



How to submit IMAGE genotype data

Authors: Jun Fan and Peter Harrison (EMBL-EBI)

Version: 2020-01-28

Introduction:

Genetic variation is the difference in DNA sequences between individuals within a population. The genetic variations could be represented at different levels, from silent mutations to observation of phenotype variations and are treated as a tool to study evolution.

The European Variation Archive (EVA; https://www.ebi.ac.uk/eva/) is a public archive which accommodates genetic variations from all species and has been selected as the archive for IMAGE genotype data submissions.

Submission method

For Genotype data to be part of the IMAGE project and be included in the IMAGE data portal (https://www.image2020genebank.eu/) the samples from which the data was generated must already:

- Meet the IMAGE metadata standards
 https://github.com/cnr-ibba/IMAGE-metadata/blob/master/docs/image_sample_metadata.md
 which can be validated within the IMAGE InjectTool
- Have already been submitted via the IMAGE InjectTool (https://inject.image2020genebank.eu/) to the EMBL-EBI BioSamples archive (https://www.ebi.ac.uk/biosamples/).

Version: 2020-01-28







Figure 1: The IMAGE InjectTool after a successful BioSamples submission. The samples listed here are ready to be used in EVA submission as they a) meet the IMAGE sample metadata standard and b) have been assigned BioSamples accessions. The Export function of this page is useful for assigning the correct identifiers to your genotype submission to EVA.

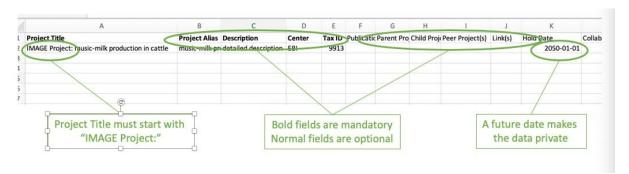


Figure 2. The Project sheet of the EVA submission template file.

Version: 2020-01-28 2





Navigate to the European Variation Archive and follow their submission instructions https://www.ebi.ac.uk/eva/?Submit-Data being aware of the below key points.

Key points for IMAGE genotype submissions:

- To assist in the preparation of your genotype data you can use the IMAGE InjectTool to export a table containing your original sample IDs and their corresponding EMBL-EBI BioSamples accessions which start with SAMEA followed by a unique number. (Figure 1)
- 2. Only VCF files will be accepted.
- 3. In the EVA submission template
 - a. on the Project sheet you **must always** start your "Project Title" with "IMAGE project: " otherwise your data will not be automatically loaded into the IMAGE data portal.
 - b. on the Project sheet if a future date is set for "Hold date", this submission will be kept private until that date. (Figure 2)
 - c. on the Sample sheet: "Sample Accession" is the place to include your BioSamples accessions (starting SAMEA...) and "Sample ID" is where you include your original sample ids. The values in the relationship table exported from InjectTool can be copied and pasted here, see points 1. (Figure 3)
 - d. on the Analysis sheet: please use the exact same value defined in the Project Title field of the Project sheet. Similarly, the same analysis alias should be used in both Sample and Analysis sheets which defines the relationship between analyses and samples
- 4. For assistance with your submission please contact the EVA helpdesk eva-helpdesk@ebi.ac.uk

Version: 2020-01-28 3





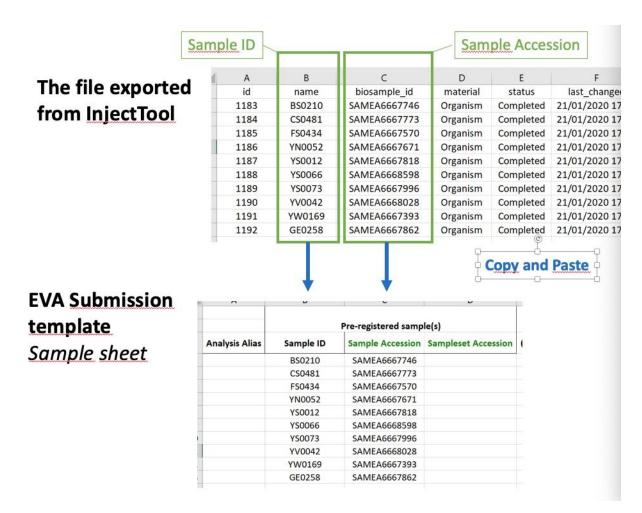


Figure 3 The Sample sheet of the EVA submission template file requires the correct BioSample and local identifiers for each record. "Sample Accession" is the place to include your BioSamples accessions (starting SAMEA...) and "Sample ID" is where you include your original sample ids. You can easily obtain these identifiers from the IMAGE InjectTool, see figure 1.

Checking your submission

After successful submission, you will receive a study accession beginning with PRJEB followed by a unique set of numbers. This study accession can be viewed in the ENA (https://www.ebi.ac.uk/ena), EVA (https://www.ebi.ac.uk/eva/) and IMAGE data portal (https://www.image2020genebank.eu/) websites.

Version: 2020-01-28 4