**Implementing Reference SNP Identifiers (rsIDs) to Foster Adoption of FAIR Genetic Variation Data Standards in Agriculture**

## Objective

FAIR-compliant genetic variation data is essential for accelerating progress in plant breeding, genomics and achieving sustainable agriculture. This guide enables researchers in adopting the standardized practices for generating, archiving and sharing FAIR (Findable, Accessible, Interoperable, and Reusable)-compliant genetic variant data.

### The need for standardization of plant variation data

Despite growing volumes of genotyping and sequencing data, the lack of standardized identifiers for genetic variants especially in plants presents significant challenges, including limited data interoperability, inconsistent functional annotation, and difficulty in comparing genetic information across individuals within the same species. Such fragmentation hinders the generation of meaningful biological insights and reusability of data across platforms and research efforts.

To address this gap, the European Variation Archive (EVA) has enabled the assignment of reference SNP identifiers (rsIDs) to non-human variation data, including crops and livestock. An rsID represents a reference SNP cluster ID that uniquely identifies a variable genomic locus. It should be unique, persistent, and stable across genome assemblies and crop varieties. These identifiers serve as essential anchors, connecting variant data to other valuable resources such as population allele and genotype frequencies, Genome-Wide Association Study (GWAS) and Quantitative Trait Locus (QTL) datasets, and trait associations, including both clinical and agronomic traits. The adoption of rsIDs not only enhances the utility of genetic variation data in plants but also lays the foundation for their broader implementation across agricultural research and breeding programs.

### Purpose of this guide

This guide enables researchers in adopting the best practices to produce FAIR-compliant variant datasets that are reusable, traceable, interoperable and ready for integration in global genomic resources. This guide answers key questions encountered during SNP data preparation and submission, including:

- How to verify whether existing variants already have assigned rsIDs in the EVA
- How to submit new SNP data to EVA, including scenarios where an accessioned reference genome exists for the species in the International Nucleotide Sequence Database Collaboration (INSDC), and how to submit one if it does not.
- How to ensure that the variant data is provided in a standardized, validated format such as Variant Call Format (VCF) including guidance on VCF metadata requirements (e.g. ##fileformat, ##reference_ac, ##contig) and the use of globally recognized sample

identifiers from major germplasm repositories e.g. BioSamples or GRIN-Global to ensure traceability and interoperability.

By following this recipe, researchers can adopt best practices for producing high quality, FAIR-compliant annotated variation datasets that support sustainable reuse, enable global interoperability and contribute to the growing ecosystem of open and accessible genomic data.

## Who should use this guide?

- Agriculture sciences researchers
- Bioinformaticians handling variant data
- Data curators and repository submitters
- Genomics researchers

## Inputs Required

- *VCF file*: Contains genotyping results; should include or be annotated with rsIDs
- *Reference genome*: Accessioned at one of the INSDC databases (*i.e.*, the National Center for Biotechnology Information (NCBI), the European Nucleotide Archive (ENA) or the DNA Data Bank of Japan (DDBJ) with GenBank/RefSeq (GCA/GCF) genome assembly ID); chromosome names must match those in the VCF
- *Sample metadata*: MIAPPE-compliant, registered via BioSamples
- *Sequencing data*: Registered and submitted to ENA
- *Project metadata*: Submission metadata (Excel template) for EVA

## Tools & Resources

- *EVA Submission Portal*: A web-based interface for submitting the variant data and metadata to EVA. Suitable for users handling small scale datasets.
- *EVA Submission CLI* (Command-Line Interface): Tool for the validation and submission of large datasets to EVA via command-line offers users automated batch processing and scripting capabilities.
- *Annotation Tools:*
    - *SnpEff:* A fast and versatile tool to annotate and predict the effects of genetic variants (e.g., missense, nonsense, synonymous).
    - *bcftools*: A command-line toolkit for manipulating VCF and BCF files, used for filtering, indexing, merging, and basic variant annotation.
    - *Ensembl Variant Effect Predictor (VEP)*: A comprehensive tool for annotating variants with information such as consequence type, gene name, transcript impact, regulatory features, and known variant databases. Often used for large-scale or highly detailed variant annotation.

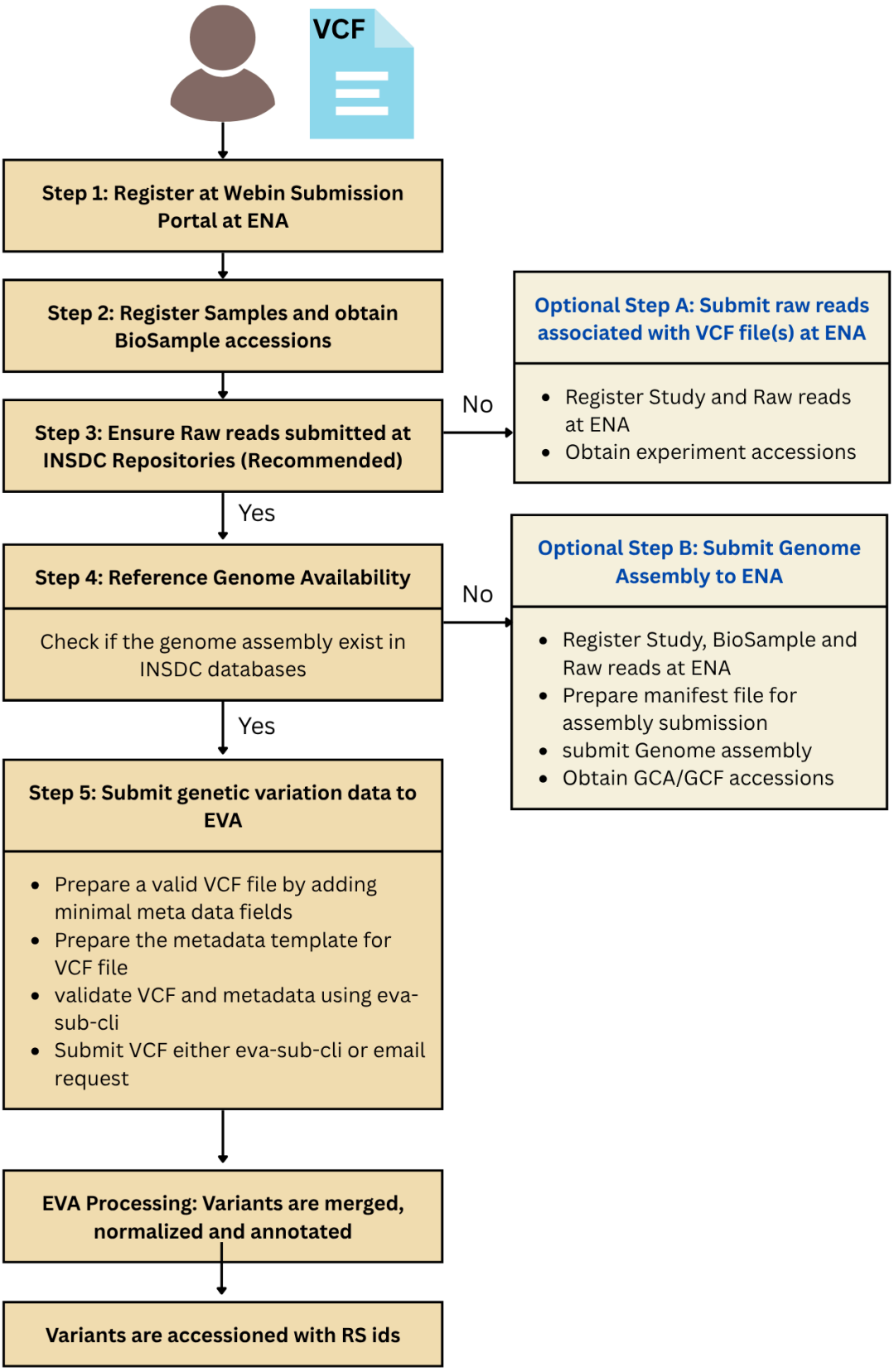


Figure 1. Overview of submission process

## Step-by-Step Guide

The European Variation Archive (EVA) is the 'one-stop-shop' for sharing all types of genetic variation data including SNPs, indels and structural variants and responsible for issuing and sustainable maintenance of variant accessions for all non-human species. Users can submit, access and download the genetic variation data at EVA. The EVA is based on direct user submission and variants must be described in the Variant Call Format (VCF; Versions v4.1, v4.2, & v4.3 accepted). Submissions to EVA consist of VCF file(s), associated data file(s) and detailed metadata describing samples(s), experiments(s) and analysis. Follow the steps below to submit your genetic variation data to EVA and associated data at ENA:

### Step 1: Register and create your user account at Webin Submission Portal

Webin Submission portal offers a range of submission services as well as reports at ENA for researchers to submit biological sequence data and associated metadata. Register and create your account at Webin Submission Portal by filling a simple registration form. After approval you can login to your account with Webin credentials.

### Step 2: Register a Sample

Samples represent the biological source of your sequencing data and are essential for organizing your reads, assemblies, or analyses. Here's how to register them easily using the Webin Submission Portal.

- Log in to your Webin Submission Portal and click on 'Register Samples'.
- Select "Download spreadsheet to register samples" to open the checklist menu.
- Browse and pick the appropriate checklist (e.g., "Plant and Fungi Sample Checklist").
- Customize your checklist by adding relevant fields and click 'Next' followed by downloading the TSV template spreadsheet.
- Fill out the template spreadsheet in excel or google sheets.

  **See the example below:**

| | |
|---|---|
| scientific_name | Pennisetum glaucum |
| collection_date | 2025-01-15 |
| geographic_location | India: Rajasthan |
| isolate | ICMB 123 |

| | |
|---|---|
| host | *None* (if not applicable) |
| environment_biome | *agricultural field* |
| description | *Pearl millet sample used for genome assembly* |

- To submit the spreadsheet, go back to the 'Register Sample' section and click on 'Upload filled spreadsheet' to register samples. Upload the file and click 'Submit Completed Spreadsheet'. Upon successful upload you'll receive sample accession numbers (e.g., ERS####### and SAMEA#######).

**Step 3: Ensure the submission of associated raw reads to ENA (Recommended)**

It is recommended that any associated raw reads should be submitted to INSDC repositories (ENA, NCBI or DDBJ) to ensure traceability, reproducibility, and compliance with FAIR data principles. Submitting raw reads allows downstream users to verify and re-analyze variant calls, enhances the visibility of your dataset, and enables EVA to link variant accessions to primary experimental data. Refer to *optional step A* for submitting raw reads to ENA.

**Step 4: Check if the reference genome assembly is available at ENA (or another INSDC database).**

The foremost requirement to submit a variant data to EVA is that all variants are submitted with an asserted position on an INSDC accessioned reference sequence. An INSDC sequence refers to any assembled biological sequence, including but not limited to: genome assemblies, transcriptomes, and gene sequences.

Search for your genome assembly at ENA: https://www.ebi.ac.uk/ena/browser.

If the genome assembly is available, proceed to *Step 5*: VCF files to EVA . If the reference genome is not available at ENA, you must submit your genome assembly to an INSDC repository (ENA, GenBank, or DDBJ) as described in Optional *Step B*: Submit a genome assembly to ENA.

Ensure your reference genome has a **GCF/GCA accession**.
Verify:
Chromosome names match those in ENA.
VCF includes: `##reference=GCA_XXXXXXX.X`

Note: Researchers from countries like India, China, and Korea can submit their data through their respective national data centres, provided these centres broker submissions to one of the INSDC member repositories (ENA, GenBank, or DDBJ). For example, the Indian Nucleotide Data Archive (INDA), hosted by the Indian Biological Data Centre (IBDC), brokers its open-access submissions in real time to the ENA. In this way, any submission will receive both Indian as well as INSDC accessions.

## Step 5: Submit the VCF files to the EVA

### *5.1 Preparation of a valid VCF data file by adding minimal list of metadata fields:*

To ensure your VCF file is FAIR-compliant and meets the standards outlined by Beier et al. (2022), following additional fields must be included. These metadata fields improve interoperability, data reuse, and submission compatibility with EVA

- Obligatory meta-information line:

  ##fileformat : file format

  *Example:*
  ##fileformat=VCFv4.3

- Recommended meta-information lines:
  - ##fileDate (Date): creation date of the VCF in the basic form without separator: YYYYMMDD.

  - ##bioinformatics_source (DOI, URL or URI): analytic approach used for creating the VCF file
    *Example:*
    ##bioinformatics_source="github.com/gramarga/tassel4-poly"

  - **##reference_ac (assembly_accession):** accession number along with the version of the concerned reference genome.
    *Example:*
    ##reference_ac= https://www.ebi.ac.uk/ena/browser/view/GCA_000003195.3"GCA_000003195.3

  - **##reference_url (DOI)**: a URL (or URI/DOI) for downloading of the concerned reference genome assembly, preferably from one INSDC archive.
    *Example:*
    ##reference_url="ftp.ncbi.nlm.nih.gov/genomes/all/GCA/902/498/975/GCA_902498975.1_Morex_v2.0/GCA_902498975.1_Morex_v2.0_genomic.fna.gz"

  - **Contig field format**: ##contig (<ID=ctg1, length=sequence_length, assembly=gca_accession, md5=md5_hash, species=NCBITaxon:id>); detailed description of individual reference genome sequence
    *Example:*
    ##contig=<ID=chr1H,length=522466905,assembly=GCA_902498975.1,md5=8d21a35cc68340ecf40e2a8dec9428fa,species=NCBITaxon:4513>

  - **Sample field format**: ##SAMPLE(<ID=url, ext_ID=$registry:identifier>); description of the concerned sample material present in the VCF file.

*Example*:
##SAMPLE=<ID=SAMEA104646767,DOI="doi.org/10.25642/IPK/GBIS/7811152”

Please read the VCF specification when converting data to VCF in order to ensure a valid format.

### *5.2 Prepare the metadata associated with VCF file*

Next, the user needs to provide the metadata details describing the samples and analysis of the EVA submission for accurate understanding, enhanced visibility and effective use of the data in future. This file will also be validated alongside of VCF files with eve-sub-cli Download the metadata template from the following link:

https://www.ebi.ac.uk/eva/?Submit-Data&src=wizard&wiztype=quicklink.

The form is divided into four sections, Submitter Details, Project, Analysis, Sample. Appropriate metadata for your submission should look similar to the example given below:

| **Submitter Details** | Information about User, contact email and its affiliation |
|---|---|
| **Project** | |
| Project Title | Whole-genome SNP analysis of Pennisetum glaucum cultivar ICMB 123 |
| Description | This study reports high-confidence SNPs identified through whole-genome sequencing of the ICMB 123 cultivar to support pearl millet genomics and breeding. |
| Centre | Centre_Name |
| Publications (DB:ID Format) | PubMed:37298320 |
| Taxonomy ID (as per NCBI taxonomy) | 4544 (Pennisetum glaucum) |
| Parent Project (Accession of parent project) | |
| Child Projects | |
| Peer Projects | |
| Links | |
| Hold date | |
| Collaborations | |
| Broker | |

| **Analysis** | |
|---|---|
| Analysis Title | Variant annotation of ICMB 123 using VEP |
| Analysis Alias | icmb123_vep_analysis_v1 |
| Description | SNPs were functionally annotated using Ensembl VEP with reference genome GCA_902498975.1. |
| Experiment type | Whole Genome Sequencing (WGS) |
| Reference | GCA_902498975.1 (Pennisetum glaucum reference genome) |
| Reference Fasta Path | /path/to/GCA_902498975.1_genomic.fna.gz |
| Platform | Illumina HiSeq 2500, Oxford Nanopore MinION |
| Software | GATK v4.3, bcftools, Ensembl VEP v107 |
| Pipeline Description | reads aligned with BWA-MEM; variants called using GATK HaplotypeCaller and annotated using VEP |
| Imputation | Not Performed |
| Phasing | Not Performed |
| Centre | EBI |
| Date | 2025-09-15 |
| Link(s) | |
| Run Accession(s) | ERR1234567, ERR1234568 |
| **Sample- Mandatory fields** | |
| Analysis Alias | ICMB123_SNPanalysis_2025 |
| Sample Name | ICMB123_leaf_sample |
| **Fields for Pre-Registered Sample** | |
| Sample Accession | SAMEA12345678 |

If you have not registered your samples, please provide details as per the metadata sample field for novel samples in the metadata template. Otherwise you can first register the samples separately in Biosample and then proceed with VCF submission.

*5.3 Submission of a VCF file to EVA*

After the preparation of a valid VCF file and metadata template, you are good to go with the submission of your VCF file to EVA following the instructions. Prior to submission, the files can be validated using EVA VCF Validation suite on GitHub. Please go through the ""Getting Started" guide" for detailed instructions.

The eva-sub-cli tool is a command line interface for the data validation and submission.

- Install eva-sub-cli using either of the two ways:
    1. Conda

    Use following command to install eva-sub-cli in a new conda environment called **eva**

    ```
    conda create -n eva -c conda-forge -c bioconda eva-sub-cli
    ```

    ```
    conda activate eva
    ```

    ```
    eva-sub-cli.py --help
    ```

    2. Source using Docker

    Ensure you have Python 3.8+ and Docker installed. Use following command to install eva-cli-cli

    ```
    pip install eva-sub-cli
    ```

    Test the installation using help message

    ```
    eva-sub-cli.py -h
    ```

- Prepare a unique submission directory for each submission and do not reuse directories to avoid data loss during the validation process.
- Metadata files can be provided in either as spreadsheets or as JSON and passed using the option `--metadata_xlsx or --metadata_json.`
- Provide a VCF file and the reference FASTA can be provided using the `--vcf_files` and `--reference_data` options. It can also be provided directly in the metadata file directly.
- Use following commands for validation:

```
eva-sub-cli.py --metadata_xlsx metadata_spreadsheet.xlsx --submission_dir submission_dir --tasks VALIDATE
```

For Docker users

```
eva-sub-cli.py --metadata_xlsx metadata_spreadsheet.xlsx --submission_dir submission_dir --tasks VALIDATE --executor docker
```

**Note:** For large VCF files, users can add `--shallow` argument to the command to validate only the first 10,000 lines in each VCF.

- After validation user can submit the vcf data and metadata using following command:

```
eva-sub-cli.py --metadata_xlsx metadata_spreadsheet.xlsx
--submission_dir submission_dir --tasks VALIDATE
```

For docker users

```
eva-sub-cli.py --metadata_xlsx metadata_spreadsheet.xlsx
--submission_dir submission_dir --tasks VALIDATE --executor
docker
```

For validation as well as submission

```
eva-sub-cli.py --metadata_xlsx metadata_spreadsheet.xlsx \
                        --vcf_files vcf_file1.vcf vcf_file2.vcf
--reference_fasta assembly.fa --submission_dir submission_dir
```

- Submission *via* email: Users can also email to eva-helpdesk@ebi.ac.uk to request submission and be required to follow the instructions sent by the EVA support team to complete the submission.

**Step 6: Post-submission normalization, annotation, and statistical analysis at EVA:**

All the non-human variants are accessioned by EVA including plants and EVA brokers the human variants to dbSNP. To make all the submitted data uniform post submission, EVA follows a few basic steps to normalize variants similar to dbSNP.

- Each variant shifted to left aligned.
- Each variant is processed to keep the start and end position specific to the variation only.

Details of the normalization process can be found here**.**

After the normalization process, variants are annotated using Ensembl's VEP.

The EVA defines allele frequency (AF) as the proportion of a specific allele at a genetic locus within a given population. AF values are study-specific, meaning the same variant can have different AFs across studies.

**Step 7: Access the accessioned variants and its future use:**

To access the accessioned variants at the EVA, you can use the following ways :

- **Web Browser Search**: Visit the EVA website at https://www.ebi.ac.uk/eva/ and use the search bar to look up variants by rsID, study accession (e.g., PRJEBxxxxxx), or species.
- **Study Browser**: Navigate to the "Browse EVA Studies" section to locate datasets by organism, study ID, or project title. You can download VCF files or explore annotated variant tables directly.
- **REST API**: EVA also offers a public REST API for programmatic access to variant data, enabling automated querying by rsID, genomic region, or gene.
- **FTP Access**: For bulk downloads, EVA provides access via FTP at ftp://ftp.ebi.ac.uk/pub/databases/eva/, where VCF files and metadata are organized by project.

## Optional step A: Register and submit sequencing reads and samples

### *A.1 Register the study*

Each submission must be associated with a pre-registered study and sample. Users can register to study (also called Project) via the Webin Submission Portal.

- Register your account at the Webin Submission Portal and login with your credentials.
- Click on the 'Register Study' button and fill in all required fields in the form and submit.
- After submission, a study will receive a study accession number (*ERP#####, PRJEB#####*) which will be needed for submitting genome assembly and raw reads. These accessions would be cited in publications involving the data.

**Example Submission:**

| | |
|---|---|
| Release Date | 2025-12-01 (Set to future date upto 2 years if data needs to remain private for a while) |
| Study Name | GenomeAssembly_ICMB123_India_2025 |
| Will you provide functional genome annotation? | Yes |
| Short Descriptive Study Title | Draft Genome Assembly of Pearl Millet Cultivar ICMB 123 |
| Detailed Study Abstract | This study presents the draft genome assembly of ICMB 123, a widely cultivated pearl millet variety in India. Sequencing was performed using Illumina HiSeq and Oxford Nanopore platforms, and hybrid assembly was generated using Flye and polished with Pilon. This dataset is expected to support future breeding and climate resilience research. |
| PubMed Citations | PMID: 37298320 (Add if already published; otherwise leave blank) |
| **Study Attributes:** | |
| Tag | Geographic location |
| Value | India |
| Tag | Target species |

| | |
|---|---|
| Value | Pennisetum glaucum |
| Locus Tag Prefixes (required only if you checked the box for functional annotation) | PMI (for *Pearl Millet India*)<br>(should contain only alphanumeric characters of 3 to 12 characters long, starts with a letter and all letters should be in uppercase) |
| | |

### *A.2 Register a Sample*

Samples represent the biological source of your sequencing data and are essential for organizing your reads, assemblies, or analyses. Here's how to register them easily using the Webin Submission Portal.

- Log in to your Webin Submission Portal and click on 'Register Samples'.
- Select "Download spreadsheet to register samples" to open the checklist menu.
- Browse and pick the appropriate checklist (e.g., "Plant and Fungi Sample Checklist").
- Customize your checklist by adding relevant fields and click 'Next' followed by downloading the TSV template spreadsheet.
- Fill out the template spreadsheet in excel or google sheets.

  **See the example below:**

| | |
|---|---|
| scientific_name | Pennisetum glaucum |
| collection_date | 2025-01-15 |
| geographic_location | India: Rajasthan |
| isolate | ICMB 123 |
| host | *None* (if not applicable) |
| environment_biome | *agricultural field* |
| description | *Pearl millet sample used for genome assembly* |

- To submit the spreadsheet, go back to the 'Register Sample' section and click on 'Upload filled spreadsheet' to register samples. Upload the file and click 'Submit Completed Spreadsheet'. Upon successful upload you'll receive sample accession numbers (e.g., ERS####### and SAMEA#######).

### *A.3 Submit associated raw reads*

- Go to the ENA Webin Submission Portal and click "Submit Reads."
- Expand "Download spreadsheet template for Read submission."
- Choose the appropriate file type for your data: Single-end FASTQ, Paired-end FASTQ, CRAM files, BAM files, HDF5 or FAST5 files
- Review the required and optional metadata fields (e.g., instrument model, insert size).
- Click "Next" and then download your customized .tsv spreadsheet template.
- Fill out the template spreadsheet.

  **See the example below:**

| | |
|---|---|
| sample | SAMEA12345678 (Your BioSamples accession) |
| study | PRJEB12345 (Your ENA Study accession) |
| instrument_model | Illumina HiSeq 2500 |
| library_name | ICMB123_RNASeqLib1 |
| library_source | GENOMIC (Permitted: GENOMIC, TRANSCRIPTOMIC, etc.) |
| library_selection | RANDOM (Permitted: RANDOM, PCR, RANDOM PCR, etc.) |
| library_strategy | WGS (Permitted: WGS, RNA-Seq, WXS, etc.) |
| library_layout | PAIRED |
| forward_file_name | ICMB123_R1.fastq.gz |
| forward_file_md5 | 1a79a4d60de6718e8e5b326e338ae533 (Generated via md5sum) |
| reverse_file_name | ICMB123_R2.fastq.gz |
| reverse_file_md5 | 4e07408562bedb8b60ce05c1decfe3ad (Generated via md5sum) |

- Return to the 'Submit Reads Section' in the Webin Submission Portal, expand "Upload filled spreadsheet template for Read submission," use the Browse button or drag-and-drop your file, and click "Submit Completed Spreadsheet."
  **Note:**

All reads must be de-multiplexed before submission.
Ensure that your files are already uploaded to your Webin FTP/SFTP account.

- If your metadata passes validation, you will immediately receive experiment (e.g., ERX123456) and run (e.g., ERR123456) accession numbers, after which ENA will check your uploaded read files and notify you of any errors via email, while allowing you to track submission progress through the Run Reports tab in the Webin Submission Portal. Link them in the RUN_REF field of the manifest file in *Step 2.4*.

### Optional Step B: Submit a genome assembly to ENA

Genome assemblies can be submitted to ENA via Webin command line submission interface with '-context genome' option. The detailed submission instructions are available at Submitting Genome Assemblies of Individuals or Cultured Isolates — ENA Documentation 1 documentation.

1. *Refer to Optional Step A to submit the sample and raw reads associated with the genome assembly*
2. *Prepare the files for genome assembly:*

- Create the manifest file to specify the set of files that are part of submission. The file is identified in the Webin-CLI command using the -manifest <filename> option. The table summarises the list of files required based on assembly level:

| **Assembly Types** | **Required Files** |
|---|---|
| Contig-Level | 1 manifest file<br>1 FASTA or flat file<br>Example: GCA_000003085 |
| Scaffold-Level | 1 manifest file<br>1 FASTA **or** flat file<br>Optional AGP file (for scaffold structure)<br>Example: GCA_902705575 |
| Chromosome-Level | 1 manifest file<br>1 FASTA **or** flat file<br>1 chromosome list file<br>Optional: AGP + unlocalised list<br>Example: GCA_000237925 |

- Add the metadata and filenames as per the assembly type in the manifest file. This text file has only two columns 'Field name' and 'Field value' as depicted in the example below:

| | |
|---|---|
| STUDY | PRJEB12345 |
| SAMPLE | SAMEA123456 |
| ASSEMBLYNAME | ICMB123_assembly_v1 |
| ASSEMBLY_TYPE | clone or isolate |
| COVERAGE | 80 |
| PROGRAM | Flye v2.9 + Pilon |
| PLATFORM | Oxford Nanopore, Illumina |
| MOLECULETYPE (optional) | genomic DNA |
| FASTA (optional) | icmb123.fasta.gz |
| MINGAPLENGTH (optional) | 200 |
| DESCRIPTION | Draft genome of ICMB 123 pearl millet cultivar |

- Install Webin-CLI from ENA docs. Please refer to the Webin command line submission interface documentation for full information about the submission process.
- Run the Webin-CLI validation command by specifying your credentials and path to your manifest file:

```
webin-cli -username Webin-XXXXX -password YYYYYYY -context genome -manifest manifest.txt -submit
```

- Now run the Webin-CLI submission command:

```
webin-cli -username Webin-XXXXX -password YYYYYYY -context genome -manifest manifest.txt -validate
```

- Upon successful submission genome assembly has been assigned with an analysis accession number (ERZxxxxxx) immediately on the interface. These are temporary accessions and should not be cited in publications.
- After the internal validation process ENA assigns stable public accessions to the assembly (GCA_123456789.1).

## References


1. Cezard, T. et al. The European Variation Archive: a FAIR resource of genomic variation for all species. *Nucleic Acids Res.* **50**, D1216–D1220 (2022).

2. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. VCF specification and best practices for variant data interoperability. *VCFv4.3 Specification*. Samtools/HTS-specs (2022). https://samtools.github.io/hts-specs/VCFv4.3.pdf

3. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).

4. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).

5. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

6. EMBL-EBI. European Variation Archive (EVA) Submission Portal. https://www.ebi.ac.uk/eva/?Submit-Data (2025).

7. EMBL-EBI. EVA Submission CLI (eva-sub-cli). GitHub repository. https://github.com/EBIvariation/eva-sub-cli (2025).

8. EMBL-EBI. VCF Validator Tool. GitHub repository. https://github.com/EBIvariation/vcf-validator (2025).

9. EMBL-EBI. ENA Webin Submission Portal. https://www.ebi.ac.uk/ena/submit/webin/login (2025).

10. EMBL-EBI. Submitting Genome Assemblies of Individuals or Cultured Isolates. ENA Documentation. https://ena-docs.readthedocs.io/en/latest/submit/assembly/genome.html (2025).