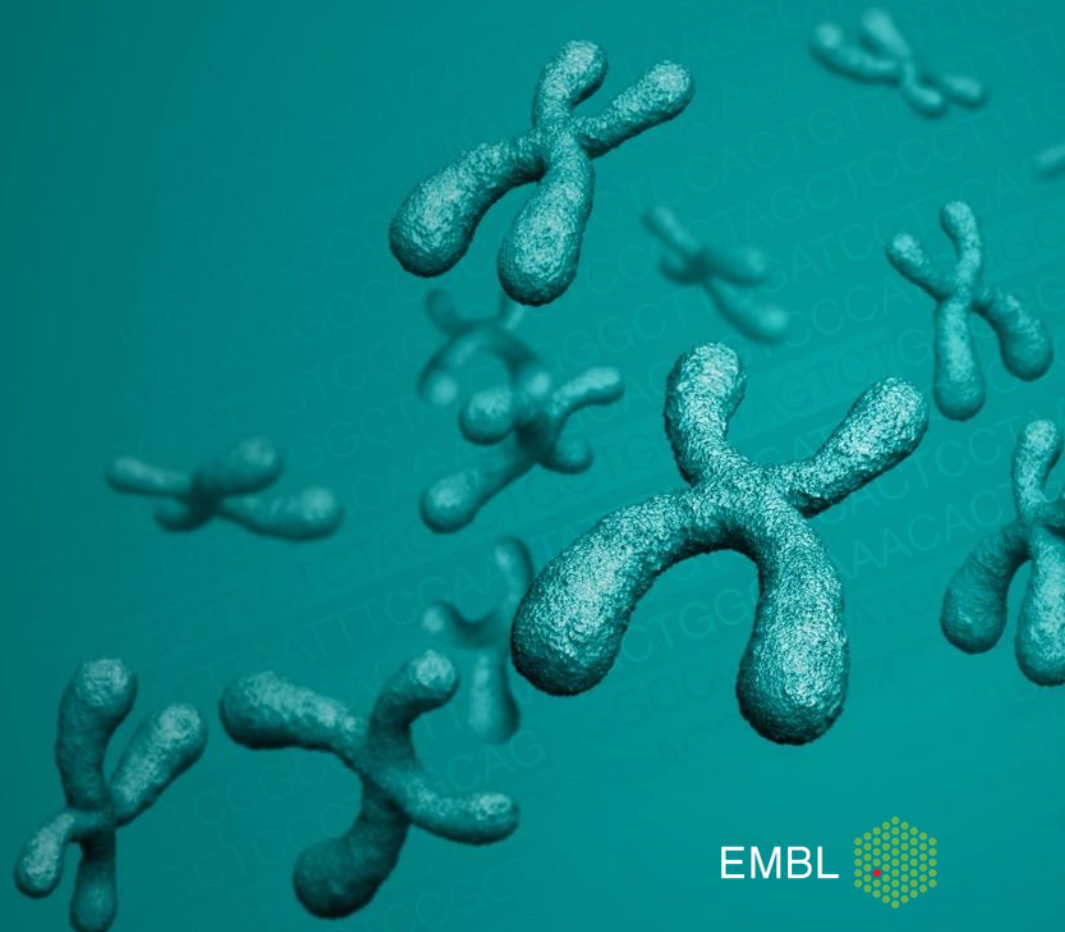
The European Nucleotide Archive

Who We Are And How We Can Help

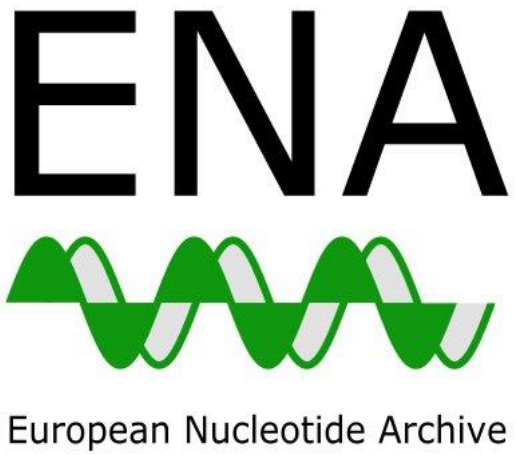
Sam Holt



EMBL

Contents

- ENA Background
  - What We Do
  - Why Submit Data
- Data and Metadata Models
- Data Submission
  - Submission Options
- Metagenomic topics
  - Metagenomic Standards
  - Metagenomic Submission



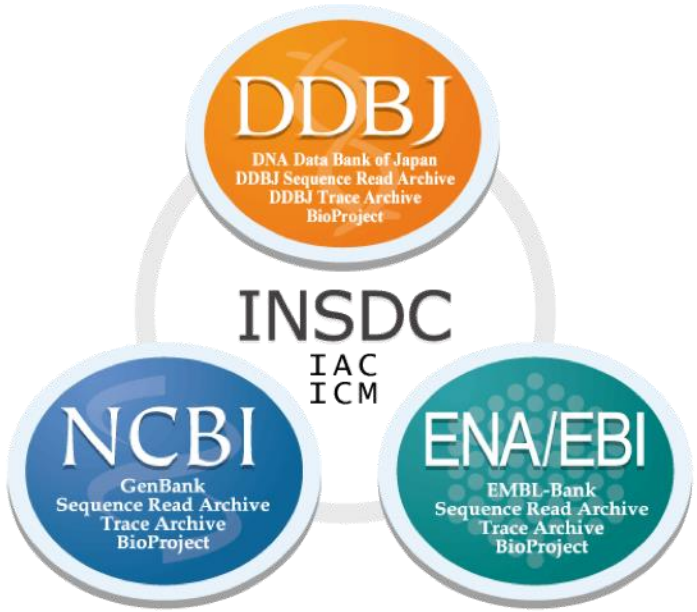
ENA Background

What We Do

How We Structure Our Data

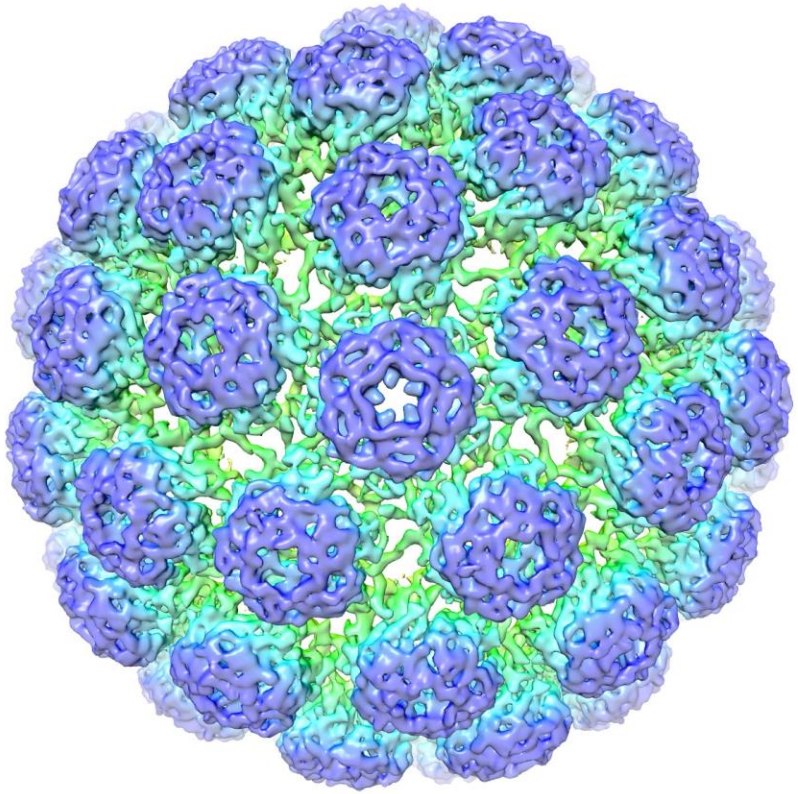
EMBL

ENA Background: What Is ENA?



- A repository of the world's nucleotide data
- Creators of tools for submission and retrieval
- European node of INSDC
- A basis for other tools

ENA Background: New Discoveries From Old Data



Cryo-EM structure of BK polyomavirus, PDBj

- Polyomavirus – *“many tumours”*
- Buck *et al.* report discovery of new polyomaviruses in fish, cows, and sheep
- Sequence searches against INSDC data identified more new species in the genomes of 7 invertebrates
- Evidence that polyomavirus existed in the last common arthropod-vertebrate ancestor

Submitting Data: Why?

- All data in the ENA is submitted by members of the research community
- What motivates people to submit?
  - Open data
    - Reproducibility
    - Reusability
  - 3rd party access
  - Archival
  - **Publication**
  - MGnify

Data and Metadata Models

How We Structure Our Data

ENA Background: The Data Model

- ENA stores huge amounts of data

  … from many users

    … with samples from many taxa

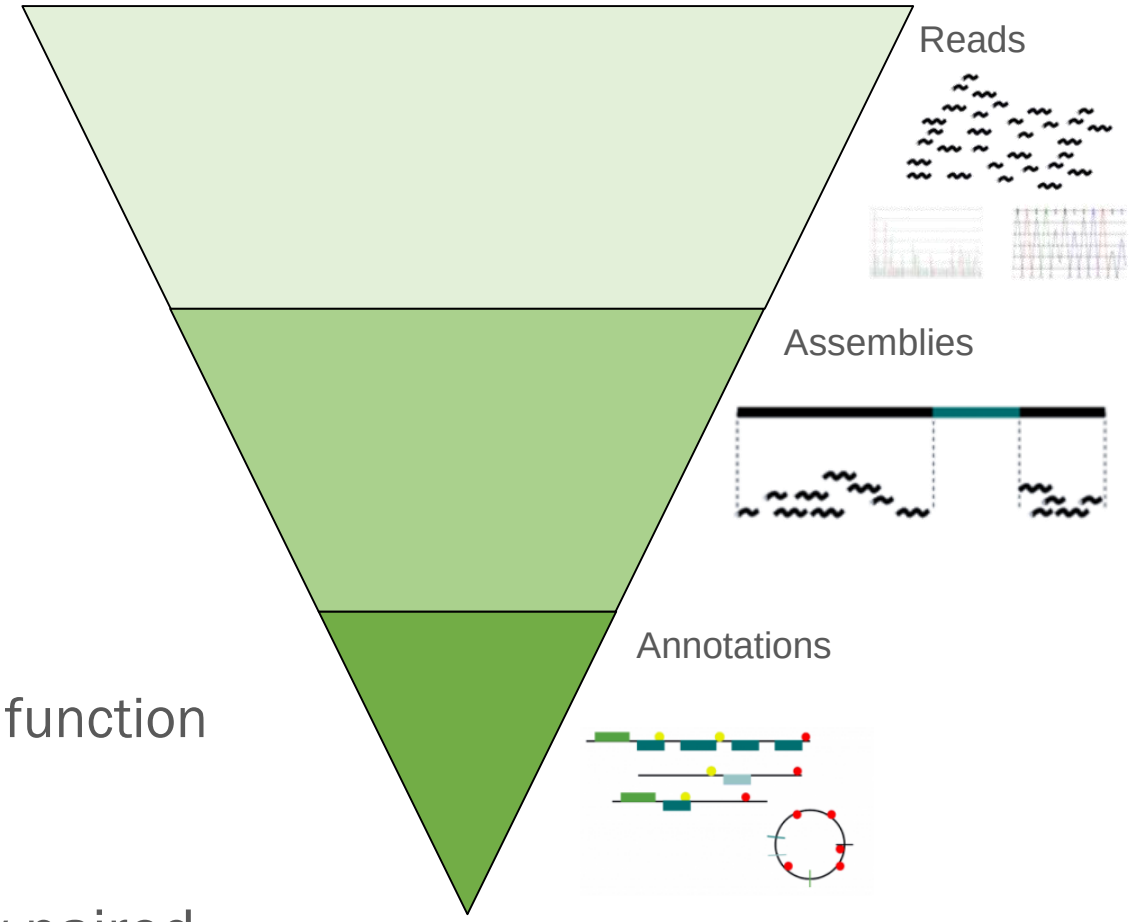
      … who use many different techniques

        … and sequence on many different platforms

- But we need to display data in a consistent manner
- A robust data model is the first step in achieving this

ENA Background: The Data Model

- Sequence data is organised into tiers:
  - **Reads:** raw sequence data
  - **Assemblies:** reconstructions of actual replicons, full or partial
  - **Annotations:** interpretations of biological function
- **Annotation** and **assembly** are frequently paired



ENA Background: The Data Model

A FASTQ file is an example of data from the read tier:

```
@SRR6033657.1 1/1
GACCATCTAGCGACCTCCACCTCATCCGGTAGAGCGAATGATTATATCCCTTGTTTTCTAAACTACCTCAACCTATTCTC
TACCTCCAACTAGTTGAGTACCCTGTCTTTCTTTCTTTATGAATCCTTTGTGTTTCGGTTCATATTGCCCC
+
--AB8CEF,,C::CFEFGFGGFGDGGG>+CC,,,<@++8,,<C,,,,,<;CC,,,,6<,,,,;,,:9:C,,
9:C,:CC@C,,,:9C@,,,:,,,9,,,,,:99?,,959?@@E,,,::,:,,,,55,,,,,,99AA+++44,,,,9,44,+
@SRR6033657.2 2/1
TTTGATGATGATTCCTTTCTCTTTCATTCGATGATCCCATCTGATTCTAATCCATTATTCCATTCAATCCCATTTTATGA
AAATTCCATTCGATTCCTTTCAATGTGGTTGTCGCTAGTTGGTCAGGTTTTGGTGGAGTTCGCTAGATGGT
+
ACC@<E,,C<9EGFGGGGGGGGGGGGGGGGFG99EFFFFEFG9,CCEC,,<CE,CC,CEEC,CCE,,<C@C,CCE,,<,,
,,,<CEE,CCE,,<AE@FBEE,,<,:,,,,,8C,47,+,:,,8A,,,9,9D,+9,,,,,99,4?,+,9,,9
@SRR6033657.3 3/1
CACCATCTAGCGACCTCCACCTCTCTGCGTTTGGTTCATCCCACATCCCCAGTTCTTCTTACCTATATTGCCCCTCTTTC
CTCTCTCTTCCAACTTCTCTTTTCTTCATCATCCATTCCCTCCACACATCTCACCCCTTTTCTTTTTTATT
+
--8A,CEEA@E:@CFGFGGGGGGFGGCFCFFE,,CCEDFECC8C,,;,;:,,;<6C,;CC,;;,,,,<9,,:,:,8C@,,
:,99CC,,6,<,,,,:<6:,9,,95:@,:,,59,,,,:,,::4,,,,,5,9?,9AE+495,,,,449+,,,
@SRR6033657.4 4/1
GTTTGTCCCTCAAACCTCCCCAAAAACTAGGGAAGCTAGCTAGGCACCACCTCTTGCCATACTTACTACACCCACTTTCA
CTTATCCATTCCATTCCCCTCCACTCACTTTTCTCATCAATCACTTTTTTATGTCGCTTTATGTTCATATC
+
-ACB<FCCCCF8E-CDFGFGGGGFGCGGDFF8,,,CF9,,,,,,,,,:,:,,,,,,,;,<,,<,,,,,<,,:
8,8,:99,,<9,66<,996,,,6<<@@6@@,,5:,,,,,959,,9,,,9,,,,,,5,+,,,99,+4++,4,,99,,,94
@SRR6033657.5 5/1
ACCATTCTGAGGGAACCTTTGGGCGCCTCCGTTACCTTTTAGGAGGCTACCGCCCCAGTCAAACTTCCCGTCAGACACTG
TCTCCGATAGCCATCACCTATCTGGGTTACAGTGGCCATAACACAAGGGTAGTATCCCATCCTCCTCTCCT
+
@CC9CEFGF,-,B@CFGFGGAFGGGGGGGFGGGGEGGGGEC,C,,,6,,;6+:67@,,:9,,,::,<C@@C@,,,:
6CC,<C6?C++8+,,,,::5BA=,?A5,,4:?7,,,:,,49,AB,A,A,+,,+9,+E,?A<A,,8,74,?=4?:,
```

But is it interesting?

The Metadata Model

- Data without any context has no value
- Metadata tells us how sequence data was produced
- Makes it possible to compare datasets:

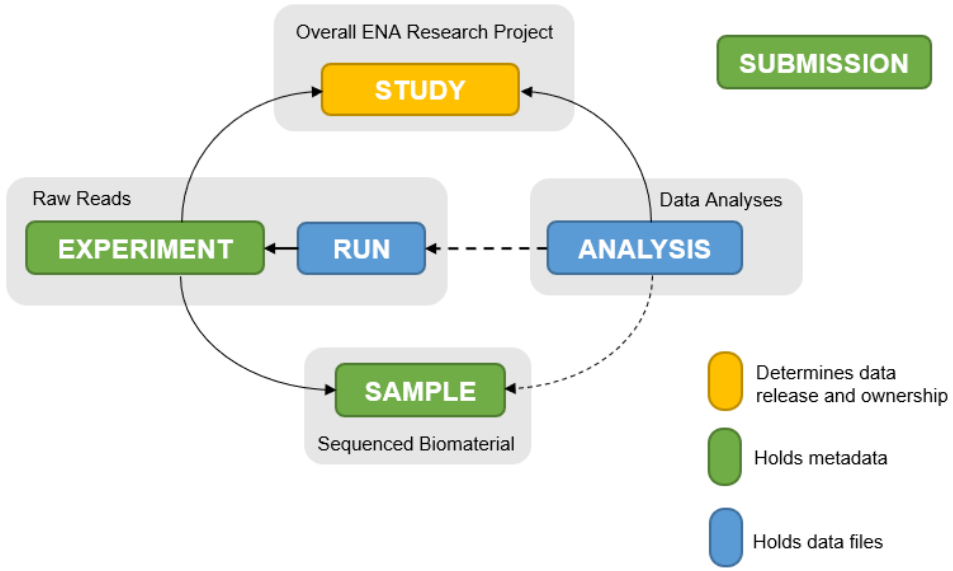
"I want to see data from bacteria …

… in the Atlantic Ocean …

… sampled between 50-100m …

… between April and July …

… compared with the same from the Indian Ocean"



ENA Background: The Metadata Model

CC0 Public Domain

Viruses have a range of effects on bees, but this is little studied outside one bee species (*Apis mellifera*) in N. America and Europe.

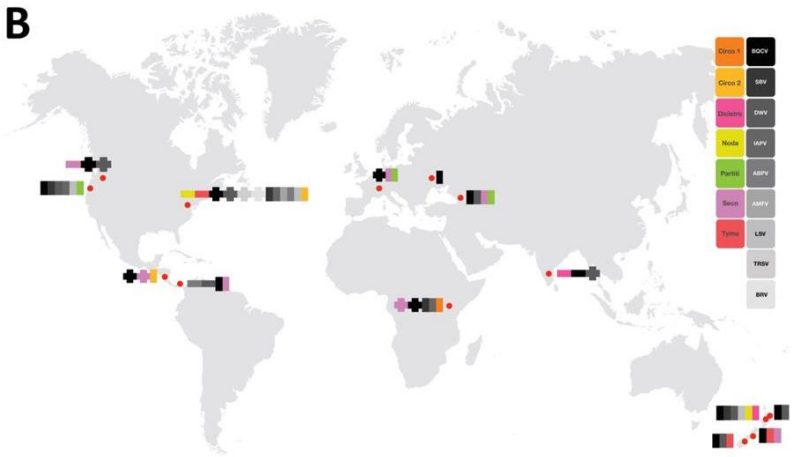
Galbraith *et al*. sequenced viral metagenomes of 11 bee species in 9 countries.

Developed a pipeline to assemble contigs from the data and identify viruses.

Galbraith *et al.* 2018, article available at: https://rdcu.be/bb3G0

ENA Background: The Metadata Model

ENA Background: The Metadata Model



Collaborators across 4 continents sampled foraging bees

Details including the species and GPS coordinates were logged. Similar bees were homogenised to give 37 separate samples.

This information is recorded in the database.

ENA Background: The Metadata Model

STUDY

SAMPLE
Location: India
Host: *Apis florea*

SAMPLE
Location: Nicaragua
Host: *Apis mellifera*

SAMPLE
Location: Switzerland
Host: *Bombus impatiens*

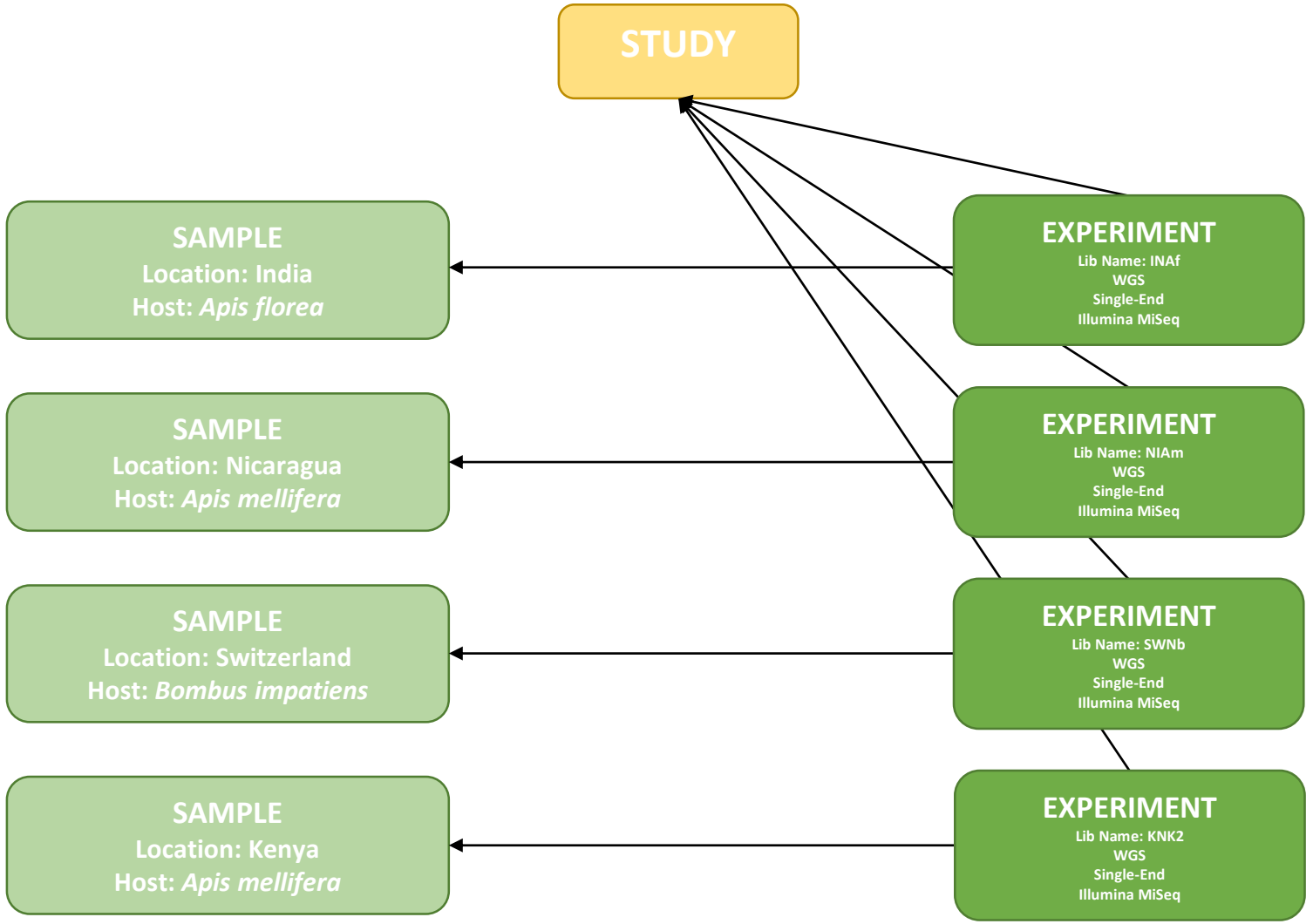
SAMPLE
Location: Kenya
Host: *Apis mellifera*

ENA Background: The Metadata Model

Viruses were isolated and their DNA/RNA extracted.

Following random, unbiased amplification, the material was sequenced on an Illumina MiSeq in 37 separate single-end experiments.

ENA Background: The Metadata Model



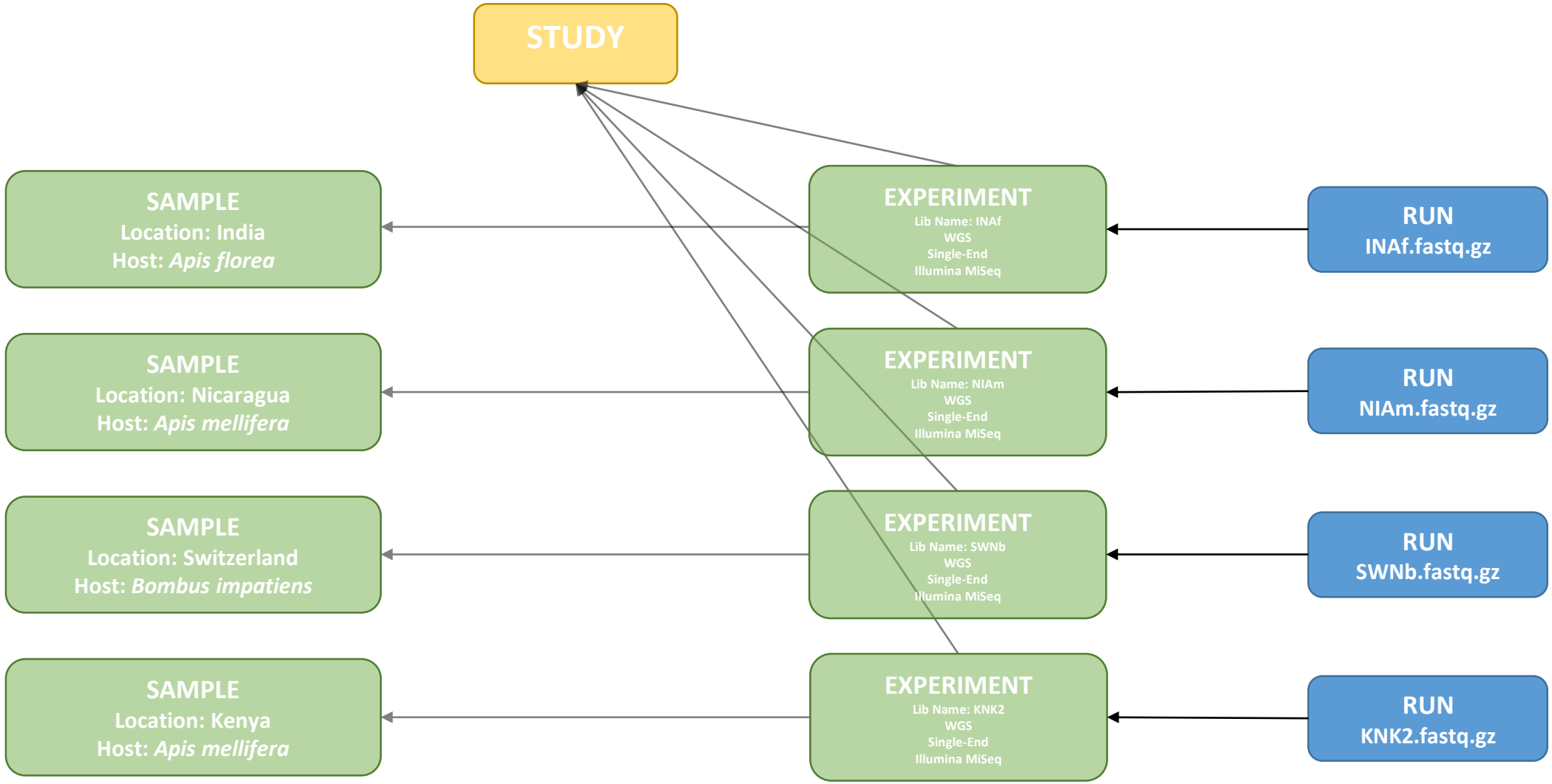
ENA Background: The Metadata Model

```
@SRR6033657.1 1/1
GACCATCTAGCGACCTCCACCTCATCCGGTAGAGCGAATGATTATATCCCTTGTTTTCTAAACTACCTCAACCTATTCTC
TACCTCCAACTAGTTGAGTACCCTGTCTTTCTTTCTTTATGAATCCTTTGTGTTTCGGTTCATATTGCCCC
+
--AB8CEF,,C::CFEFGFGGFGDGGG>+CC,,,<@++8,,<C,,,,,<;CC,,,,6<,,,,;,,:9:C,,
9:C,:CC@C,,,:9C@,,,:,,,9,,,,,:99?,,959?@@E,,,::,:,,,,55,,,,,,99AA+++44,,,,9,44,+
@SRR6033657.2 2/1
TTTGATGATGATTCCTTTCTCTTTCATTCGATGATCCCATCTGATTCTAATCCATTATTCCATTCAATCCCATTTTATGA
AAATTCCATTCGATTCCTTTCAATGTGGTTGTCGCTAGTTGGTCAGGTTTTGGTGGAGTTCGCTAGATGGT
+
ACC@<E,,C<9EGFGGGGGGGGGGGGGGGFG99EFFFFEFG9,CCEC,,<CE,CC,CEEC,CCE,,<C@C,CCE,,<,,
,,,<CEE,CCE,,<AE@FBEE,,<,:,,,,,8C,47,+,:,,8A,,,9,9D,+9,,,,,99,4?,+,9,,9
@SRR6033657.3 3/1
CACCATCTAGCGACCTCCACCTCTCTGCGTTTGGTTCATCCCACATCCCCAGTTCTTCTTACCTATATTGCCCCTCTTTC
CTCTCTCTTCCAACTTCTCTTTTCTTCATCATCCATTCCCTCCACACATCTCACCCCTTTTCTTTTTTATT
+
--8A,CEEA@E:@CFGFGGGGGGFGGCFCFFE,,CCEDFECC8C,,;,;:,,;<6C,;CC,;;,,,,<9,,:,:,8C@,,
:,99CC,,6,<,,,,:<6:,9,,95:@,:,,59,,,,:,,::4,,,,,5,9?,9AE+495,,,,449+,,,
@SRR6033657.4 4/1
GTTTGTCCCTCAAACCTCCCCAAAAACTAGGGAAGCTAGCTAGGCACCACCTCTTGCCATACTTACTACACCCACTTTCA
CTTATCCATTCCATTCCCCTCCACTCACTTTTCTCATCAATCACTTTTTTATGTCGCTTTATGTTCATATC
+
-ACB<FCCCCF8E-CDFGFGGGGFGCGGDFF8,,,CF9,,,,,,,,,:,:,,,,,,,;,<,,<,,,,,<,,:
8,8,:99,,<9,66<,996,,,6<<@@6@@,,5:,,,,,959,,9,,,9,,,,,,5,+,,,99,+4++,4,,99,,,94
@SRR6033657.5 5/1
ACCATTCTGAGGGAACCTTTGGGCGCCTCCGTTACCTTTTAGGAGGCTACCGCCCCAGTCAAACTTCCCGTCAGACACTG
TCTCCGATAGCCATCACCTATCTGGGTTACAGTGGCCATAACACAAGGGTAGTATCCCATCCTCCTCTCCT
+
@CC9CEFGF,-,B@CFGFGGAFGGGGGGGFGGGGEGGGGEC,C,,,6,,;6+:67@,,:9,,,::,<C@@C@,,,:
6CC,<C6?C++8+,,,,::5BA=,?A5,,4:?7,,,:,,49,AB,A,A,+,,+9,+E,?A<A,,8,74,?=4?:,
```

The result of these experiments was a collection of 37 separate FASTQ files.

These were compressed and uploaded to a database, where they underwent processing and waited to be made public.

ENA Background: The Metadata Model

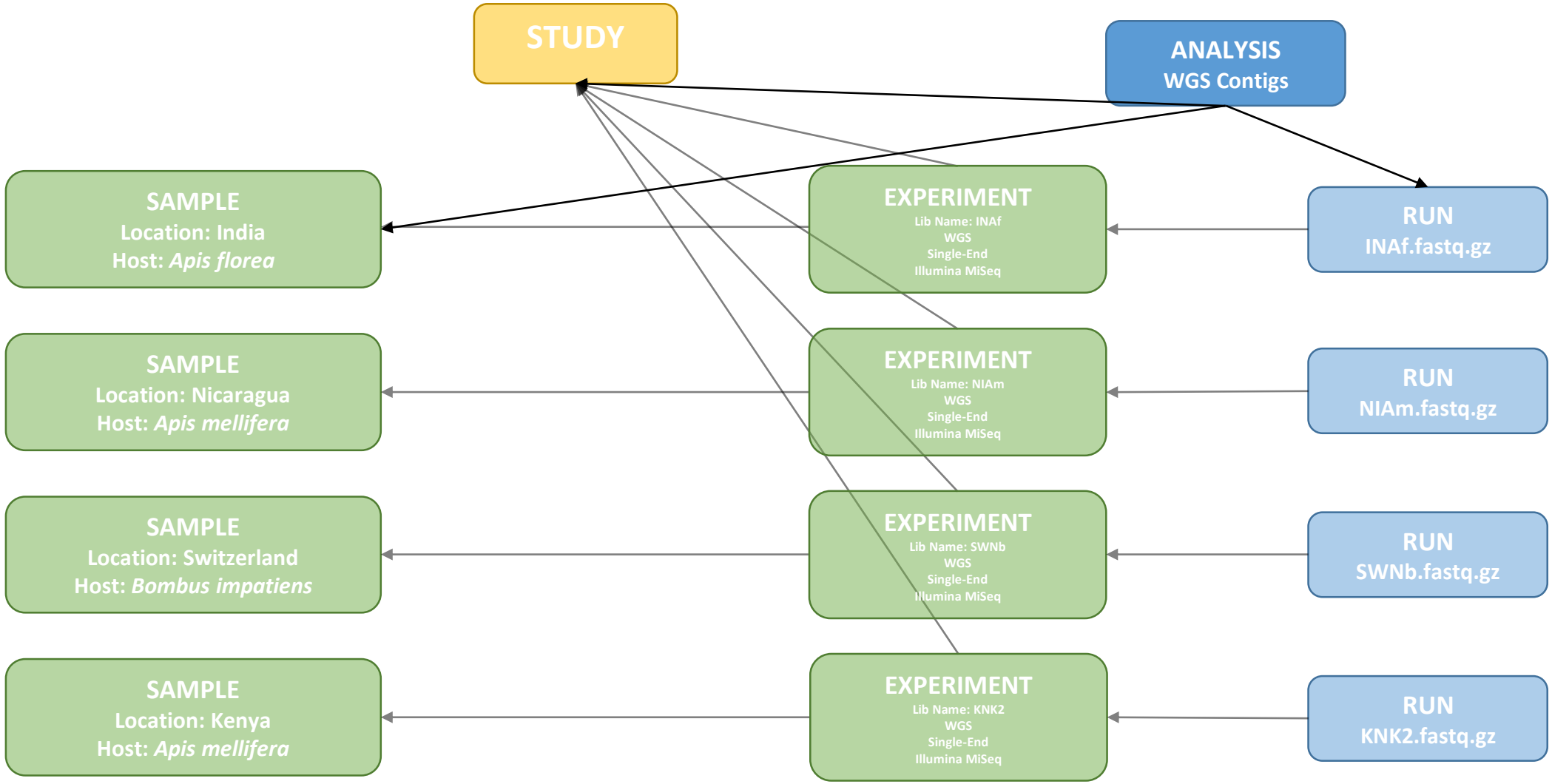


ENA Background: The Metadata Model

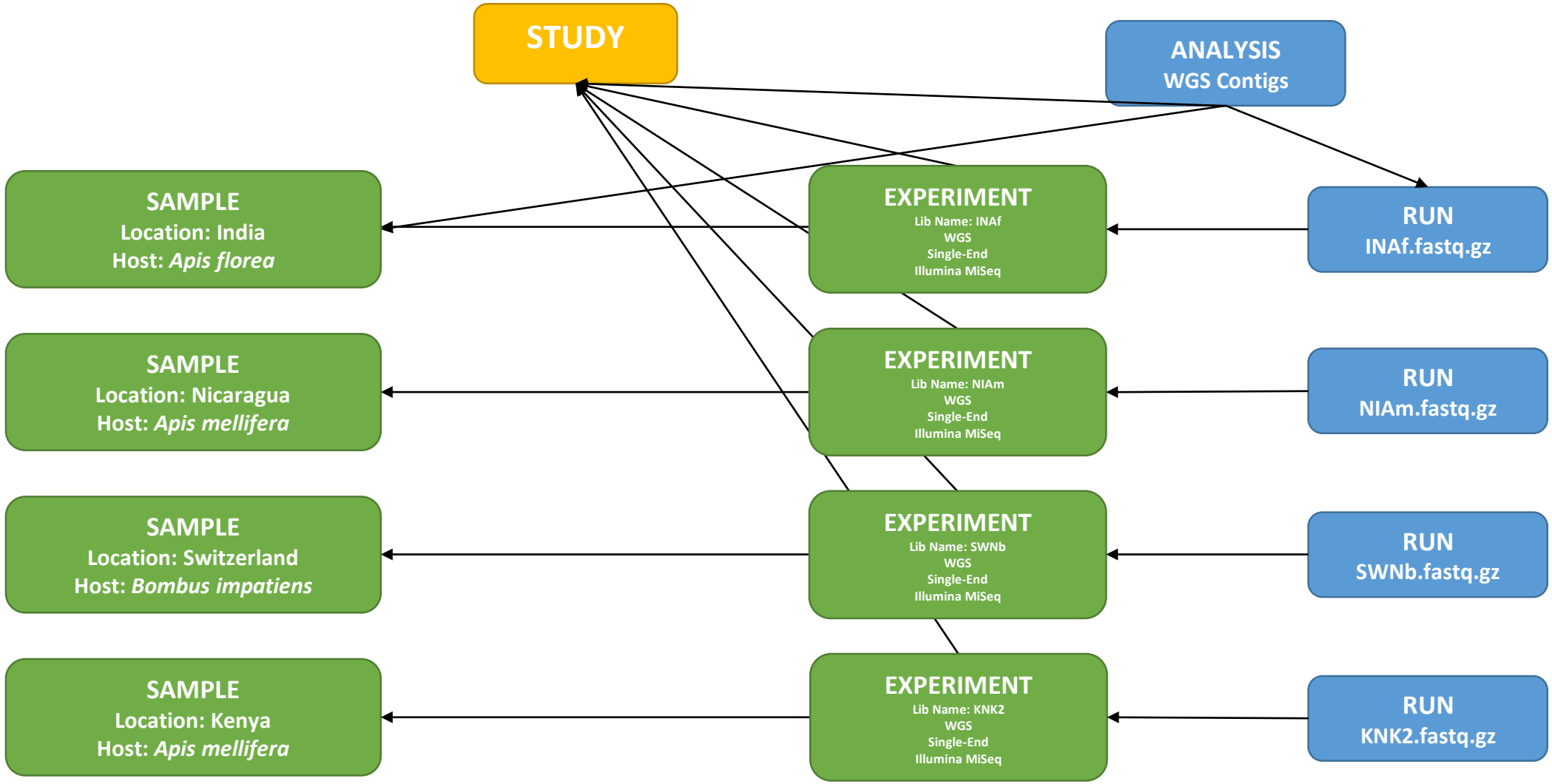
```
>ENA|PEHZ01000001|PEHZ01000001.1 Insect metagenome contig_0_374_CA-Am, whole genome shotgun sequence.
GATACGTGTGCAGGTGCAAGGGTATGTGCATGGAGGGGTGCGTGTATAGGTGCGAGGATG
TGTGGGGTGTGTAGGTGGAAGTGTGTGTGCATGGAGGATACGTGTGCAGGTGCAAGGGTA
TGTGCATGGAGGGGTGCGTGTATAGGTGCGAGGATGTGTGGGGTGTGTAGGTGTGAGGGT
GTGTGTAGGGGGTGTGTAGGTGTGAGTGTGTTATGTGCATGGAGGGGTGTGCAGGTGCAA
GGGTATGTGCATGGAGGGGTGCGTGTATAGGTGCGAGGATGTGTGGGGTGTGTAGGTGTG
AGGGTGTGTGTAGGGGGTGTGTAGGTGTGAGTGTGTTATGTGCATGGAGGGGTGTGCAGG
TGCAAGGGTATGTG
>ENA|PEHZ01000002|PEHZ01000002.1 Insect metagenome contig_2_497_CA-Am, whole genome shotgun sequence.
GACCATCTAGCGACCTCCACATACTAGGGTTAAAATACCCTAAAGTAGAAGCAAAAGTTA
ATATATTAACGCATAACTATGAGATACTATTTCTCTGTTATGAAAATGATTAAGGATTAG
TATGAGAATCTACGGGTTTTCTATTTCATCTGAGTTTATGCTGGAGTTCTATTATATCTA
TTTAAACGAGAATAAAGAAACCCTAGATACTTTTATAACACAAAGAGATATTAACAATTT
CTATTCACGTTTATTTAGTAGGTATGAATTAGAATACTATCTAAGATTCATTAATTGCAT
AATGAAAGAAAATGAATGGTTAGAAAAAGAAATAAAAAACTTTCATTTACTTCTAGAATC
TGGTAGTAACCTAATAGAAGATGATAATTTGATTATCATGCCAAAGATTATAGCGGAGTT
AAAATAGTGAGTAAAATTGACCAGTTTCAGTGGAGGTCGCTAGATGGTCCGGGTTAGGTG
GAGGTCGCTAGATGGTC
>ENA|PEHZ01000003|PEHZ01000003.1 Insect metagenome contig_4_469_CA-Am, whole genome shotgun sequence.
GACCATCTAGCGACCTCCACATCATTCGCTAAATGCAGTTCTGGATGCTTGTTTGGACTT
TTCCATGAAGTGTATGTTTCTCCATCATCGACAGTCAATGTTGTACGCGAATCAGGAATC
GTAGTCGATTGACTCACCGAGTCATCATCATCGGGCTCGACTGTCGATGAACTAATCGTG
TCTGATGCATGAAAGGTACTGGGCCATTGTGTCGACCGAATTTGAAAAGTACGCCGTACC
TCAGGAATCATCGGTGTGGGAACTTTGTAAAACATCAGTGTGTCATTTATCAAGACAGCT
TTGAAAGGGCGCCATCCGCGCCCAGGGTCTTTAAGCTCAGTATTTGCCACAGTATCAGAA
AACCAGGCAATACTAAACAGACCCGTGCTGCGTTTGATAAATCCTTCCAACAAGGCCCCT
TCTTCGTCATTTGATCGAAATGGGGCCTGGTGGAGGTCGCTAGATGGTC
>ENA|PEHZ01000004|PEHZ01000004.1 Insect metagenome contig_8_628_CA-Am, whole genome shotgun sequence.
GACCATCTAGCGACCTCCACTCAAACCCTCCAAGGTGATTCGTAGCTCTAAAATAACCAC
TATCGCTATCGTAATGAATTTCATCTTGAATCTCACTACCTGGTTTTTCAGCGGCTTTAA
```

Read data for one of the samples was put through an assembly program (SPAdes) to produce a set of contigs which could be searched for evidence of viral origins.

ENA Background: The Metadata Model



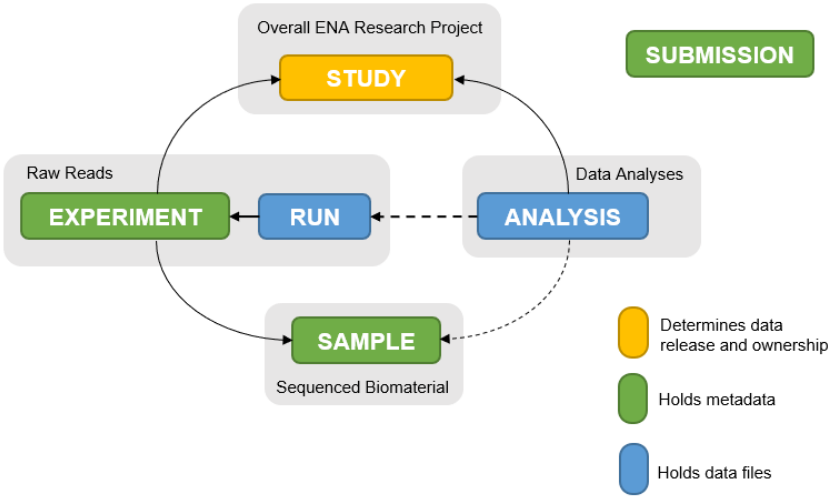
ENA Background: The Metadata Model



ENA Background

- What object might the following attributes belong to?

Library Source:
Metatranscriptome



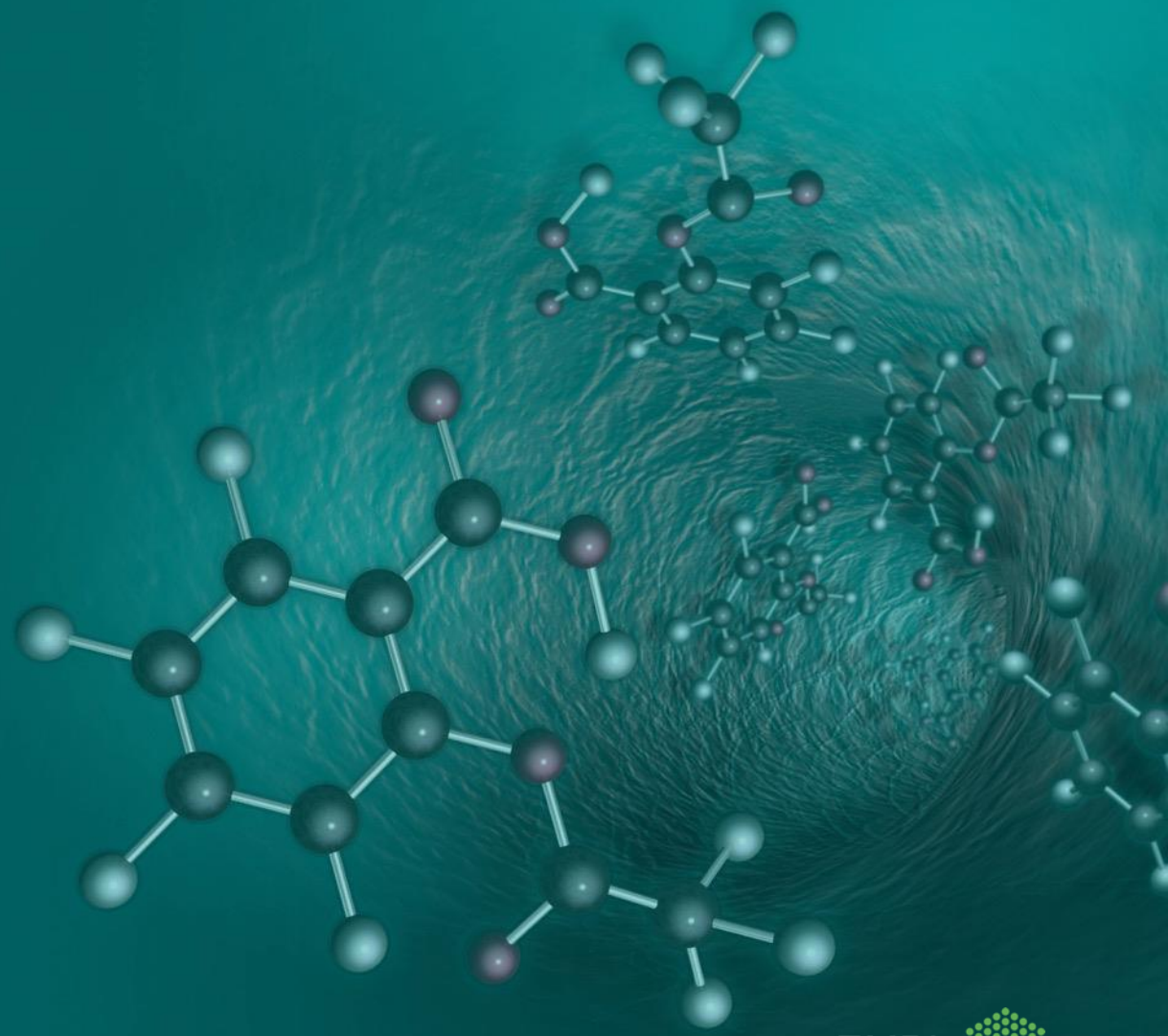
Assembly Program:
SPAdes

Species:
Bos taurus

Collection Date:
13/02/2017

Submitting Data

How You Can Join In



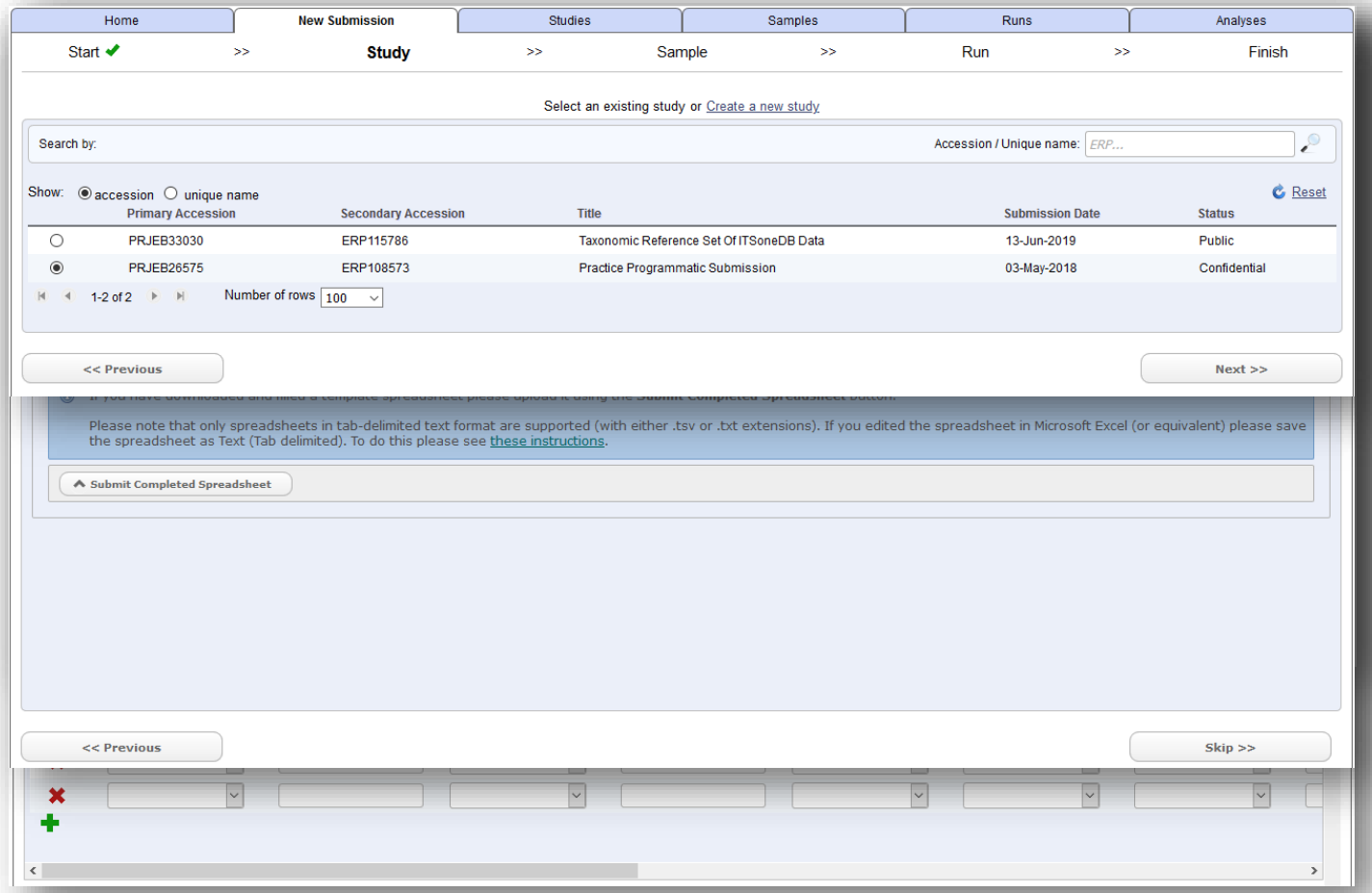
EMBL

Submitting Data: How It's Done

- There are three submissions routes
- '*Interactive Submission*':
  - Use your browser to fill out web forms describing your work
- '*Programmatic Submission*':
  - Describe your work in XML documents, submit them to us using cURL
- '*Webin-CLI*':
  - Smart new submission interface, made in-house

Submitting Data: The Interactive Route

- Register your objects using your browser
- Familiar and largely accessible
- Prepare spreadsheets for bigger submissions



Submitting Data: The Programmatic Route

- Prepare an XML file describing your submission
- Send this to us via HTTPS
- Example cURL command:

```
curl -u username:password \
        -F "SAMPLE=@sample.xml" \
        -F "SUBMISSION=@submission.xml" \
        "https://wwwdev.ebi.ac.uk/ena/submit/drop-box/submit/"
```

```
<SUBMISSION>
    <ACTIONS>
        <ACTION>
            <ADD/>
        </ACTION>
    </ACTIONS>
</SUBMISSION>
```

```
<SAMPLE_SET>
<SAMPLE alias="SWAm">
    <TITLE>SWAm</TITLE>
    <SAMPLE_NAME>
        <TAXON_ID>1234904</TAXON_ID>
        <SCIENTIFIC_NAME>insect metagenome</SCIENTIFIC_NAME>
    </SAMPLE_NAME>
    <SAMPLE_ATTRIBUTES>
        <SAMPLE_ATTRIBUTE>
            <TAG>collection_date</TAG>
            <VALUE>01-Aug-2015</VALUE>
        </SAMPLE_ATTRIBUTE>
        <SAMPLE_ATTRIBUTE>
            <TAG>host</TAG>
            <VALUE>Apis mellifera</VALUE>
        </SAMPLE_ATTRIBUTE>
        <SAMPLE_ATTRIBUTE>
            <TAG>isolation_source</TAG>
            <VALUE>Whole body homogenate</VALUE>
        </SAMPLE_ATTRIBUTE>
        <SAMPLE_ATTRIBUTE>
            <TAG>lat_lon</TAG>
            <VALUE>46.5197 N 6.6323 E</VALUE>
        </SAMPLE_ATTRIBUTE>
        <SAMPLE_ATTRIBUTE>
            <TAG>geo_loc_name</TAG>
            <VALUE>Switzerland</VALUE>
        </SAMPLE_ATTRIBUTE>
        <SAMPLE_ATTRIBUTE>
            <TAG>samp_size</TAG>
            <VALUE>10</VALUE>
        </SAMPLE_ATTRIBUTE>
        <SAMPLE_ATTRIBUTE>
            <TAG>BioSampleModel</TAG>
            <VALUE>Metagenome or environmental</VALUE>
        </SAMPLE_ATTRIBUTE>
    </SAMPLE_ATTRIBUTES>
</SAMPLE>
</SAMPLE_SET>
```

Submitting Data: Webin-CLI

- Submit your data in a single step
- Pre-submission validation
  - Confidence that your submission has worked
- Describe your submission in a manifest file:

```
NAME reads_01
PLATFORM Illumina
INSTRUMENT NextSeq 500
INSERT_SIZE 200
LIBRARY_NAME library_01
LIBRARY_STRATEGY RNA-Seq
LIBRARY_SOURCE Transcriptomic
LIBRARY_SELECTION  RT-PCR
SAMPLE   ERS3194300
STUDY    ERP108573
FASTQ   reads_01.fastq.gz
```

```
@SRR2960126.1 1/1
GCCAGCTATATCAGTTTCCTTTGTGATGGGGCGTTTATGTAAGGGATGGGTATGGAACTGCACCCTGGCTAGAACGAAAAGACTTGCTAT
TGTGAAGTTA
+
=;?AD=D?FDB???ACGHGGEHCEEHC@GDG>GDGHFF@9?C<38BFEACHIID>EEC>A3@D;?=;ABCA?
@@A8=8=B>89>ACCCACCADC3@4@@A
@SRR2960126.2 2/1
GTGGGTAACATTTGGAACAGAAAAACAAAAAGTAGAATGGTGGGTGGAGGAGAACAATAAAAATGATACTTCATTTAATTATCATTTAGA
GATGTATATC
+
CCCFFDFFHHHHHIJIJIJJJJJJJJJJIJII?
DGIJJJJBFHIFGGGIJIJIJIJJJHHHHHFFFFFFFDEEEEDFEDEEEDEEEEEDDDDDEEFFDD
@SRR2960126.3 3/1
CTTTTAACAAAGATTTTTTTAAAAAATGCTGTGCTGTAATTCACTAGCAAAAACCTAATTTGCTAGTTGCATTGGAAAGATATTGATACT
TGACAATTTA
+
CCCFFFFFHDDFFGHHHJIIIJJIIJIJJJIJIGHIIJJJJIJIJIIJJJIJJJJJJIGJIHGHHHHFFFFFFFEDCCADCDDEEEDEDDD
DDDDDDDDDD
@SRR2960126.4 4/1
GCTTTTTTTCGGCAACGATGGAATCGATCAATTGTTATTGTGACAGCTAAATTGATAAAACAAAATGATACAGTTTCAAACTGATCAAAA
ATTTGCAGCA
+
@=?DDDBDFBHD<FBHAHBD@FGIEDHGH??BB9BFGHIDF@===FCGEGGGH3AEHC;@@DDF@C>;A@AC:
5-5;>@C@CCC::@ACCBA@C:4>:(2
```

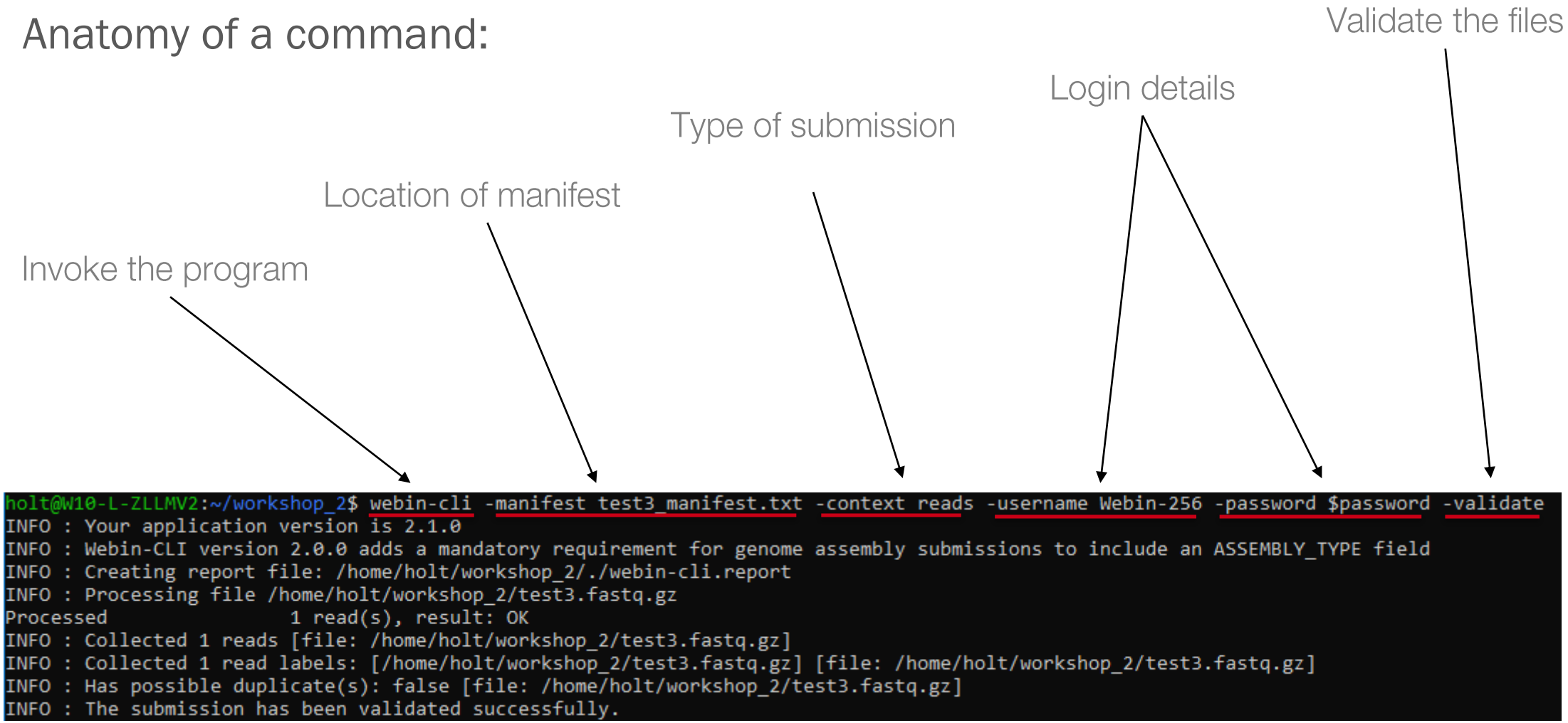
- Only way to submit assembly data

Submitting Data: Webin-CLI

- Required options:
  - -context - the type of data to be submitted
  - -manifest - location of file describing the submission
  - -username - your Webin username
  - -password - your Webin password
- Other options
  - -submit - instruction to submit the data
  - -validate - instruction to just validate without submitting
  - -test - use the test server, submission is forgotten after 24 hours

Submitting Data: Webin-CLI

- Anatomy of a command:



Submitting Data: Practical Exercise

- Use the Interactive and Programmatic interfaces as well as Webin-CLI to submit a dataset to the ENA test service

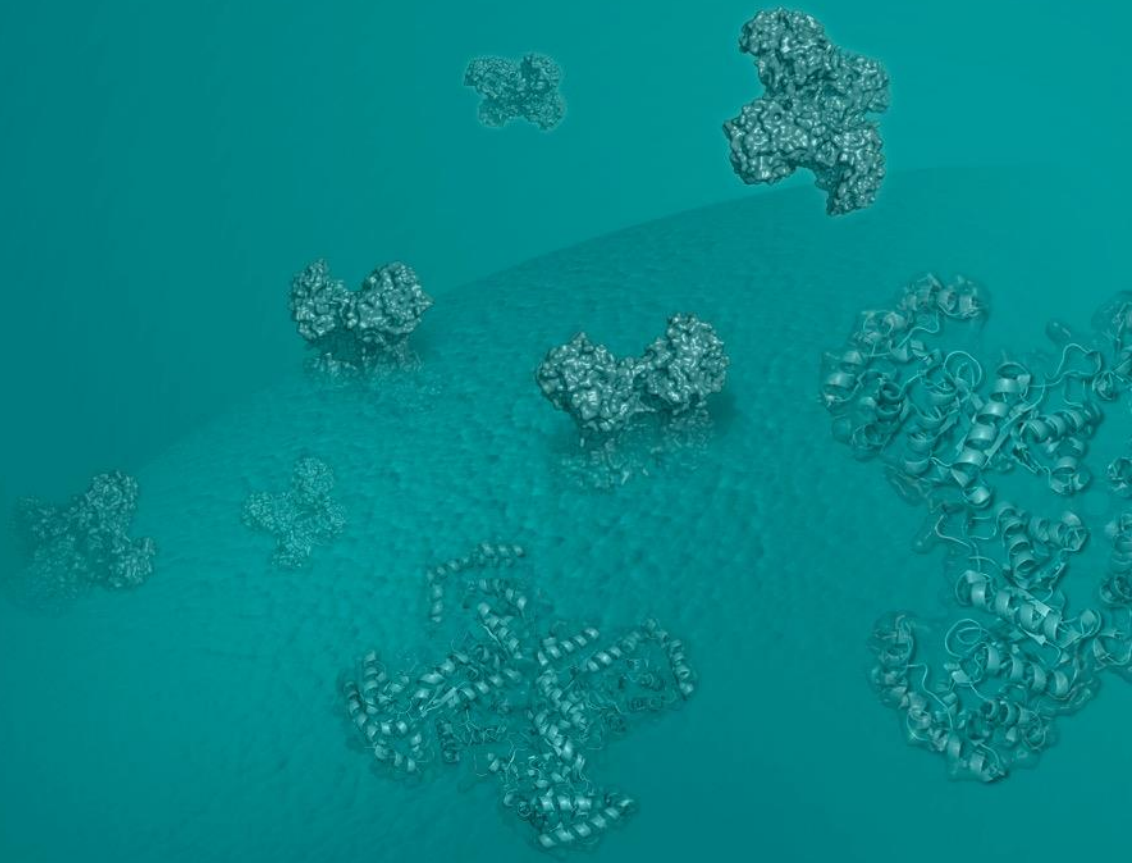
```
$ ssh student<??>@gdcsrv2.ethz.ch   # replace ?? with your student number
```

- Let me know if you have any questions or need any help with the practical exercise

- Future comments, questions and concerns to our helpdesk: https://www.ebi.ac.uk/ena/browser/support

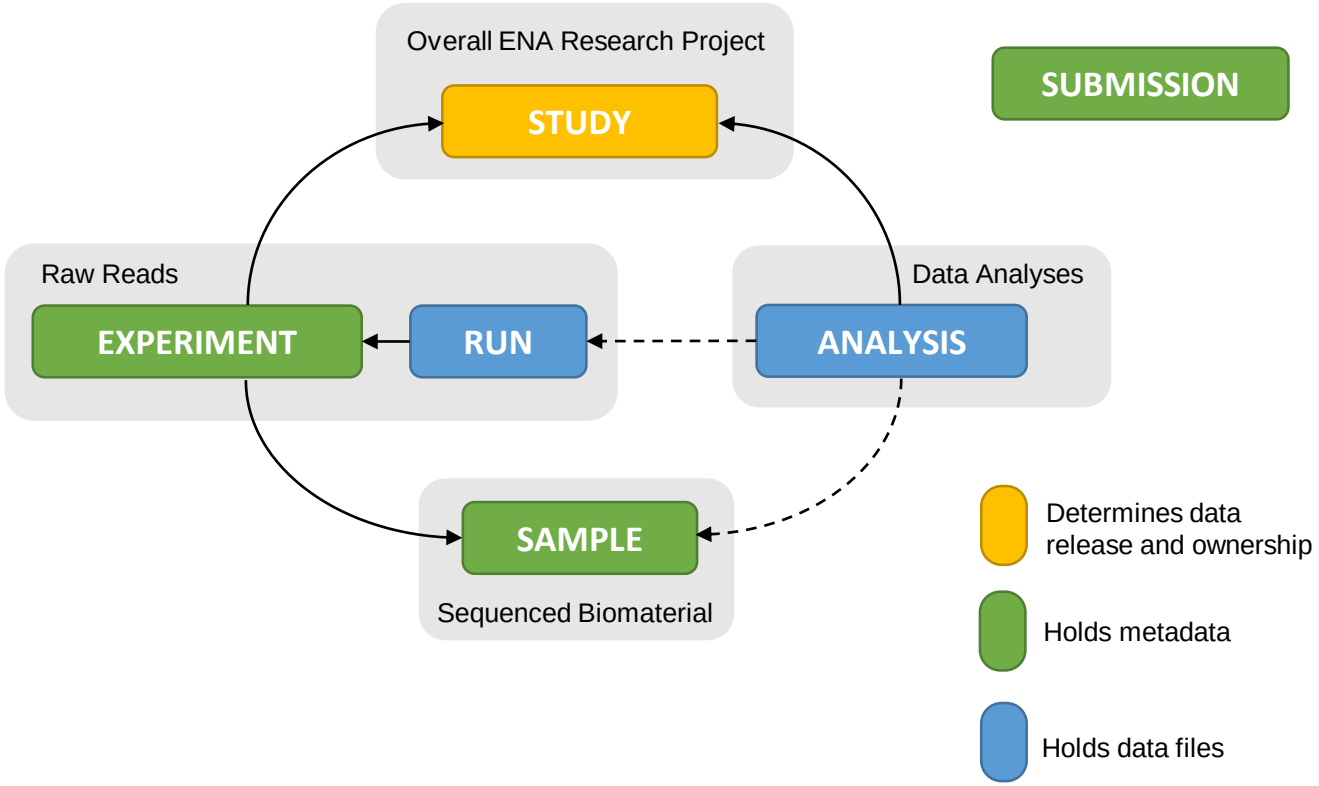
ENA Metagenomic Standards

Understanding And Exploiting Them



EMBL

ENA Metagenomic Standards – Metadata Recap



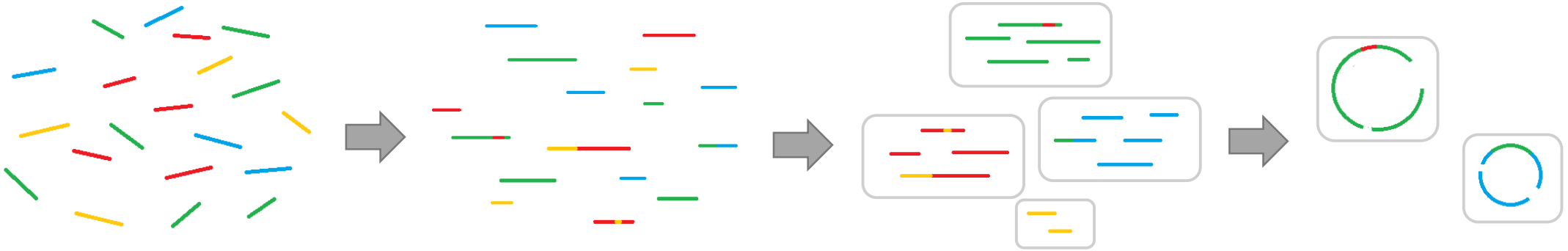
ENA Metagenomic Standards – Why Are They Different?

Raw reads are derived from an 'environmental sample'

All the contigs from the biome are assembled together

Contigs are binned into taxonomic groups

Contigs are further assembled to generate high quality genomes



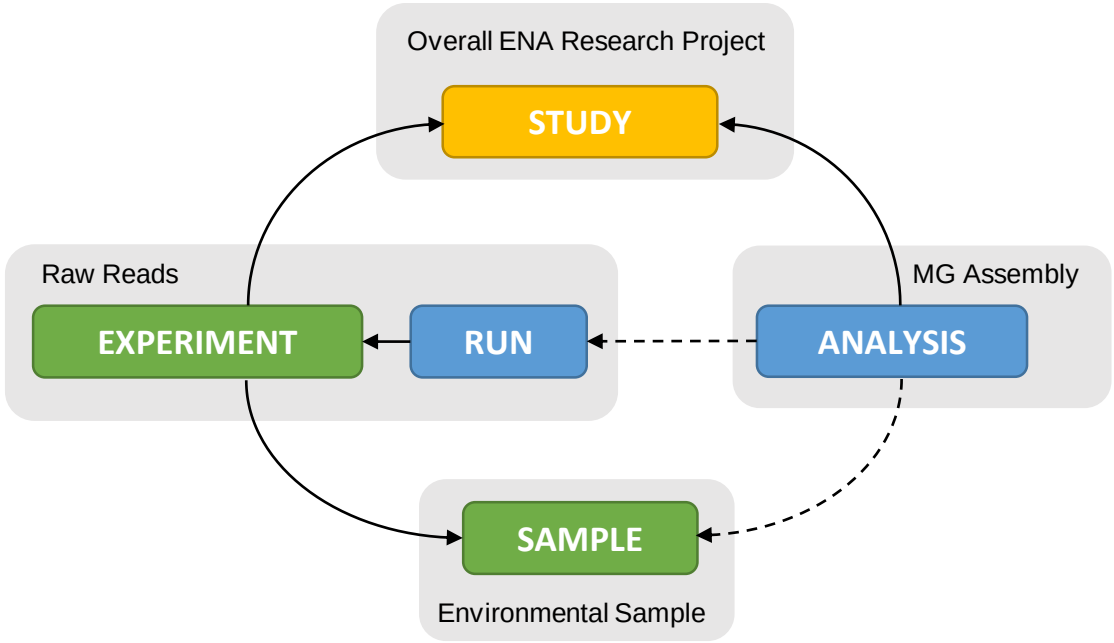
Raw Reads

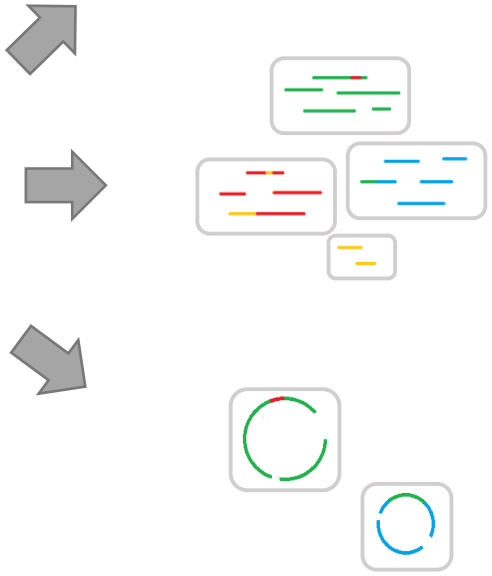
Primary Assembly

Binned Assembly

Metagenome-Assembled Genome (MAG)

ENA Metagenomic Standards – Submitting MG Assemblies



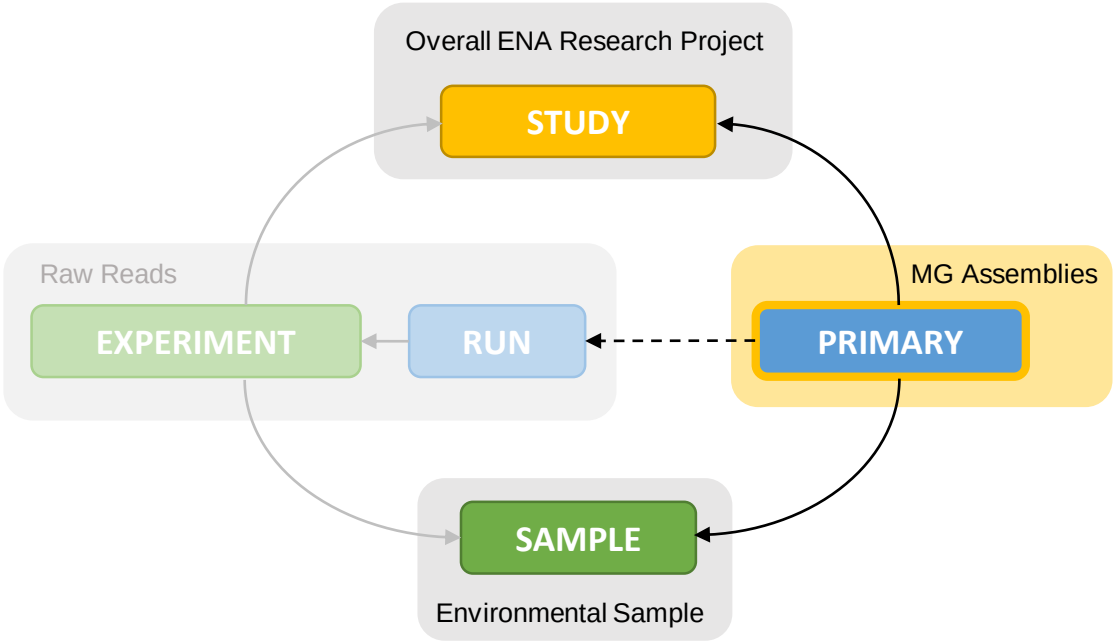


Primary assembly?

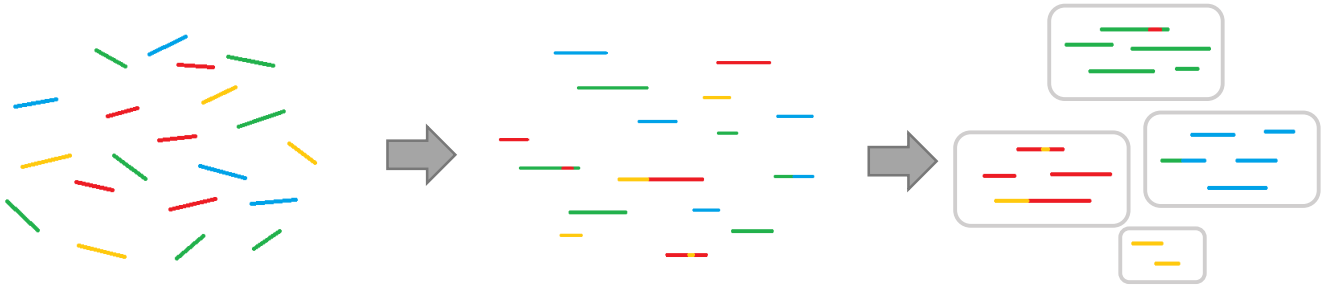
Binned assembly?

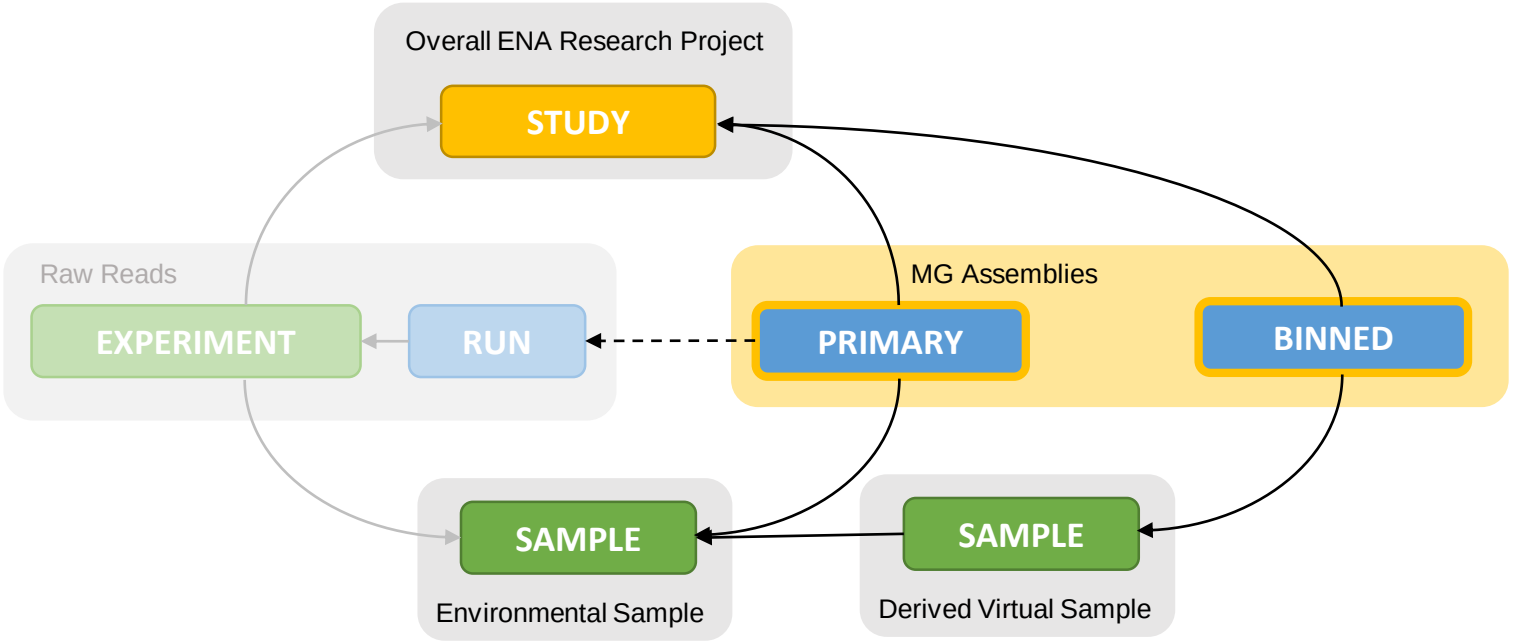
Metagenome-Assembled Genome (MAG)?

ENA Metagenomic Standards – Submitting MG Assemblies

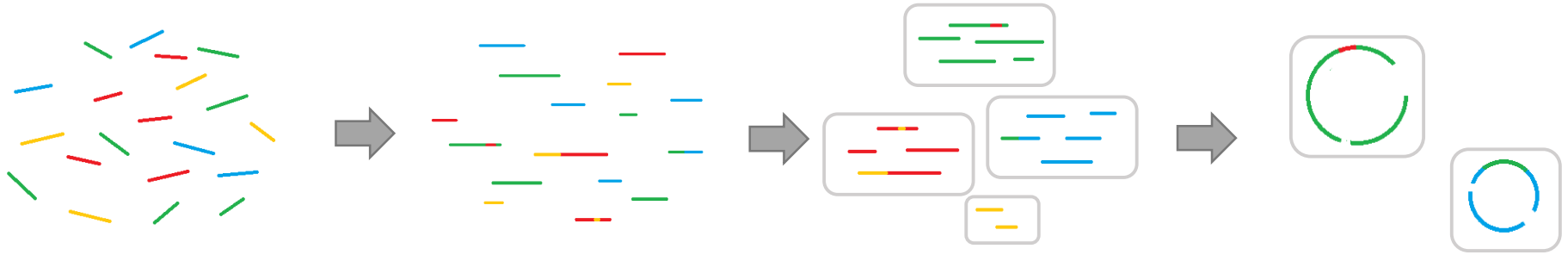


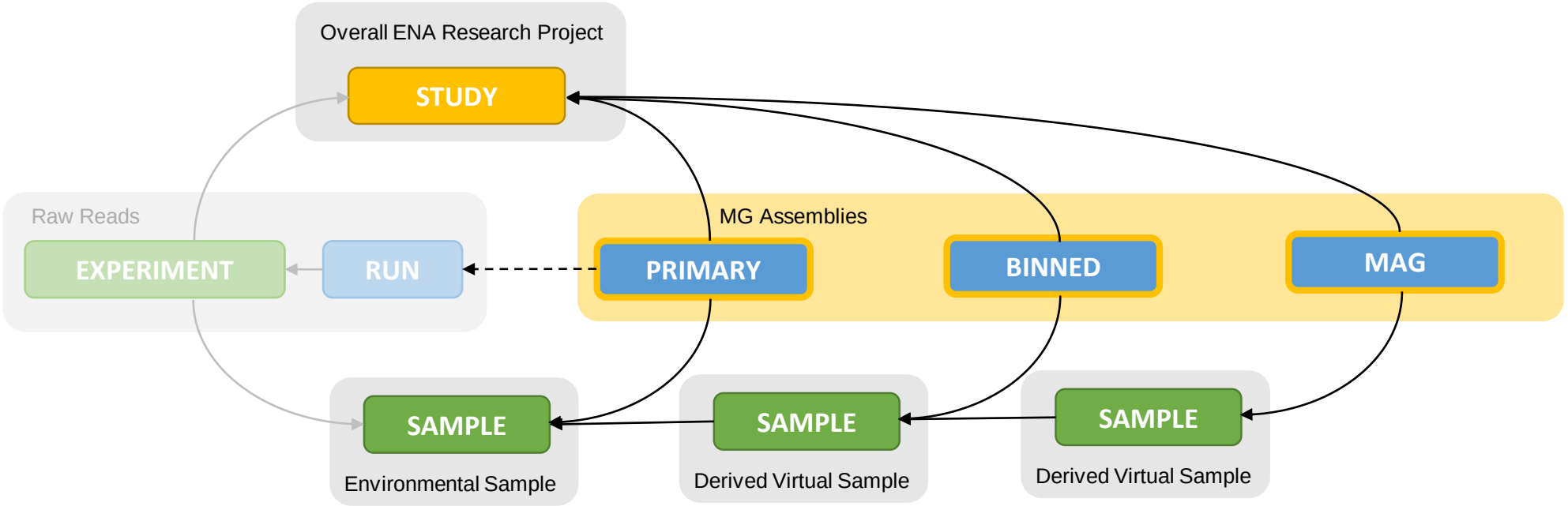
ENA Metagenomic Standards – Submitting MG Assemblies





ENA Metagenomic Standards – Submitting MG Assemblies



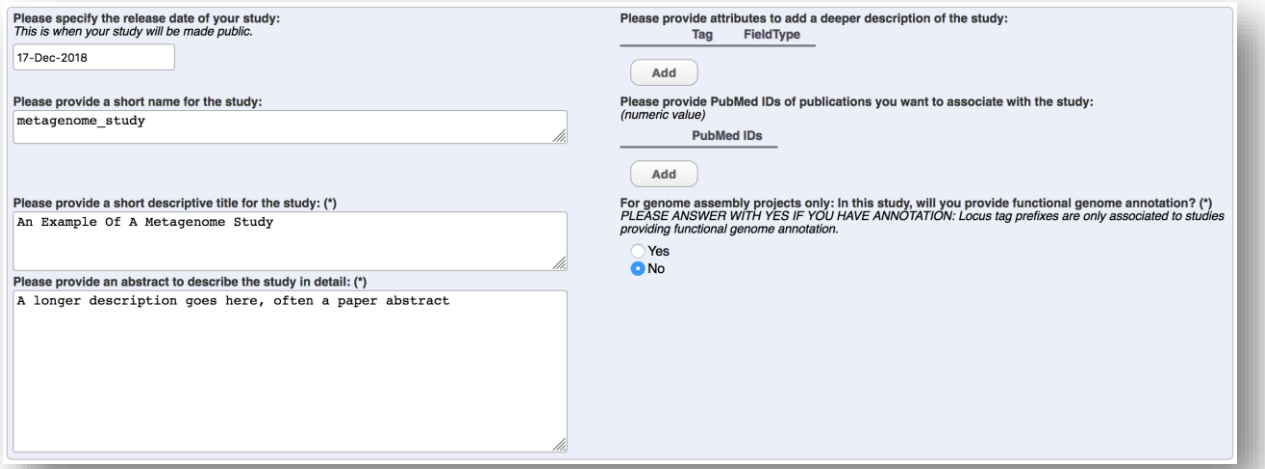


ENA Metagenomic Standards – Data Submission Recap

- There are three submissions routes
- '*Interactive Submission*':
  - Use your browser to fill out web forms describing your work
- '*Programmatic Submission*':
  - Describe your work in XML documents, submit them to use using cURL
- '*Webin-CLI*':
  - Smart new submission interface, made in-house

ENA Metagenomic Standards – Study Registration

Use Webin Interactive to fill out a form in your browser:

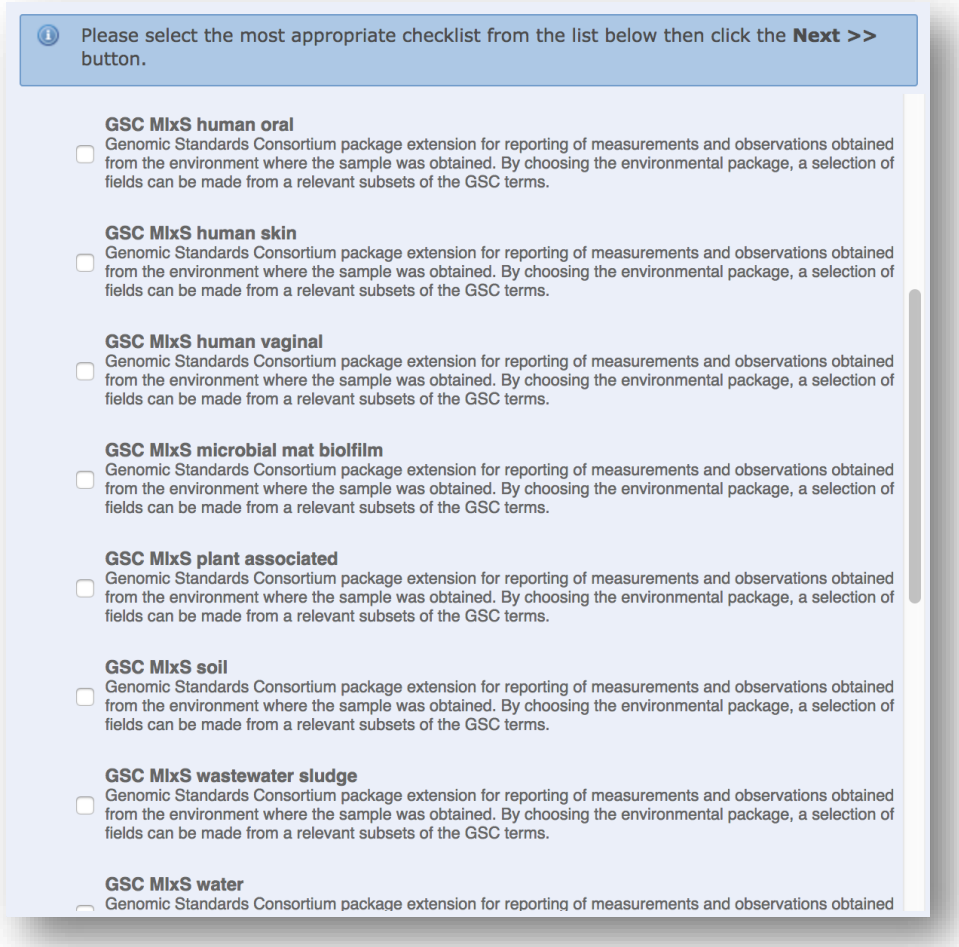


Or submit an XML via REST:

```
<SUBMISSION>
    <ACTIONS>
        <ACTION>
            <ADD/>
        </ACTION>
    </ACTIONS>
</SUBMISSION>
```

```
<PROJECT_SET>
    <PROJECT alias="metagenome_study">
        <NAME>metagenome_study</NAME>
        <TITLE>An Example Of A Metagenomic Study</TITLE>
        <DESCRIPTION>A longer description goes here, often a paper abstract</DESCRIPTION>
        <SUBMISSION_PROJECT>
            <SEQUENCING_PROJECT/>
        </SUBMISSION_PROJECT>
        <PROJECT_LINKS>
            <PROJECT_LINK>
                <XREF_LINK>
                    <DB>PUBMED</DB>
                    <ID>28043580</ID>
                </XREF_LINK>
            </PROJECT_LINK>
        </PROJECT_LINKS>
    </PROJECT>
</PROJECT_SET>
```

ENA Metagenomic Standards – Environmental Samples



Sample checklists ensure data standards are met and that metadata is consistent between submitters

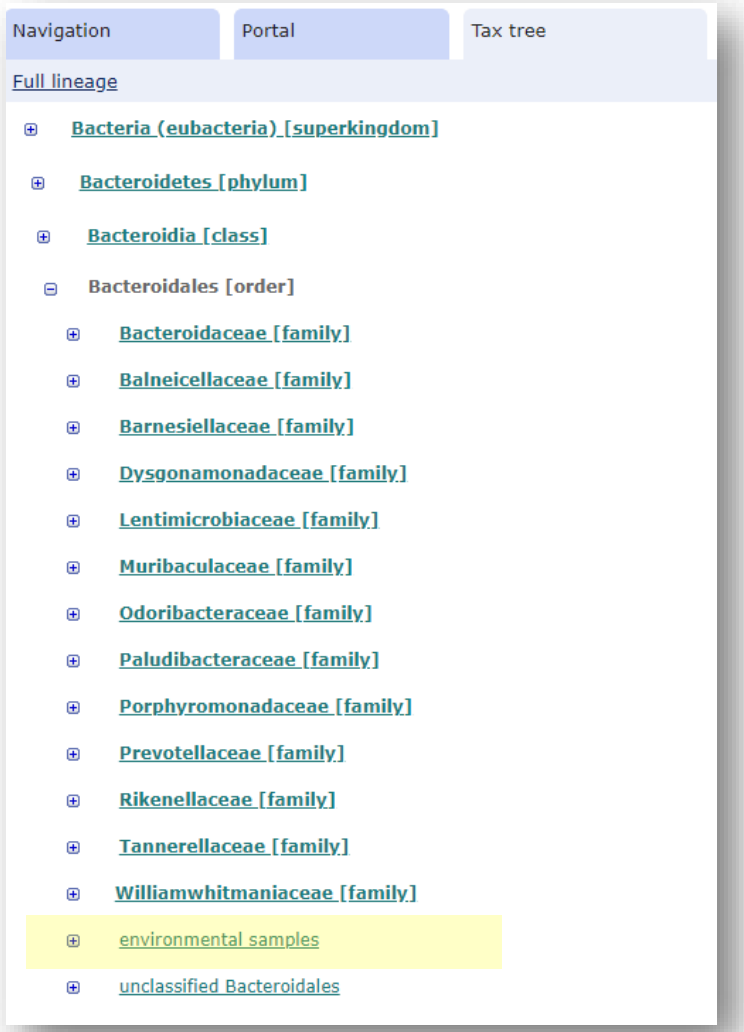
Make sure you:

- Pick the most appropriate checklist
- Use environmental taxonomy

✓ **human gut metagenome**
Taxonomy ID: 408170
Scientific name: **human gut metagenome**
Inherited blast name: **metagenomes**
Rank: **species**

✘ **Homo sapiens**
Taxonomy ID: 9606
Scientific name: ***Homo sapiens*** Linnaeus, 1758
Inherited blast name: **primates**
Rank: **species**

ENA Metagenomic Standards – Binned And MAG Samples



Binned and MAG samples are virtual samples that contain information on binning and assembly methods. They also define the taxonomy of the assembly.

Make sure you:

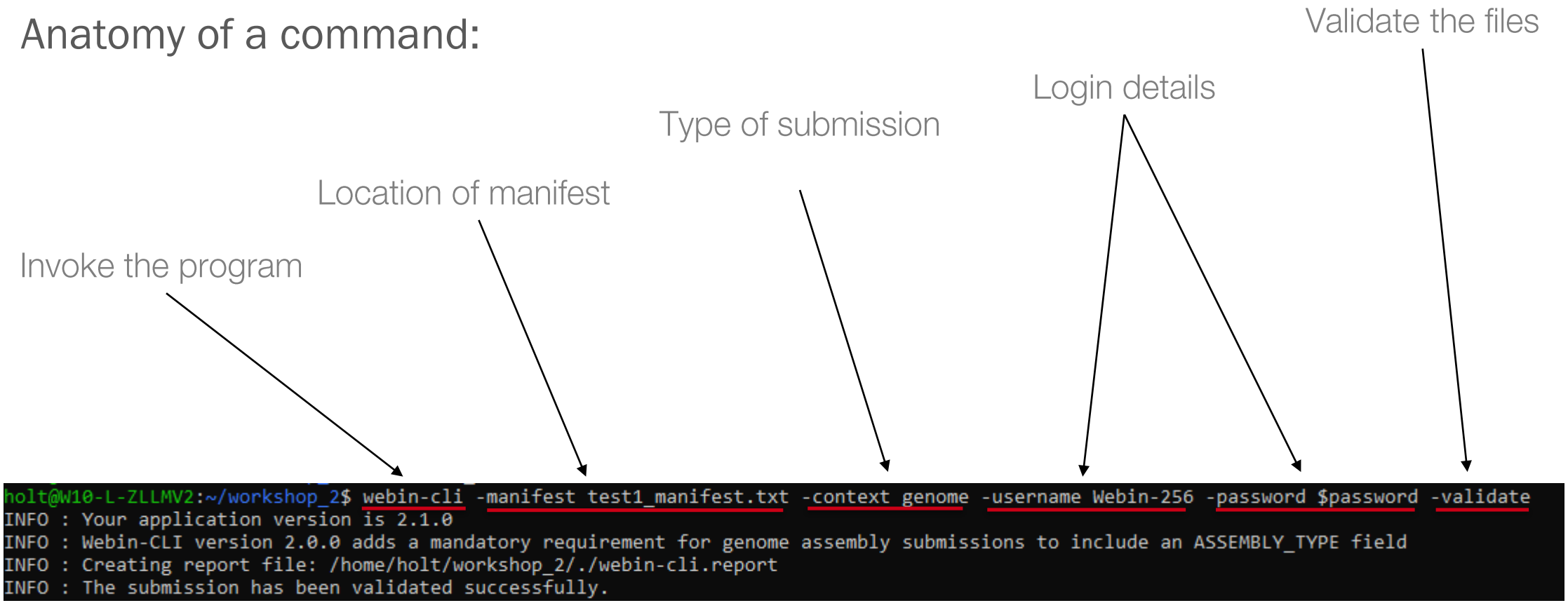
- Pick the **GSC MIMAG** checklist for MAG samples
- Pick the **ENA binned metagenome** checklist for binned samples
- Use uncultured taxonomy

✓ **uncultured Bacteroidales bacterium**
Taxonomy ID: 194843
Scientific name: **uncultured Bacteriodales bacterium**
Inherited blast name: **CFB group bacteria**
Rank: **species**

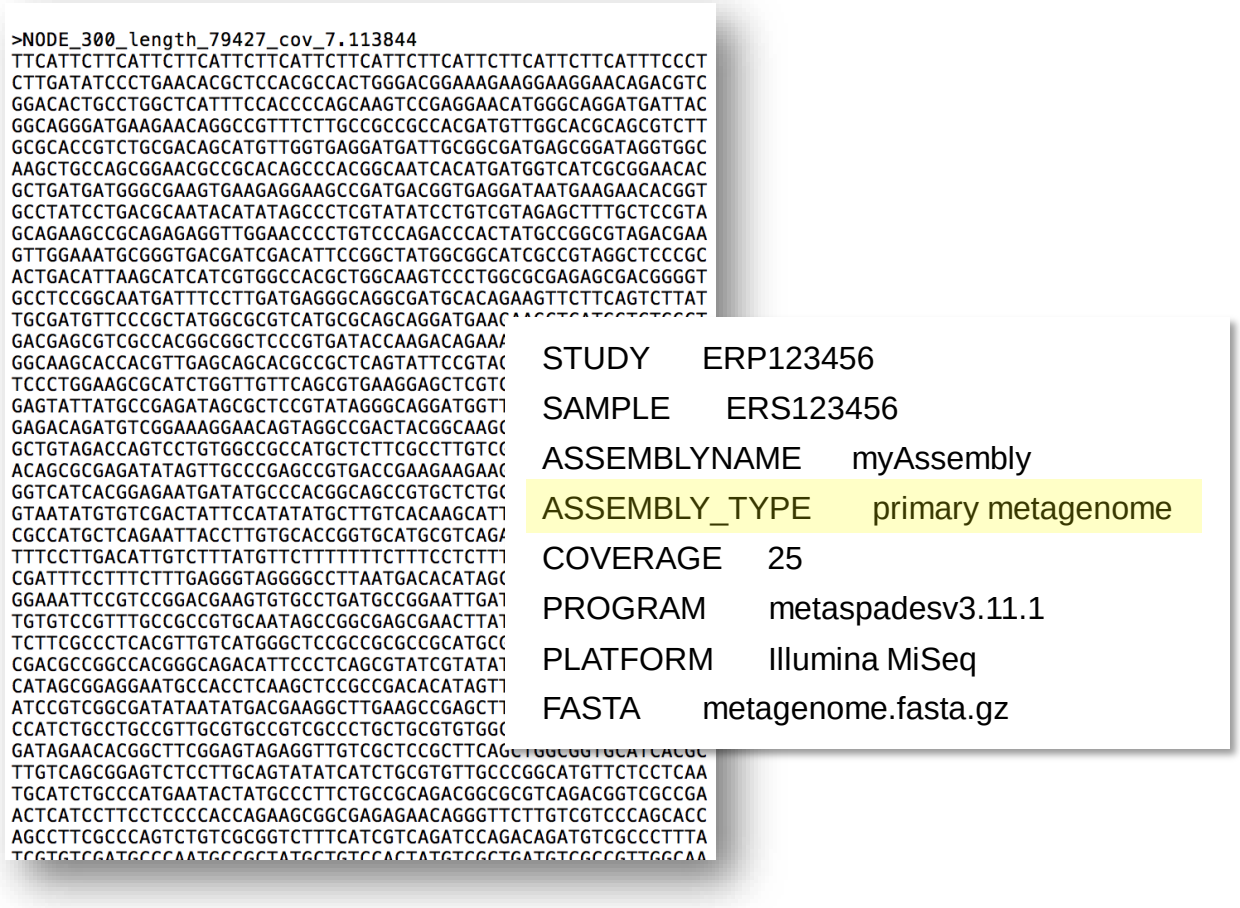
EMBL

Submitting Data: Webin-CLI

- Anatomy of a command:



ENA Metagenomic Standards – Assembly Types



- Prepare your sequence files
- Prepare a manifest file
  - Information on methods
  - Sample/study reference
  - File names
- Send these to ENA with one command

webin-cli -context genome -manifest lib_01_manifest.txt -submit -userName "Webin-1234" -password XXXX

Submitting Data: Practical Exercise

- Use the Programmatic interface as well as Webin-CLI to submit a binned metagenome dataset to the ENA test service

```
$ ssh student<??>@gdcsrv2.ethz.ch   # replace ?? with your student number
```

- Let me know if you have any questions or need any help with the practical exercise

- Future comments, questions and concerns to our helpdesk: https://www.ebi.ac.uk/ena/browser/support