



NASA GeneLab: Open Science for Life in Space



Kristen Peach³, San-Huei Lai Polo², Samrawit Gebre¹

¹Space Biosciences Division, NASA Ames Research Center, ²KBR, NASA Ames Research Center, ³Bionetics, NASA Ames Research Center

Genelab.nasa.gov

INTRODUCTION

GeneLab is an open access database containing fully coordinated and curated 'omics' (genomics, transcriptomics, proteomics, metabolomics) data, and detailed metadata for a variety of model organisms. GeneLab enables the exploration of molecular network responses to space environments using a systems biology approach. To explore the data, the GeneLab Data Systems (GLDS) provides users various tools for data analysis, collaborative workspace for file storage and sharing, a visualization portal, and training programs, GL4HS and GL4U, to train the next generation of space biologists.

Our designs for the data repository continue to be guided by the FAIR principles. GeneLab has emerged as an omics data archive that is accepted by the space biology community. Furthermore, recent NASA Research Announcements require deposition of data from proposals selected for funding into the GeneLab database.

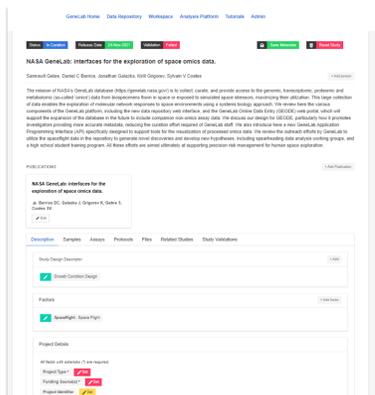
CURATION FOR SPACE OMICS

What Makes GeneLab Unique

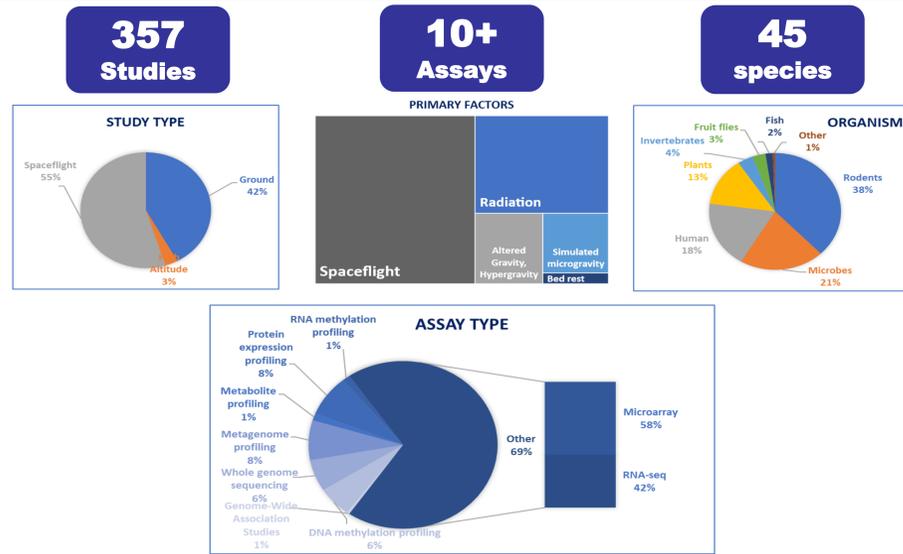
- A repository of *space-related* omics data that is extremely rare and valuable.
- Collect and publish spaceflight environmental data: space radiation dosimetry.
- Identify unique space factors in addition to treatment: enable search across repository for comparisons and data mining
- Apply metadata standard specific for spaceflight missions such as hardware, diet and recovery time on earth post-flight. (if applicable)
- Using standardized pipelines, analyze and make available high-order data (e.g. DGE) along with detailed processing scripts.
- Curators with expertise in various model organisms and multi-omics data types.
- Community-guided metadata standards (Analysis Working Group, AWGs)

GeneLab Environment for Online Data Entry (GEODE)

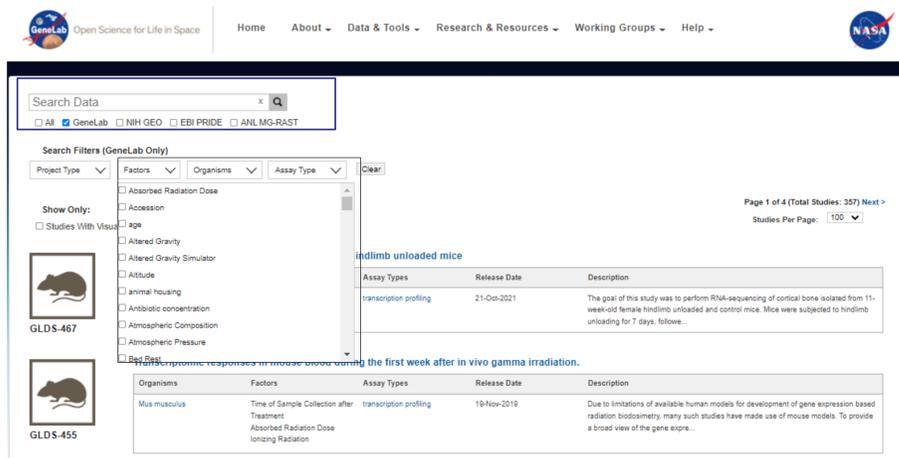
- A web-based submission portal which allows PIs and scientists to enter their own data and curate metadata while generating data
- Provides an accession number (GLDS), and DOI as soon as a study is created



DATA REPOSITORY



Use the keyword search bar or filters to search for data. In addition, you can search across several federated databases by selecting NCBI GEO, EBI PRIDE, MG-RAST, or all.



Each study contains detailed metadata, raw and processed data files.

STUDY RECORD

DESCRIPTION

FACTORS

ORGANISM

ASSAY TYPE

SAMPLES

Source Name	Sample Name	Characteristics: Group	Characteristics: Sex	Characteristics: Age	Characteristics: Diet	Characteristics: Time	Characteristics: Treatment	Characteristics: Assay Type	Factor Value	Product REF	Factor ID
MMBL_01234567	MMBL_01234567	Mus musculus	OTX:W/O	Neon-Bioscience	W/O Type	Female	Human Upprotein Protein Fold (UpP)	all:Brain	28 weeks	Animal:Rodent	SpaceType
MMBL_01234567	MMBL_01234567	Mus musculus	OTX:W/O	Neon-Bioscience	W/O Type	Female	Human Upprotein Protein Fold (UpP)	all:Brain	28 weeks	Animal:Rodent	SpaceType
MMBL_01234567	MMBL_01234567	Mus musculus	OTX:W/O	Neon-Bioscience	W/O Type	Female	Human Upprotein Protein Fold (UpP)	all:Brain	28 weeks	Animal:Rodent	SpaceType

DATA MODEL & STANDARDS

Data Model

ISA (Investigation-Study-Assay) experiment model for tagging data files with metadata. Framework provides rich description of the experimental metadata so that the resulting data and discoveries are reproducible and reusable.

- Leverages community-developed concept vocabularies, ontologies
- Support linkage of data values
- Supports standards guidelines for assays (MIAME, MIAPE, MIxS)

Data Files: open-source file formats and common exchange formats

FAIRness and Usability

- Persistent IDs (**GLDS, DOIs**)
- Retrievable by their ID using a standard protocol (**NASA Open data policy, APIs**)
- Use formal, broadly applicable languages, **JSON-LD**, use standard vocabularies (**ontologies**)
- Rich, accurate metadata, **DOI-enabled citation in research papers**



WHAT'S NEXT?

- Continue metadata harmonization across repository
- Creating sample-level metadata templates specific for plants, microbes and mammals to streamline data submission and reduce curation time
- Create a metadata aggregator to enhance search functionality
- Programmatic federation with NIH Databases such as SRA
- Integrate with NASA Ames Life Sciences Data Archive (ALSDA) – phenotype and non-omics database for space life sciences
- Integrate large data transfer protocol using open-source software
- Continue to curate with FAIR principles in mind and consider a more quantitative method for FAIR checking our datasets to track our improvement

ACKNOWLEDGMENTS

We acknowledge the NASA GeneLab Team. Research funding provided by the GeneLab Project at NASA Ames Research Center, through NASA's Biological and Physical Sciences Division of the Science Mission Directorate.