



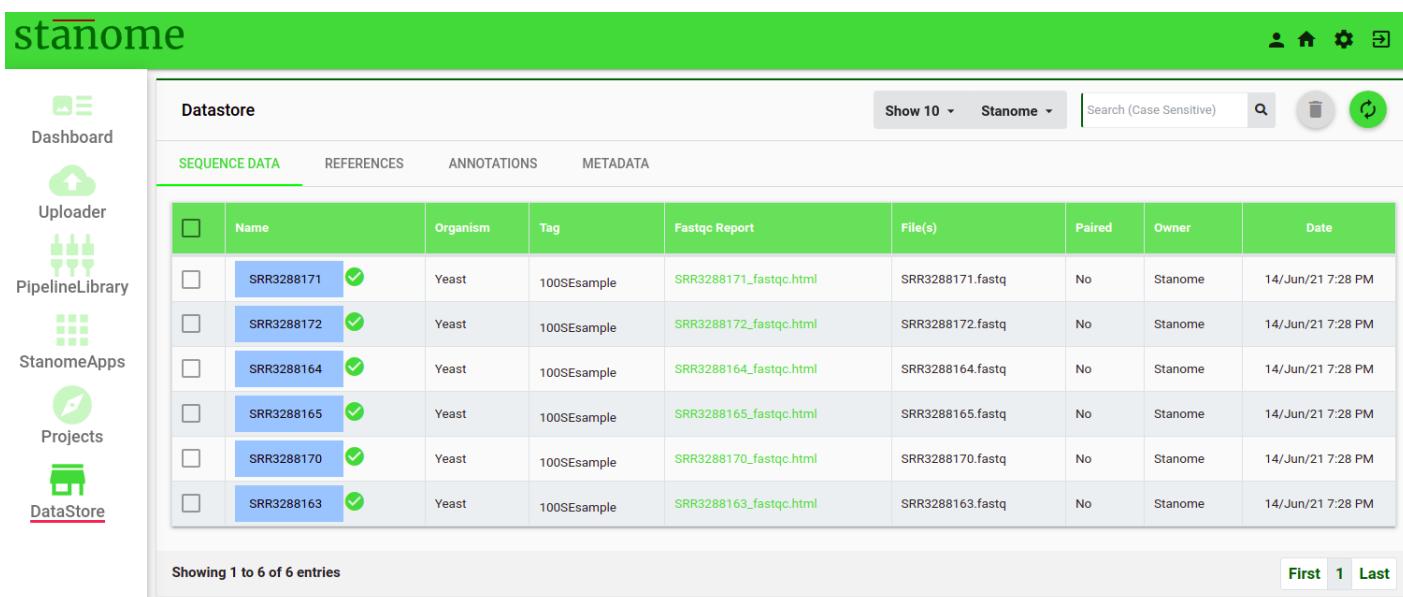
# Data Store

Data Store represents four different data types, i.e. Sequence Data, References, Annotations, and Metadata.

- Sequence Data
- References
- Annotations
- Metadata

# Sequence Data

**Sequence Data** stores the sequencing data files uploaded by the users and displays them in reverse chronological order. Click the  icon in the Utilities menu to access the **Sequence Data** window (Fig. 1). The paginated window displays 10 samples, by default, per page and can be increased to show 20, 40, 60, 80, 100 samples. The search box (Fig. 1) on the top right corner allows users to filter samples with tags, organism names, or pairing information. Validation failed samples are greyed out and displayed with a “Validation Failed” message. The failed samples are not available for selection and should be deleted from the platform by the user. Validation passed samples are shown with a green check . The quality (fastqc) report is available for the validated samples (Fig. 1).



The list view shows eight fields :



- **Name:** Derived sample name from the file name by the platform

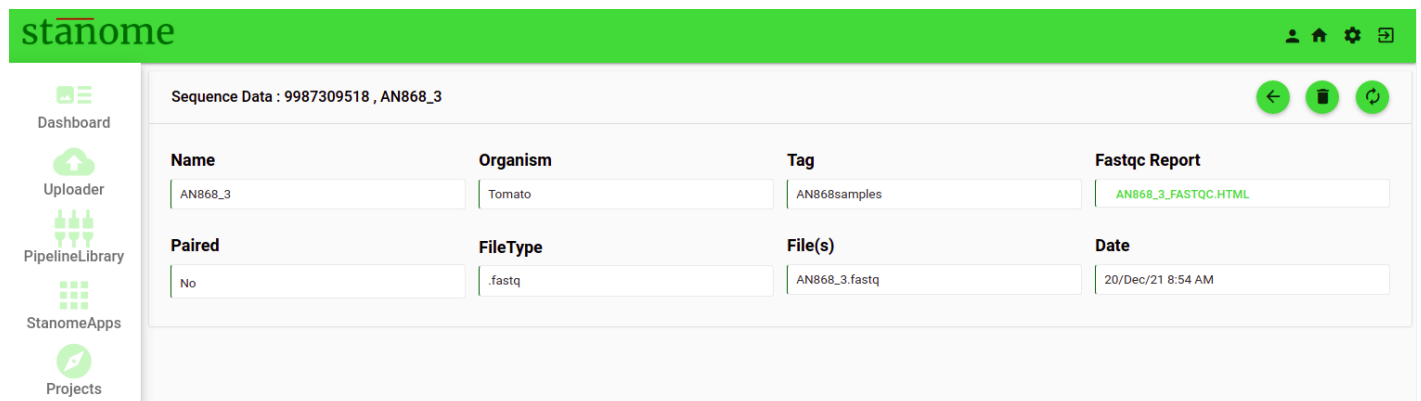
*Example:*

Sequencing Type	File Name	Sample Name
Single End	<i>seq_file.fastq</i>	<i>seq_file</i>
Paired End	<i>seq_file_1.fastq</i> <i>seq_file_2.fastq</i>	<i>seq_file</i>

- **Organism:** Organism name, provided by the user during upload

- **Tag:** Tag name, provided by the user during upload
- **Fastqc Reports:** The FastQC quality report of each sample
- **File(s):** The actual file name(s) associated with the sample name
- **Paired:** Sample pairing information, provided by the user during upload
- **Owner:** Name of the user, who uploaded the samples
- **Date:** Date of upload


Additional sample details can be accessed by clicking on the sample name. Click the  icon on the top right corner of the sample window (Fig. 2) to delete the selected sample. Click the  icon to go back to the **Sequence Data** window.

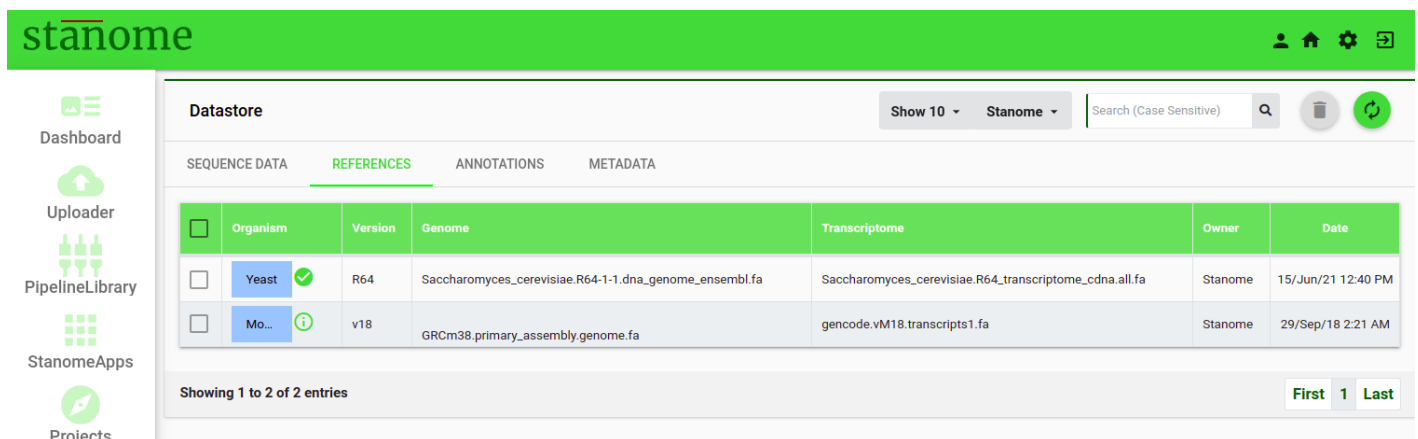




The screenshot shows the Stanome web interface. The top navigation bar is green with the 'stanome' logo on the left and user, home, settings, and help icons on the right. A left sidebar contains icons for Dashboard, Uploader, PipelineLibrary, StanomeApps, and Projects. The main content area displays 'Sequence Data : 9987309518 , AN868\_3' with three action icons (back, delete, refresh) on the right. Below this is a table with the following data:

Name	Organism	Tag	Fastqc Report
AN868_3	Tomato	AN868samples	AN868_3_FASTQC.HTML
Paired	FileType	File(s)	Date
No	.fastq	AN868_3.fastq	20/Dec/21 8:54 AM

# References



Click the **References** tab after clicking the  icon to access the **References** window (Fig. 1). The **References** window shows the list of validated genomes and transcriptomes for several model organisms. Validated references are also pre-indexed for commonly used tools such as Bowtie2, and Samtools and are shown as “Stanome” owned (Fig. 1). The account username is the owner of the custom references. The paginated window displays 10 references, by default, per page and can be increased to show 20, 40, 60, 80, 100 references. The search box (Fig. 1) on the top right corner of the page allows users to filter references using organism name, reference version, owner, or reference file name. Custom references are validated for file integrity, format, and content. Validation failed references are greyed out and displayed with a “Validation Failed” message. The failed references are not available for selection and should be deleted from the platform by the user.



<input type="checkbox"/>	Organism	Version	Genome	Transcriptome	Owner	Date
<input type="checkbox"/>	Yeast 	R64	Saccharomyces_cerevisiae.R64-1-1.dna_genome_ensembl.fa	Saccharomyces_cerevisiae.R64_transcriptome_cdna.all.fa	Stanome	15/Jun/21 12:40 PM
<input type="checkbox"/>	Mo... 	v18	GRCm38.primary_assembly.genome.fa	gencode.vM18.transcripts1.fa	Stanome	29/Sep/18 2:21 AM

The list view shows six columns:

- **Organism:** The organism name
- **Owner:** Validated and indexed references are owned by stanome whereas the custom references are owned by users
- **Version:** The version number of the reference file(s). Users can choose the version name for the custom references
- **Date:** Date of upload
- **Genome:** Name of the genome file
- **Transcriptome:** Name of the transcriptome file

Additional reference details (Fig. 2) can be accessed by clicking on the organism name. Click the  icon on the top right corner to delete a reference. Click the  icon to go back to the **References** window.


-  Dashboard
-  Uploader
-  PipelineLibrary
-  StanomeApps

References : 7904413154 , Tomato\_forGATK



Organism *	Version *	Owner	Date
<input type="text" value="Tomato"/>	<input type="text" value="forGATK"/>	<input type="text" value="kshama.goyal"/>	<input type="text" value="8/Jan/22 6:56 AM"/>
Genome		Transcriptome	
<input type="text" value="Solanum_lycopersicum.SL2.50.dna.toplevel_GATK_new.fa"/>		<input type="text" value="NA"/>	

# Annotations

Click the **Annotations** tab after clicking the  icon to access the **Annotations** window (Fig.). The **Annotations** window shows the list of the annotation files, which are utilized in the functional annotations applications, such as

- Gene Models (GFF/GTF);
- Pathway (GMT);
- Gene Ontology (GO);
- Antibiotic Resistance (ABR); and
- Variant Effect Predictor (VEP).
- Variations

The paginated window displays 10 annotations, by default, per page and can be increased to show 20, 40, 60, 80, 100 annotations. The search box (Fig.) on the top right corner of the page allows users to filter annotations with file names, type, organism name, or version. During upload, the annotation files undergo a rigorous validation process for file format, integrity, and content.

Examples:

- Gene Models should be compatible with the reference genome version.
- Pathway and Gene Ontology files are validated against the corresponding Gene Models.

Validation failed annotations are greyed out and displayed with a “Validation Failed” message. The failed annotations are not available for selection and should be deleted from the platform by the user.

**Datastore** Show 10 ▾ Stanome ▾  🔍 🗑️ 🔄

SEQUENCE DATA REFERENCES **ANNOTATIONS** METADATA

<input type="checkbox"/>	Name	Organism	Version	Validation	File Type	Owner	Date
<input type="checkbox"/>	Saccharomyces_cerevisiae_reactome_p...	Yeast	Reference : R64 GTF : v1	v1	GMT	Stanome	15/Jun/21 1:07 PM
<input type="checkbox"/>	yeast_new.gtf	Yeast	Reference : R64	v1	GTF	Stanome	15/Jun/21 12:53 PM


Showing 1 to 2 of 2 entries First 1 Last

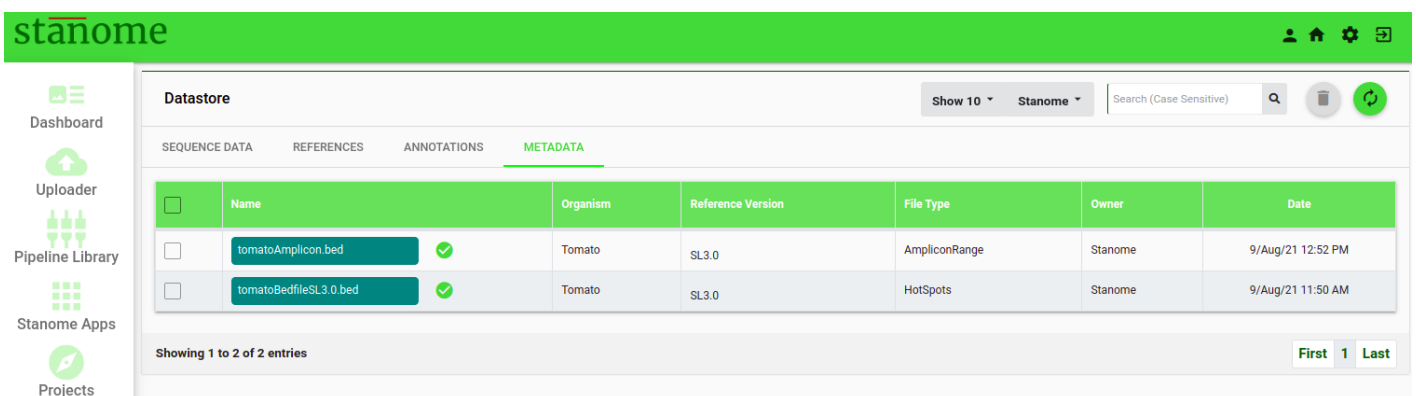
The list view shows seven fields :

- **Name:** Filename
- **Organism:** Organism to which annotation file corresponds to
- **Version:** Reference version to which the annotation file is associated
- **Validation:** Version of the annotation file
- **FileType:** Type of the file (Example: Pathway, GO, or VEP)
- **Owner:** Owner of the file
- **Date:** Displays the file upload date

# Metadata



Click the **Metadata** tab after clicking the  icon to access the **Metadata** window (Fig.). The **Metadata** window shows the list view of Gene lists, Hotspots and Amplicon ranges uploaded by stanome or the user. These files are used as metadata during analysis. The window is restricted to display 10 (or more) Metadata files per page. The search box (Fig.) on the top right corner of the page allows users to filter files with name(s), type (i.e. gene list or hotspots), platform, or organism name. Metadata files are also checked for format and file integrity and their correspondence to a reference genome version. Only the validated files are available for consumption on the platform, and the remaining will be greyed out with the “Validation Failed” message.



The screenshot shows the stanome interface with the Metadata window open. The window displays a table of metadata files with the following columns: Name, Organism, Reference Version, File Type, Owner, and Date. Two files are listed, both with a green checkmark indicating successful validation.

	Name	Organism	Reference Version	File Type	Owner	Date
<input type="checkbox"/>	tomatoAmplicon.bed	Tomato	SL3.0	AmpliconRange	Stanome	9/Aug/21 12:52 PM
<input type="checkbox"/>	tomatoBedfileSL3.0.bed	Tomato	SL3.0	HotSpots	Stanome	9/Aug/21 11:50 AM

Showing 1 to 2 of 2 entries

Navigation: First 1 Last

The list view shows following six fields :

- **Name:** Filename
- **Organism:** Name of the organism to which metadata belongs to
- **Reference Version:** Reference version to which metadata file correspond to
- **FileType:** Type of the file e.g. Gene list, Hotspots, or Amplicon range
- **Owner:** Owner of the metadata file
- **Date:** Date of upload