here, the data is exported to GenBank and GBIF.
3. Exporting data to GenBank is done by creating a request in Specify and then uploading the table and sequences to GenBank through batch upload.
4. Exporting sequence data to GBIF is done for all sequences open for publication in the database after receiving the corresponding numbers in GenBank. The export of sequences is done through a query template, exporting to a .csv table, and then importing the table into the Integrated Publishing Toolkit portal, where each sequence is linked to the corresponding specimen in the collection dataset.

## Troubleshooting

## Safety warnings

❗ При внесении данных в Specify, старайтесь периодически сохранять карточку, чтобы не потерять внесенные данные.

## Ethics statement

Протокол не требует согласования с этическим комитетом.

## Before start

Для выполнения протокола необходимо:

- 1) Установить программное обеспечение для анализа хроматограмм,
- 2) Зарегистрировать аккаунт и получить доступ к соответствующим инструментам в ИС "Specify 7" (<http://specify.ugrasu.ru/>),
- 3) Зарегистрировать аккаунт на портале GenBank (<https://www.ncbi.nlm.nih.gov/>),
- 4) Зарегистрировать аккаунт и получить доступ к соответствующим инструментам в IPT (<http://ipt.ugrasu.ru:8080/>),
- 5) Получить доступ к лабораторному журналу и хранилищу сиквенсов лаборатории.

## Подготовка сиквенсов

- 1 Первая стадия подготовки сиквенсов включает их первичную обработку (обрезание концов, ассемблинг, и другие операции по необходимости), сохранение готовой отредактированной последовательности и комментарии в лабораторном журнале.
- 1.1 Создать проект для первичного редактирования сиквенсов в *Codon Code Aligner* или другом программном обеспечении. Загрузить исходные последовательности с исходными именами (номерами пробирок).
- 1.2 Провести анализ и необходимое редактирование сиквенсов:
  1. Если сиквенс слишком короткий, грязный или не отвечает другим параметрам качества, напротив этого сиквенса в лабораторном журнале (поле SeqQualRemarks) ставится пометка "плохой" и в дополнительном поле комментариев (поле REMARKS) - рекомендация (сделать с нуля, поставить вторично секвенсовую, поставить обратную секвенсовую и др). Такой сиквенс не загружается в Specify.
  2. Если сиквенс не достаточной длины, но хорошего качества, в лабораторном журнале (поле SeqQualRemarks) ставится пометка "хороший" и в дополнительном поле комментария (поле REMARKS) - "сделать обратный". Такой сиквенс редактируется и загружается в Specify, оставляя поле сиквенса пустым (оно будет заполнено после получения обратной последовательности).
  3. Если сиквенс достаточно хороший, его редактируют, проверяют по базе NCBI утилитой BLAST и сохраняют в формате FASTA, в первой строке .fas файла после знака > вносят номер образца с акронимом гербария и через нижнее подчеркивание указывают локус .

ФайлИзменитьПросмотр

>YSU-F-12495\_ITS|AACAGCTATGACTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTAGTCCTACCTGATTTGAGGTCAGAGGTCGGGTCTTTGTACGGAGACGGCTGATGAGCGGAAAGGCGAACGCCCCCTCTCCGACGCAAGGGTTTATCACGTCTCGCAGGGGGCGGAGGCTGTTCCGCTAATGCGTTTGAAGACGAGCCTACGCGGGGTCGAGACCCTCGCCGTCGGCAGACGTCCAGATCCGAAAGCCGCCACCCACCGAAACGGAAGGCGCGGCTCCGAGATAACTTCACGACACTCAAAACAGGCATGCTCCTCGGAATGCCGAGGAGCGCAAGATGCGTTCAAAGATTCGATGATTCACGAATTCTGCAATTCACATTACGTTTCGCGCTTTCGCGGCGTTCTTCATCGATGCGAAAGCCAAGAGATCCGTTGTTGAAAGTTGTGTCAAGTTTTTCGACTTTTTCACACACGCTCGATTAACAGAAGGGTTTCGTATGAACAGAGGGGGGAAAGGAAGCCCGAGGCTCCCTTCCCTCACGGGTGCACGTGGTGTGGTTTGTGAATTGAAAAGGGGAAATCCGGAAGACTCCCCCTCTAATCGGCAATGATCCTTCCGCAGGTTACCTACGGAACCTGTTACGACTTTTACTTCTCTAAATGACCAAGACTGGCCGTCGTTTTACAA

Строка 1, стол719 символов100%Windows (CRLF)UTF-8

Пример подготовленной последовательности сиквенса

4. Результат проверки по базе NCBI вносят в соответствующие поля в лабораторном журнале (поля BLAST\_closest\_species, genSequence FASTA, BLAST\_closest\_ID, TYPE\_specimen), а затем загружается в Specify в полном объеме (см. ниже).

	A	B	C	D	E	F	G	H	I	J	K	L
	№ пр об ир ки	№ в ко лл ек ци и	Вид в ко лл ек ци и	BLA ST бли жай шее сход ство	Имя исходног о файла	По сл ед ова тел ьн ост ь	% BL AS T	BL AS T clo se st ID	TY PE sp eci me n	Ge nB an kID	Seq _Qu al_R ema rks	Work _rem arks
	17 3	YS U-F-126 19	Lenti nus cast oreu s	Lenti nellu s subli neol atus	173_ITS_F _A09_01_2 022-10-06-23-24-52.ab1	seq	99. 4	NR _11 95 05. 1	TE NN :05 93 07	OQ 45 03 97	хор оши й	коро ткий, сде лать обра тный
	17 3	YS U-F-126 19	Lenti nus cast oreu s	Lenti nellu s subli neol atus	173_ITS_R _A08_01_2 023-02-15-16-27-33.ab1						хор оши й	

	A	B	C	D	E	F	G	H	I	J	K	L
	17 4	YS U- F- 126 28	Hyp sizy gus ulma rius	Hyps izyg us mar more us	174_ITS_F _B09_02_ 2022-10- 06-23- 24-52.ab1	seq	99. 0	OP 98 08 86. 1		OQ 45 03 98	хор оши й	коро ткий, сдел ать обра тный
	17 4	YS U- F- 126 28	Hyp sizy gus ulma rius	Hyps izyg us mar more us	174_ITS_R _B08_02_ 2023-02- 15-16-27- 33.ab1						хор оши й	
	17 5	YS U- F- 125 94	Corti nariu s adus tore mos us	Corti nariu s adus torim osus	175_ITS_F _C09_03_ 2022-10- 06-23- 24-52.ab1	seq	99. 0	NR _17 231 9.1	PC: R. He nry 88 3	OQ 45 03 91	хор оши й	коро ткий, сдел ать обра тный
	17 5	YS U- F- 125 94	Corti nariu s adus tore mos us	Corti nariu s adus torim osus	175_ITS_R _A07_01_2 023-02- 14-17-16- 01.ab1						хор оши й	

Пример фрагмента таблицы лабораторного журнала с анализом качества полученных последовательностей и оценкой ближайшего сходства

## Загрузка сиквенсов в БД Specify 7

- 2 Вторая стадия - собственно загрузка сиквенсов, их метаданных и исходных файлов в БД Specify 7. База данных является основным местом хранения сиквенсов на долгую перспективу. Отсюда данные экспортируются в GenBank и GBIF. Внесение данных на этом этапе требует внимательного отношения.
- 2.1 Войти в БД Specify 7 под своим логином. Сделать поиск и открыть карточку соответствующего образца коллекции.
- 2.2 В нижней части формы, **открыть кнопку DNA sequence**. Внести данные в поля:
  1. Genbank Accession Number - обязательное поле. Если номера еще нет, внести название фаста-файла без знака > (например, "YSU-F-12619\_LOCUS")
  2. Molecule type - по умолчанию "DNA" (или выбрать иное)
  3. Sequencer - выбрать ФИО ответственного за секвенирование
  4. BOLD Barcode ID - добавить уникальный номер BOLD, если имеется

5. Target Marker - по умолчанию "ITS", можно выбрать соответствующий локус
6. Collection object - заполняется автоматически при сохранении
7. Make public - поставить галочку, если сиквенс открыт для публикации в GenBank и GBIF, открывается после получения номера генбанка
8. Assembled - поставить галочку, если получена консенсусная последовательность из прямого и обратного (или других) ридов
9. Nucleotide Sequence - вставить последовательность в формате FASTA. См. пример ниже
10. Отредактировать последовательность так, чтобы в названии последовательности был номер коллекции в формате ">YSU-F-12619\_LOCUS", без кавычек, вместо LOCUS соответствующий маркер ДНК
11. Открыть следующую форму: DnaR (DNA Sequencing Runs), в которой будет внесена информация об исходных файлах.

DNA Sequences + -

Genbank Accession Number: OQ450397

Molecule Type: DNA

Sequencer: Rudykina, Elena

BOLD Barcode Id:

Target Marker: ITS

Collection Object: 12619

Make public: ☒

Assembled: ☐

Nucleotide Sequence

```
>YSU-F-12619_JTS
GTAGGTGAACCTGGGGAAGGATCATATCGTAAACAAGAGGCCGGGTACGGCTGTCGCTGGCCCCCTCGGGGGCATGTGCACGCCGGTGCATCCTTCACACCCCTGTGCACCTCTGGGTGG
GTTCGTGGCTTGGCCCTTCGAGCCCGTCCCTTCTACACACACCTTTGTATGTCTCAGAATGTCAACATGGGATAAAAGCATATACAACTTCAACAACGGATCTCTGGCTCTGCGCATGATG
AAGAAGCGCAGCGAAATGCGATAAGTAATGTGAATTGCGAGAATCAGTGAATCATCGAATCTTTGAACGCACCTTGACCCCTTGATTCCGAGGGGTACGCTGTCTGAGTGTGCGAAATCTCAACC
CCACCCCTTTTGGAGGGGTGTCGGGATTGGAATGGACTTGGANGCTTTCGGGAACCCGGTGTGCCCTCCCTCTTGGGGGTGGCTGTGTCGGGATCGGCTCTCTCAAAGGCATTANCAAGSACCCC
TCTGCGGCTCGGTGTGATAATTGTCTA
```

Close

Пример формы для заполнения данных полученного сиквенса (таблица DNA Sequence)

## 2.3 Открыть кнопку **DnaR (DNA sequence runs)**. Внести данные в поля:

1. Trace File Name - название .ab1 файла исходного сиквенса
2. Run Date - скопировать дату из названия .ab1
3. Seq Primer Name - название праймера секвенсовой реакции (в нашем случае при использовании адаптеров используется 2 значения: M13F, M13R). При использовании для ПЦР праймеров без адаптеров, секвенирование производится с праймерами для ПЦР. Соответственно в этом случае указывается соответствующий праймер
4. DNA Primer - название пары праймеров, использованных во время ПЦР (выбор из выпадающего списка)

### Note

В базу данные внесено большое число используемых праймеров и их метаданные. Ввод новых пар понадобится в исключительных случаях использования новых праймеров. В этом случае нужно нажать значок "+" рядом с полем DNA Primer и внести данные в открывшуюся форму

5. Seq Type - тип секвенирования (по умолчанию "Sanger")
6. Если исходных праймеров несколько (например, прямой и обратный), то заполнить следующий файл, нажав "+" в левом верхнем углу формы
7. После заполнения всех форм, нажать "закрыть".

Пример формы для заполнения данных исходных файлов (таблица DNA Sequencing Runs)

Пример формы для заполнения данных праймеров (таблица DNA Primer)

## 2.4 Прикрепить **исходные .ab1 файлы сиквенсов**:

1. Прикрепить исходные файлы сиквенсов можно в форме **DNA Seq Runs** или **DNA Sequences**
2. Мы пользуемся DNA Sequence Attachments
3. Открыть форму, нажать "+", перетащить соответствующий файл, и т.д. (метаданные файла заполняются автоматически)
4. По завершении нажать "Close" формы DNA Sequences

- 3 Добавить **новую детерминацию таксона** на основе поиска ближайших последовательностей в NCBI:
1. Добавить новую детерминацию к уже существующей истории определений (нажать "+" рядом с заголовком формы Determinations)
  2. Добавить новый таксон
  3. Добавить ФИО того, кто выполнил поиск в BLAST
  4. Добавить дату (когда выполнен поиск, или когда был получен сиквенс)
  5. В поле "Remarks" добавить комментарий по шаблону:
    - Closest type specimen sequence (NR\_119505) 99.4%
    - Closest voucher specimen sequence (TD119503) 99.1%

Пример заполнения детерминации таксона на основе молекулярных признаков (поиск ближайшей последовательности в NCBI BLAST)

### 3.1

#### Note

Алгоритм выбора таксона для новой детерминации:

1. При обнаружении в NCBI последовательности **типового образца** таксона (последовательности ITS курируемой базы типовых образцов грибов в GenBank обычно начинаются с NR\_), максимально перекрывающейся с нашей и сходной более чем на 99%, внести в карточку **детерминацию до вида**, к которому принадлежит данный типовой образец.
2. В случае, если сходство с типовым образцом составляет менее 99% и родовые названия ближайших последовательностей соответствуют морфологическому определению нашего образца - сделать новую детерминацию только **до уровня рода**.
3. Во всех остальных случаях произвести ревизию образца и последовательности, текущую детерминацию в карточке не менять до принятия решения о новой детерминации.

## Note

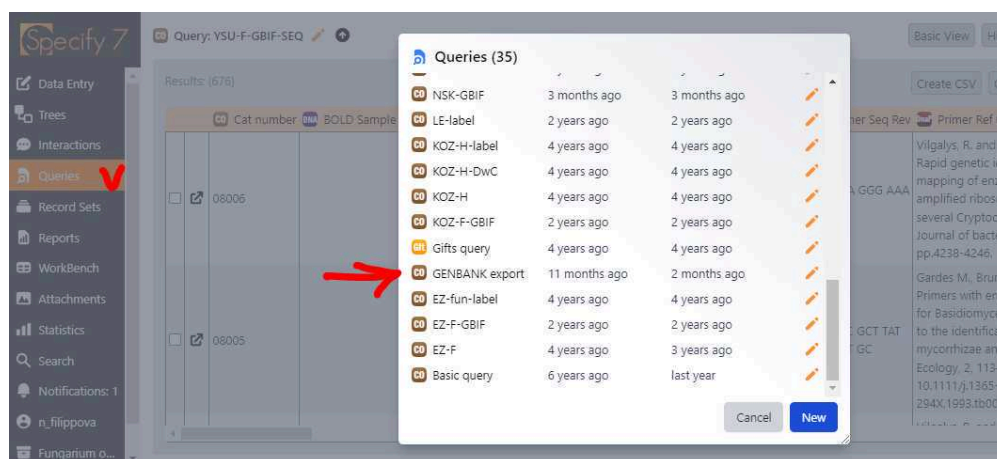
НЕ ЗАБУДЬТЕ **нажать Save** перед закрытием формы образца (Collection Object Form), иначе внесенные сиквенсы и определения не сохранятся.

Сохранение данных после заполнения форм сиквенсов

## Экспорт данных в GenBank

4 Экспорт данных в GenBank осуществляется через создание запроса в Specify и последующую загрузку таблицы и сиквенсов пакетным образом в GenBank.

4.1 **Войти в БД Specify 7** под своим логином. Зайти в раздел Queries и выбрать запрос под названием GENBANK export



Последовательность действий при выборе шаблона запроса для экспорта данных сиквенсов в GenBank

- 4.2 В открывшихся настройках запроса в поле Cat Number с включенным оператором "In" ввести необходимые образцы для экспорта в GenBank (например: 12619, 12628, 12594).

#### Note

В большинстве случаев в результатах поиска получается дублирование записей из-за наличия прямого и обратного сиквенса в таблице Dna Sequencing Runs (откуда происходит запрос праймеров). Избавиться от дублирования можно удалением дубликатов в итоговой таблице.

- 4.3 Экспортировать созданный запрос в виде таблицы .csv.

#### Note

Пакетный импорт сиквенсов и метаданных в GenBank требует строгого соблюдения формата данных. Шаблоны с примерами можно скачать **в последней секции "Шаблоны"**.

- 4.4 Подготовьте файл с метаданными сиквенса (модификаторами). Для этого скачайте шаблон таблицы формата .tsv с портала GenBank или создайте файл с соответствующими названиями полей:

A	B
<b>Sequence_ID</b>	Идентификатор сиквенса, должен совпадать с названием сиквенса в фаста файле (YSU-F-12594)
<b>Organism</b>	Название таксона на латыни (Cortinarius adustorimosus)
<b>Specimen-Voucher</b>	Номер ваучерного образца (YSU-F-12594)
<b>Country</b>	Страна (Russia)
<b>Latitude-Longitude</b>	Широта и долгота в формате (61.04119 N 69.31318 E)
<b>Collection-Date</b>	Дата сбора образца (03.10.2023)
<b>Collected-By</b>	Автор находки (Bulyonkova, Tatiana)
<b>Fwd-Primer-Name</b>	Название прямого праймера (ITS1-F)
<b>Rev-Primer-Name</b>	Название обратного праймера (ITS4)



A	B
<b>Fwd-Primer-Seq</b>	Последовательность прямого праймера
<b>Rev-Primer-Seq</b>	Последовательность обратного праймера

Пример заполнения полей модификаторов (метаданных сиквенса) при загрузке последовательностей в GenBank пакетным образом

Перенесите данные из скачанной из Specify таблицы в шаблон. Отредактируйте формат данных в поле географических координат, Страна и Дата. Сохраните шаблон с данными в виде .tsv файла.

#### Note

В случае, если загружаются последовательности, которые были собраны (assembled) и не собранные, на этом шаге нужно разделить таблицу и сиквенсы на 2 части и последовательно загрузить их в GenBank двумя партиями.

- 4.5 Подготовьте файл с сиквенсами в формате Fasta: скопируйте последовательности **из столбца экспортированной таблицы Gene Sequence** в текстовый файл. Убедитесь, что между названием сиквенса и его последовательностью есть перенос строки или добавьте переносы. Сохраните файл с расширением .fasta или .txt.
- 4.6 Зайдите на портал NCBI (<https://www.ncbi.nlm.nih.gov/>) под своим логином.
- 4.7 Пройдите последовательно **шаги ввода данных** для отправки сиквенса:
1. Submit
  2. GenBank
  3. New Submission
  4. Submission Type (ribosomal RNA (rRNA) or rRNA-ITS; Eukaryotic nuclear rRNA or rRNA-ITS; contains rRNA-ITS region)
  5. Submitter (Group 1 - настройте группу соавторов для каждого случая пакетного импорта)
  6. Sequencing technology (Sanger dideoxy sequencing)
  7. Assembly state (Unassembled sequence reads или Assembled sequences)
  8. Sequences (на этом шаге загружается подготовленный файл с сиквенсами в формате fasta. В случае, если появляются предупреждения об ошибках, файл можно отредактировать и загрузить еще раз).

9. Source info (Specimen - Voucher)
10. Source modifiers (на этом шаге выбираем "upload a tab-delimited table" и загружаем подготовленный файл с модификаторами)
11. References (выбираем авторов сиквенсов, публикацию/проект и ее авторов из ранее созданных групп, или создаем заново)
12. Review and Submit (проверяем результат загрузки и **нажимаем Submit**).

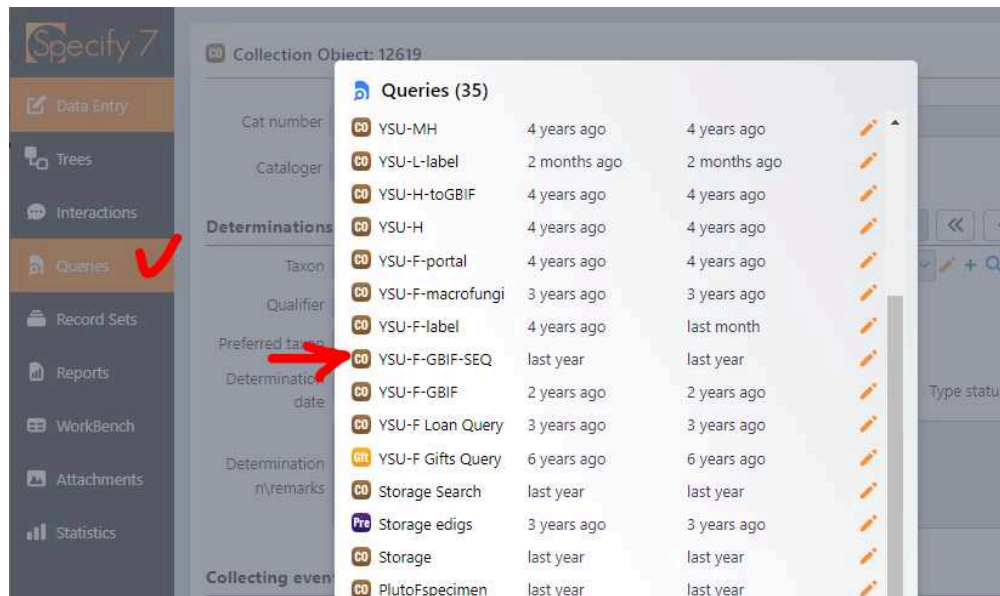


Последовательность ввода данных сиквенсов на портале GenBank (стрелками отмечены этапы, когда загружаются подготовленные шаблоны с данными для пакетной публикации)

- 4.8 После получения уникальных номеров сиквенсов в GenBank, зайдите в базу данных Specify и добавьте номера в соответствующее поле в шаблоне DNA Sequence.

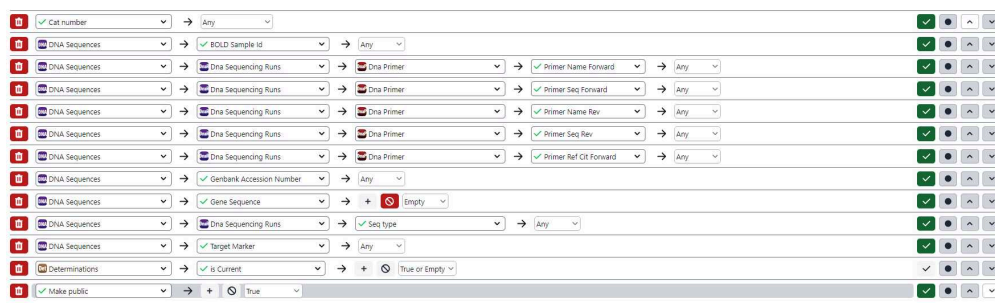
## Экспорт данных сиквенсов в GBIF

- 5 Экспорт данных сиквенсов в GBIF осуществляется для всех открытых для публикации сиквенсов в базе данных после получения соответствующих номеров в GenBank. Публикация осуществляется регулярно раз в месяц (в первых числах каждого месяца). Экспорт сиквенсов осуществляется через шаблон запроса, экспорт в таблицу .csv, и последующий импорт таблицы на портал Integrated Publishing Toolkit, где каждый сиквенс получает привязку к соответствующему образцу набора данных коллекции.
- 5.1 **Войти в БД Specify 7** под своим логином. Зайти в раздел Queries и выбрать запрос под названием YSU-F-GBIF-Seq



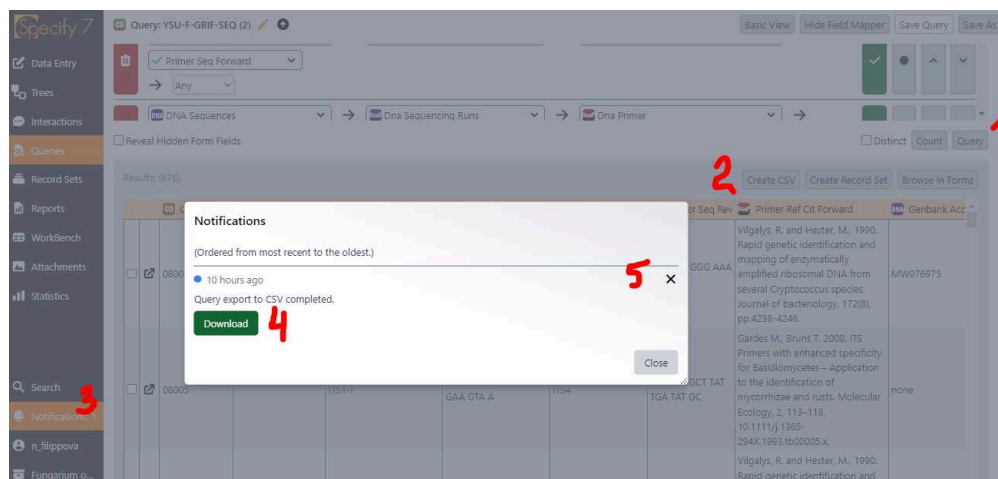
Последовательность действий при выборе шаблона запроса для экспорта данных сиквенсов в GBIF

- 5.2 В открывшейся форме запроса можно настроить редактирование на экспорт по заданным параметрам, или использовать поиск всех имеющихся последовательностей:
1. CatNumber - номер образца в коллекции (используя оператор "In" можно перечислить интересующие номера через запятую)
  2. BoldSampleID - номер сиквенса в BOLD
  3. Primer Name Forward - название прямого праймера (ПЦР)
  4. Primer Name Reverse - название обратного праймера (ПЦР)
  5. Primer Seq Forward - последовательность прямого праймера
  6. Primer Seq Reverse - последовательность обратного праймера
  7. Genbank Accession Number - номер в генбанке
  8. Gene Sequence - используя оператор "Not Empty" настроен экспорт только заполненного поля с последовательностью сиквенса. В противном случае поиск выдаст все 10K образцов в базе данных
  9. Seq type - тип секвенирования
  10. Target Marker - маркер
  11. Make Public - True (запрос только тех сиквенсов которые разрешены для публичной публикации)
  12. Determinations | is Current - поле "текущая детерминация" с оператором "True or Empty" помогает отфильтровать только последние детерминации, если их было несколько у одного образца. В противном случае поиск выдаст примерно в 2 раза больше записей, номера образцов при этом будут дублироваться.



Скриншот настройки запроса для выбора диапазона сиквенсов из базы данных

### 5.3 Создать запрос (Query) и экспортировать его в таблицу (Create CSV) и скачать ее на компьютер



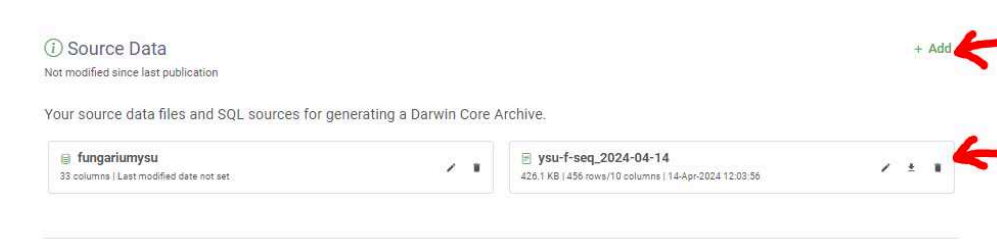
Последовательность действий при скачивании данных созданного запроса: Query > Create CSV > Notifications > Download > Close

### 5.4 Загруженная таблица может потребовать доработки: добавления нулей в поле номеров коллекции (5-значный номер), или другие погрешности при вводе данных (лучше редактировать сразу в базе данных и повторить экспорт).

## Загрузка таблицы сиквенсов на портал IPT и публикация в GBIF

- 6 Публикация данных в GBIF происходит через размещение исходных данных на портале Integrated Publishing Toolkit. В случае загрузки сиквенсов, используется специальное расширение DNA-derived-data, которое позволяет связать таблицу сиквенсов с таблицей загруженных ранее образцов коллекции.

- 6.1 **Зайти на портал IPT** под своим логином (<http://ipt.ugrasu.ru:8080/>). Открыть интересующий набор данных на портале IPT (в нашем случае The Fungarium of Yugra State University), начать редактирование набора данных.
- 6.2 **В разделе Source Data** удалить (если раньше была загружена) и добавить новую таблицу сиквенсов.



Начало загрузки таблицы сиквенсов в IPT: Source Data

- 6.3 В разделе **Darwin Core Mappings** добавить новое картирование и связать поля таблицы сиквенсов с полями IPT. Отредактировать и сохранить картирование и перейти к редактированию набора данных.  
! связывание полей пройдет автоматически, если в исходной таблице названия столбцов переименовать в названия полей в IPT.

A	B
Название поля в Specify	Название для импорта в IPT
Cat number	occurrenceID
Primer Name Forward	pcr_primer_forward
Primer Seq Forward	pcr_primer_reverse
Primer Name Rev	pcr_primer_name_forward
Primer Seq Rev	pcr_primer_name_reverse
Primer Ref Cit Forward	pcr_primer_reference
Genbank Accession Number	url
Gene Sequence	DNA_sequence
Seq type	seq_meth
Target Marker	target_gene

Шаблон для переименования заголовков столбцов таблицы сиквенсов для импорта в IPT

- 6.4 Обновить метаданные набора данных: если идет обновление только сиквенсов, то обновить информацию об общем числе полученных сиквенсов и прошедших баркодинг образцов в Абстракте и Методике работы. Нажать **Publish** и подождать некоторое время, пока пройдет обновление набора данных в GBIF.
- 6.5 Проверить несколько загруженных образцов в наборе данных в GBIF. Таблица с сиквенсами появится в нижней части карточки образца в разделе Dna Derived Data.

Url	Seq_meth	Dna Sequence	Pcr Primer Name Forward	Pcr Primer Forward	Pcr P
PP27729.6	Sanger	>YSU-F-13828 GTACAGGTTTCCGTAGG TGAACCTGCGGAGGATCATTTATGAGTT GAATGCGGGTTGTCCCTGGCCTACCAAGG CATGTTGCGGCGCTTTGCACTTCTAATTCA TCTTCCTTTTGTGCACTCTTTGGCGTAGTTC AAGGCTGACCCCTTTTCAAGGTTTTTGCAAC TTTGSTTTTGAAGCTGCGA...more	ITS1-F	CTT GGT CAT TTA GAG GAA GTA A	TCC

Пример таблицы сиквенса и его метаданных для образца, опубликованного в наборе данных коллекции Фунгария ЮГУ:  
<https://www.gbif.org/occurrence/4605346470>

- 6.6 **Источники информации** о формате данных DNA-Derived-Data и публикации данных последовательностей в GBIF:

#### Citation

dnaDerivedData (Invalid date). DNA derived data. Darwin Core Extension .

[https://rs.gbif.org/extension/gbif/1.0/dna\\_derived\\_data\\_2021-07-05.xml](https://rs.gbif.org/extension/gbif/1.0/dna_derived_data_2021-07-05.xml)<sup>LINK</sup>



### Citation

Abarenkov K, Andersson AF, Bissett A, Finstad AG, Fossøy F, Grosjean M, Hope M, Jeppesen TS, Kõljalg U, Lundin D, Nilsson RN, Prager M, Provoost P, Schigel D, Suominen S, Svenningsen C & Frøslev TG  
(Invalid date). Publishing DNA-derived data through biodiversity data platforms, v1.3..


<https://doi.org/10.35035/doc-vf1a-nr22>


LINK

## Шаблоны

7

 GENBANK\_SEQ\_template.fas 1KB

 GenBank\_modifiers\_template.csv 0B

 GBIF\_seq\_template.xlsx 15KB

## Protocol references

Abarenkov K, Andersson AF, Bissett A, Finstad AG, Fossøy F, Grosjean M, Hope M, Jeppesen TS, Kõljalg U, Lundin D, Nilsson RN, Prager M, Provoost P, Schigel D, Suominen S, Svenningsen C & Frøslev TG (2023) Publishing DNA-derived data through biodiversity data platforms, v1.3. Copenhagen: GBIF

Secretariat. <https://doi.org/10.35035/doc-vf1a-nr22>.

Filippova N, Ganasevich G, Filippov I, Meshcheryakova A, Lapshina E, Karpov D (2022) Yugra State University Biological Collection (Khanty-Mansiysk, Russia): general and digitisation overview. Biodiversity Data Journal 10: e77669. <https://doi.org/10.3897/BDJ.10.e77669>.



## Citations

### Step 6.6

dnaDerivedData. DNA derived data

[https://rs.gbif.org/extension/gbif/1.0/dna\\_derived\\_data\\_2021-07-05.xml](https://rs.gbif.org/extension/gbif/1.0/dna_derived_data_2021-07-05.xml)

### Step 6.6

Abarenkov K, Andersson AF, Bissett A, Finstad AG, Fossøy F, Grosjean M, Hope M, Jeppesen TS, Kõljalg U, Lundin D, Nilsson RN, Prager M, Provoost P, Schigel D, Suominen S, Svenningsen C & Frøslev TG. Publishing DNA-derived data through biodiversity data platforms, v1.3.

<https://doi.org/10.35035/doc-vf1a-nr22>