

Uploader

Data upload is the first step in the analysis. Uploader is used to upload different data types on the platform.

- Overview
- [UPLOAD SEQUENCE DATA](#)
- [UPLOAD REFERENCES](#)
- [UPLOAD ANNOTATIONS](#)
- [UPLOAD METADATA](#)

Overview

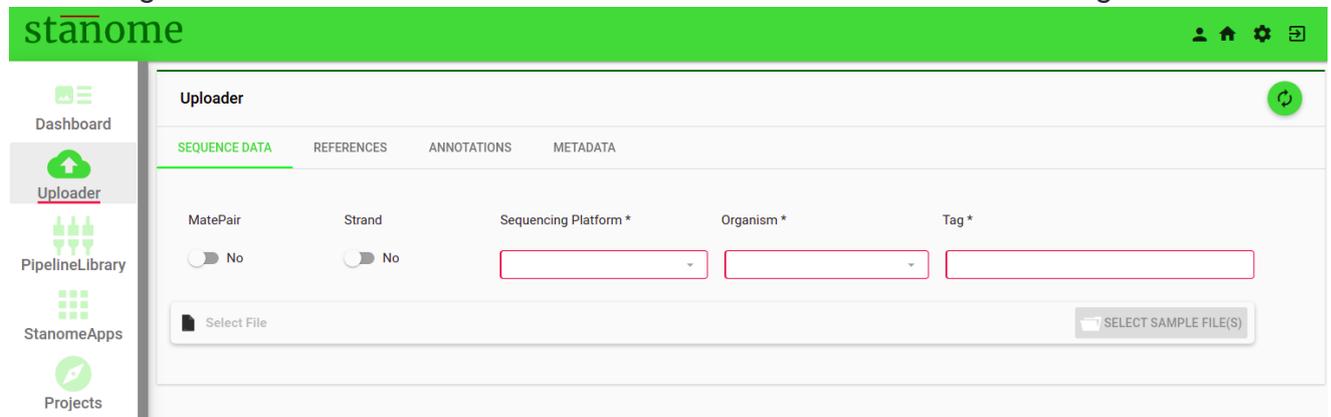
Data upload is the first step in the analysis. Generally, four data types are required:

1. Sequencing data (read files)
2. Reference genomes
3. Annotation files
4. Metadata

Click the  icon in the Utilities menu to access the **Uploader** window. Each data type is uploaded from a separate tab and follows the following three high-level steps:

1. Define the data types with correct tags
2. Verify and check for its existence in users account, format, and name conflicts
3. Upload

Pre-configured references and annotation files are available for the model organisms.



The screenshot shows the Stanome web application interface. At the top is a green header with the 'stanome' logo and navigation icons. A left sidebar contains menu items: Dashboard, Uploader (highlighted), PipelineLibrary, StanomeApps, and Projects. The main content area is titled 'Uploader' and has four tabs: SEQUENCE DATA (active), REFERENCES, ANNOTATIONS, and METADATA. The SEQUENCE DATA tab contains form fields for 'MatePair' (toggle set to 'No'), 'Strand' (toggle set to 'No'), 'Sequencing Platform *' (dropdown menu), 'Organism *' (dropdown menu), and 'Tag *' (text input). Below these fields is a 'Select File' button and a 'SELECT SAMPLE FILE(S)' button.

UPLOAD SEQUENCE DATA

Sequencing files can be uploaded through the **SEQUENCE DATA** (Fig. 1) tab. Complete all the fields in the form before selecting the files.

- **MatePair**: Select appropriate mate-pair information, if available. This is an optional parameter (Default: No).
- **Strand**: Select appropriate strand information, if available. This is an optional parameter (Default: No).
- **Sequencing Platform***: Select the sequencing instrument used to generate the data (check with the provider).
- **Organism***: Select the organism from the drop-down menu.
 -  **HINT** - Select **Other** if the organism is not listed.
- **Tag***: Provide a tag name for the sample files that can be used to filter the data later.

Caution - Tags should contain alphanumeric characters only.

- The user can select sample files through the selection dialog box.

 **HINT** - Currently allowed data file formats are fq, fastq, bam, sam, ubam, cram, hdf5, and their zipped versions.

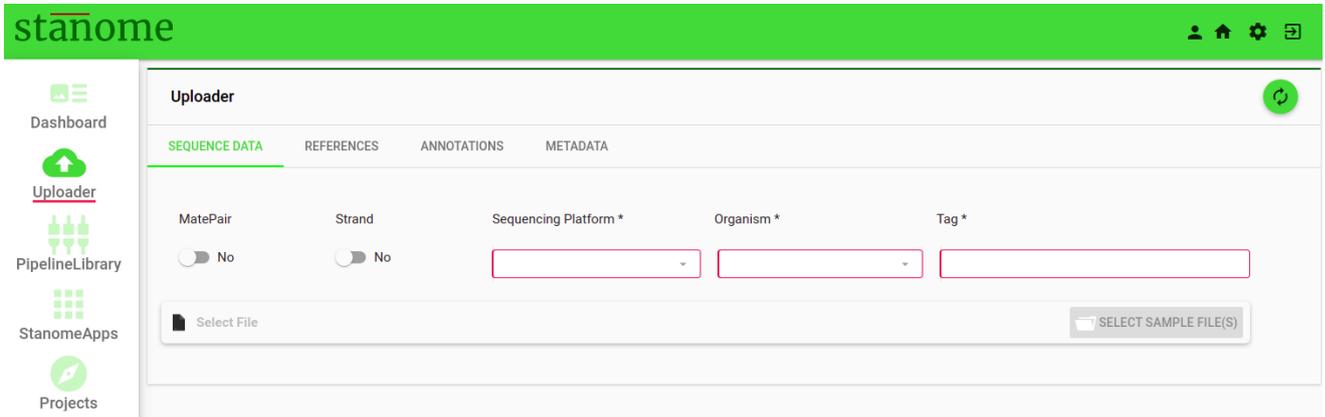
Ensure that each sample file is compressed separately with the same name as the name of the sample file itself

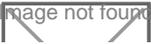
 **HINT** - Sequencing data from other sources (NCBI SRA) should be converted to fastq/fq format.

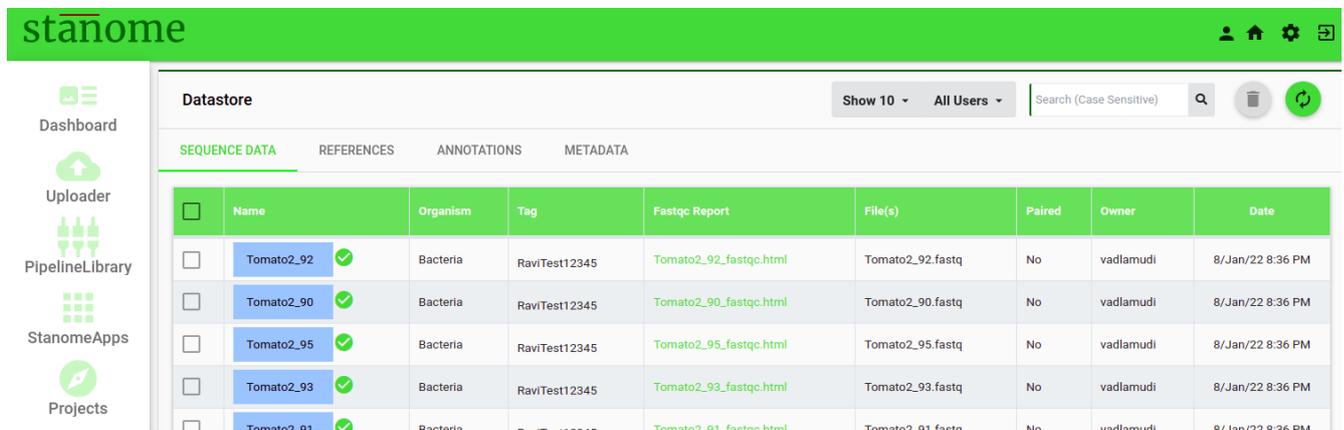
 **HINT** - If sequencing is multiplexed, demultiplex the data before uploading.

- Click  to verify file types, check file existence in the user's account, and pair files (not applicable to Single-End data). File pairing information (i.e. forward/reverse sample) is retained for PE sample data.

Pairing is done if forward and reverse samples have the suffix combinations: `_1/_2`; `_F/_R`; `_f/_r`; `-1/-2`; `-F/-R`; `-f/-r`; `_R1/_R2`.



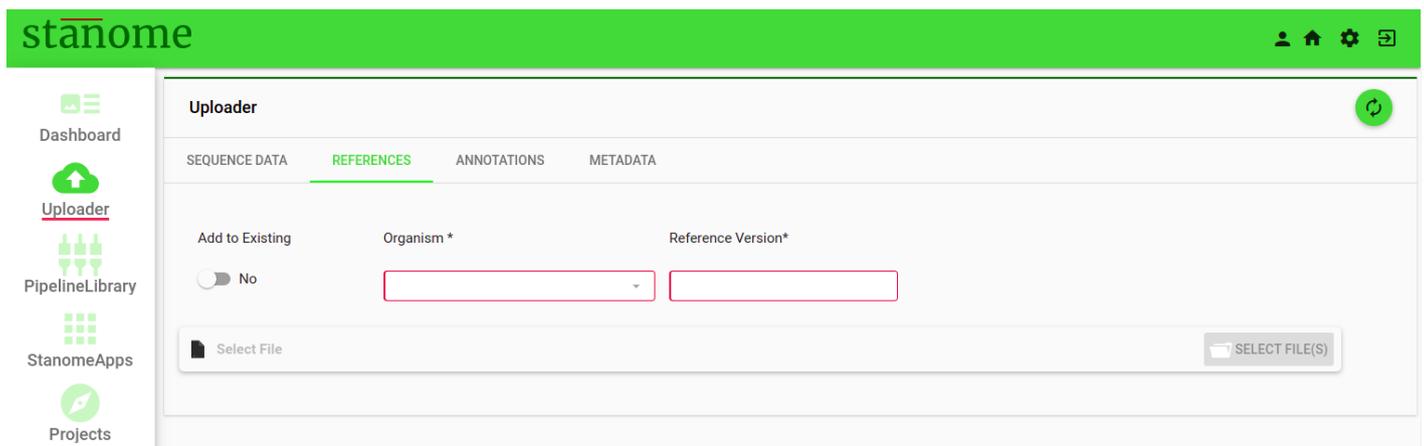
- Upon confirming the Data Store check, click  to begin the file transfer (the progress bar indicates the upload status for each file). Following four actions are performed (failed action errors will be displayed, if any) after a successful upload.
 - Uncompression of the files
 - Data quality check with FastQC
 - Registration into the **SEQUENCE DATA**
 - Status notification through email
- Go to the **SEQUENCE DATA** to access the uploaded files and quality report (Fig. 2). Occasionally, files are not visible instantly - a window refresh is required.
- Click on a sample name to access its additional details.
- Use the  icon on the upper right corner of the sample details window to delete samples.



UPLOAD REFERENCES

The latest versions of reference genomes and transcriptomes are available on the platform for eight model organisms. Pre-configured references can be explored through **REFERENCES** (Fig.). The *owner* column helps to identify the pre-configured references (owned by Stanome) from the custom references (owned by the users).

Click the **REFERENCES** tab on the **Upload** window to upload genome and transcriptome files.



The screenshot shows the Stanome web interface. The top navigation bar is green with the 'stanome' logo on the left and user, home, settings, and help icons on the right. A left sidebar contains icons for Dashboard, Uploader, PipelineLibrary, StanomeApps, and Projects. The main content area is titled 'Uploader' and has four tabs: SEQUENCE DATA, REFERENCES (which is selected and highlighted in green), ANNOTATIONS, and METADATA. Below the tabs, there are three input fields: 'Add to Existing' with a toggle switch set to 'No', 'Organism *' with a dropdown menu, and 'Reference Version*' with a text input field. At the bottom, there is a 'Select File' button and a 'SELECT FILE(S)' button.

Complete all the fields in the form before selecting the files.

1. **Add to Existing:** Helps to add additional files such as transcriptome/genome. to an existing genome/transcriptome. This is an optional parameter (Default: No).
2. **Organism*:** Select the organism name.
3. **Version*:** Provide *Genome Build* (version) name or the number of the reference file(s). This will help to select the correct reference version during analysis.

 **HINT** - Allowed formats for genome or transcriptome files are fa, fasta, fna, and their compressed formats. Each file should be compressed separately.

Following actions are performed on the uploaded references:

- Uncompression of the files
- Validation of the format and integrity of the file
- Registration into the **REFERENCES**
- Status notification through email

Successfully uploaded reference files are stored in the **REFERENCES**. Custom references can be deleted using  on the reference details window.

UPLOAD ANNOTATIONS

Pathways, gene ontology (GO) terms, ABR genes, and VEP (Variant Effect Predictor) files are classified as annotations and can be uploaded through the **ANNOTATIONS** (Fig.) tab.

ABR and GO_OBO upload has two fields:

1. **Organism:** Name of the organism
2. **Tag:** The version number of the annotation file

Gene Model, VEP, and Variations upload have three fields:

1. **Organism:** Name of the organism
2. **Reference version:** Version of the reference file
3. **Tag:** Version of the annotation file

GO and Pathway upload has four fields:

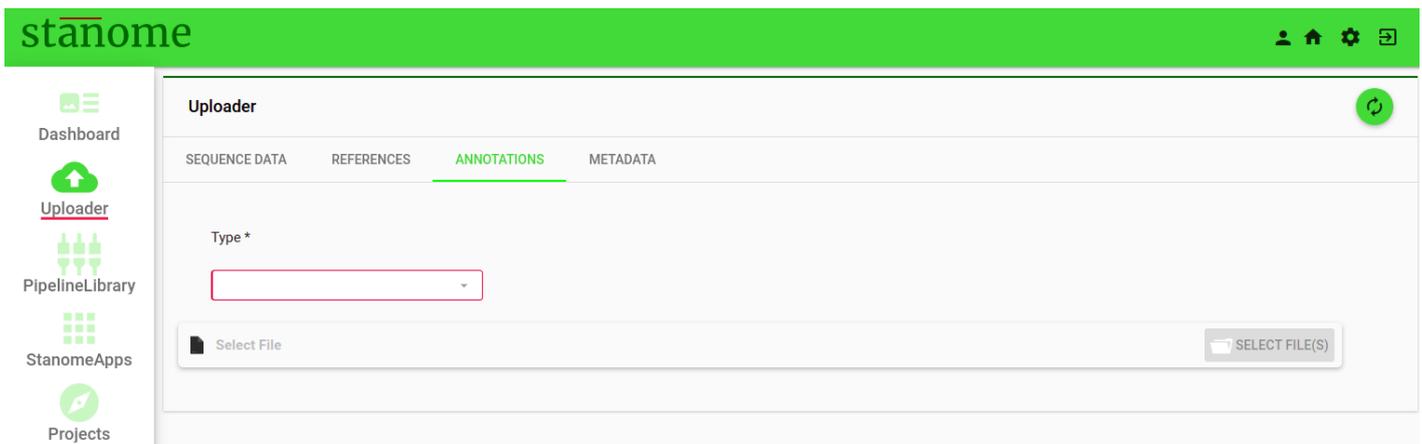
1. **Organism:** Name of the organism
2. **Reference version:** Version of the reference file
3. **GTF version:** Version of the gene model file
4. **Tag:** Version

Different types of annotations and their allowed formats are explained in Table below.

Tag	Data Type	Format	Details
Gene models	Gene annotations	Gtf, Gff3	Source: Ensembl. Should correspond to the Ensembl genome versions
Pathway	Pathways	Tab, Gmt	Source: Wiki pathways
Gene Ontology	Gene Ontology associations	Obo, Tab, Txt	GO terms

VEP	Variant annotations	Custom	Source: Ensembl. Should correspond to the Ensembl genome versions
Variations	GATK variants	VCF	Source: Ensembl

Table. Annotation file formats. Compressed files are allowed as long as each file is compressed separately.



Similar to Sequence and reference file uploads, the following actions are performed on the uploaded annotation files:

- Uncompression of the files
- Validation of the format and integrity of the file and its compatibility to genome/gene annotation file.
- Registration into the **ANNOTATIONS**
- Status notification through email

Successfully uploaded files are stored in **ANNOTATIONS** and can be accessed while executing the pipelines.

UPLOAD METADATA

Any files associated with an experiment (excluding sequencing files, references, and annotations) can be uploaded through the **METADATA** (Fig.) tab. During upload, the metadata files should always be associated with an organism and tagged appropriately, as explained in Table below.

Data Type	Tag Name	Format	Details
List of genes	Gene List	Tab, CSV, TXT	Ensembl Gene IDs only
Target markers	Hotspots	BED, VCF	SNPs, MNPs, INDELS
Amplicon ranges	Amplicon Range	BED, VCF	Target region with start and ends
Variants/genotypes	Genotypes	BED, VCF	Called variants

Table. Metadata file formats and associated upload tags.

Similar to Sequence and reference file uploads, the following actions are performed on the uploaded metadata files:

- Uncompression of the files
- Validation of the format and integrity of the file and its compatibility to the reference genome.
- Registration into the **METADATA**
- Status notification through email

- Dashboard
- Uploader**
- PipelineLibrary
- StanomeApps
- Projects

Uploader

SEQUENCE DATA REFERENCES ANNOTATIONS **METADATA**

Type

Select File

SELECT FILE(S)

Successfully uploaded metadata files are stored in **METADATA** and can be accessed while executing the pipelines.