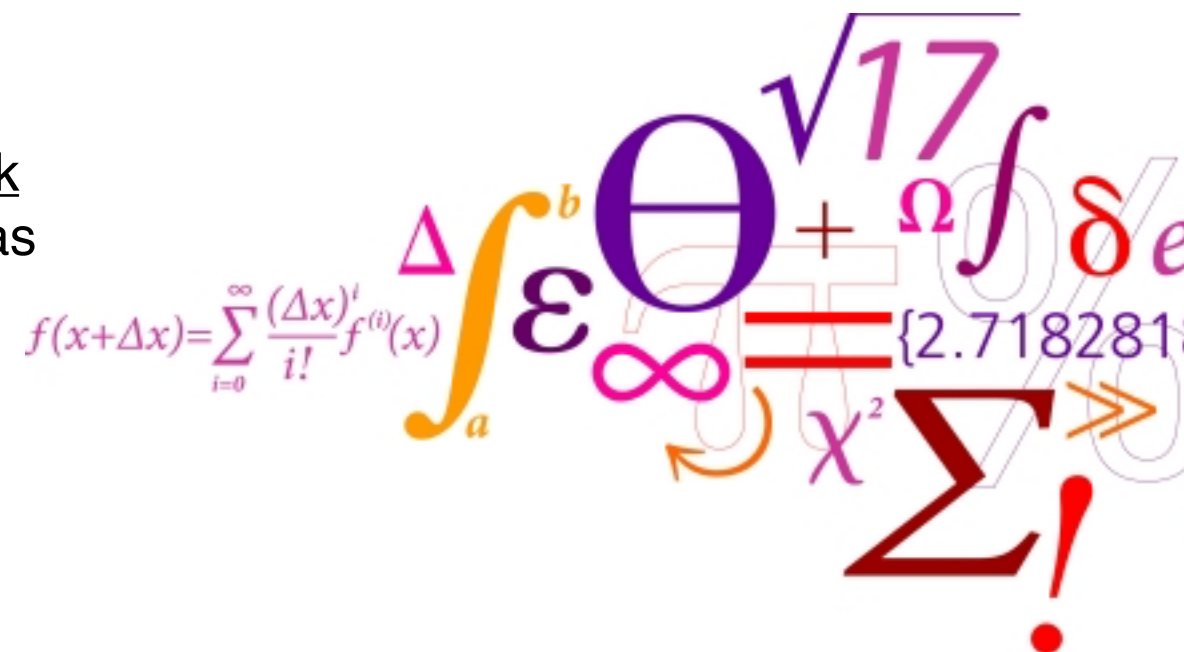


Sharing sequence data to enable collective action against AMR

Pimlapas Leekitcharoenphon (Shinny)
Research Group of Genomic Epidemiology, DTU-Food

WHO Collaborating Centre for Antimicrobial Resistance in Foodborne Pathogens and Genomics
European Union Reference Laboratory for Antimicrobial Resistance (EURL-AMR)

pile@food.dtu.dk
 @ShinnyPimlapas

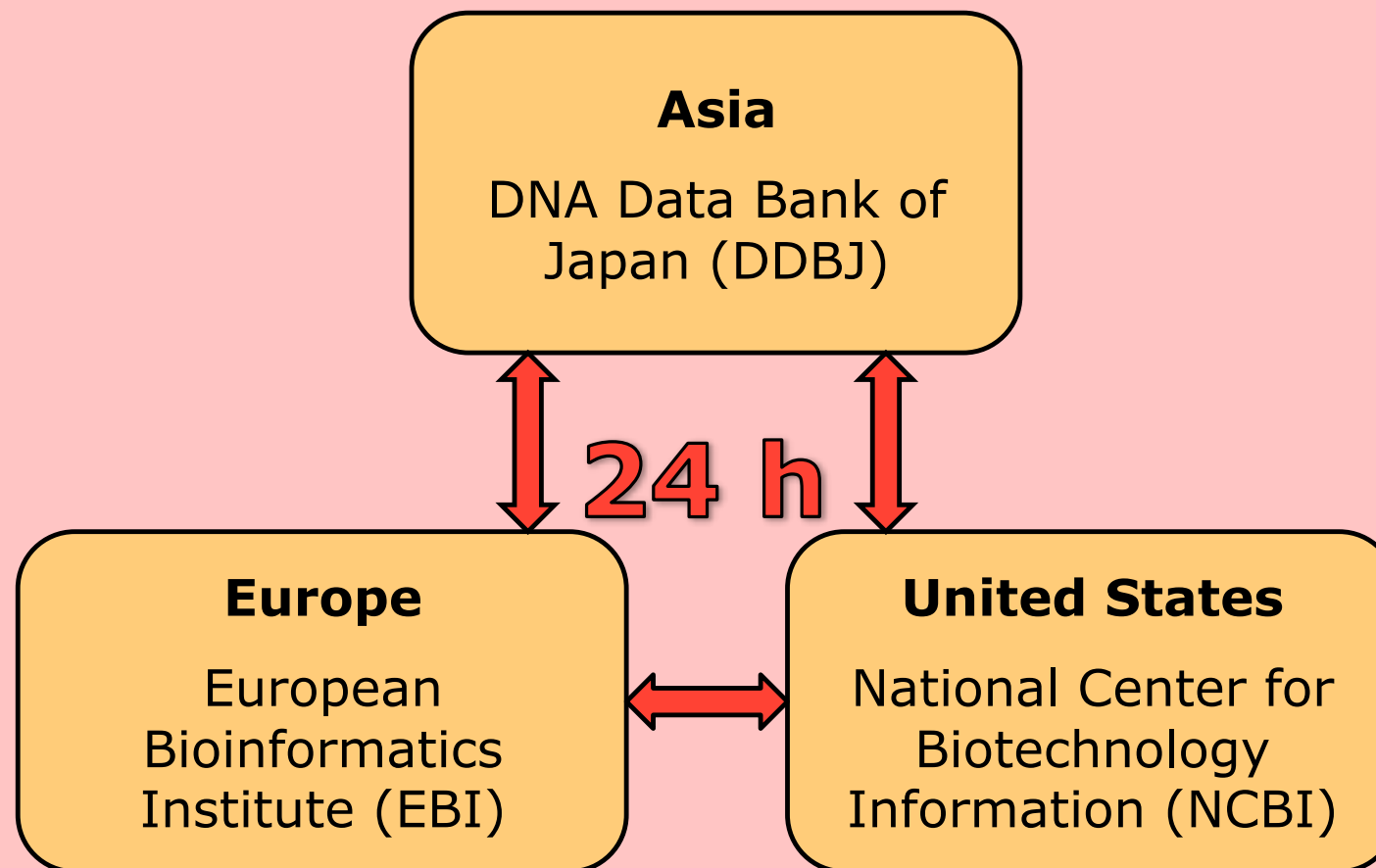


Topics

- Submitting data via ENA
- Sharing data via ENA private hub
- Accessing sequence data through ENA private hub
- Accessing WGS results through ENA private hub


Data storage & Access


International Nucleotide Sequence Database Collaboration (INSDC)



Data storage & Access

European Bioinformatics Institute (EBI)

EMBL-EBI  <http://www.ebi.ac.uk/ena> Services Research Training About us

 **ENA**
European Nucleotide Archive

Examples: [BN000065](#), [histone](#) [Advanced](#)
[Sequence](#)

Home Search & Browse Submit & Update Software About ENA Support

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

Text Search

Examples: [BN000065](#), [histone](#)

[Advanced search](#)

Sequence Search

Enter or paste a nucleotide sequence or accession number

[Advanced search](#)

Popular

- [Submit and update](#)
- [Sequence submissions](#)
- [Genome assembly submissions](#)
- [Submitting environmental sequences](#)
- [Citing ENA data](#)
- [Rest URLs for data retrieval](#)
- [Rest URLs to search ENA](#)

Latest ENA News

24 Sep 2014: [ENA Release 121](#)
Release 121 of ENA's assembled/annotated sequences now available.

20 Aug 2014: [Read data through Globus GridFTP](#)
Read data can now be downloaded using [Globus GridFTP](#) through `ebi#ena` Globus Online public endpoint.

18 Aug 2014: [Changes to SRA XML 1.5](#)
Small changes to Experiment XML, Analysis XML, EGA Dataset XML, EGA DAC XMLs were deployed on 11th of August 2014.

EMBL-EBI [Find](#) [Help](#) [Feedback](#)

[Databases](#) [Tools](#) [Research](#) [Training](#) [Industry](#) [About Us](#) [Help](#) [Site Index](#)

ENA

- [ENA Home](#)
- [Search & Browse](#)
- [Submit & Update](#)
- [About ENA](#)
- [Contact](#)

NEWS AND ANNOUNCEMENTS

CRAM toolkit 0.7 released
7 Mar 2012
CRAM toolkit 0.7 has been released. More information with download, installation and usage instructions are available [here](#).

The future of sequence archiving
2 Mar 2012
Future of sequence archiving and the role of data compression is explored in a [new paper](#) from EBI to be published in [Gigascience](#).

[EBI Home](#) » [ENA Home](#) »

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation ... [more](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

ENA

European Nucleotide Archive

Text search

[Search](#)

Sequence Search [Advanced Search](#)

[Search](#)

www.ebi.ac.uk/ena/

EBI Search

PRJEB22091

Examples: [VAV_HUMAN](#) , [tpi1](#) , [Sulston ...](#)



Build Query

Help & Documentation

About EBI Search

Feedback

EMBL-EBI to be HTTPS by default from 1st October

On the 1st October the majority of services hosted on www.ebi.ac.uk will be served over HTTPS by default. Services that are becoming HTTPS by default will automatically redirect users accessing the site on insecure HTTP URLs to secure HTTPS URLs.

Users of EMBL-EBI services may wish to update links, bookmarks or API clients to use the HTTPS URLs.

Search results for **PRJEB22091**

Showing **15** results out of **129** in All results

Filter your results

Source

All results (129)

[Nucleotide sequences](#) (129)

Nucleotide sequences (129 results)

[PRJEB22091](#)

Danmap 2016

Related data

Source: Study
ID: PRJEB22091

Navigation Read Files Portal Attributes

Bulk Download Files (Please use Firefox to launch the bulk downloader app.)

Download: - of 127 results in [TEXT](#)

[Select columns](#)

Showing results 1 - 10 of 127 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRAM Index files (FTP)	CRAM Index files (Galaxy)
PRJEB22091	SAMEA104205896	ERS1864914	ERX2148098	ERR2091304	562	Escherichia coli	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	Fastq file 1 Fastq file 2	Fastq file 1 Fastq file 2	File 1	File 1		
PRJEB22091	SAMEA104205897	ERS1864915	ERX2148099	ERR2091305	562	Escherichia coli	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	Fastq file 1 Fastq file 2	Fastq file 1 Fastq file 2	File 1	File 1		
PRJEB22091	SAMEA104205898	ERS1864916	ERX2148100	ERR2091306	562	Escherichia coli	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	Fastq file 1 Fastq file 2	Fastq file 1 Fastq file 2	File 1	File 1		
PRJEB22091	SAMEA104205899	ERS1864917	ERX2148101	ERR2091307	562	Escherichia coli	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	Fastq file 1 Fastq file 2	Fastq file 1 Fastq file 2	File 1	File 1		
PRJEB22091	SAMEA104205900	ERS1864918	ERX2148102	ERR2091308	562	Escherichia coli	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	Fastq file 1 Fastq file 2	Fastq file 1 Fastq file 2	File 1	File 1		
PRJEB22091	SAMEA104205901	ERS1864919	ERX2148103	ERR2091309	562	Escherichia coli	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	Fastq file 1 Fastq file 2	Fastq file 1 Fastq file 2	File 1	File 1		
PRJEB22091	SAMEA104205902	ERS1864920	ERX2148104	ERR2091310	562	Escherichia coli	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	Fastq file 1 Fastq file 2	Fastq file 1 Fastq file 2	File 1	File 1		
PRJEB22091	SAMEA104205903	ERS1864921	ERX2148105	ERR2091311	562	Escherichia coli	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	Fastq file 1 Fastq file 2	Fastq file 1 Fastq file 2	File 1	File 1		

Submitting data in ENA

- Manual and automated
- Use Webin (interactive web submission system) for new sequencing projects, assembled sequences and annotation
- For other types of data, different channels are available (e.g. datasubs@ebi.ac.uk, FTP, RESTful web-based service)
- www.ebi.ac.uk/ena/about/submit_and_update

[Link to download sequence data for ENA submission](#)

Submitting data in ENA

- 1) Install a program to connect to FTP
- <https://filezilla-project.org>

FileZilla The free FTP solution

Home

FileZilla
Features
Screenshots
Download
Documentation
FileZilla Pro

FileZilla Server
Download

Community
Forum
Project page
Wiki

General
FAQ
Support
Contact
License
Privacy Policy
Trademark Policy

Development
Source code
Nightly builds
Translations
Version history
Changelog
Issue tracker

Other projects
libfilezilla
Octochess

Sponsors:

Promotion:
FileZilla® Pro
The Best FTP Solution
GET IN NOW >

Overview

Welcome to the homepage of FileZilla®, the free FTP solution. The *FileZilla Client* not only supports FTP, but also FTP over TLS (FTPS) and SFTP. It is open source software distributed free of charge under the terms of the GNU General Public License.

We are also offering *FileZilla Pro*, with additional protocol support for WebDAV, Amazon S3, Backblaze B2, Dropbox, Microsoft OneDrive, Google Drive, Microsoft Azure Blob and File Storage, and Google Cloud Storage.

Last but not least, *FileZilla Server* is a free open source FTP and FTPS Server.

Support is available through our [forums](#), the [wiki](#) and the [bug and feature request trackers](#).

In addition, you will find documentation on how to compile FileZilla and nightly builds for multiple platforms in the development section.

Quick download links

Download FileZilla Client
All platforms

Download FileZilla Server
Windows only

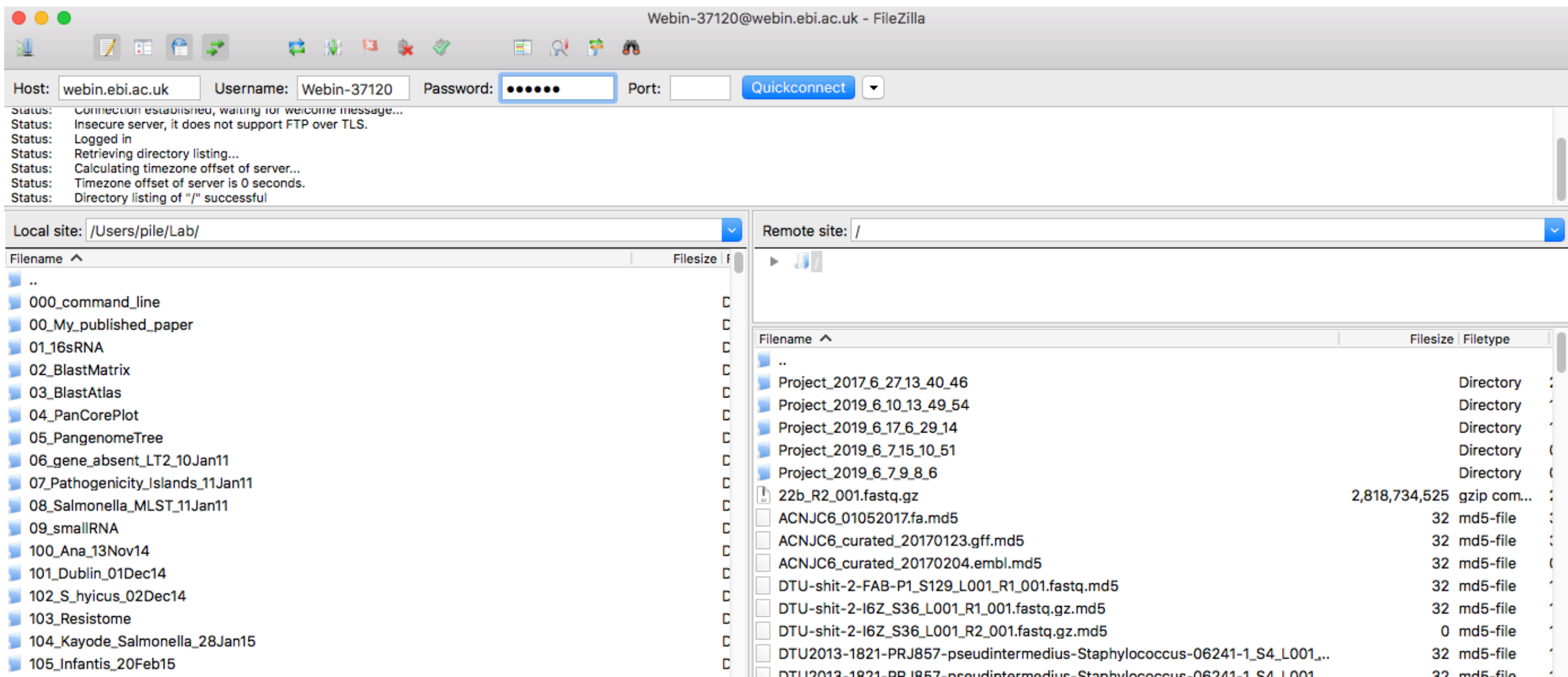
Pick the client if you want to transfer files. Get the server if you want to make files available for others.

News

- ◆ **2019-08-15 - FileZilla Client 3.44.2 released**
Bugfixes and minor changes:
 - MSW: Fixed a crash if using predefined sites through fzdefaults.xml
 - Fixes to protocol selection glitches in the Site Manager
 - Increase maximum length of response lines when using FTP
- ◆ **2019-08-09 - FileZilla Client 3.44.1 released**
Bugfixes and minor changes:
 - Fixed a regression introduced in 3.44.0-rc1 breaking support for insecure servers
- ◆ **2019-08-09 - FileZilla Client 3.44.0 released**
Bugfixes and minor changes:
 - Fixed export in context menu of Site Manager to handle multiple selected items
- ◆ **2019-07-29 - FileZilla refocuses on security after participation in EU bug bounty program**

Advertisement:
Carbon
BuySellAds
Reach developers effortlessly through this ad banner with Carbon.

- 2) Connect to FTP
- Host: webin.ebi.ac.uk
- Username: Your own Webin ID
- Password: Your own password for your ENA account. Then click 'Connect'



The screenshot shows the FileZilla FTP client interface. The title bar reads "Webin-37120@webin.ebi.ac.uk - FileZilla". The connection fields are: Host: , Username: , Password: , Port: , and a "Quickconnect" button. The status bar shows: Status: Connection established, waiting for welcome message...; Status: Insecure server, it does not support FTP over TLS.; Status: Logged in; Status: Retrieving directory listing...; Status: Calculating timezone offset of server...; Status: Timezone offset of server is 0 seconds.; Status: Directory listing of "/" successful.

The local site is and the remote site is . The local site listing shows a directory structure with files like "000_command_line", "00_My_published_paper", "01_16sRNA", "02_BlastMatrix", "03_BlastAtlas", "04_PanCorePlot", "05_PangenomeTree", "06_gene_absent_LT2_10Jan11", "07_Pathogenicity_Islands_11Jan11", "08_Salmonella_MLST_11Jan11", "09_smallRNA", "100_Ana_13Nov14", "101_Dublin_01Dec14", "102_S_hyicus_02Dec14", "103_Resistome", "104_Kayode_Salmonella_28Jan15", and "105_Infantis_20Feb15".

The remote site listing shows a directory structure with files like "Project_2017_6_27_13_40_46", "Project_2019_6_10_13_49_54", "Project_2019_6_17_6_29_14", "Project_2019_6_7_15_10_51", "Project_2019_6_7_9_8_6", "22b_R2_001.fastq.gz" (2,818,734,525 gzip com...), "ACNJC6_01052017.fa.md5" (32 md5-file), "ACNJC6_curated_20170123.gff.md5" (32 md5-file), "ACNJC6_curated_20170204.embl.md5" (32 md5-file), "DTU-shit-2-FAB-P1_S129_L001_R1_001.fastq.md5" (32 md5-file), "DTU-shit-2-I6Z_S36_L001_R1_001.fastq.gz.md5" (32 md5-file), "DTU-shit-2-I6Z_S36_L001_R2_001.fastq.gz.md5" (0 md5-file), "DTU2013-1821-PRJ857-pseudintermedius-Staphylococcus-06241-1_S4_L001_..." (32 md5-file), and "DTU2013-1821-PRJ857-pseudintermedius-Staphylococcus-06241-1_S4_L001_..." (32 md5-file).

- 3) Transfer the genome files to ENA FTP

Webin-37120@webin.ebi.ac.uk - FileZilla

Host: webin.ebi.ac.uk Username: Webin-37120 Password: Port: Quickconnect

Status: Resolving address of webin.ebi.ac.uk
 Status: Connecting to 193.62.192.11:21...
 Status: Connection established, waiting for welcome message...
 Error: Connection closed by server
 Error: Could not connect to server
 Status: Disconnected from server
 Status: Delaying connection for 5 seconds due to previously failed connection attempt...

Local site: /Users/pile/Lab/ Remote site: /

Filename	Filesize	File
..		Dire
000_command_line		Dire
00_My_published_paper		Dire
01_16sRNA		Dire
02_BlastMatrix		Dire
03_BlastAtlas		Dire
04_PanCorePlot		Dire
05_PangenomeTree		Dire
06_gene_absent_LT2_10Jan11		Dire
07_Pathogenicity_Islands_11Jan11		Dire
08_Salmonella_MLST_11Jan11		Dire
09_smallRNA		Dire
100_Ana_13Nov14		Dire
101_Dublin_01Dec14		Dire
102_S_hyicus_02Dec14		Dire
103_Resistome		Dire
104_Kayode_Salmonella_28Jan15		Dire
105_Infantis_20Feb15		Dire
106_source_attribution		Dire

22 files and 381 directories. Total size: 97,370,680 bytes

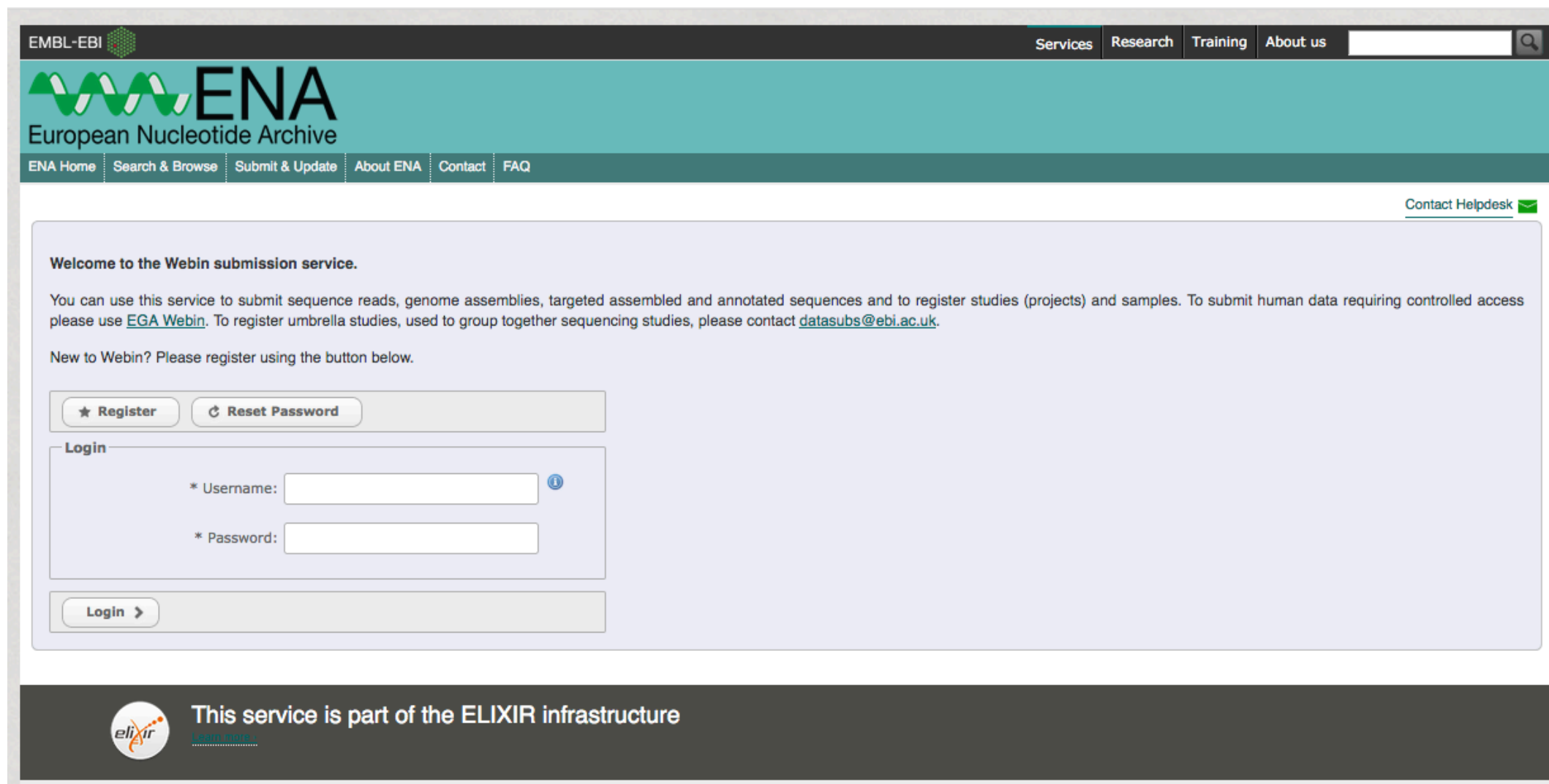
Filename	Filesize	Filetype	La
..			
Project_2017_6_27_13_40_46		Directory	28/
Project_2019_6_10_13_49_54		Directory	11/
Project_2019_6_17_6_29_14		Directory	17/
Project_2019_6_7_15_10_51		Directory	07/
Project_2019_6_7_9_8_6		Directory	07/
22b_R2_001.fastq.gz	2,818,734,525	gzip com...	23/
ACNJC6_01052017.fa.md5	32	md5-file	30/
ACNJC6_curated_20170123.gff.md5	32	md5-file	30/
ACNJC6_curated_20170204.embl.md5	32	md5-file	04/
DTU-shit-2-FAB-P1_S129_L001_R1_001.fastq.md5	32	md5-file	19/
DTU-shit-2-I6Z_S36_L001_R1_001.fastq.gz.md5	32	md5-file	19/
DTU-shit-2-I6Z_S36_L001_R2_001.fastq.gz.md5	0	md5-file	19/
DTU2013-1821-PRJ857-pseudintermedius-Staphylococcus-06241-1_S4_L001_...	32	md5-file	13/
DTU2013-1821-PRJ857-pseudintermedius-Staphylococcus-06241-1_S4_L001_...	32	md5-file	13/

151 files and 5 directories. Total size: 60,712,631,420 bytes

Server/Local file	Direction	Remote file
Webin-37120@webin.ebi.ac.uk		
/Users/pile/Meetings/175_EURL_September2018/presentation/Exercise_3/Fastq_files/Ecoli_036_R1.fq.gz	-->>	/Ecoli_036_R1.fq.gz
00:00:00 elapsed --:--:-- left		
/Users/pile/Meetings/175_EURL_September2018/presentation/Exercise_3/Fastq_files/Ecoli_036_R2.fq.gz	-->>	/Ecoli_036_R2.fq.gz
00:00:20 elapsed 02:38:46 left 9.5% 8,233,328 bytes (8.0 KiB/s)		

Submitting data in ENA

- 4) log in ENA with your own webin account
- <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 is a navigation bar with links for 'Services', 'Research', 'Training', and 'About us'. Below this is the ENA logo and the text 'European Nucleotide Archive'. A secondary navigation bar includes links for 'ENA Home', 'Search & Browse', 'Submit & Update', 'About ENA', 'Contact', and 'FAQ'. On the right side of the main content area, there is a 'Contact Helpdesk' link with a green checkmark icon.

The main content area features a welcome message: 'Welcome to the Webin submission service.' followed by a paragraph explaining 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 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.' Below this, it says 'New to Webin? Please register using the button below.'

There are two buttons: '★ Register' and '↻ Reset Password'. Below these is a 'Login' section with two input fields: '* Username:' and '* Password:'. A 'Login >' button is positioned below the password field.

At the bottom of the page, there is a footer with the Elixir logo and the text 'This service is part of the ELIXIR infrastructure'.



ENA

European Nucleotide Archive

Home

New Submission

Studies

Samples

Runs

Analyses

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 register umbrella studies, used to group together sequencing studies, please contact datasubs@ebi.ac.uk.

[Help about submitting read data](#)

▶ My account details



This service is part of the ELIXIR infrastructure

[Learn more](#)

- 5) click 'New Submission' and add 'Center Name'. Then click 'Next'

EMBL-EBI Services Research Training About us

ENA

European Nucleotide Archive

ENA Home Search & Browse Submit & Update About ENA Contact FAQ

[Contact Helpdesk](#) [Webin-37120](#) [Logout](#)

Home **New Submission** Studies Samples Runs Analyses

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

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.

Please select the type of submission you would like to make:

Submit sequence reads and experiments

We recommend that Fastq, BAM, and CRAM read files are submitted using [Webin-CLI](#).

When using this interface instead of [Webin-CLI](#), raw sequences must be [uploaded](#) in one of the supported [data formats](#) before they can be submitted. All data submitted in a single submission will be associated with the same study. Data for different studies must be submitted in separate submissions. The study and the sequenced samples can be either pre-registered or registered during the submission process. Please note that each individual study and sample should be registered only once. In addition, you will be asked to provide information about the sequencing libraries and instruments. Please quote the study accession number (ERP*) when citing data submitted to ENA.

Read [here](#) for more information on how submit your raw reads to ENA.

Register study (project)

Taxonomy Check/Request

Register samples

Submit genome assemblies

Submit other assembled and annotated sequences [formerly EMBL-Bank]

The first step of your submission is to [upload data files](#). Data files can be uploaded using FTP or Aspera, or using the [Webin File Uploader](#). If you have already uploaded your data files into your Webin upload area please proceed directly to the next step. Please note that unsubmitted files that are older than 2 months will be deleted as explained in our [Fair Use Policy](#).

Specify the name of the center you are submitting for.

Centre Name: (*)

[Next >>](#)

- 6) Click 'Create a new study'

EMBL-EBI Services Research Training About us

ENA

European Nucleotide Archive

[ENA Home](#) | [Search & Browse](#) | [Submit & Update](#) | [About ENA](#) | [Contact](#) | [FAQ](#)

[Contact Helpdesk](#) [Webin-37120](#) [Logout](#)

Home | New Submission | Studies | Samples | Runs | Analyses

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

Select an existing study or [Create a new study](#)

Search by: Accession / Unique name:

Show: accession unique name [Reset](#)

	Primary Accession	Secondary Accession	Title	Submission Date	Status
<input type="radio"/>	PRJEB33587	ERP116393	Whole genome sequencing of non-typhoidal Salmonellae serovars isolated from humans and poultry in Ma [...]	19-Jul-2019	Public
<input type="radio"/>	PRJEB33169	ERP115939	EFSA_EURL-AR_Confirmatory_Testing_2017	24-Jun-2019	Confidential
<input type="radio"/>	PRJEB32144	ERP114785	Formalin-fixed paraffin-embedded (FFPE) ring trial	12-Apr-2019	Public
<input type="radio"/>	PRJEB32083	ERP114714	Bacterial alpha diversity trajectories and immune-microbial associations of the gut, oral, and nasal [...]	08-Apr-2019	Confidential
<input type="radio"/>	PRJEB31650	ERP114226	Metagenomic assessment of the effect of sample processing on different microbiomes	11-Mar-2019	Public
<input type="radio"/>	PRJEB30991	ERP113496	The ESBL/AmpC resistance pool in Escherichia coli in pigs and pig farmers Vietnam	28-Jan-2019	Confidential
<input type="radio"/>	PRJEB30894	ERP113384	The gut, oral, and nasal microbiota in pediatric allogeneic HSCT and microbial marker candidates for [...]	21-Jan-2019	Confidential
<input type="radio"/>	PRJEB30892	ERP113382	Food Metagenomic ring trial - smoked salmon spiked with a microbial mock community	21-Jan-2019	Confidential
<input type="radio"/>	PRJEB30604	ERP113078	Genomic of Vibrio Cholera from Tanzania	07-Jan-2019	Confidential
<input type="radio"/>	PRJEB27955	ERP110095	Comamonas testosteroni	27-Jul-2018	Public

⏪ ⏩ 1-10 of 67 ⏪ ⏩ Number of rows

<< Previous Next >>

- 7) Set release date, provide title, and abstract

The screenshot shows the EMBL-EBI ENA submission interface. At the top, there is a navigation bar with 'Services', 'Research', 'Training', and 'About us'. Below this is the ENA logo and a secondary navigation bar with 'ENA Home', 'Search & Browse', 'Submit & Update', 'About ENA', 'Contact', and 'FAQ'. A progress bar at the top indicates the current step: 'Start' (checked), '>> Study', '>> Sample', '>> Run', '>> Finish'. The main content area is titled 'Select an existing study or Create a new study'. Below this, there is a paragraph explaining the submission process and a note about umbrella studies. The form is divided into two columns. The left column contains: 'Please specify the release date of your study: This is when your study will be made public.' with a text input field containing '30-Nov-2019'; 'Please provide a short name for the study:' with a text input field; 'Please provide a short descriptive title for the study: (*)' with a text input field containing 'E.coli'; and 'Please provide an abstract to describe the study in detail: (*)' with a large text area containing 'E.coli'. The right column contains: 'Please provide attributes to add a deeper description of the study:' with a table header 'Tag FieldType' and an 'Add' button; 'Please provide PubMed IDs of publications you want to associate with the study: (numeric value)' with a 'PubMed IDs' header and an 'Add' button; and '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.' with radio buttons for 'Yes' and 'No' (selected).

- 8) 'Select Checklist'

EMBL-EBI Services Research Training About us

ENA

European Nucleotide Archive

ENA Home Search & Browse Submit & Update About ENA Contact FAQ Contact Helpdesk Webin-37120 Logout

Home **New Submission** Studies Samples Runs Analyses

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

Start building your submission

i 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.

Select Checklist >

Upload a submission completed using a template spreadsheet

i If you have downloaded and filled a template spreadsheet please upload it using the **Submit 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](#).

Submit Completed Spreadsheet

- 9) Select 'Pathogens Checklists' and then select 'ENA GMI_MDK:1.1'

The screenshot shows the ENA (European Nucleotide Archive) submission interface. At the top, there is a navigation bar with 'EMBL-EBI' and 'Services', 'Research', 'Training', 'About us'. Below this is the ENA logo and a secondary navigation bar with 'ENA Home', 'Search & Browse', 'Submit & Update', 'About ENA', 'Contact', 'FAQ'. A progress bar at the top indicates the current step: 'Start' (checked), '>>', 'Study' (checked), '>>', 'Sample' (current), '>>', 'Run', '>>', 'Finish'. Below the progress bar, an information box states: 'Please select the most appropriate checklist from the list below then click the **Next >>** button.' The main content area lists several checklist categories: 'Environmental Checklists', 'Marine Checklists', and 'Pathogens Checklists'. Under 'Pathogens Checklists', the 'ENA Global Microbial Identifier reporting standard checklist GMI_MDM:1.1' is selected with a checked checkbox. Other unselected checklists include 'COMPARE-ECDC-EFSA pilot food-associated reporting standard', 'COMPARE-ECDC-EFSA pilot human-associated reporting standard', 'ENA parasite sample checklist', 'ENA prokaryotic pathogen minimal sample checklist', 'ENA virus pathogen reporting standard checklist', and 'ENA Influenza virus reporting standard checklist'. At the bottom, there are three buttons: '<< Previous', 'Skip >>', and 'Next >>'.

- 10) You have an option to download metadata as excel template or you just click 'Next' to fill out metadata via website

EMBL-EBI [Services](#) [Research](#) [Training](#) [About us](#)

ENA

European Nucleotide Archive

[ENA Home](#) [Search & Browse](#) [Submit & Update](#) [About ENA](#) [Contact](#) [FAQ](#)

[Contact Helpdesk](#) [Webin-37120](#) [Logout](#)

Home **New Submission** Studies Samples Runs Analyses

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

Please select any additional optional fields. Mandatory and recommended fields are selected by default. You may add any of the optional fields. You may also add custom fields.

Filter fields...
Add your own custom f **+ Add**

- + Human surveillance data
- + Collection event information
- + sample collection
- + Organism characteristics
- + host disorder
- + host description
- + General collection event information
- + Host association
- + Pointer to physical material

14 of 46 fields selected

+ Expand **- Collapse**

When you have selected the fields click the **Next >>** button to begin entering your data. Alternatively, download a template spreadsheet using the **Download Template Spreadsheet** button. Once you have filled the spreadsheet please restart the submission process and upload the spreadsheet using the **Upload Completed Spreadsheet** button.

Download Template Spreadsheet

<< Previous Skip >> Next >>

- 11) Add 'title', search 'Escherichia coli' (you have to wait until the system give you organism name to choose), add 'collected_by', add 'collection date'

[Contact Helpdesk](#) [Webin-37120](#) [Logout](#)

Home **New Submission** Studies Samples Runs Analyses

Start >> Study >> **Sample** >> Run >> Finish

Template Basic Details

Unique Name Prefix:

* Title:

Description:

Organism Details

If your organism is not found please go [here](#) and email datasubs@ebi.ac.uk with the required details listed on the page in order for us to request a taxon Id for your organism.

Search:

* Tax Id:

* Scientific Name:

Common Name:

Collection event information

* collected_by:

* collection date:

<< Previous Skip >> Next >>

- 12) Add 'country', 'latitude', 'longitude', 'environmental_sample', 'host health state', 'host scientific name', 'host associated' and 'isolate'

[Contact Helpdesk](#) [Webin-37120](#) [Logout](#)

Home
New Submission
Studies
Samples
Runs
Analyses

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

Please complete any fields that you would like to apply to all samples. This will act as a template for all samples.

* geographic location (country and/or sea):

* geographic location (latitude):

* geographic location (longitude):

sample collection

* environmental_sample:

host description

* host health state:

* host scientific name:

Host association

* Is the sequenced pathogen host associated?:

Infraspecies information

* isolate:

serovar:

<< Previous
Skip >>
Next >>

- 13) add number of sample. In this case, it is only 1 sample to add (click '+Add')

Home **New Submission** Studies Samples Runs Analyses

Start >> Study >>> **Sample** >> Run >> Finish

Please add samples to the submission. Multiple samples can be created by increasing the number by the add button

+ Add 1 samples

1

⏪ ⏩ 1-1 of 1 ⏪ ⏩

Please submit by clicking the **Submit** Button. Alternatively, download your data as a spreadsheet using the **Download Spreadsheet** button. Once you have filled the spreadsheet please restart the submission process and upload the spreadsheet using the **Upload Completed Spreadsheet** button.

Download Spreadsheet

There are no samples to submit

Basic Details

< Previous Sample **Next Sample >**

* Unique Name: 1

* Title: Ecoli

Description:

Organism Details

If your organism is not found please go [here](#) and email datasubs@ebi.ac.uk with the required details listed on the page in order for us to request a taxon Id for your organism.

Search:

* Tax Id: 562

* Scientific Name: Escherichia coli

Common Name:

Collection event information

* collected_by: DTU

<< Previous **Skip >>** **Next >>**

- 14) Add 'Unique Name' for the submitted sample. In this case, try to add 'Ecoli_001'

[Contact Helpdesk](#) [Webin-37120](#) [Logout](#)

Home **New Submission** Studies Samples Runs Analyses

Start >> Study >> **Sample** >> Run >> Finish

Please add samples to the submission. Multiple samples can be created by increasing the number by the add button

+ Add 1 samples

Ecoli_001

1-1 of 1

Please submit by clicking the **Submit** Button. Alternatively, download your data as a spreadsheet using the **Download Spreadsheet** button. Once you have filled the spreadsheet please restart the submission process and upload the spreadsheet using the **Upload Completed Spreadsheet** button.

Download Spreadsheet

There are no samples to submit

Basic Details

* Unique Name:

* Title:

Description:

Organism Details

If your organism is not found please go [here](#) and email datasubs@ebi.ac.uk with the required details listed on the page in order for us to request a taxon Id for your organism.

Search:

* Tax Id:

* Scientific Name:

Common Name:

Collection event information

* collected_by:

<< Previous Skip >> Next >>

- 15) Click type of submitted files. In this case is 'Two Fastq files (Paired)'

[Contact Helpdesk](#) [Webin-37120](#) [Logout](#)

Home **New Submission** Studies Samples Runs Analyses

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 (*)	Des
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

- 16) Add required metadata (*)

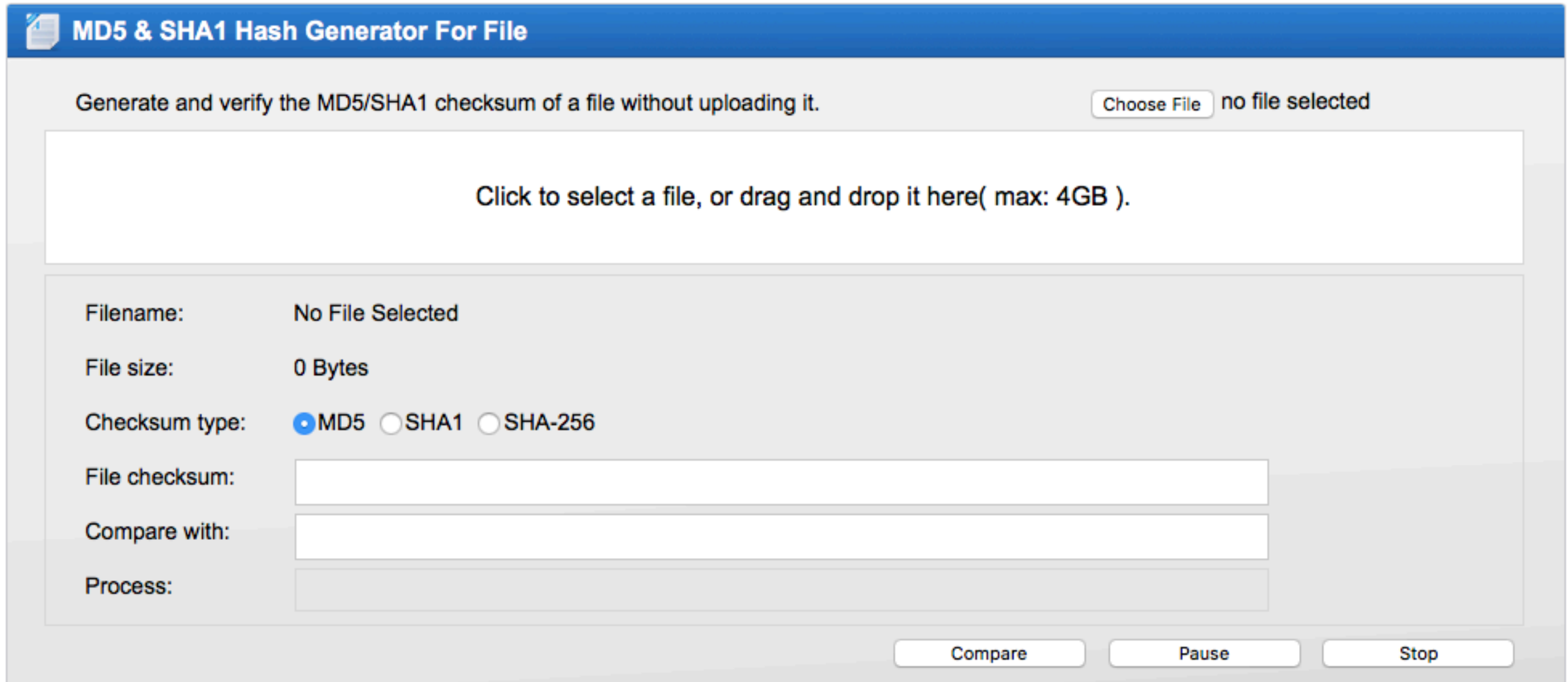
[Sample reference suggestions] ⓘ	Sample reference (*) ⓘ	Instrument Model (*) ⓘ	Library Name ⓘ	Library Source (*) ⓘ	Library Selection (*) ⓘ	Library Strategy (*) ⓘ	Des
✖ Ecoli_001	Ecoli_001	Illumina MiSeq		GENOMIC	RANDOM	WGS	
+							

Mandatory fields are denoted by (*).

Library Strategy (*) ⓘ	Design description ⓘ	Library construction protocol ⓘ	Insert Size (*) ⓘ	First File Name (*) ⓘ	First MD5 checksum ⓘ	Second File Name (*) ⓘ	Second MD5 checksum ⓘ
			300	Ecoli_036_R1.fq.gz		Ecoli_036_R2.fq.gz	

- 17) MD5 checksum is also required

<http://onlinemd5.com>



The screenshot shows a web application titled "MD5 & SHA1 Hash Generator For File". The interface includes a header with the title and a sub-header with the instruction "Generate and verify the MD5/SHA1 checksum of a file without uploading it." and a "Choose File" button indicating "no file selected". A large central area contains the text "Click to select a file, or drag and drop it here(max: 4GB).". Below this, there are input fields for "Filename:" (displaying "No File Selected"), "File size:" (displaying "0 Bytes"), and "Checksum type:" with radio buttons for "MD5" (selected), "SHA1", and "SHA-256". There are also empty input fields for "File checksum:", "Compare with:", and "Process:". At the bottom right, there are three buttons: "Compare", "Pause", and "Stop".

MD5 & SHA1 Hash Generator For File

Generate and verify the MD5/SHA1 checksum of a file without uploading it.

[Choose File](#) no file selected

Click to select a file, or drag and drop it here(max: 4GB).


Filename: Ecoli_036_R1.fq.gz

File size: 87,524,866 Bytes

Checksum type: MD5 SHA1 SHA-256

File checksum: E434CAE6777C250EBE45C0DBC34E8B1C

Compare with:

Process:  100.00%

[Compare](#)

[Pause](#)

[Stop](#)

MD5 & SHA1 Hash Generator For File

Generate and verify the MD5/SHA1 checksum of a file without uploading it.

Choose File no file selected

Click to select a file, or drag and drop it here(max: 4GB).


Filename: Ecoli_036_R2.fq.gz

File size: 86,096,901 Bytes

Checksum type: MD5 SHA1 SHA-256

File checksum: 95C4F0A13B3AF36C9AE0EE570266C253

Compare with:

Process:  100.00%

Compare

Pause

Stop

```
Fastq_files — -bash — 130x45
pile@co... [screen... ... [screen... ... pile@co... ... pile@co... ... [screen... ... [screen... ... [screen... ... pile@co... >> +
Last login: Fri Sep 13 15:33:57 on ttys011
[foodly-al0001:Fastq_files pile$ cd /Users/pile/Meetings/175_EURL_September2018/presentation/Exercise_3/Fastq_files
[foodly-al0001:Fastq_files pile$ ls -lh
total 339112
-rwx----- 1 pile  staff   83M 22 Sep  2018 Ecoli_036_R1.fq.gz
-rwx----- 1 pile  staff   82M 22 Sep  2018 Ecoli_036_R2.fq.gz
[foodly-al0001:Fastq_files pile$ md5 Ecoli_036_R*
MD5 (Ecoli_036_R1.fq.gz) = e434cae6777c250ebe45c0dbc34e8b1c
MD5 (Ecoli_036_R2.fq.gz) = 95c4f0a13b3af36c9ae0ee570266c253
foodly-al0001:Fastq_files pile$ █
```

- 18) The last step is to click 'Submit'. If something goes wrong, the system will give you error message.
- **DO NOT click 'Submit' in this exercise**

Mandatory fields are denoted by (*).

Strategy (*)	Design description	Library construction protocol	Insert Size (*)	First File Name (*)	First MD5 checksum	Second File Name (*)	Second MD5 checksum
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text" value="300"/>	<input type="text" value="Ecoli_036_R1.fq.gz"/>	<input type="text" value="e434cae6777c250ebe45c"/>	<input type="text" value="Ecoli_036_R2.fq.gz"/>	<input type="text" value="95c4f0a13b3af36c9ae0e"/>

[Disable Edit mode](#)

[<< Previous](#)

[Submit](#)

Topics

- Submitting data via ENA
- **Sharing data via ENA private hub**
- Accessing sequence data through ENA private hub
- Accessing WGS results through ENA private hub

ENA private hub

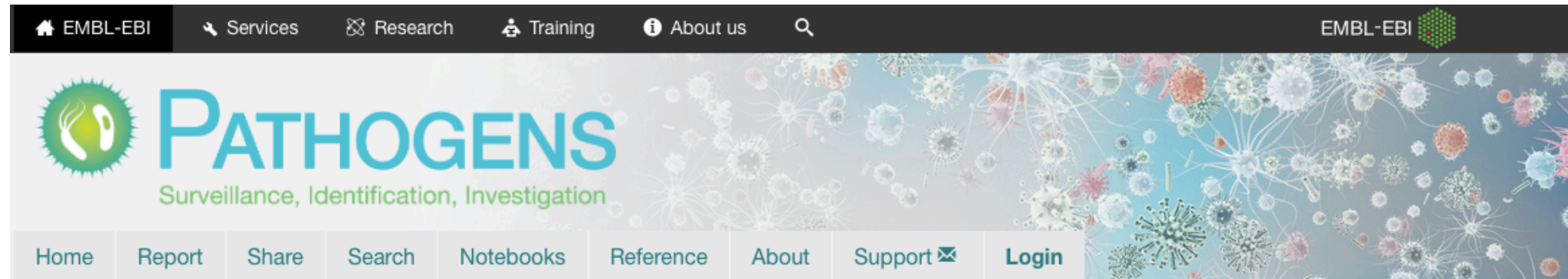
- a private repository for submitting and downloading of public and private genomic and phenotypic data for AMR
- A **data provider** - able to upload data to ENA, share data to the hub and download metadata and sequences
 - Require to have ENA account (Webin ID)
 - <https://www.ebi.ac.uk/ena/submit>
 - Ask ENA to link your ENA account to the private hub
- A **data consumer** - able to download metadata and sequences
 - No ENA account required

ENA private hub

- Submitting data to ENA (only for data provider)
- Share sequence and metadata in the hub (only for data provider)
- Access metadata
- Download sequences

ENA private hub

- 1) Login to the pathogen website using your ENA account
- <https://www.ebi.ac.uk/ena/pathogens/login>



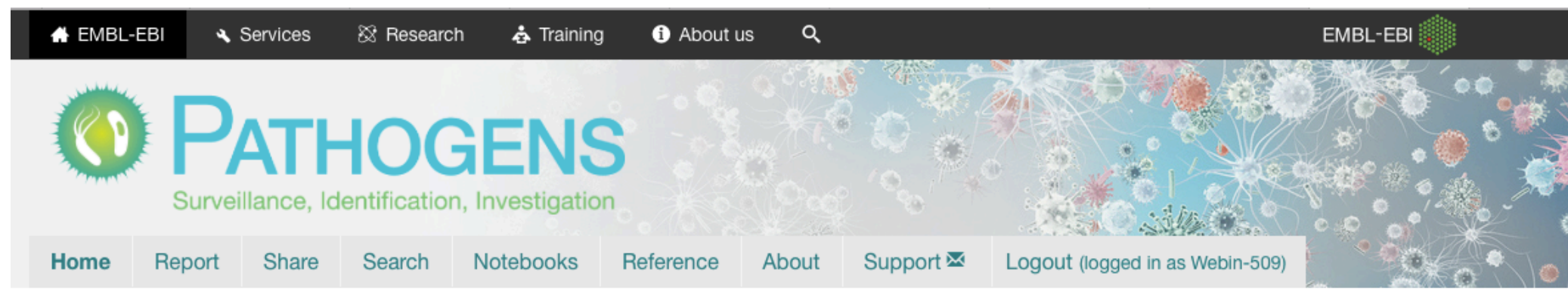
Login

Username

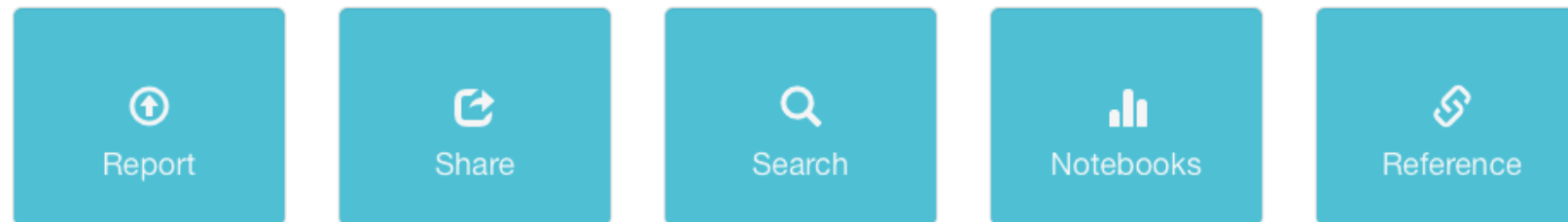
Password

Login

- 2) Click 'Share'. You will see one the columns is 'dcc_bromhead' which is the name of our EURL hub



Welcome to Pathogens



- 3) You can share any of your study by clicking at a button under our hub column (dcc_bromhead)

Choose studies to share with data hubs

Type to filter by study accession or title

Study Accession	Study Title	dcc_broadbent	dcc_bromhead	dcc_cole	dcc_liszt	dcc_schubert	dcc_strauss
PRJEB14086	reference testing for ENGAGE E.coli (P160095)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
PRJEB18618	EURL reference testing	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

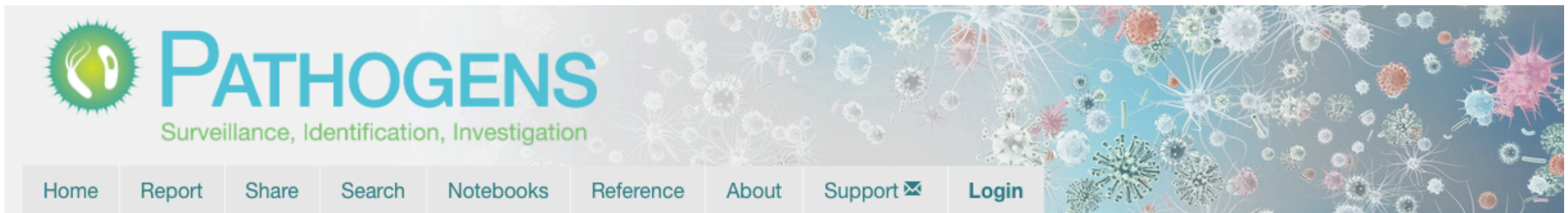
- 4) Please allow 24 hours for system to link the shared data to the hub

Topics

- Submitting data via ENA
- Sharing data via ENA private hub
- **Accessing sequence data through ENA private hub**
- Accessing WGS results through ENA private hub

Accessing metadata in ENA hub

- 1) Search through web portal using ‘dcc_bromhead’ (ENA hub)
- <https://www.ebi.ac.uk/ena/pathogens/login>



Login

Username

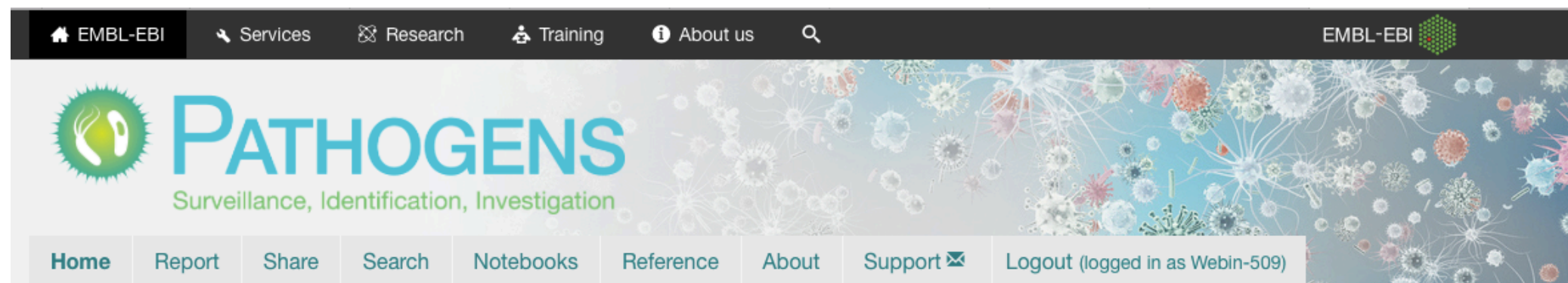
dcc_bromhead

Password

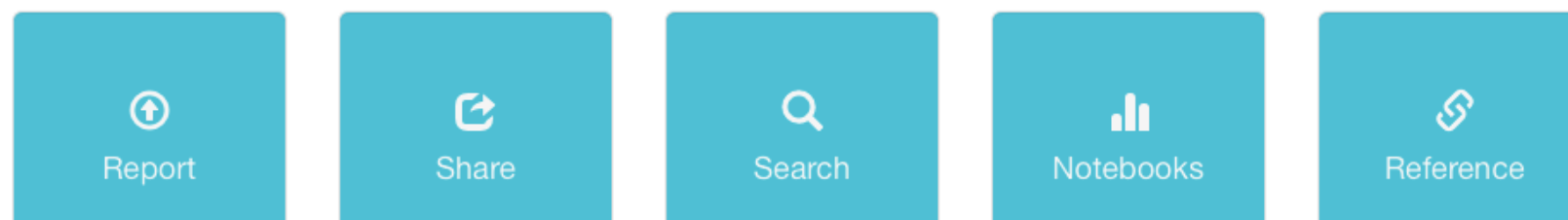
.....

Login

- 2) After login, go to 'Search' tab, select data type e.g. 'read run' or 'analysis'
- 3) On the next few pages, you can refine your search and select fields manually



Welcome to Pathogens





PATHOGENS

Surveillance, Identification, Investigation

- Home
- Report
- Share
- Search**
- Explore
- Reference
- About
- Support ✉
- Logout (logged in as dcc_bromhead)

Advanced Search



Data type:

Nucleotide sequence analyses from reads

- Copy Curl Request
- Next
- Search



PATHOGENS

Surveillance, Identification, Investigation

- Home
- Report
- Share
- Search**
- Explore
- Reference
- About
- Support
- Logout (logged in as dcc_bromhead)

Advanced Search

DATA TYPE

QUERY

FIELDS

DATA FILTERS

RESULTS

Default fields Manually select fields

Fields:

collection_date,country

Select and order Fields:

Available Fields

- checklist
- collected_by
- collecting_institute
- dev_stage
- environmental_sample
- first_public
- host_common_name
- host_scientific_name
- host_sex
- host_status
- host_subject_id
- host_tax_id



Selected Fields

- collection_date
- country

Back

Copy Curl Request

Next

Search

- 4) Specify 'Dcc data only' before performing the search

The screenshot displays the 'PATHOGENS' web application interface. The header includes the logo and the text 'PATHOGENS Surveillance, Identification, Investigation'. A navigation menu contains links for Home, Report, Share, Search, Explore, Reference, About, Support, and Logout (logged in as dcc_bromhead). The main content area is titled 'Advanced Search' and features a progress indicator with five steps: DATA TYPE, QUERY, FIELDS, DATA FILTERS, and RESULTS. The 'DATA FILTERS' step is currently active. Below the progress indicator, there are three filter options: 'Limit (number of records):' with an empty input field, 'DCC data only:' with a checked checkbox, and 'Include Metagenomes:' with an unchecked checkbox. At the bottom right, there are three buttons: 'Back', 'Copy Curl Request', and 'Search'.

- 5) Search results/reports can be viewed and downloaded. You can download report by clicking at Download report either in JSON or TSV. The TSV format can be imported and viewed in Excel

Advanced Search

Progress indicator: DATA TYPE (●), QUERY (●), FIELDS (●), DATA FILTERS (●), RESULTS (●)

Type to filter results [Download report:](#) [JSON](#) [TSV](#)

Analysis Accession	Description
ERZ390162	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390163	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390164	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390165	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390166	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390167	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390168	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390169	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390170	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)
ERZ390171	Antibiotic sensitivity test for ERP015694: reference testing for ENGAGE E.coli (P160095)

Showing 1 to 10 of 166 rows. Rows per page: 10

Accessing sequence data in ENA hub

The access to the ENA data hub is given via password protected FTP. Below you find FTP specifications, the folder organization to download the fastq files and instructions to set up FTP access via Mac or Windows.

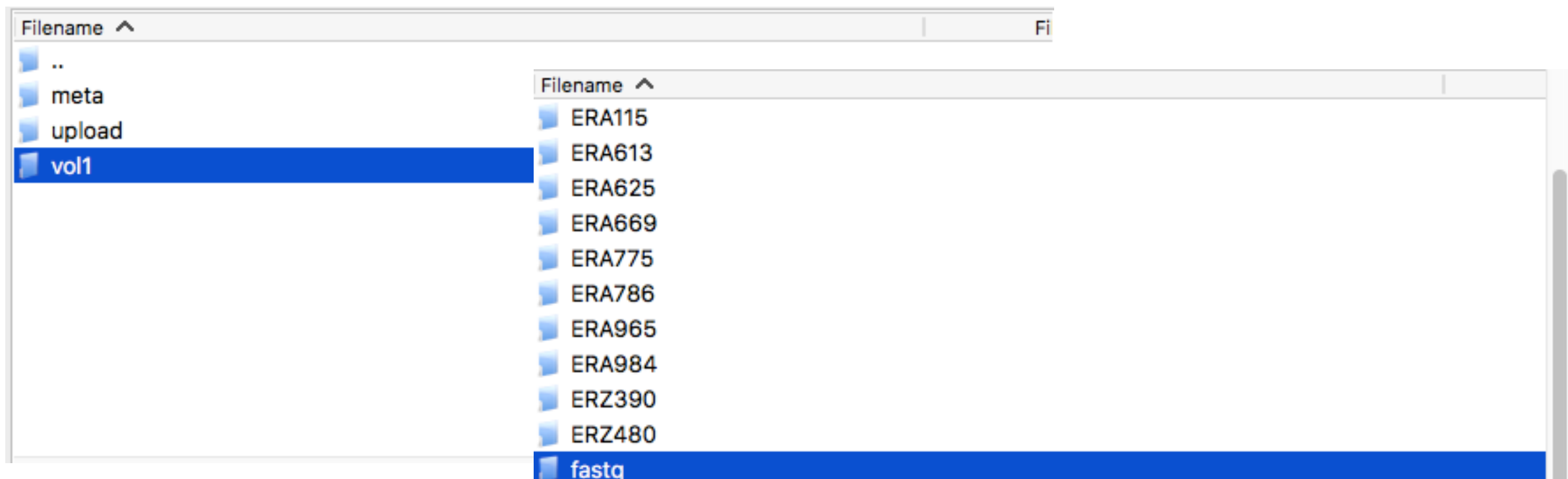
Specifications

Host: ftp://ftp.dcc-private.ebi.ac.uk

Username: dcc_bromhead

Password:

Sequences in FASTQ are under folder: vol1/fastq



Topics

- Submitting data via ENA
- Sharing data via ENA private hub
- Accessing sequence data through ENA private hub
- **Accessing WGS results through ENA private hub**

WGS results via ENA hub

- 1) Go to pathogen web portal using 'dcc_bromhead' (ENA hub)
- <https://www.ebi.ac.uk/ena/pathogens/login>

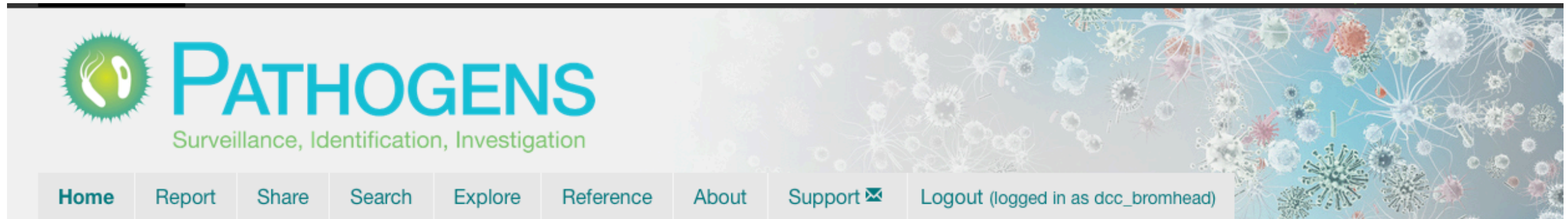


Login

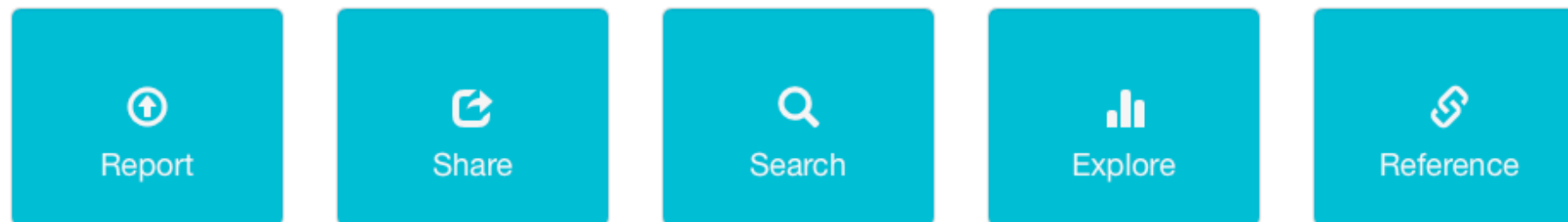
Username

Password

- 2) After login, go to 'Explore' tab



Welcome to Pathogens



- 3) Click 'View' to view report. Click 'Download' to download interactive report



Home Report Share Search **Explore** Reference About Support ✉ Logout (logged in as dcc_bromhead)

Data Hubs

A data hub is a platform that includes workflows for structured data storage and sharing of sequencing data and its analysis interpretations. For more details please see:

<https://www.biorxiv.org/content/10.1101/555938v1>

To see which Data Hubs and analysis workflows are available currently, please see [here](#).

Notebook reports

The basic view contains a summary of the notebook report and can be viewed in most web browsers. The full view with dynamic controls is currently supported in Firefox, Safari, Chrome, and MS Edge. You can also download the full report (as a zip archive) to view in any browser. After downloading, please extract the zip file and open the index.html file within.

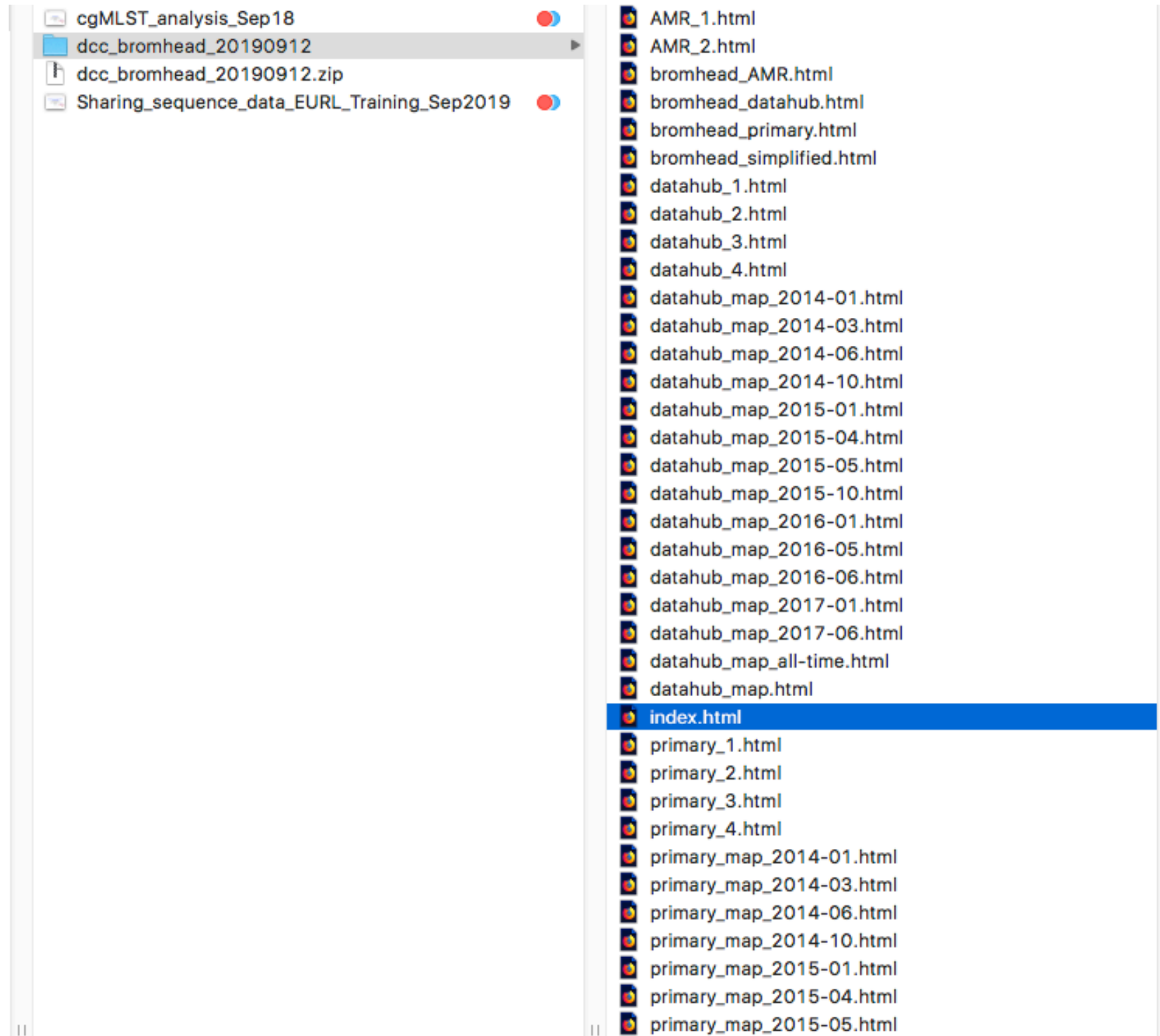
[View Demo](#)

Data Hub	Reports (by date)	View Basic Report	View Full Report	Download Full Report (zip)
dcc_benoit	20190913 (latest) ▾	View	View	Download
dcc_bromhead	20190912 (latest) ▾	View	View	Download

Showing 1 to 2 of 2 rows.

[Link to download interactive report for dcc_bromhead](#)

- 4) To view interactive report; unzip the file and click 'index.html'



Browser window showing file:///Users/pile/Meetings/204_EURL_training_Sep19/presentation/dcc_bromhead_20190912

Data Hub content Primary Analysis AMR

dcc_bromhead database state overview

This notebook was using the database state as of 20190912.

Table 1.) All the submitted reads to the datahub

[pop out]

center_name	country	study_accession	Totals
ANIMAL AND PLANT HEALTH AGENCY	NA	PRJEB13576	3
		PRJEB26940	10
		PRJEB26942	41
Center for Genomic Epidemiology	France	PRJEB21546	1
	Germany	PRJEB21546	2
	Austria	PRJEB18618	2
	Belgium	PRJEB18618	9
	Bulgaria	PRJEB18618	4
	Croatia	PRJEB18618	2

Browser tabs: /Users/pile/Meetings/204_EURL_tra X +

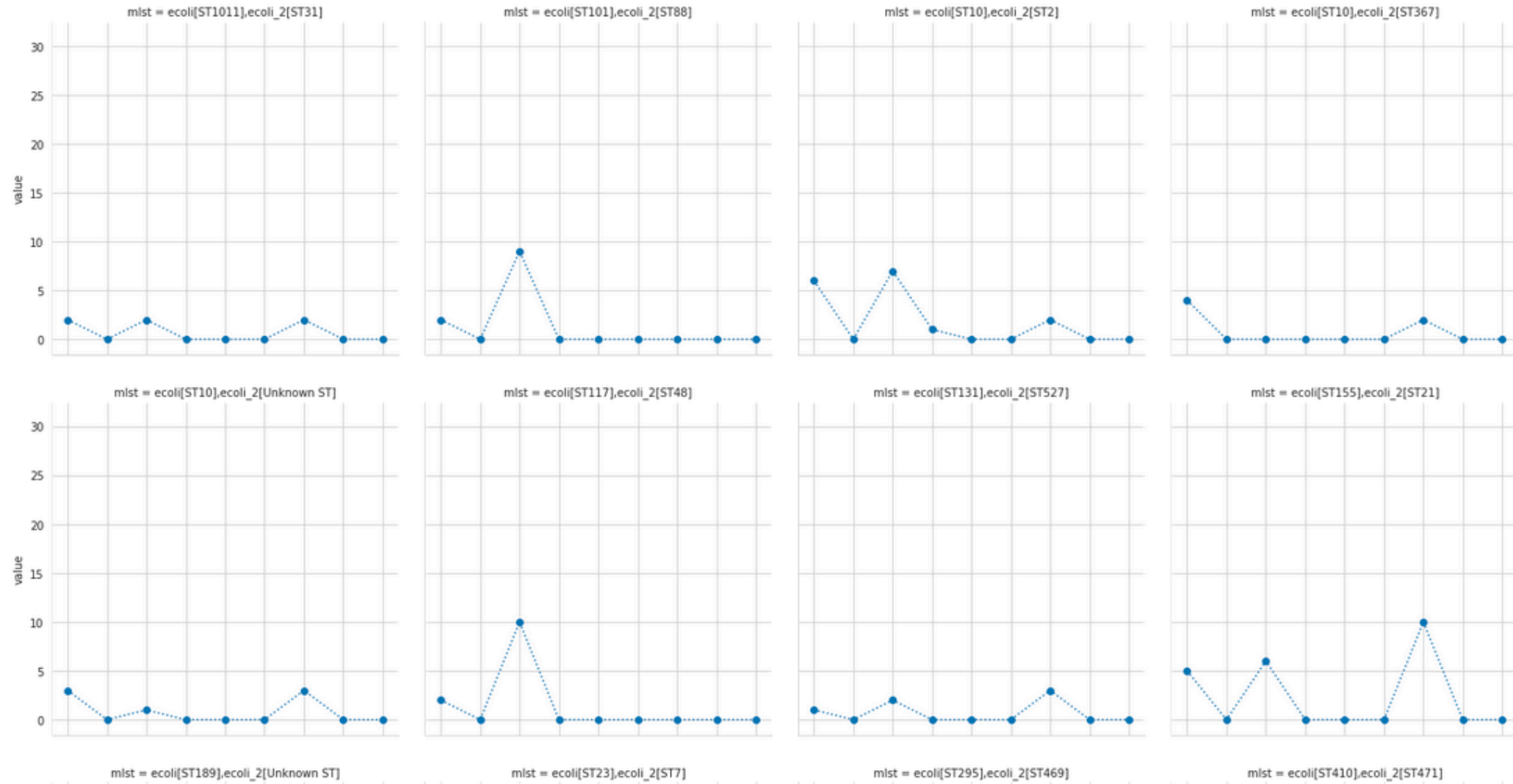
Address bar: file:///Users/pile/Meetings/204_EURL_training_Sep19/presentation/dcc_bromhead_2019091... Search

Bookmarks: Most Visited, Getting Started, Apple, iCloud, Facebook, Twitter, Wikipedia, Yahoo, News, Popular, http://rest.pubmlst....

Data Hub content Primary Analysis AMR

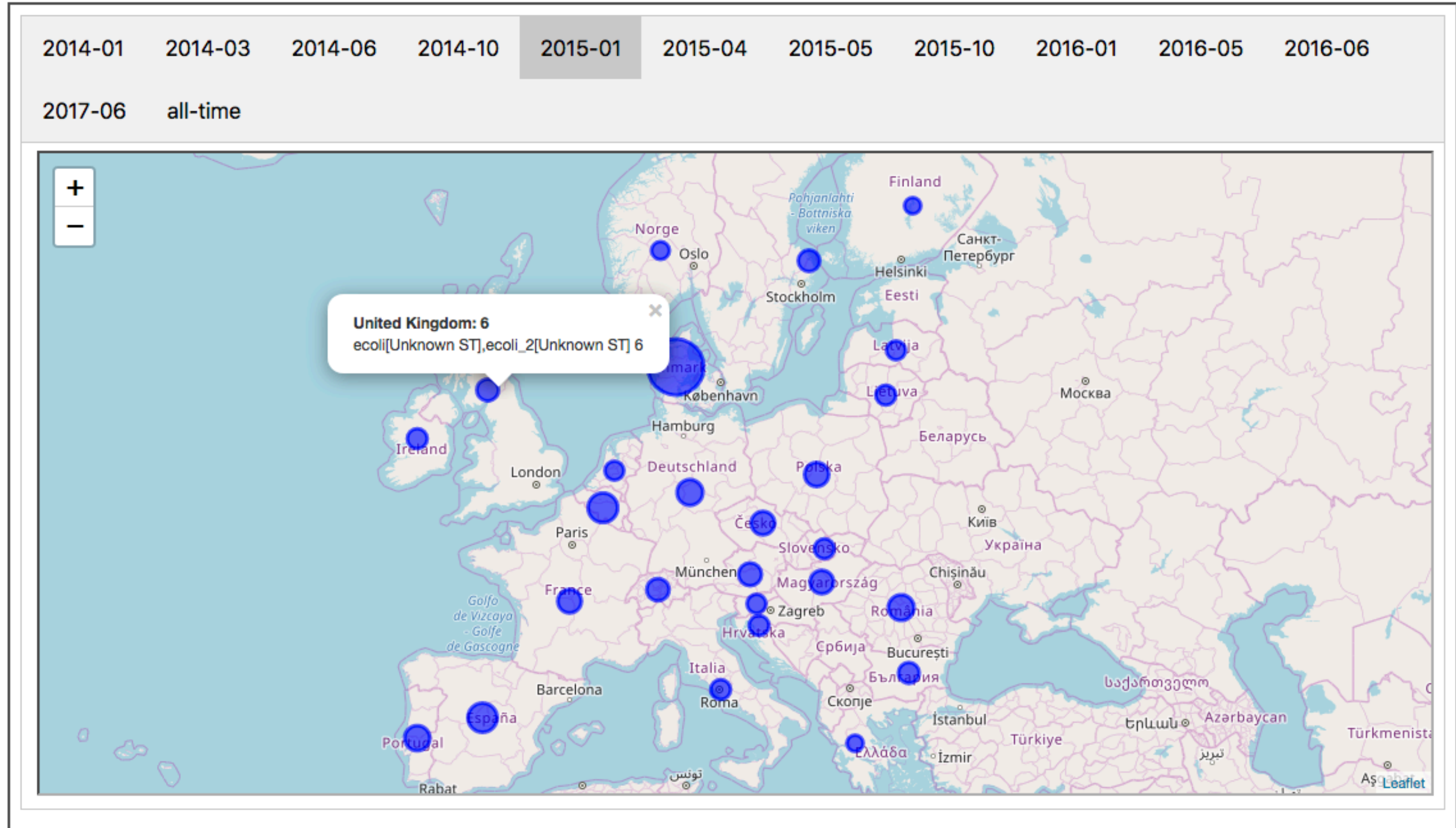
Plot shows MLST appearance frequency

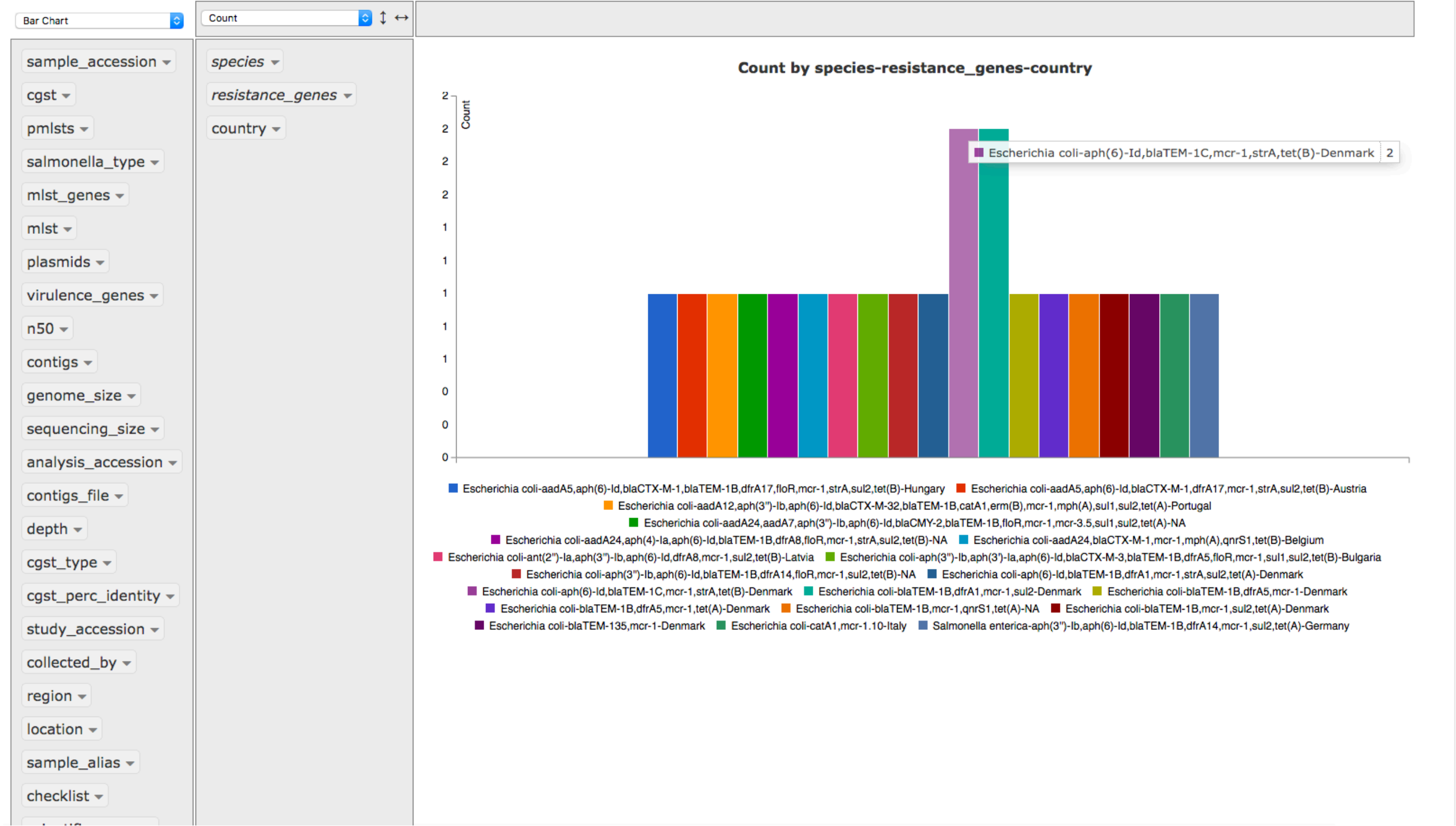
MLSTs shown only when there are more than 5 occurrence after 2014-01-01



Map shows the samples with different MLST types by submission date

monthly plots are available only after 2014-01-01





AMR data - dcc_bromhead datahub content overview

This notebook was using the database state as of 20190912.

Table 1.) Samples submitted to the datahub with AMR_ANTIBIOGRAM tag

[\[pop out\]](#)

Table		Count	scientific_name						
study_accession	analysis_type		analysis_type	checklist	center_name	country	scientific_name	Escherichia coli	Totals
isolation_source	checklist		AMR_ANTIBIOGRAM	ERC000011	DTU-GE	Denmark		41	41
	center_name					Denmark		88	88
	country					France		24	24
				ERC000029	DTU-GE	Germany		12	12
						Poland		1	1
							Totals	166	166

Table 2.) MIC / Disk diffusion test's results in the datahub

Thank you for your attention



Pimlapas Leekitcharoenphon (Shinny), PhD
Research Group Genomic Epidemiology
WHO Collaborating Centre for Antimicrobial Resistance in Food borne Pathogens
and Genomics
European Union Reference Laboratory for Antimicrobial Resistance
National Food Institute, Technical University of Denmark
pile@food.dtu.dk



@ShinnyPimlapas