

EukBank Submission Guidelines

Please login to your Webin account or register for an account if you do not have one:

<https://www.ebi.ac.uk/ena/submit/sra/#home>

The screenshot shows the ENA (European Nucleotide Archive) Webin submission service interface. At the top, there's a navigation bar with links for Services, Research, Training, and About us. Below this is the ENA logo and a secondary navigation bar with links for ENA Home, Search & Browse, Submit & Update, About ENA, Contact, and FAQ. A 'Contact Helpdesk' link is also visible in the top right corner. The main content area has a light purple background and contains a welcome message: 'Welcome to the Webin submission service. You can use this service to submit sequence reads, genome assemblies, targeted assembled and annotated sequences and to register studies (projects) and samples. To submit human data requiring controlled access please use EGA Webin. To register umbrella studies, used to group together sequencing studies, please contact datasubs@ebi.ac.uk. New to Webin? Please register using the button below.' Below the welcome message is a blue banner with a link to subscribe to the ena-announce mailing list. Underneath, there are buttons for 'Register' and 'Reset Password'. A 'Login' section follows, featuring input fields for 'Username' and 'Password', each with a small information icon to its right. A 'Login >' button is positioned below the password field.

1) Select the “New Submission” tab and then select the “Register study (project)” option.

Please see documentation here: http://ena-docs.readthedocs.io/en/latest/mod_02.html

The screenshot displays the 'New Submission' page within the ENA Webin submission service. The top navigation bar is identical to the previous screenshot. Below it, the ENA logo and secondary navigation bar are also present. The main content area has a light blue background. At the top of this area, there's a progress bar with four tabs: 'Home', 'New Submission', 'Studies', and 'Sample Groups'. The 'New Submission' tab is currently selected. Below the progress bar, there's a section titled 'Start' followed by a '>>' button, then the word 'Study', another '>>' button, and finally the word 'Finish'. A paragraph of text explains the service: 'You can use this service to submit [sequence reads](#), [genome assemblies](#), [targeted assembled and annotated sequences](#) and to register [studies \(projects\)](#) and [samples](#). To register umbrella studies, used to group together sequencing studies, please contact datasubs@ebi.ac.uk.' Below this text is a section titled 'Please select the type of submission you would like to make:'. It contains four radio button options: 'Submit sequence reads and experiments', 'Register study (project)' (which is selected), 'Taxonomy Check/Request', and 'Register samples'. A sub-section for 'Register study (project)' explains that 'Studies (projects) can be pre-registered before submitting data into ENA.' Below the radio buttons, there's a section titled 'Specify the name of the center you are submitting for.' with a label 'Centre Name: (*)' and an input field. At the bottom right, there's a 'Next >>' button. At the bottom left, there's a 'Restart Submission' link with a circular arrow icon.

Fill in the description details and then click on “Add” under “Tag FieldType”. Under “Tag” add “Study keyword” and under “FieldType” add “UniEuk_EukBank” as shown in the screenshot below.

Please also ensure you set the correct Release Date as this is when your study will become public. The default is two months but this can be extended up to two years and can be changed at any time prior to the release date. Then click Submit.

The screenshot shows the 'New Submission' page for the European Nucleotide Archive (ENA). The page has a navigation bar with tabs: Home, New Submission (active), Studies, Sample Groups, Samples, Experiments, Runs, and Assemblies. Below the navigation bar is a progress bar with steps: Start (with a green checkmark), Study (with a double arrow), and Finish. The main content area contains instructions and form fields. On the left, there are fields for 'Please specify the release date of your study: This is when your study will be made public.' (with a date picker set to 15-May-2017), 'Please provide a short name for the study:', 'Please provide a short descriptive title for the study: (*)', and 'Please provide an abstract to describe the study in detail: (*)'. On the right, there is a section 'Please provide attributes to add a deeper description of the study:' with a table for 'Tag' and 'FieldType'. The table has one row with 'Study Keyword' in the 'Tag' column and 'UniEuk_EukBank' in the 'FieldType' column. Below the table is an 'Add' button. There is also a section for 'Please provide PubMed IDs of publications you want to associate with the study: (numeric value)' with an 'Add' button. At the bottom right, there is a question 'For genome assembly projects only: In this study, will you provide functional genome annotation? (*)' with radio buttons for 'Yes' and 'No' (selected). The page has a footer with a '<< Previous' button, a 'Restart Submission' button with a green circular arrow icon, and a 'Submit' button.

Home New Submission Studies Sample Groups Samples Experiments Runs Assemblies

Start >> Study >> Finish

You can use this page to submit a new study into the European Nucleotide Archive (ENA).

Please answer the questions below. Mandatory fields are denoted by (*).

If you need to create umbrella studies to group several studies together, please contact us at datasubs@ebi.ac.uk. You can then link your umbrella study by using the 'Studies' report tab, click on the 'Edit' button of the study you want to edit and use the 'Edit ENA links' tab.

Please specify the release date of your study:
This is when your study will be made public.

15-May-2017

Please provide a short name for the study:

Please provide a short descriptive title for the study: (*)

Please provide an abstract to describe the study in detail: (*)

Please provide attributes to add a deeper description of the study:

Tag	FieldType
Study Keyword	UniEuk_EukBank

Add

Please provide PubMed IDs of publications you want to associate with the study:
(numeric value)

PubMed IDs

Add

For genome assembly projects only: In this study, will you provide functional genome annotation? (*)
PLEASE ANSWER WITH YES IF YOU HAVE ANNOTATION: Locus tag prefixes are only associated to studies providing functional genome annotation.

☐ Yes
☒ No

<< Previous Restart Submission Submit

Your study (project) has now been created.

2) Next you need to register samples:

Please see documentation here: http://ena-docs.readthedocs.io/en/latest/mod_03.html

Select the “New Submission” tab again and select the “Register samples” option.
Then click on “Select Checklist” -> “Environmental Checklists” -> **ENA UniEuk_EukBank Checklist** and select Next

and then “+Expand”

All mandatory fields will be checked and need to be completed, please also select any optional fields for which you have the metadata and click Next.

The following are mandatory fields and the ones highlighted in red should contain the same value for all sample submissions within EukBank so please ensure this is the case:

tax_id 100272
scientific_name uncultured eukaryote
environmental_sample yes
collection_date (in format: YYYY-MM-DD)
geographic location (country and/or sea)
geographic location (latitude)
geographic location (longitude)
environment (biome)
environment (feature)
environment (material)
environmental_sample
target_gene 18S
target_subfragment V4
pcr primers

Also, there is an optional field called “Further Details, please use this to add the link to the UniEuk project as follows:

Further details <http://unieuk.org/project/>

In the Search box under “Organism Details” please type “uncultured eukaryote” which will automatically fill in the Tax ID box with “100272”.

The screenshot shows the UniEuk submission interface. At the top, there are tabs for Home, New Submission, Studies, Sample Groups, Samples, Experiments, Runs, and Assemblies. The 'New Submission' tab is active, and the 'Sample' sub-tab is selected. The form is divided into two main sections: 'Template Basic Details' and 'Organism Details'. The 'Template Basic Details' section includes fields for 'Unique Name Prefix', 'Title', and 'Description'. The 'Organism Details' section includes a search bar, a 'Tax Id' field (filled with 100272), a 'Scientific Name' field (filled with uncultured eukaryote), and a 'Common Name' field. A blue box on the left contains instructions: 'Please complete any fields that you would like to apply to all samples. This will act as a template for all samples.' At the bottom, there are buttons for '<< Previous', 'Restart Submission', and 'Next >>'.

Please also fill out any other fields which are common to all your samples and click Next.

On the left hand side of the screen please click “+Add” to add the number of samples you have and then select “Download Spreadsheet” which will open up a tab delimited text file with all the mandatory fields and optional fields you have selected (see screenshot below).

Please note you can submit multiple samples simultaneously using this method (bulk submission). Please ensure you fill in any fields that are empty but do not alter the table in any other way. If fields are left empty the submission will fail. In the first column below “#units” please add the sample alias for each sample (shown in red in screenshot below). This can be any name such as an internal identifier but must be unique for each sample.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	#checklist_access	ERC000040											
2	#unique_name_prefix												
3	sample_alias	tax_id	scientific_name	common	sample_title	sample_dest	target	pcr	prime	collection	geograph	geograph	geograph
4	#template	100272	uncultured eukaryote		EukBank Test Submission	18S	V4						
5	#units											DD	DD
6	sample_cam_01	100272	uncultured eukar	unculture	EukBank Test	sample extr	18S	V4	TTTACGA	2016-04-	United Ki	52.2053	0.1218
7	sample_cam_02	100272	uncultured eukar	unculture	EukBank Real	sample extr	18S	V4	CAGCAGC	2015-97-	France	48.8566	2.3522

Once you are ready to submit your spreadsheet please go back to the New Submission sample page by clicking on “Previous” 4 times.

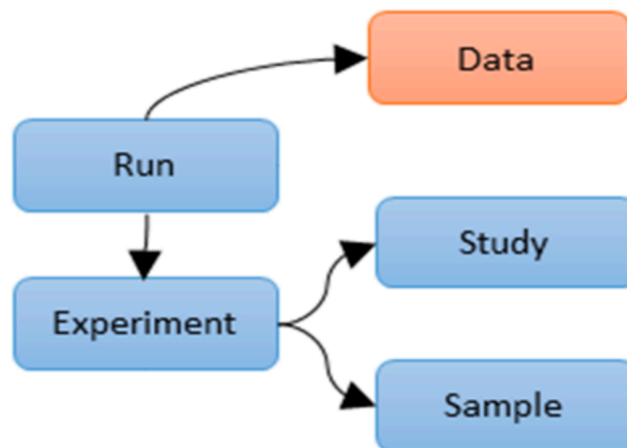
Click on “Upload Completed Spreadsheet” and add your spreadsheet. Click Submit. Your samples are now registered and you should get an email confirming this.

3) Adding read files:

Data files must be uploaded to your ENA ftp directory before they can be submitted either using an applet embedded into Webin or by using FTP or Aspera. For FTP/Aspera please calculate the md5 sum for each file that you upload. Please see here for more details:

<http://www.ebi.ac.uk/ena/submit/uploading-data-files>

The next step is registering the files that you have uploaded to your personal ENA ftp directory. We need to wrap each file or pair of files into a run object, point that run to an experiment object, and point that experiment object to the correct sample.



Please see documentation here: http://ena-docs.readthedocs.io/en/latest/mod_04.html

Please go to “New Submission” and select “Submit sequence reads and experiments” and click Next.

Select the project you are using and click Next

The next page is the sample page and as you have already registered your samples please click on “Skip”

The screenshot shows a submission workflow interface with the following components:

- Progress Bar:** A horizontal bar at the top with steps: **Start** (checked), **Study** (checked), **Sample** (active), **Run**, and **Finish**. Arrows indicate the flow between steps.
- Main Content Area:**
 - Start building your submission:** A section with a blue header containing an information icon and text: "We use checklists to help provide required information in a standard format. You will be guided through the following steps:
 - Selecting a checklist
 - Selecting optional fields in addition to mandatory ones
 - Entering your data directly into this application
 Alternatively, after selecting the checklist and fields you will be able to download a template spreadsheet. You can then enter your data in the spreadsheet and upload it." Below this is a button labeled "Select Checklist >".
 - Upload a submission completed using a template spreadsheet:** A section with a blue header containing an information icon and text: "If you have downloaded and filled a template spreadsheet please upload it using the **Upload Completed Spreadsheet** button. Please note that only spreadsheets in tab-delimited text format are supported (with either .tsv or .txt extensions). If you edited the spreadsheet in Microsoft Excel (or equivalent) please save the spreadsheet as Text (Tab delimited). To do this please see [these instructions](#)." Below this is a button labeled "Upload Completed Spreadsheet".
- Navigation:** At the bottom, there are two buttons: "<< Previous" on the left and "Skip >>" on the right.

Please select “Two Fastq files (Paired)”

In the Sample reference box please enter your sample ID (ERSXXXXXXX). You can look under the samples tab to get these IDs or at the email you received at the time of submission.

As shown in the screenshot below please add the following:

Under Instrument model please select **Illumina MiSeq**

Under Library Source please select **METAGENOMIC**

Under Library Selection please select **PCR**

Under Library Strategy please select **AMPLICON**

Please complete the Insert Size box

Home new submission Studies Sample Groups Samples Experiments Runs Assemblies

Start ✓ >> Study ✓ >> Sample ✓ >> Run >> Finish

Please provide library, instrument and data file details by uploading a spreadsheet or by editing the table below.

Please select the file format. If you have files of different types please submit them in separate submissions.

☐ CRAM
☐ BAM
☐ SFF
☐ One Fastq file (Single)
☒ Two Fastq files (Paired)

Two fastq files containing *paired reads* are submitted for each run. All technical sequences including adaptor sequences, linker sequences and barcode sequences must be removed from the reads before submission. *The first reads must be in the first Fastq file and the second reads must be in the second Fastq file ordered in the same order as in the first file.*

☐ Complete Genomics
☐ PacBio HDF5
☐ Oxford Nanopore

Mandatory fields are denoted by (*).

Download Template Spreadsheet Upload Completed Spreadsheet Download Spreadsheet

[Sample reference suggestions]	Sample reference (*)	Instrument Model (*)	Library Name	Library Source (*)	Library Selection (*)	Library Strategy (*)	Design
✖	ERSXXXXXX	Illumina MiSeq		METAGENOMIC	PCR	AMPLICON	

<< Previous Submit

If you have more than one sample please click on Download Spreadsheet and complete this offline by adding the file names for the forward and reverse read files, as well as the MD5 checksums and then Upload Completed Spreadsheet. Finally, click on Submit.

Your project has now been registered and contains the samples and runs you have added.

Please contact datasubs@ebi.ac.uk if you have any queries after reading the documentation and please ensure you mention you are submitting as part of the EukBank initiative in your email.

EukBank Data Analysis

Now you have submitted your data to ENA you can provide access to it to the EukBank team led by Frédéric Mahé at CIRAD, via the EukBank data hub. Please see below for details.

EukBank Data hub

EukBank data providers can share their submitted pre-publication sequencing data and associated sample metadata (see submissions guidelines above) through a data hub at the European Nucleotide Archive (EMBL-ENA) to be analysed by the EukBank team.

Initial set up

A data hub (dcc_basie) has been issued by EBI and the authoritative list of data providers and data consumers is administered at EBI.

Current list of data providers and consumers can be found below:

Data providers:

Name	Affiliation	Address	Email

Data consumers:

Name	Affiliation	Address	Email
Frédéric Mahé	CIRAD	Campus international de Baillarguet 34398 Montpellier Cedex 5 France	frederic.mahe@cirad.fr

The data providers above give their consent to all data consumers listed here to have pre-publication access to data and metadata in the data hub associated with the above projects.

Adding new data providers and consumers requires an email to be sent to datasubs@ebi.ac.uk. In this email please use the **subject line 'New EukBank data provider/consumer'**. ENA staff will return the above consent form to the requester and ask them to complete their details.

Note: Currently, membership of a data hub is not institutional but individual. Adding new data consumers at any later stage, even if they are from institutions that already have a consented data consumer in the given data hub, will require EBI to ask for another consent from the original data providers. Once the completed consent form has been returned to ENA access credentials to the data hub will be sent out to newly added data consumers (and if requested also to data providers).

Once a project is registered (see UniEuk_EukBank_submission_guide), **and the above consent form has been completed and returned to the ENA, data providers should email the project ID (PRJEBXXXXX) to datasubs@ebi.ac.uk, and ask the project to be associated with the EukBank data hub 'dcc_basie'.**

IMPORTANT: Please do not circulate access credentials, even between those working within the same institution.