



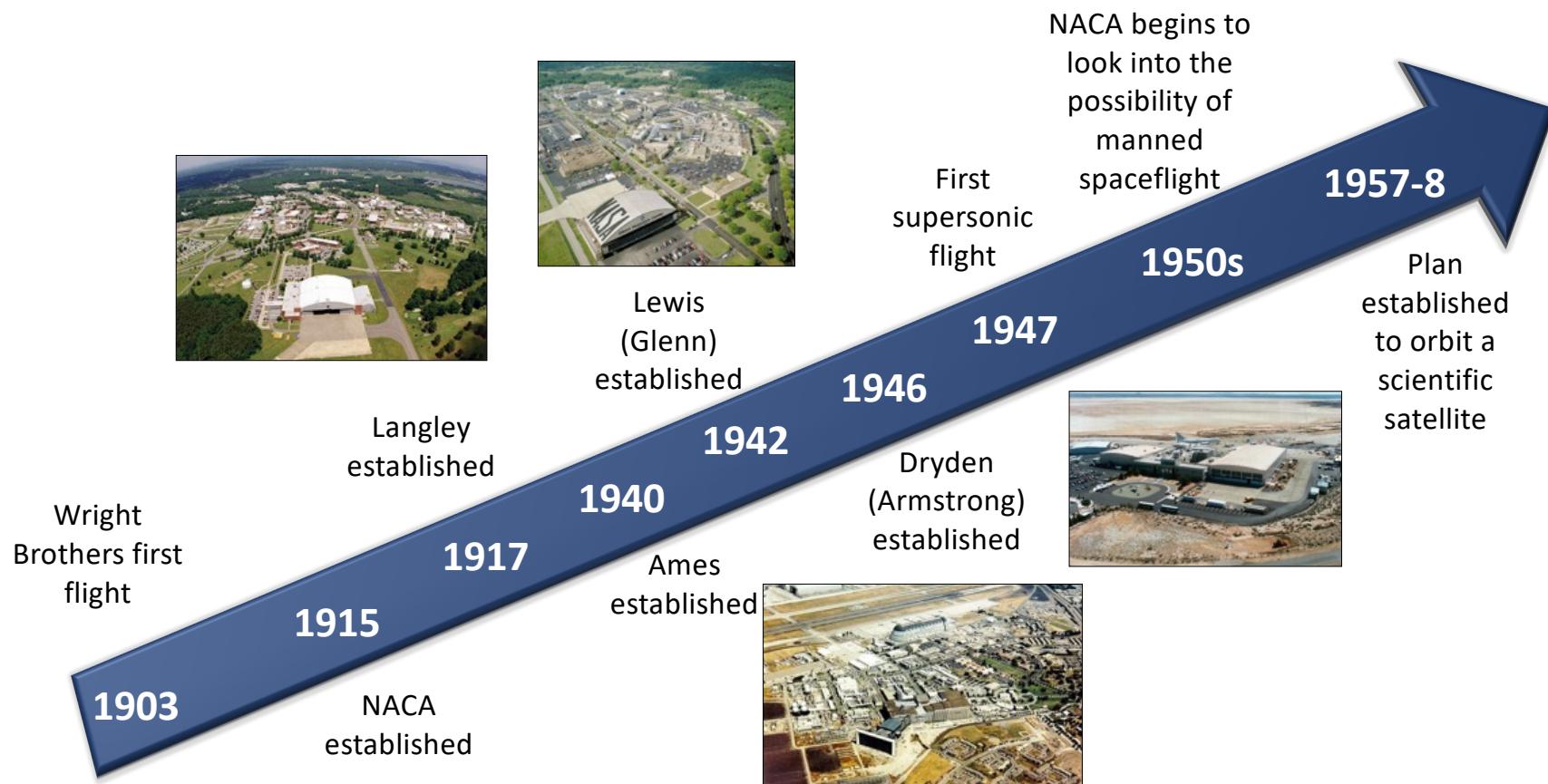
Introduction to NASA, GeneLab, and the Command Line

Biological and Physical
Sciences

Amanda M. Saravia-Butler, Ph.D.
GeneLab Data Processing Lead
NASA Ames Research Center
Contractor: KBR

Introduction to the National Aeronautics and Space Administration (NASA)

A Brief History of NACA



<https://www.nasa.gov/ames/the-national-advisory-committee-for-aeronautics>

A Brief History of NASA



'An Act to provide for research into the problems of flight within and outside the Earth's atmosphere, and for other purposes.'

Soviet Union launches Sputnik 1

1957

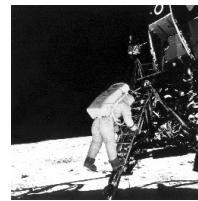
NASA established and absorbs itself into NACA

1958

US launches it's first Earth satellite

1961-3

Mercury's single astronaut program: Can humans survive in space?



Project Gemini: Two astronauts practice space operations

1965-6

Project Apollo: Explore the Moon

1968-72

Congress authorized NASA to build the Space Station

1981

Space Shuttle's first mission: STS-1



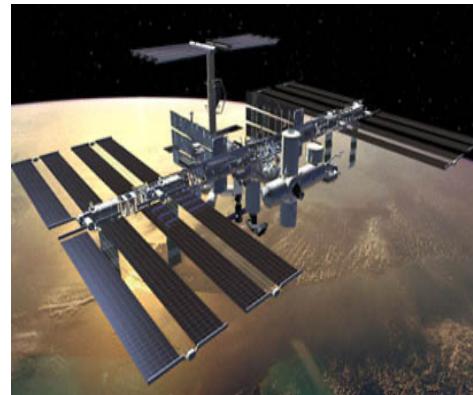
2011

Last Space Shuttle Mission: STS-135

NASA's Vision



We reach for new heights and reveal the unknown for the benefit of humankind



NASA's Mission



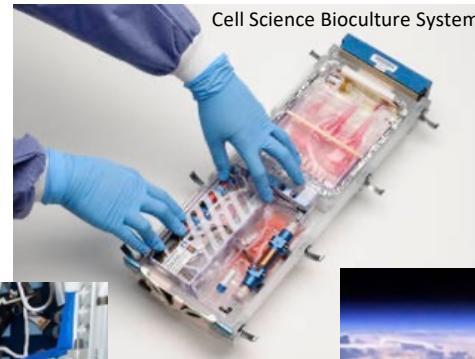
Drive advances in science, technology, aeronautics, and space exploration to enhance knowledge, education, innovation, economic vitality and stewardship of Earth



<https://www.nasa.gov/content/the-crawlers>



NASA astronaut Christina Koch



Cell Science Bioculture System



<https://www.nasa.gov/cygnss/overview>

How does NASA achieve its Mission?



- **NASA HQ** – Provides overall guidance and direction to the Agency
- **NASA Centers and Facilities** – Where the day-to-day NASA work is performed in laboratories, airfields, wind tunnels, etc.
- NASA conducts its work through four main organizations called **Mission Directorates**:

➤ Aeronautics Research (ARMD)



➤ Human Exploration and Operations (HEOMD)



➤ Science (SMD)

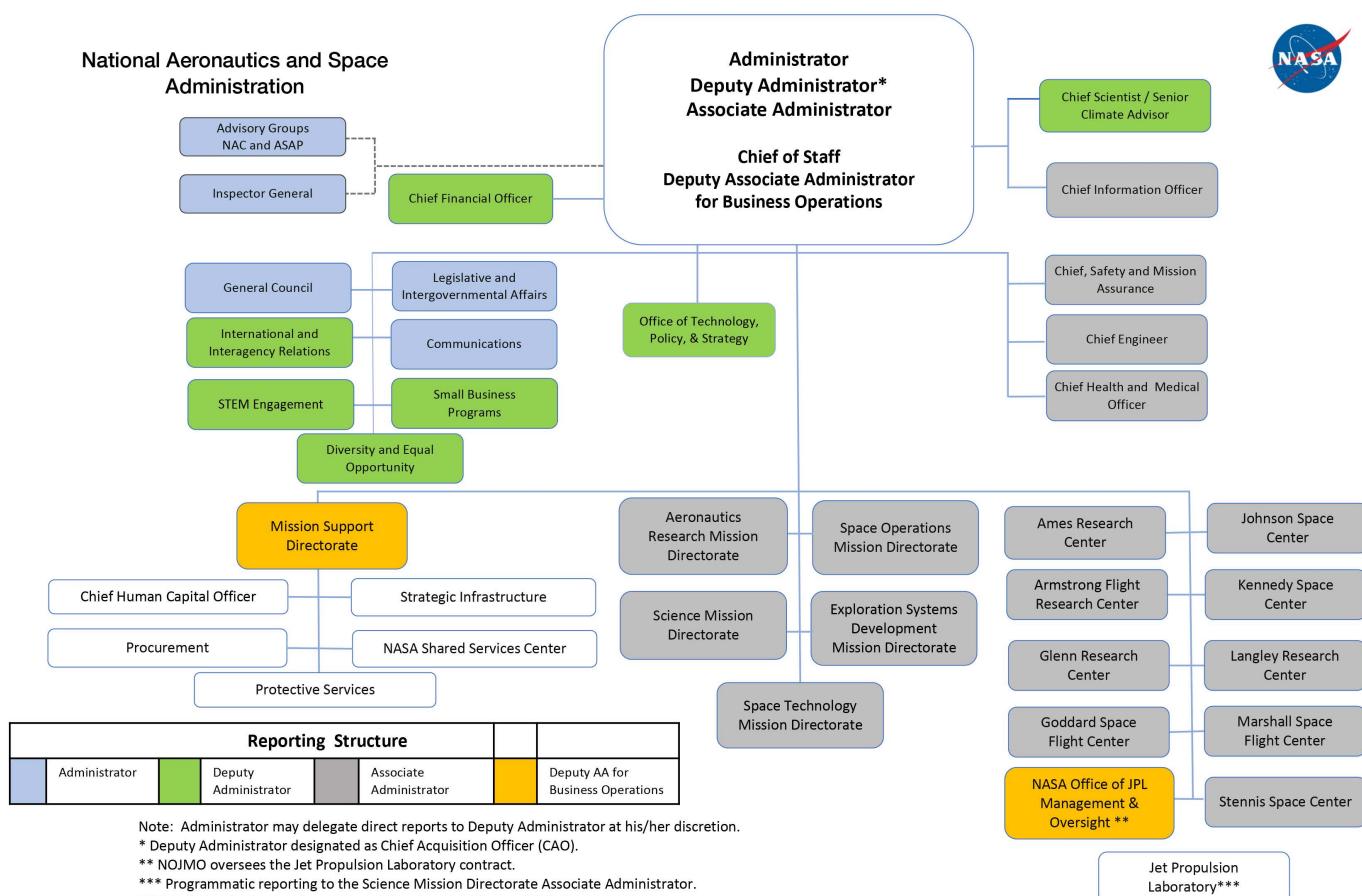


➤ Space Technology (STMD)



<https://www.nasa.gov/offices/olia/overview/index.html>

NASA Organization



NASA Leadership



Administrator: Bill Nelson

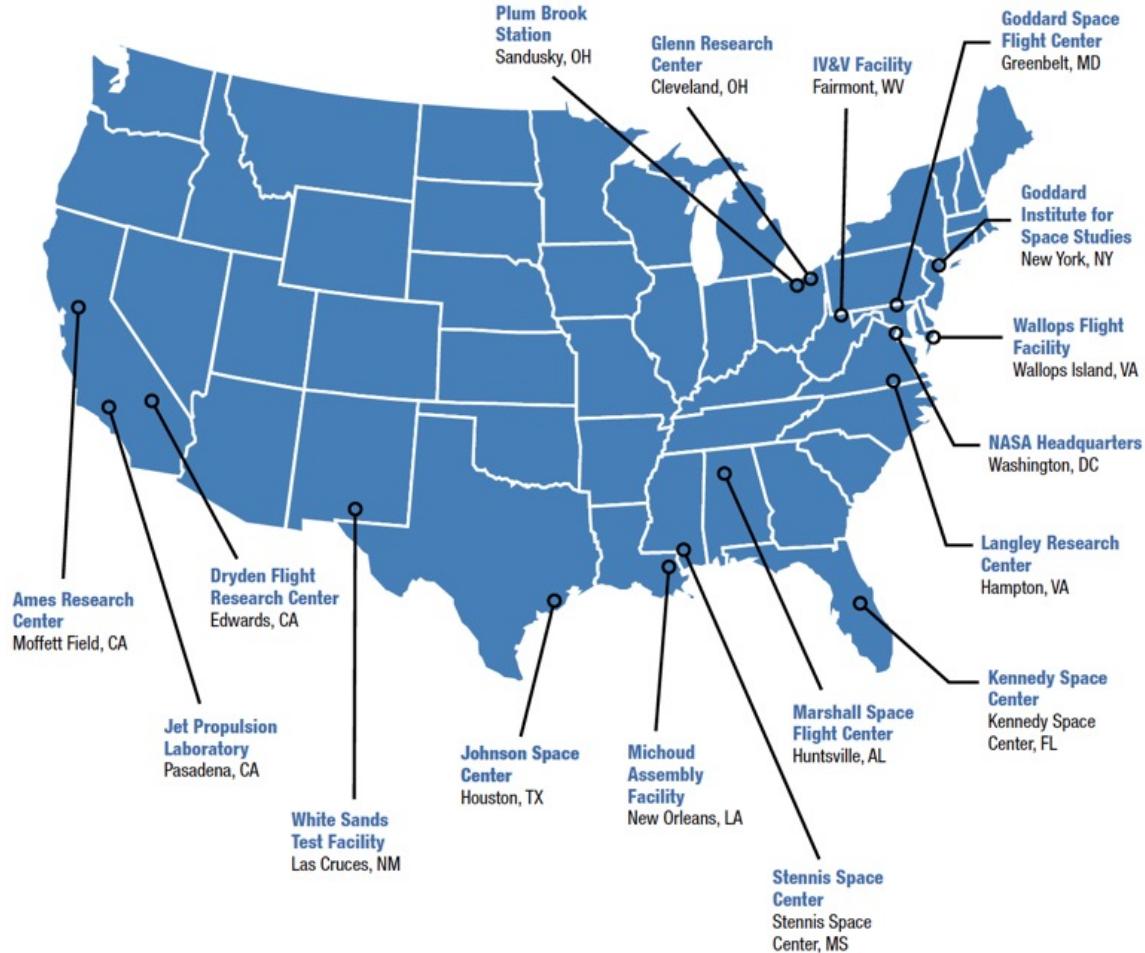
Sen. Bill Nelson was sworn in as the 14th NASA administrator on May 3, 2021, tasked with carrying out the Biden-Harris administration's vision for the agency. Nelson served in the U.S. Senate for 18 years from Florida and as a payload specialist on space shuttle mission 61-C in 1986. Bill Nelson's full bio: <https://www.nasa.gov/feature/nasa-administrator-bill-nelson/>



Deputy Administrator: Pamela Melroy

One of only two women to command a space shuttle, Melroy logged more than 38 days in space. All three of her missions were assembly missions to build the International Space Station. After serving more than two decades in the Air Force and as a NASA astronaut, Melroy took on a number of leadership roles, including at Lockheed Martin, the Federal Aviation Administration, the Defense Advanced Research Projects Agency, Nova Systems Pty, Australia, and as an advisor to the Australian Space Agency. She currently is an independent consultant and a member of the National Space Council's Users Advisory Group. Pam Melroy's full bio: https://www.nasa.gov/sites/default/files/atoms/files/melroy_pamela.pdf

NASA Centers and Facilities



NASA Centers and Facilities



NASA HQ



Ames Research Center



Armstrong Flight Research Center



Glenn Research Center



Goddard Space Flight Center



Jet Propulsion Lab



Johnson Space Center



Kennedy Space Center



Langley Research Center



Marshall Space Flight Center



Stennis Space Center



White Sands Test Facility

<https://history.nasa.gov/>

What has NASA done for us?



A Workforce Innovating To Go Farther

Space Launch System/Multi-Purpose Crew Vehicle

NASA is innovating to continue American leadership in an ambitious space program that takes us farther than ever before. The Agency is supporting technology innovation and commercial partnerships that fuel the American economy. With the development of the newest generation of space transportation, the Space Launch System (SLS) and the Orion Multi-Purpose Crew Vehicle, NASA is developing capabilities to access deep space destinations like the Moon, an asteroid, and Mars.

This capability is being built right here in the United States. Thousands of workers in over 30 states are currently building our Nation's next generation of space travel.

An exploration flight test of Orion will take place in 2014 with a follow-on integrated Orion/SLS uncrewed flight in 2017.

A crewed flight of the integrated deep space system will occur as early as 2021.



Game-Changing Innovation

International Space Station—A Laboratory for Life Improvement Today

The United States has an active astronaut program. In fact, American human space flight occurs daily on board the International Space Station (ISS).

NASA is committed to maintaining American excellence in science, technology, engineering, and mathematics (STEM) while sustaining a long-term human presence in space. The ISS provides important technology and innovation development, not only for future long-term human space missions, but to enhance life on Earth today.

Outstanding advances in microbial vaccine development, unprecedented cancer research involving treatment delivery, and improved water purification technologies are just some of the technological advances that research on board the ISS has provided to life here on Earth.

Science

NASA is committed to groundbreaking science missions today and in the future. Our missions focus on providing critical data about our home planet, unraveling the mysteries of our universe, understanding solar variability and the impacts of space weather, and exploring the diverse planetary bodies of our solar system.

Aeronautics

Aeronautics research is leading the Nation to less congested airways and a more environmentally friendly aviation industry.



Investment in American Technology

By pushing the boundaries of technology, NASA's programs keep American business and workers on the cutting edge. Investing in technology development today enables future missions of NASA while growing the economy by creating new industries, jobs, products, and services. To ensure access to the most innovative tech development, NASA partners with American companies throughout the Nation in all 50 states.

\$12.5 Billion in FY 2011 Obligated to American Industry

Inspiring the Next Generation

Over the past year, NASA's K-12 education projects reached more than 1 million students through STEM programs and initiatives. The Summer of Innovation Program engaged over 45,000 students and over 5,500 educators through camps and activities in 46 states, the District of Columbia, and Puerto Rico.

\$941 Million in FY 2011 Obligated to Educational Institutions Throughout the Country

Investing in a New Aerospace Economy

Commercial Partners

NASA is committed to developing a new aerospace economy. The Agency is currently partnering with American industry to achieve safe and reliable access to low-Earth orbit and the ISS.



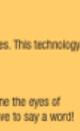
In fact, private industry began test flights in 2010, and these will continue through 2012. This progress will culminate with human access to the ISS by 2017.

Building a strong commercial space industry enables the development of a capability that will assure the Nation's future in space. Doing so will increase national revenue while positioning American companies as competitors in the global marketplace in the space economy.

NASA Technologies Americans Use Daily

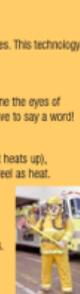
TV Satellite Dish

NASA developed ways to correct errors in the signals coming from spacecraft. This technology is used to reduce noise in TV signals coming from satellites.



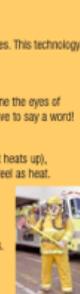
Medical Imaging

NASA developed ways to process signals from spacecraft to produce clearer images. This technology also makes photolike images of the insides of our bodies possible.



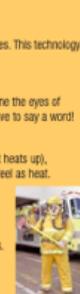
Vision Screening System

This technology uses techniques developed for processing space images to examine the eyes of children and find out quickly if they have any vision problems. The child doesn't have to say a word!



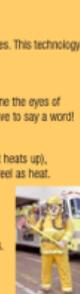
Ear Thermometer

Instead of measuring temperature using a column of mercury (which expands as it heats up), this thermometer has a lens like a camera and detects infrared energy, which we feel as heat. The warmer something is (like your body), the more infrared energy it puts out. This technology was originally developed to detect the birth of stars.



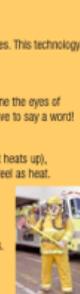
Firefighter Equipment

Firefighters wear suits made of fire-resistant fabric developed for use in spacesuits.



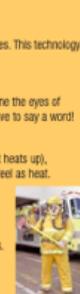
Smoke Detector

These devices were first used in the Earth-orbiting space station called Skylab (launched back in 1973) to help detect toxic vapors. Now they are used in most homes and other buildings to warn people of fire.



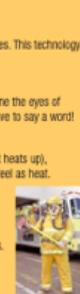
Automobile Design Tools

A computer program developed by NASA to analyze a spacecraft or airplane design and predict how parts will perform is now used to help design automobiles.



Thermal Gloves and Boots

Thermal gloves and boots have heating elements that run on rechargeable batteries worn on the inside wrist of the glove or embedded in the sole of the ski boot. This technology was adapted from a spacesuit design for the Apollo astronauts.



Advanced Plastics

Spacecraft and other electronics need very special, low-cost materials as the base for printed circuits (like those inside your computer). Some of these "liquid crystal polymers" have turned out to be very good, low-cost materials for making containers for food and beverages.

National Aeronautics and Space Administration

NASA
Reaching for New Heights

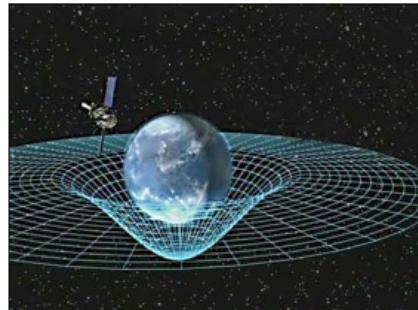
National Aeronautics and Space Administration
Headquarters
Office of Legislative and Intergovernmental Affairs
300 E Street SW
Washington, DC 20546-0001
Phone: 202-358-1055
Fax: 202-358-4340
<http://www.nasa.gov/offices/olia>

www.nasa.gov

What's in store for NASA?



1. Extend and sustain human activities across the solar system.



2. Expand scientific understanding of the Earth and the universe in which we live.



3. Create the innovative new space technologies for our exploration, science, and economic future.



4. Advance aeronautics research for societal benefit.



5. Enable program and institutional capabilities to conduct NASA's aeronautics and space activities.



6. Share NASA with the public, educators, and students to provide opportunities to participate in our mission, foster innovation and contribute to a strong national economy.

NASA SMD Science by the Numbers



RESEARCH

~10,000 U.S. Scientists Funded
~3,000 Competitively Selected Awards
~\$600M Awarded Annually

TECHNOLOGY DEVELOPMENT

~\$397M Invested Annually

EARTH-BASED OBSERVATIONS

23 Operating Missions
5 Upcoming Earth System Observatory Missions

MISSIONS

114 Missions from formulation through extended operations



Science by the
NUMBERS

SMALLSATS/ CUBESATS

49 Science Missions
29 Technology Demos

SOUNDING ROCKETS

2 Science Missions Launched
45 In Development

BALLOONS

54 Missions in Development

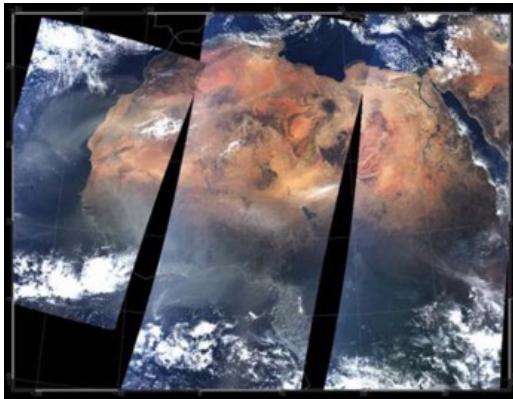
February 2021

<https://science.nasa.gov/about-us>



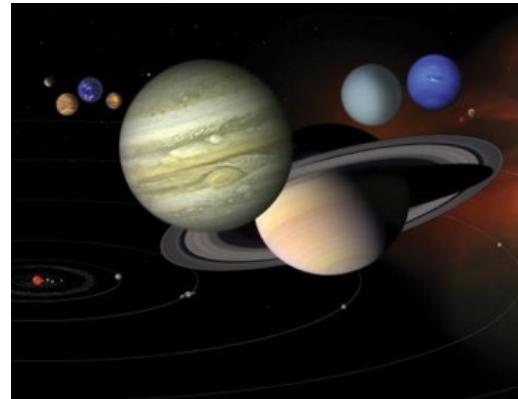
Earth Science

The study of planet Earth



Planetary Science

the study of the origin and history of the solar system and the potential for extraterrestrial life



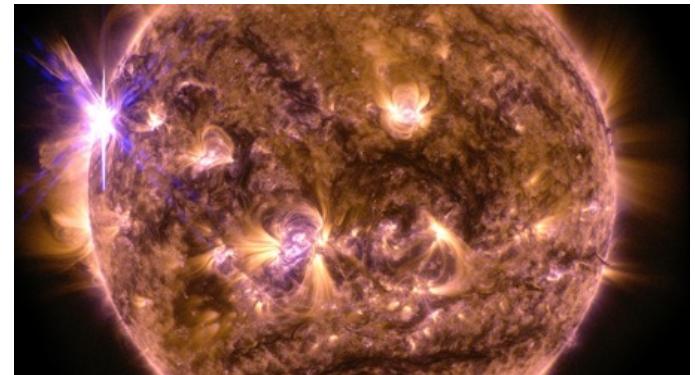
Astrophysics

The study of the origin, structure, evolution and destiny of the universe

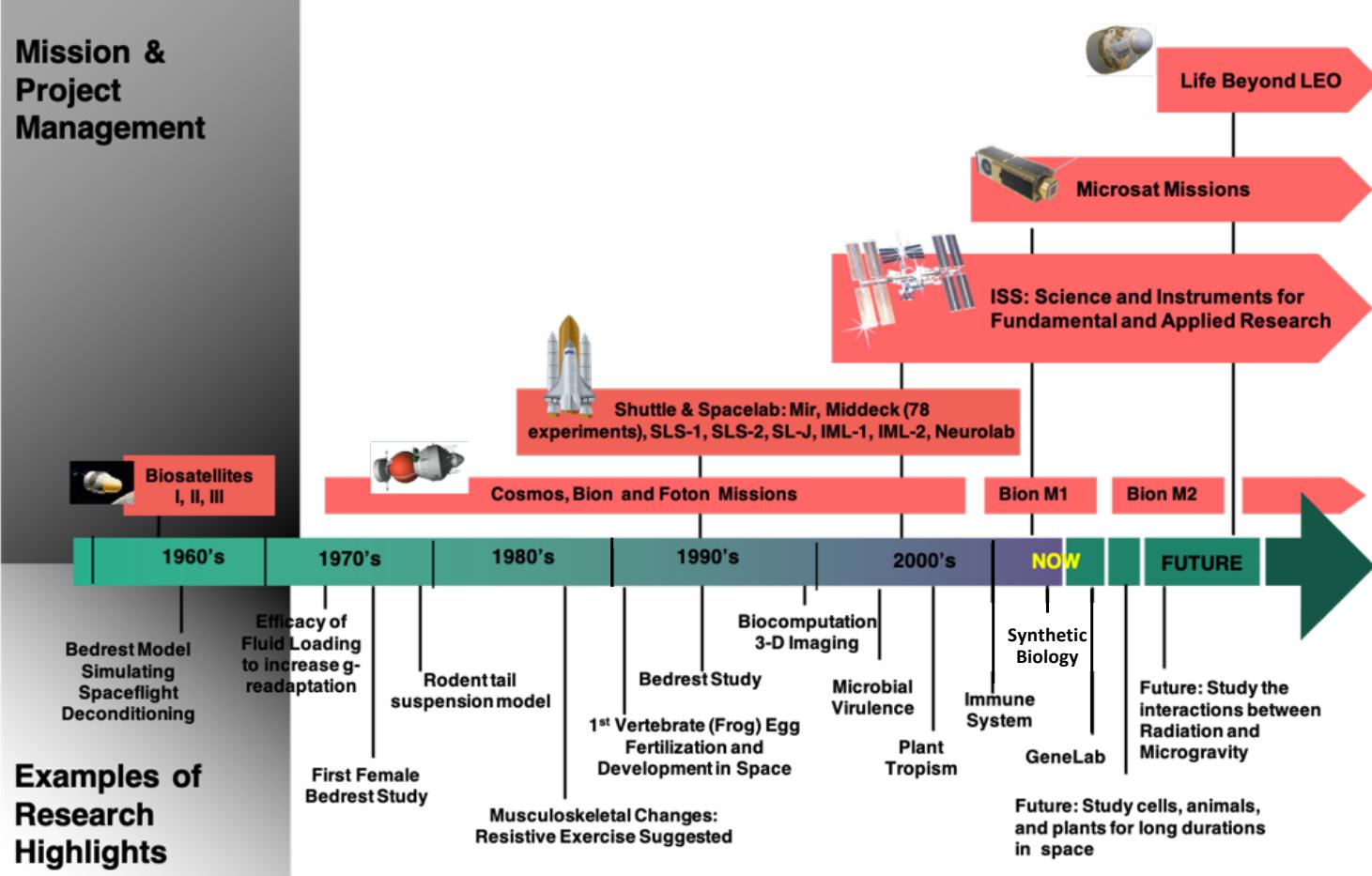
Monkey Head Nebulae

Heliophysics

The study of the sun and its effects on the solar system



5+ Decades of NASA Life Sciences Research



Exposure to Spaceflight Changes our Physiological Systems



The Musculoskeletal System

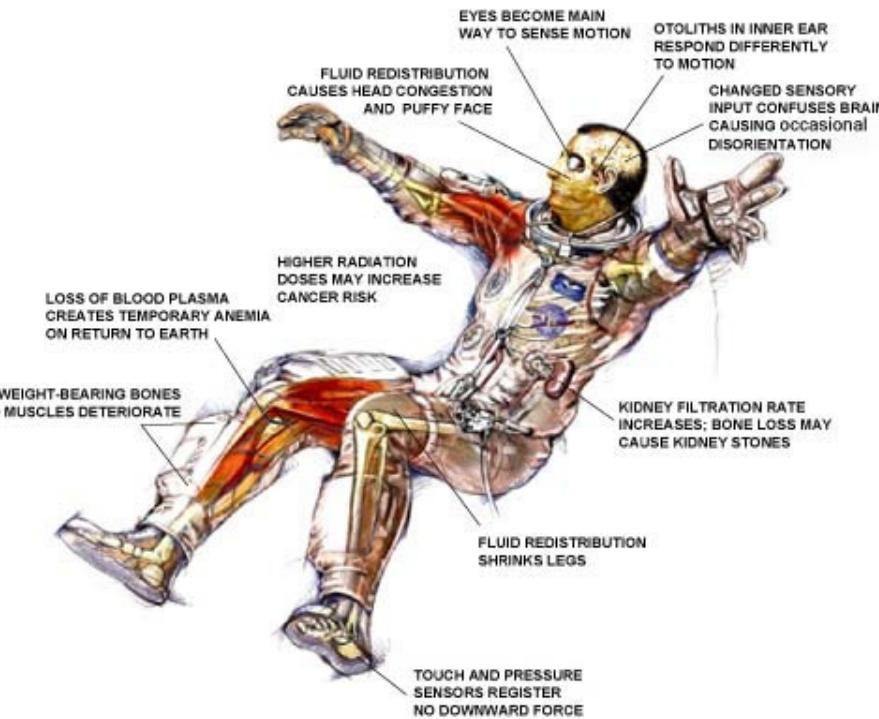
Osteoporosis
Osteoarthritis
Muscle atrophy

The Cardiovascular System

CVD
Atherosclerosis
Anemia

The Lymphatics

Lymphatic pump disorders



The Central Nervous System

Memory loss
Retinal edema

The Reproductive System

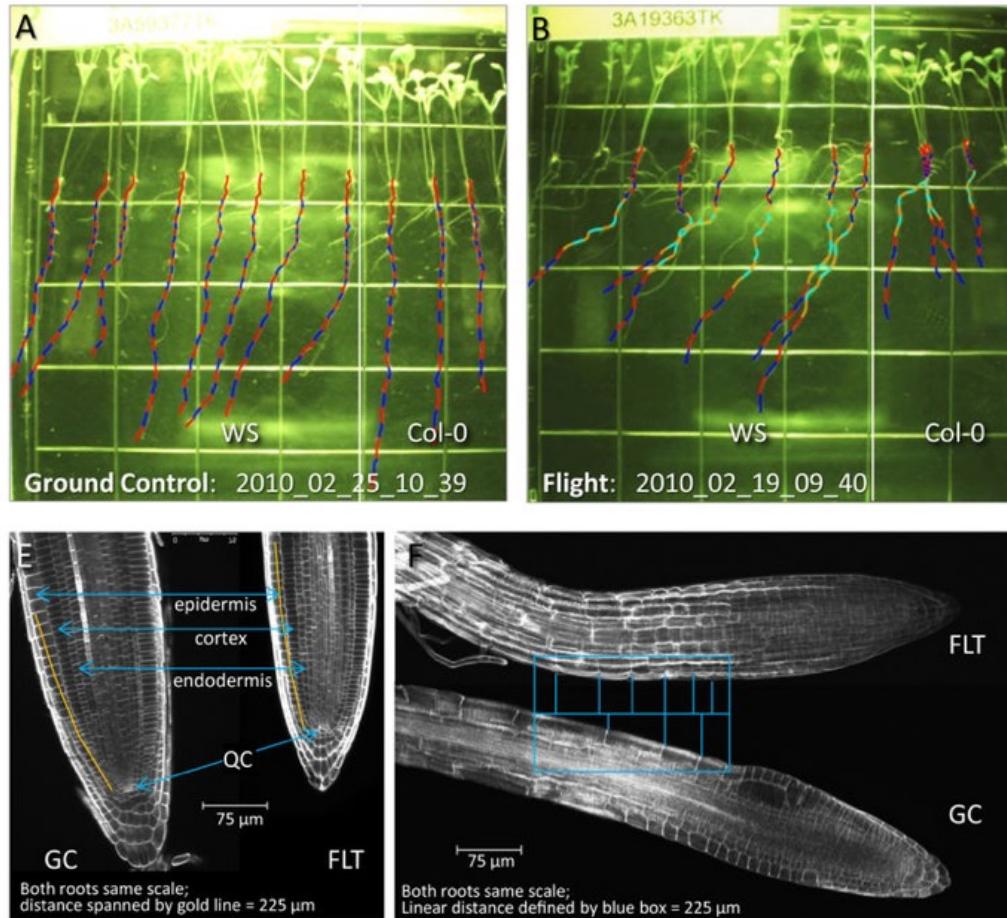
Sperm motility
Infertility
Hormonal imbalances

The Digestive System

Liver fibrosis
NAFLD
IBS
Microbiome

Astronaut image retrieved from <http://marinscienceseminar.com/space-travel-how-does-outer-space-affect-your-body/>

Exposure to Spaceflight Changes Plant Physiology Too



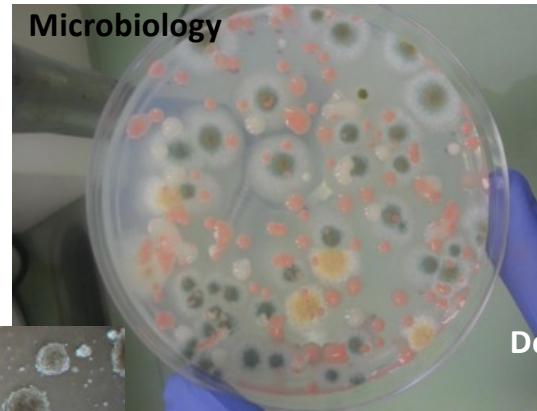
Space Biology Program



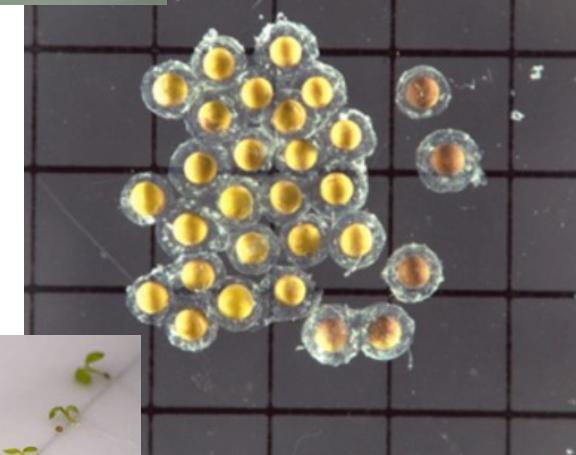
Animal Biology



Microbiology



**Developmental, Reproductive,
and Evolutionary Biology**



Plant Biology



<https://science.nasa.gov/biological-physical/programs/space-biology>

Spaceflight Hardware: Cell Culture Systems



**Bioculture
System**



CellCult



Spaceflight Hardware: Microbiology



Biological Research in Canisters (BRIC)



Fig. 1 BRIC-60 and BRIC-60M with petri dish configuration.

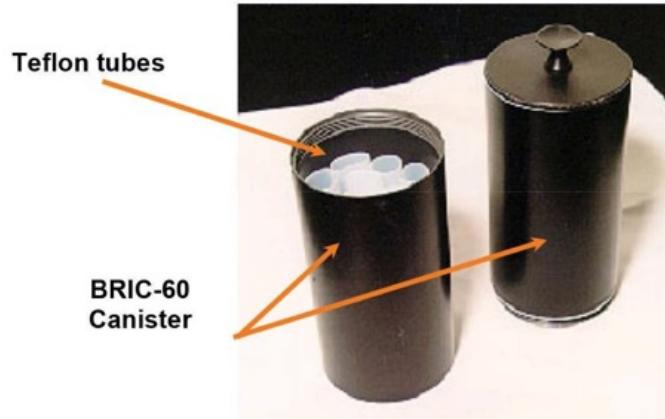


Fig. 2 BRIC-60 containing

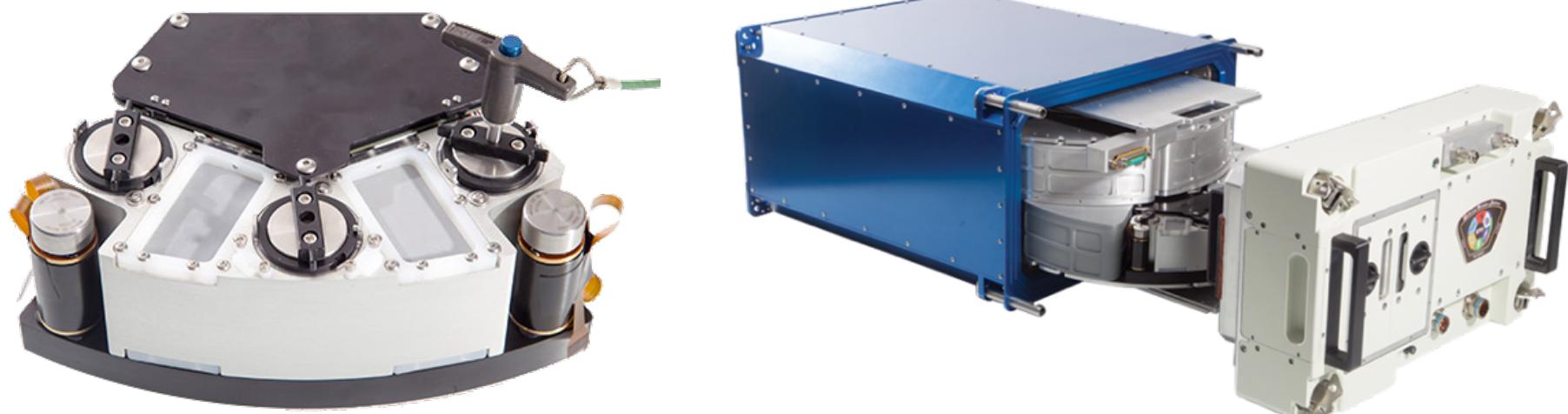


Fig. 3 BRIC-60 Canisters as flown

Spaceflight Hardware: Small Animals



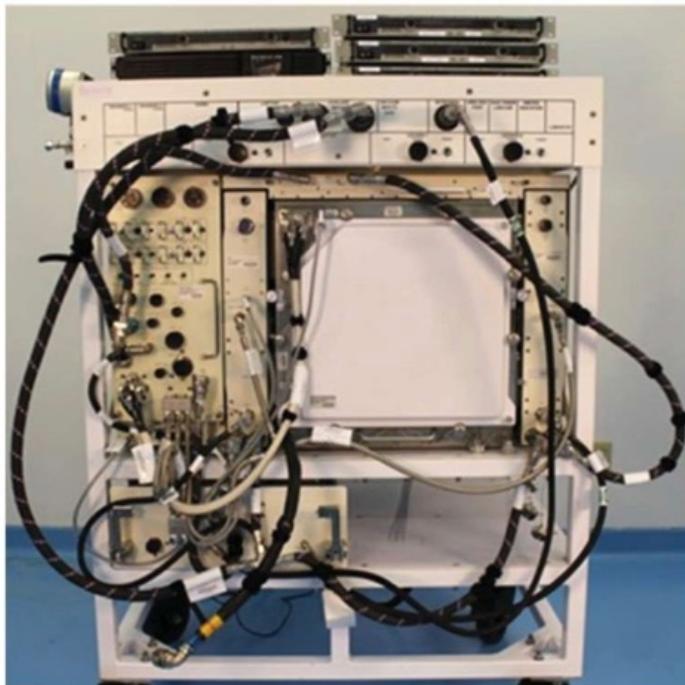
Multi-use Variable gravity Platform (MVP)



Spaceflight Hardware: Plants



Advanced Plant Habitat



APH Flight Unit in Ground Support Equipment



Dwarf Wheat Growing in APH

Spaceflight Hardware: Plants



Veggie Habitat

Veggie Configured for Growth of Lettuce



Veggie LED light bank.



Veggie plant pillow with quick disconnect fitting for adding water.



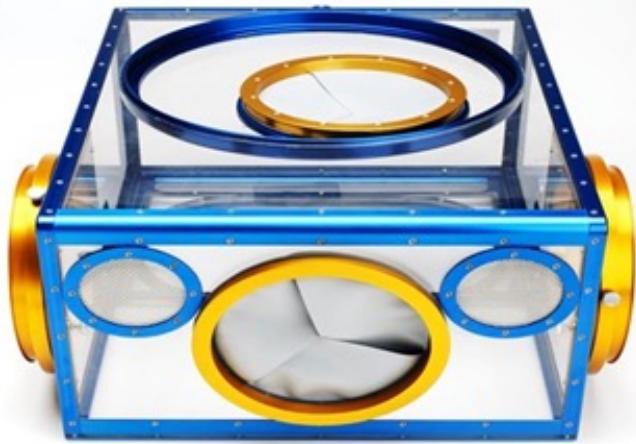
Veggie with 6 plant pillows contained within adjustable bellows.



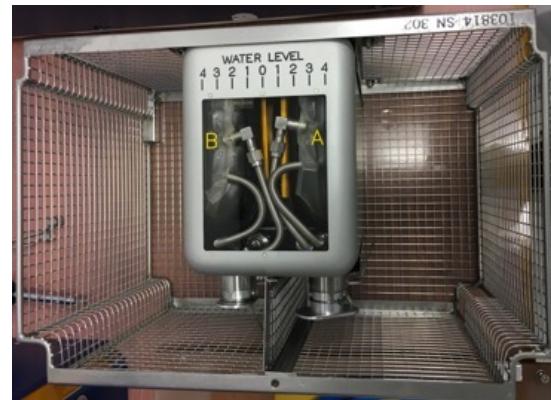
Spaceflight Hardware: Rodent Research



Animal Access Unit (AAU)



RR Habitat



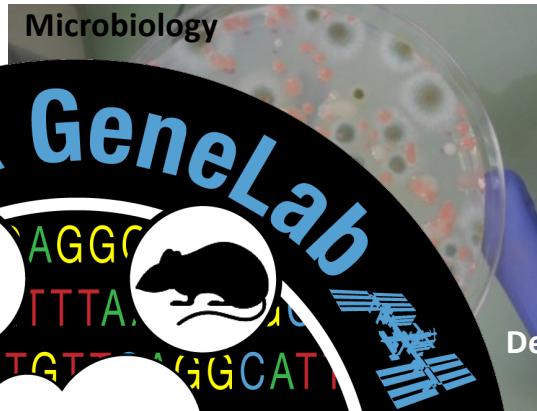
Space Biology Program



Animal Biology



Microbiology



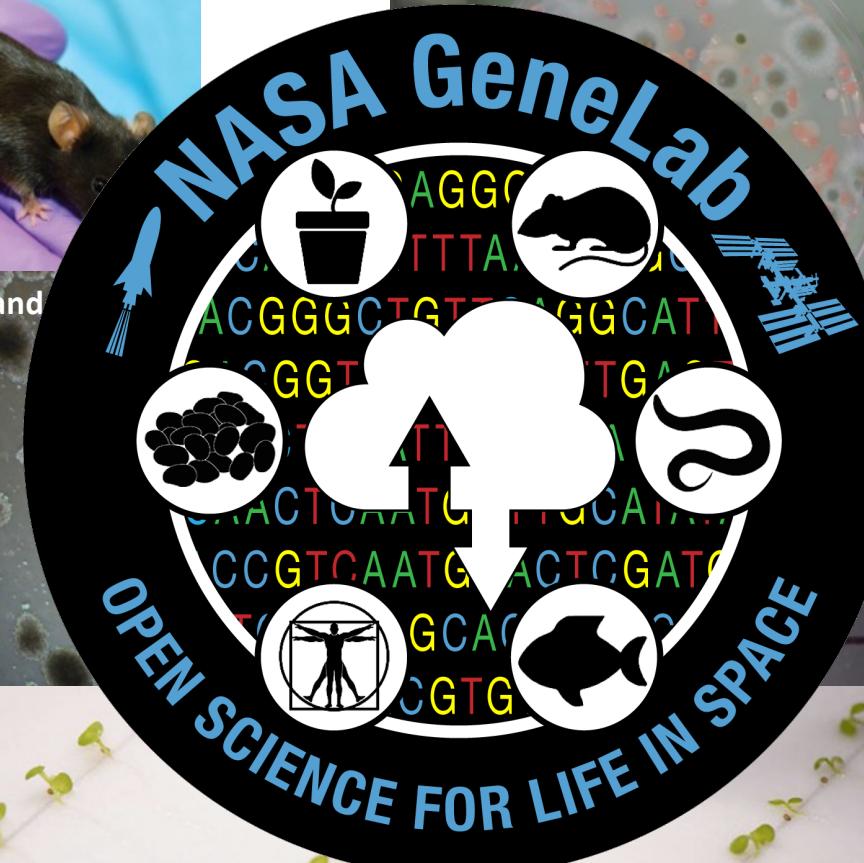
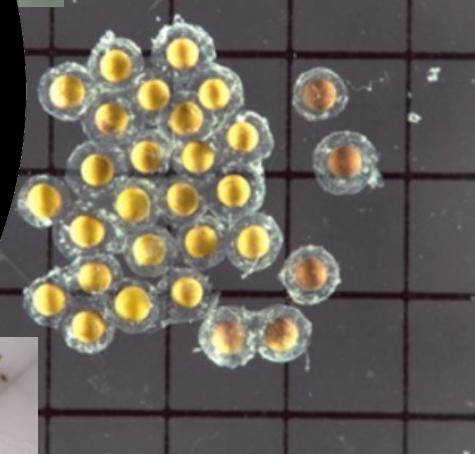
Cell and



Plant Biology

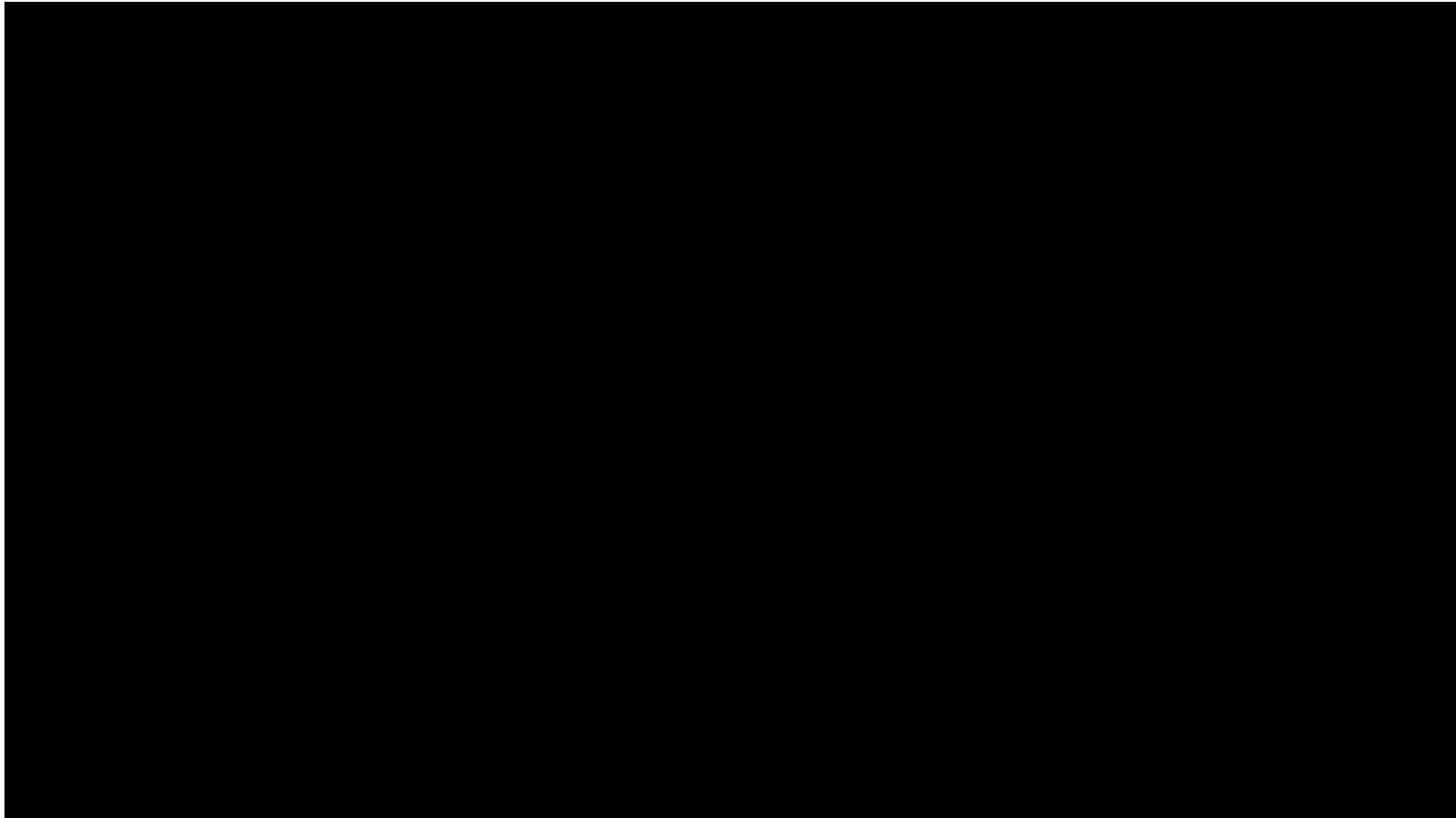


Developmental, Reproductive,
and Evolutionary Biology



<https://science.nasa.gov/biological-physical/programs/space-biology>

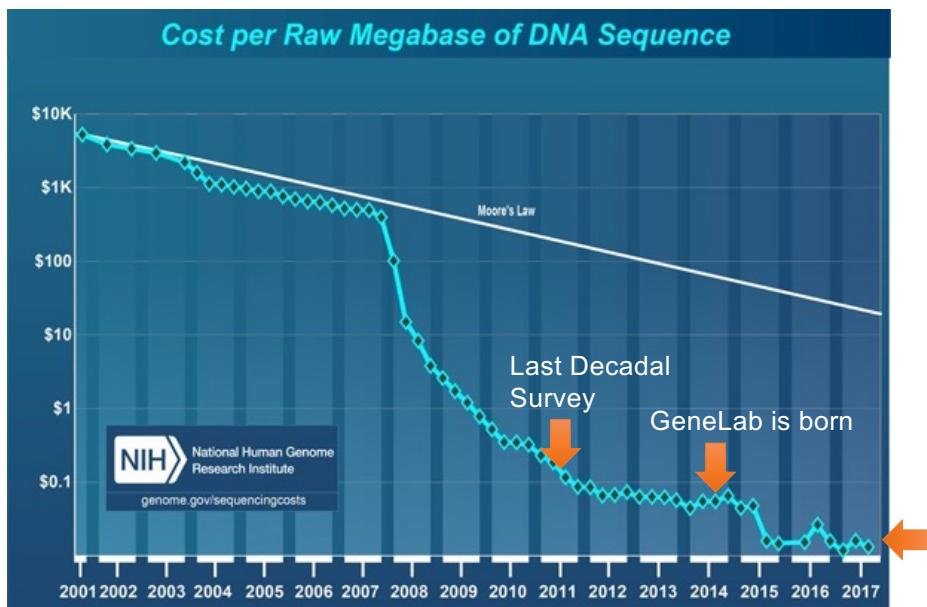
The Future of Space Exploration



https://www.nasa.gov/about/whats_next.html | <https://www.youtube.com/watch?v=mB1nAzriqRQ>

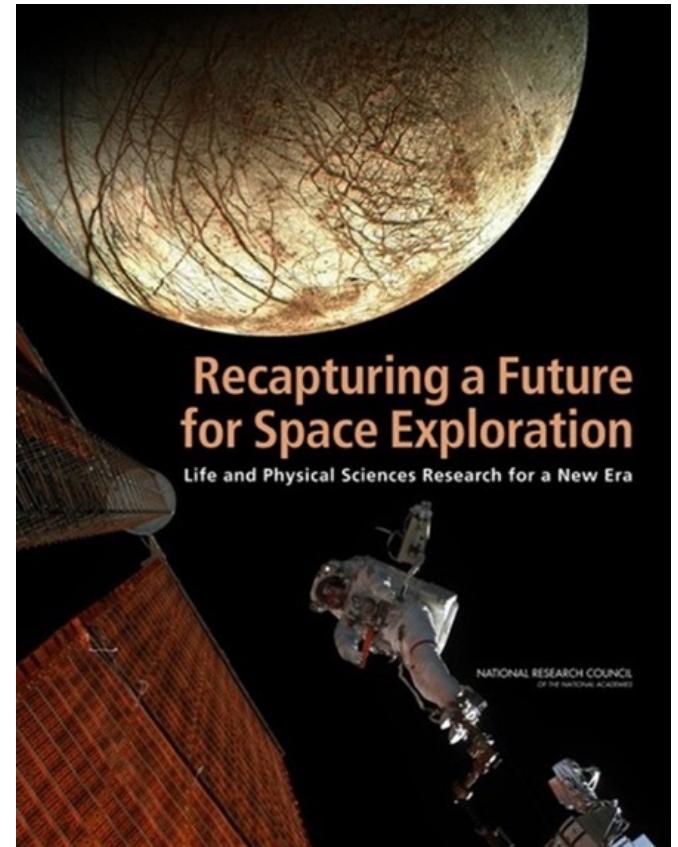
Introduction to GeneLab

2011 NRC Decadal Survey and the Sequencing Paradigm Shift

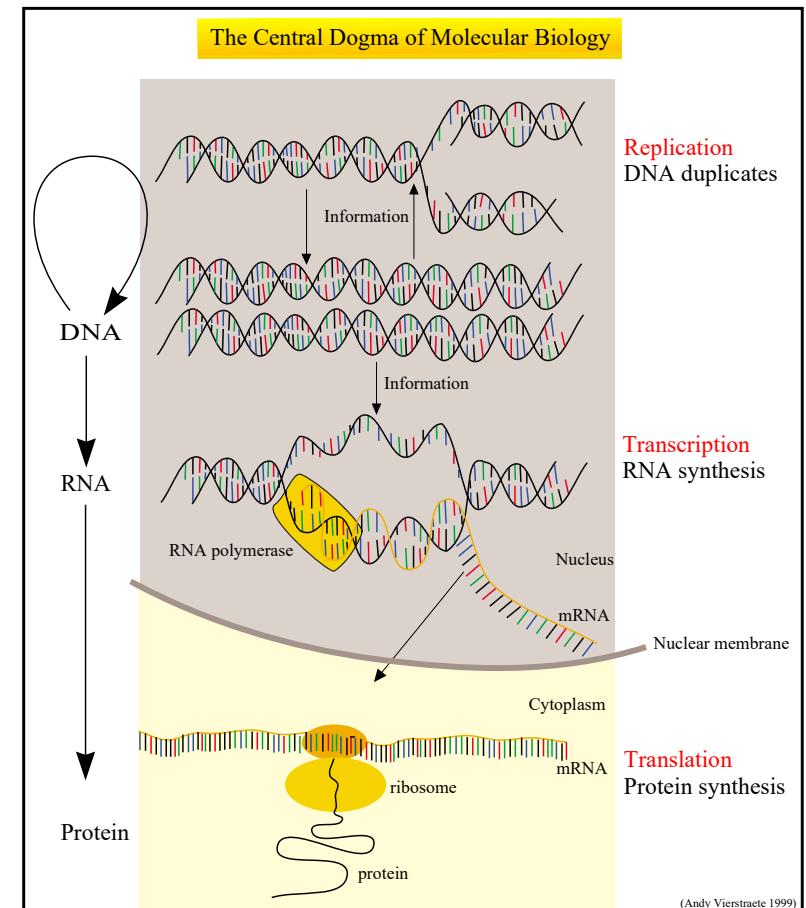
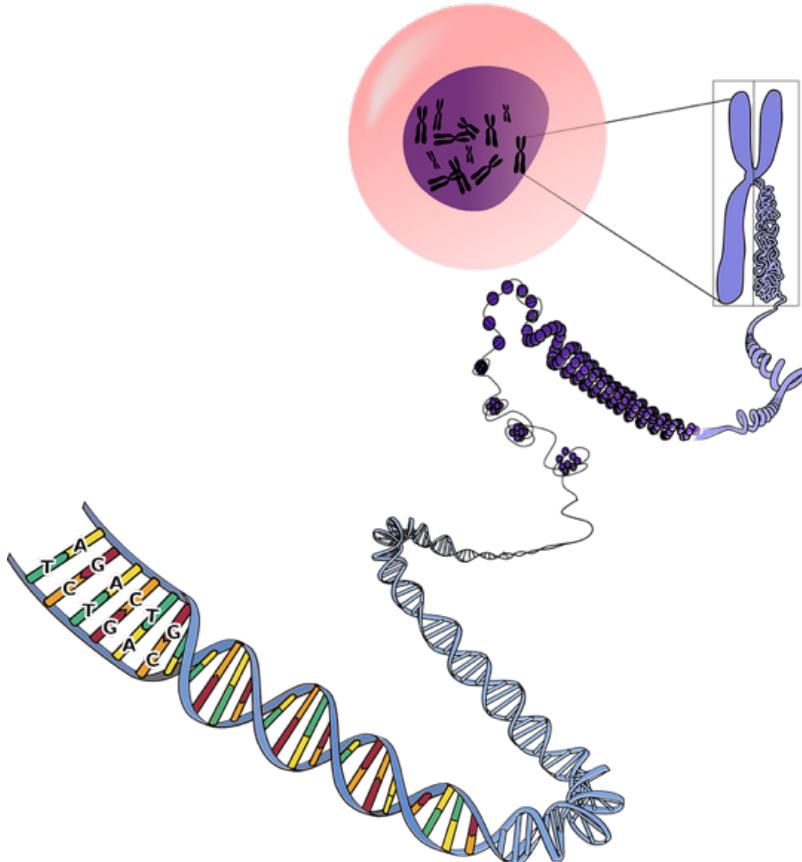


...genomics, transcriptomics, proteomics, and metabolomics offer an immense opportunity to understand the effects of spaceflight on biological systems...

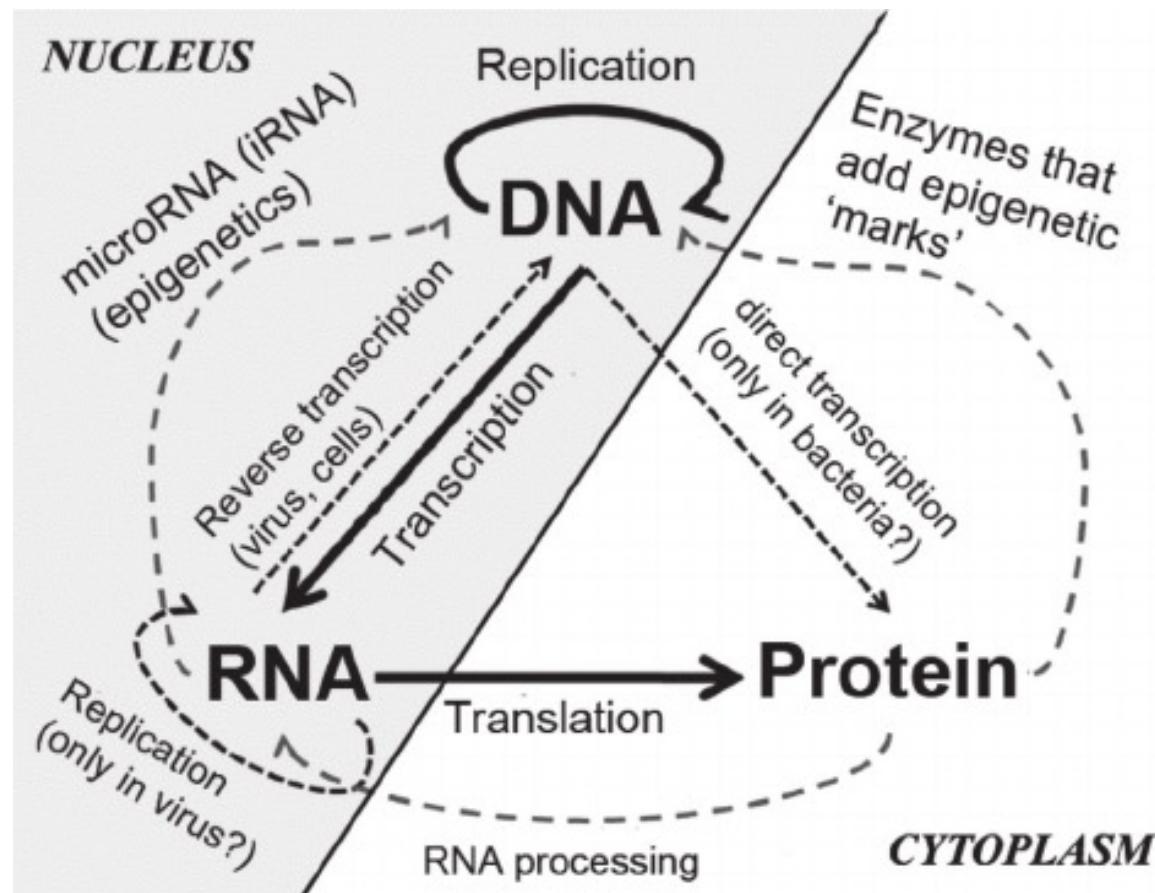
*...Such techniques generate considerable amounts of **data that can be mined and analyzed** for information by multiple researchers...*



What are 'omics?

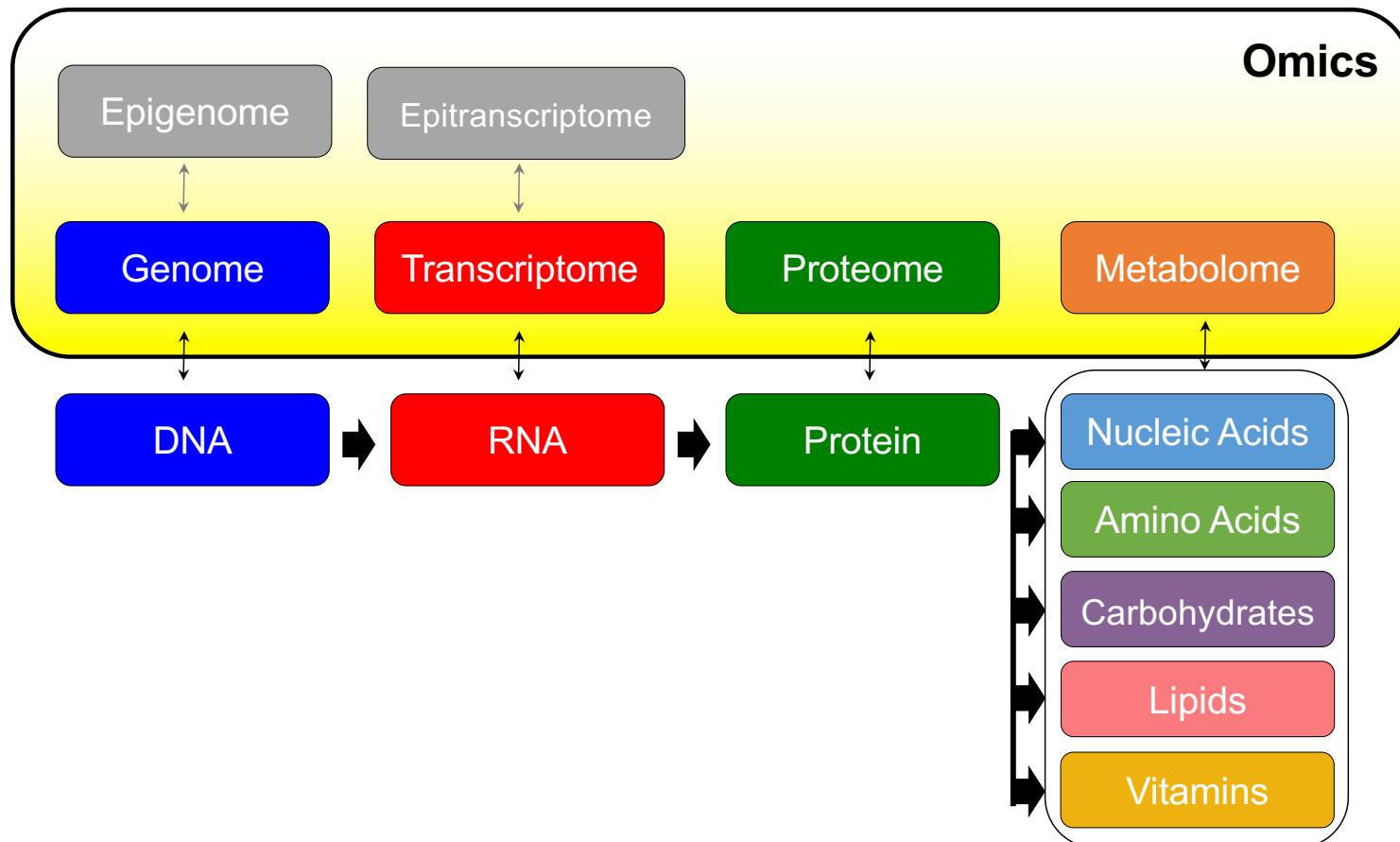


Beyond the Central Dogma



Gonzalez-Pardo and Alvarez, 2013

What are 'omics?

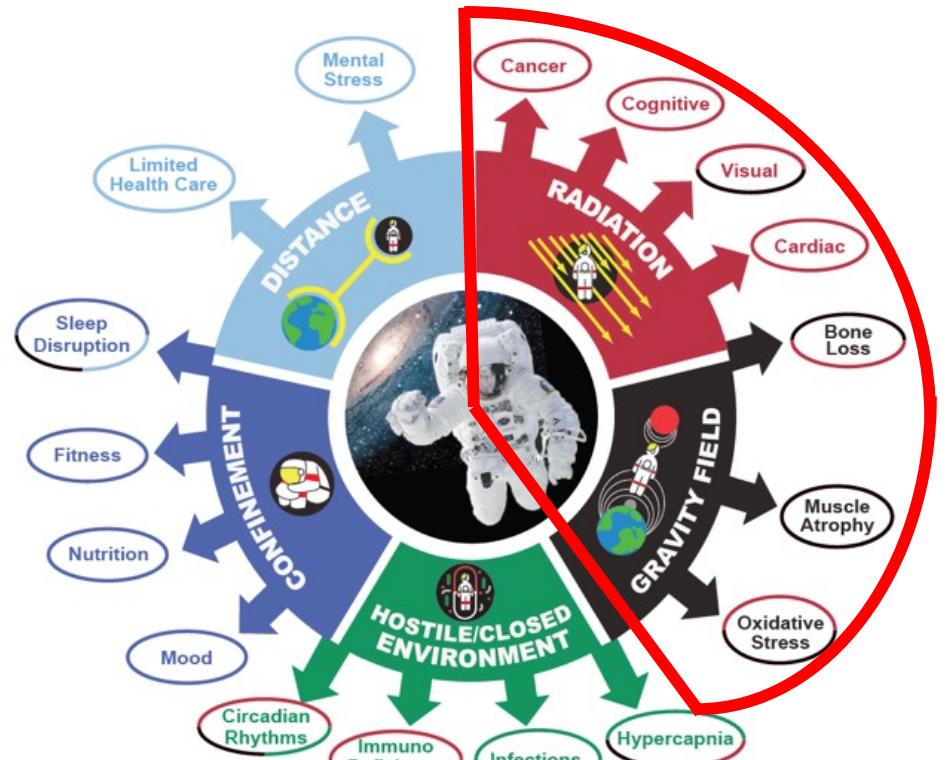


Why is studying omics important for spaceflight?



- What, when, and where genes are expressed allow for cell type diversity and enable living organisms to respond and adapt to surroundings
- Gene expression is primarily regulated by environmental factors both micro (cell's micro-environment) and macro (organism's external stimuli or stressors)
- Spaceflight alters the transcriptional patterns and molecular signaling networks within our cells, which in turn causes physiological changes
- Understanding such changes will enable development of mitigation strategies to better withstand the rigors of long-duration spaceflight

Primary Stressors of Spaceflight



GeneLab: Open Science for Life in Space (<https://genelab.nasa.gov>)



- Data Portal**
- First multi-omics space related database
 - 371 publicly available studies
 - High-order data for transcriptomics and metagenomics studies
 - Spaceflight environmental data associated with sample metadata (radiation dosimetry, temperature, humidity)
 - Interactive **visualization portal** to explore changes in gene expression



- Tools**
- User friendly submission portal to submit and publish data
 - Analysis platform to tools and workflows to analyze your own data or data from the GeneLab repository.
 - Tutorials and online resources to learn how to analyze RNA-Seq data
 - Workspace to store, share, and organize data files



- Sequencing**
- State-of-the-art sequencing facility to process spaceflight samples
 - Optimized SOPs with standard processing workflow
 - Provide sequencing service for NASA funded PI without any university overhead
 - Generate high quality data from shared tissues for open science access



- Community**
- Analysis Working Groups comprised of over 250+ scientists worldwide collaborating and analyze space omics data.
 - Education Working Group focused on providing resources to educators and students to learn about space biology and bioinformatics

GeneLab: Open Science for Life in Space (<https://genelab.nasa.gov>)



Open Science for Life in Space

[Home](#) [About](#) ▾ [Data & Tools](#) ▾ [Research & Resources](#) ▾ [Working Groups](#) ▾ [Help](#) ▾

Keywords



Welcome to NASA GeneLab - the first comprehensive space-related omics database; users can upload, download, share, store, and analyze spaceflight and spaceflight-relevant data from experiments using model organisms.

Data Repository

Search and upload spaceflight datasets

Analyze Data

Perform large-scale analysis of biological
omics data

Environmental Data

Radiation data collected during experiments
conducted in space

Collaborative Workspace

Share, organize and store files

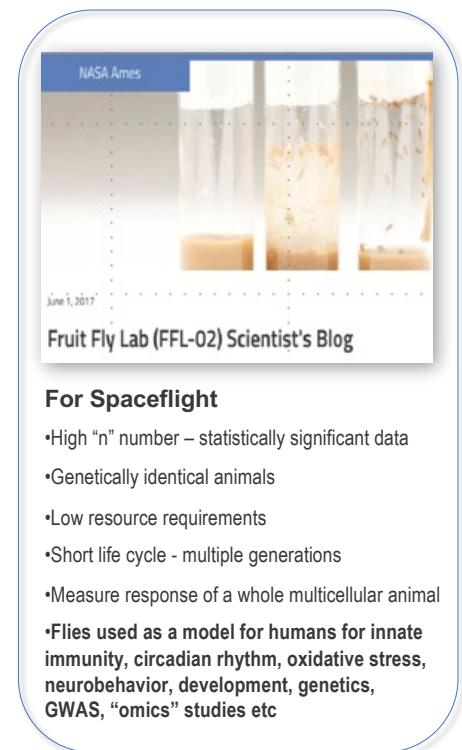
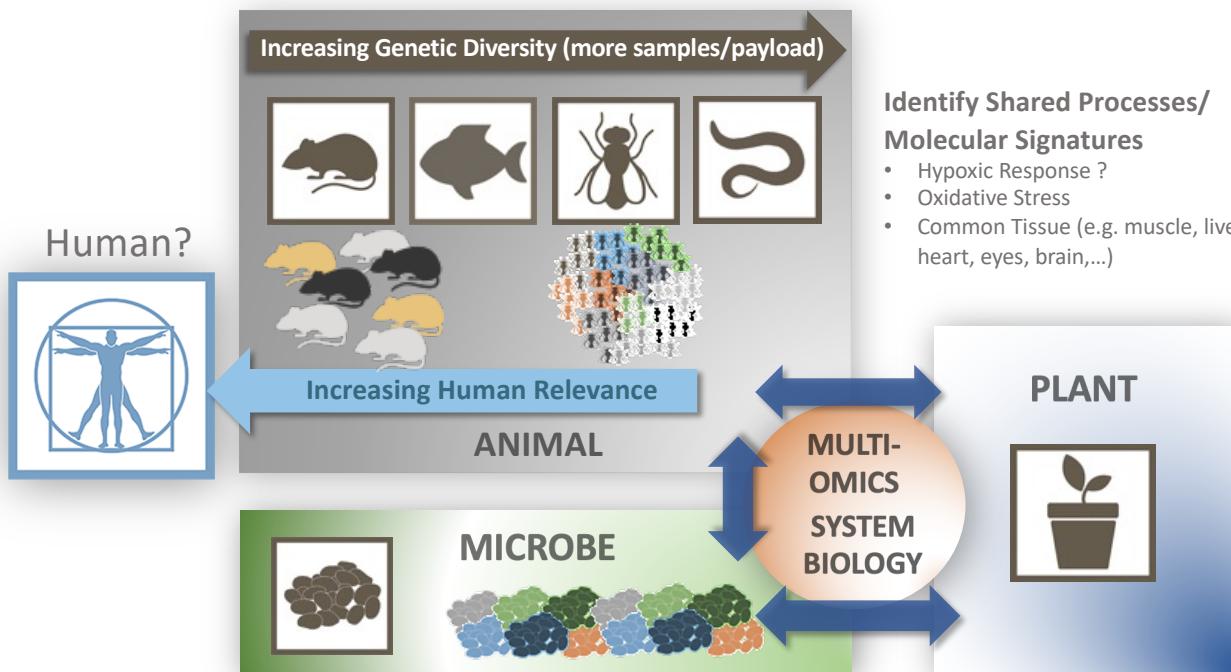
Submit Data

Have space-relevant data to submit
to GeneLab?

Visualize Data

Interact with GeneLab processed data

GeneLab ecosystem: Maximizing knowledge by bringing experiments together



NASA Biological Open Science Resources



Biospecimen Sharing Program (BSP)



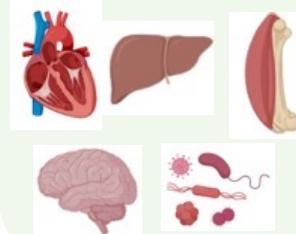
- *Dissects and preserves* rodent tissues from Flight and Ground investigations
- *Coordinates internal* tissue sharing



NASA Biological Institutional Scientific Collection (NBISC)



- *Collection of non-human* specimens and space microbial culture



Ames Life Sciences Data Archive (ALSDA)



- *Collects and curates* physiological, mission, project and imaging data



GeneLab (GL)



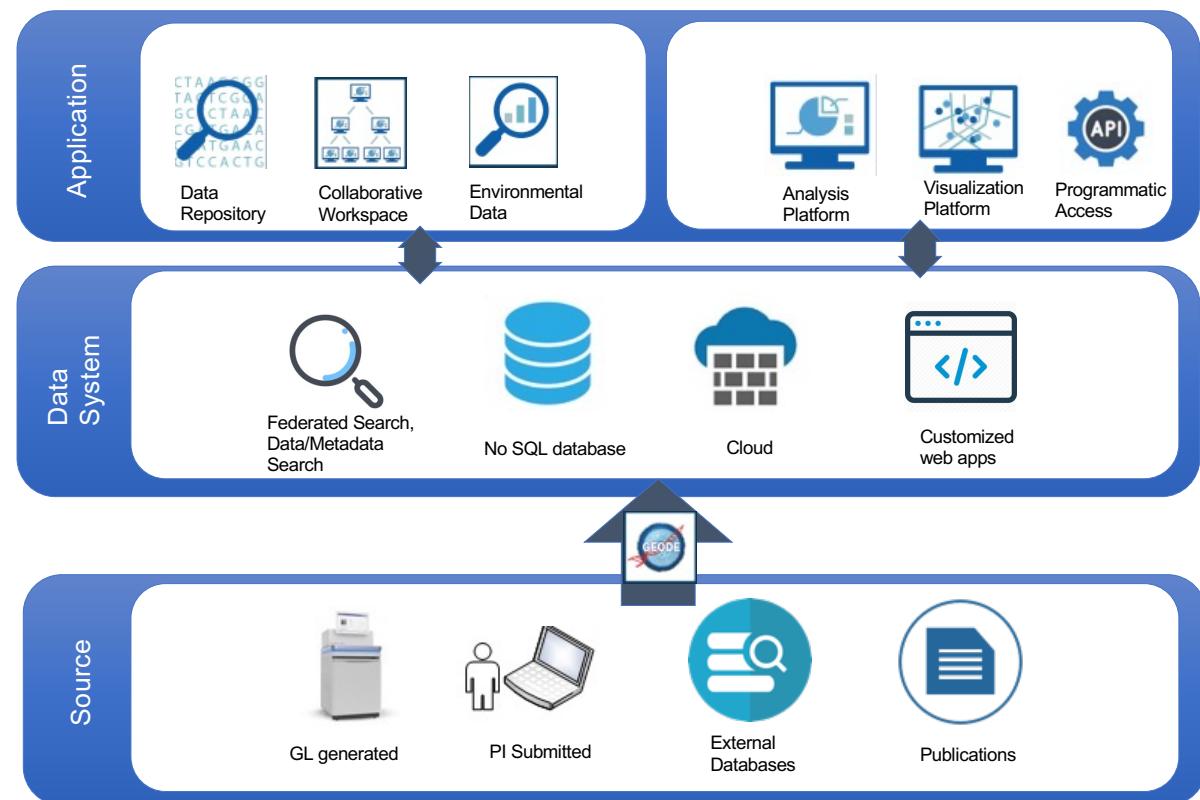
- *Collects and curates* omics data



GeneLab Data Systems



- Open access data
 - **FAIR (Findable, Accessible, Interoperable, Reusable)**
 - Controlled access tools
 - API - internal and external
 - User Friendly Interface
 - Tutorials
 - Self-service Submission Portal
-
- Federated search – GEO, PRIDE, MG-RAST, **ALSDA**
 - Database & Cloud – Scalable, easy access, fast
 - Web apps
 - Data Access & Management
 - Security
 - Operation
 - Governance and Integration
 - Open Source software – *no maintenance cost for software*
-
- Multiple data sources
 - Standard metadata organization
 - Open file formats



GeneLab: Submission Portal – GEODE (<https://genelab.nasa.gov>)



Open Science for Life in Space

[Home](#)

About ▾

Data & Tools ▾

Research & Resources ▾

Working Groups ▾

Help ▾

Keywords



Welcome to NASA GeneLab - the first comprehensive space-related omics database; users can upload, download, share, store, and analyze spaceflight and spaceflight-relevant data from experiments using model organisms.



Data Repository

Search and upload spaceflight datasets



Analyze Data

Perform large-scale analysis of biological
omics data



Environmental Data

Radiation data collected during experiments
conducted in space



Collaborative Workspace

Share, organize and store files



Submit Data

Have space-relevant data to submit
to GeneLab?



Visualize Data

Interact with GeneLab processed data

GeneLab Environment for Online Data Entry (GEODE)



Sample Name	Organism	Strain	Material Type	Protocol REF	Parameter Value		Unit	Ionizing Radiation	Time	Unit	Dose	Unit
					- Storage Temperature Setting							
Sample1	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Silicon	1	hour	1	gray	
Sample2	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Sham-irradiated	0	hour	0	gray	
Sample3	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Iron-56 Ion Irradiation	24	hour	2	gray	
Sample4	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Silicon	12	hour	3	gray	
Sample5	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Sham-irradiated	0	hour	0	gray	

Sample level data

- Space Biology PIs can now submit data through GEODE.
- Allows efficient input of sample and assay level metadata alongside data files.
- Tailored to specialized needs of spaceflight samples.
- Now exploring extending this portal for use on ALSDA submissions.

Sample Name	Protocol REF	Parameter Value - QA Instrument	Parameter Value - QA Assay	Parameter Value - QA Score	Unit	Extract Name	Protocol REF
Sample1	Nucleic Acid Extraction	BioAnalyzer	NanoChip	8	RIN	Extract 1	Library Construction
Sample2	Nucleic Acid Extraction	BioAnalyzer	NanoChip	9	RIN	Extract 2	Library Construction
Sample3	Nucleic Acid Extraction	BioAnalyzer	NanoChip	9.1	RIN	Extract 3	Library Construction

Assay level data with fields specific to that assay

The screenshot shows a tooltip for the unit 'gray'. It includes the definition: 'An absorbed dose unit which is equal to the absorption of one joule of radiation energy by one kilogram of matter.' Below the definition, there is another entry for 'gram' with its own definition: 'A mass unit which is equal to one thousandth of a kilogram or 10^-3 kg.'

Terms are connected to a controlled ontology

The screenshot displays a study preview for GLDS-528. It includes sections for Study Description, Contacts, Study Design Factors, Organisms, Assays, and Project. The 'Study Description' section provides a brief overview of the experiment. The 'Contacts' section lists a contact named Sam Odore. The 'Study Design Factors' section details the factors: 'Irradiation' (Type: X-ray), 'Time' (Value: 0.1, 1, 3, 10, 12, 24 hours), and 'Dose' (Value: Absorbed Radiation Dose). The 'Organisms' section lists 'Mus musculus'. The 'Assays' section shows 'transcription profiling' using 'Illumina Sequencing (RNA-Seq)' technology. The 'Project' section includes 'Project Identifier: JBR-22', 'Project Title: Transcription profiling of radiation effects on mouse tissues', and 'Project Type: Ground Study'.

PIs are presented with a study preview and a private link for sharing with reviewers and collaborators.

GeneLab: Data Repository (<https://genelab.nasa.gov>)



Open Science for Life in Space

[Home](#)

About ▾

Data & Tools ▾

Research & Resources ▾

Working Groups ▾

Help ▾

Keywords



Welcome to NASA GeneLab - the first comprehensive space-related omics database; users can upload, download, share, store, and analyze spaceflight and spaceflight-relevant data from experiments using model organisms.



Data Repository

Search and upload spaceflight datasets



Analyze Data

Perform large-scale analysis of biological
omics data



Environmental Data

Radiation data collected during experiments
conducted in space



Collaborative Workspace

Share, organize and store files



Submit Data

Have space-relevant data to submit
to GeneLab?



Visualize Data

Interact with GeneLab processed data

Search Data x Q All GeneLab NIH GEO EBI PRIDE ANL MG-RAST

Search Filters (GeneLab Only)

Show Only:

 Studies With VisualizationsPage 1 of 12 (Total Studies: 278) [Next >](#)Studies Per Page: **Comparative RNA-Seq transcriptome analyses reveal dynamic time dependent effects of 56Fe, 16O, and 28Si irradiation on the induction of murine hepatocellular carcinoma**

Organisms	Factors	Assay Types	Release Date	Description
Mus musculus	Ionizing Radiation Time	transcription profiling	27-Aug-2020	One of the health risks posed to astronauts during deep space flights is exposure to high charge, high-energy (HZE) ions ($Z>13$) which can lead to induction of hepatocellular carcinoma (HCC). We perfor...

**Gamma radiation and HZE treatment of seedlings in Arabidopsis**

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Genotype Time of sample collection after treatment Ionizing Radiation	transcription profiling	18-Sep-2014	Plants exhibit a robust transcriptional response to gamma radiation which includes the induction of transcripts required for homologous recombination and the suppression of transcripts that promote ce...

GeneLab: Data Repository



GLDS-137

Version 4

DESCRIPTION

PROTOCOLS

SAMPLES

ASSAYS

PUBLICATIONS

STUDY FILES

VISUALIZATION

Search Data

GLDS-137: Rodent Research-3-CASIS: Mouse liver transcriptomic, proteomic, and epigenomic data Version 4

Select a Version: 4

DOI: 10.26030/9k6w-4c28
Source Accession Number
Total Data Volume: 2.5 TB

Submitted Date: 04-Aug-2017
Release Date: 28-Aug-2017

DESCRIPTION

Study Description

The Rodent Research-3 (RR-3) mission was sponsored by the pharmaceutical company Eli Lilly and Co. and the Center for the Advancement of Science in Space to study the effectiveness of a potential countermeasure for the loss of muscle and bone mass that occurs during spaceflight. Twenty BALB/c, 18-weeks old female mice (ten controls and ten treated) were flown to the ISS and housed in the Rodent Habitat for 39-42 days. Twenty mice of similar age, sex and strain were used for ground controls housed in identical hardware and matching ISS environmental conditions. Basal controls were housed in standard vivarium cages. Spaceflight, ground controls and basal groups had blood collected, then were euthanized, had one hind limb removed, and finally whole carcasses were stored at -80 C until dissection. All mice in this data set received only the control/sham injection.

Contacts	NAME	ROLE	ORGANIZATION	E-MAIL
Ruth Globus	Project Scientist	NASA Ames Research Center	ruth.k.globus@nasa.gov	
Jonathan Galazka	Project Scientist	NASA GeneLab	jonathan.m.galazka@nasa.gov	
Rosamund Smith	Principal Investigator	Eli Lilly and Company		



GLDS-137

Version 4

DESCRIPTION

PROTOCOLS

Mission	MISSION NAME	START DATE	END DATE
	SpaceX-8	08-Apr-2016	11-May-2016

Acknowledgments

Funding for sample processing and sequencing was provided to the GeneLab project by the NASA Space Biology Program Office, Space Life and Physical Sciences Research and Applications Division. Samples from the RR-3 experiment were provided to GeneLab through the Biospecimen Sharing Plan of the NASA Life Sciences Data Archive.

Data Citation

Smith R, Cramer M, Globus R, Galazka J. "Rodent Research-3-CASIS: Mouse liver transcriptomic, proteomic, and epigenomic data", GeneLab, Version 4, <http://doi.org/10.26030/9k6w-4c28>

[Download Citation \(BibTex\)](#)

[Download Citation \(RIS\)](#)



GLDS-137

Version 4

DESCRIPTION

PROTOCOLS

SAMPLES

ASSAYS

PUBLICATIONS

STUDY FILES

VISUALIZATION

Select Export Columns

Select columns of the samples panel below to export [Export CSV](#)

<input checked="" type="checkbox"/> Source Name	<input checked="" type="checkbox"/> Sample Name	<input checked="" type="checkbox"/> Protocol REF	<input checked="" type="checkbox"/> Comment: Source description
<input checked="" type="checkbox"/> Characteristics: Organism	<input checked="" type="checkbox"/> Characteristics: Strain	<input checked="" type="checkbox"/> Comment: Animal Source	<input checked="" type="checkbox"/> Characteristics: sex
<input checked="" type="checkbox"/> Characteristics: Age at Launch	<input checked="" type="checkbox"/> Characteristics: Diet	<input checked="" type="checkbox"/> Comment: Feeding Schedule	<input checked="" type="checkbox"/> Parameter Value: Habitat
<input checked="" type="checkbox"/> Factor Value: Spaceflight	<input checked="" type="checkbox"/> Parameter Value: treatment	<input checked="" type="checkbox"/> Parameter Value: Duration	<input checked="" type="checkbox"/> Parameter Value: Euthanasia Chemical
<input checked="" type="checkbox"/> Parameter Value: Carcass Weight	<input checked="" type="checkbox"/> Parameter Value: Source Storage Temperature	<input checked="" type="checkbox"/> Comment: Initiation Date	<input checked="" type="checkbox"/> Comment: Completion Date
<input checked="" type="checkbox"/> Comment: Dissection date	<input checked="" type="checkbox"/> Protocol REF	<input checked="" type="checkbox"/> Comment: Source description	<input checked="" type="checkbox"/> Material Type
<input checked="" type="checkbox"/> Comment: Original Submitted Sample Name	<input checked="" type="checkbox"/> Comment: Sample Name Description	<input checked="" type="checkbox"/> Comment: LSDA Biospecimen Source ID	<input checked="" type="checkbox"/> Comment: LSDA Biospecimen Subject ID
<input checked="" type="checkbox"/> Parameter Value: Sample Preservation Method	<input checked="" type="checkbox"/> Parameter Value: Sample Storage Temperature		

[close](#) [select all](#) [unselect all](#)

SAMPLES							
Source Name	Sample Name	Protocol REF	Comment: Source description	Characteristics: Organism	Characteristics: Strain	Comment: Animal Source	
CB1	Mimus_BAL-TAL_LVR_BSL_Rep1_B1	mouse habituation	previously dissected carcass	Mus musculus	BALB/c	Taconic Biosciences	
CB2	Mimus_BAL-TAL_LVR_BSL_Rep2_B2	mouse habituation	previously dissected carcass	Mus musculus	BALB/c	Taconic Biosciences	

GeneLab: Data Repository



STUDY FILES

To view files, click on the folder of interest.

- ◀ All Files
 - RNA-Seq Data
 - Proteomics Data Files
 - Whole Genome Bisulfite Sequencing Data
 - GeneLab Processed RNA-Seq Files
 - Aligned sequence data
 - Trimmed sequence data/FastQC reports
 - Merged sequence data/FastQC reports
 - Aligned sequence data/Alignment logs
 - Raw counts data
 - Trimmed sequence data/MultiQC report
 - Differential expression analysis data
 - Normalized counts data
 - Merged sequence data/MultiQC report
 - Trimmed sequence data
 - Merged sequence data
 - Trimmed sequence data/Trimming reports
 - Study Metadata Files

4 files selected

Download Selected Files

FILES	FILE SIZE	RESOURCE CATEGORY	RESOURCE DESCRIPTION
<input checked="" type="checkbox"/> GLDS-137_ma_seq_Mmus_BAL-TAL_LVR_FLT_Rep5_F5_SJ.out.tab	7.26 MB	Aligned sequence data	Data processed by the GeneLab Team, providing users high order data using standardized methods.
<input checked="" type="checkbox"/> GLDS-137_ma_seq_Mmus_BAL-TAL_LVR_BSL_Rep4_B4_R1_trimmed_fastqc.zip	433.13 KB	Trimmed sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.
<input type="checkbox"/> GLDS-137_ma_seq_Mmus_BAL-TAL_LVR_GC_Rep4_G5_R2_raw_fastqc.zip	456.51 KB	Merged sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.
<input checked="" type="checkbox"/> GLDS-137_ma_seq_Mmus_BAL-TAL_LVR_FLT_Rep2_F2_R2_raw_fastqc.html	636.29 KB	Merged sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.
<input checked="" type="checkbox"/> GLDS-137_ma_seq_Mmus_BAL-TAL_LVR_GC_Rep3_G3_R1_trimmed_fastqc.zip	434.84 KB	Trimmed sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.
<input type="checkbox"/> GLDS-137_ma_seq_Mmus_BAL-TAL_LVR_GC_Rep1_G1_R1_raw_fastqc.zip	449.76 KB	Merged sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.

GeneLab: Data Repository



371

Studies

422

Datasets

45

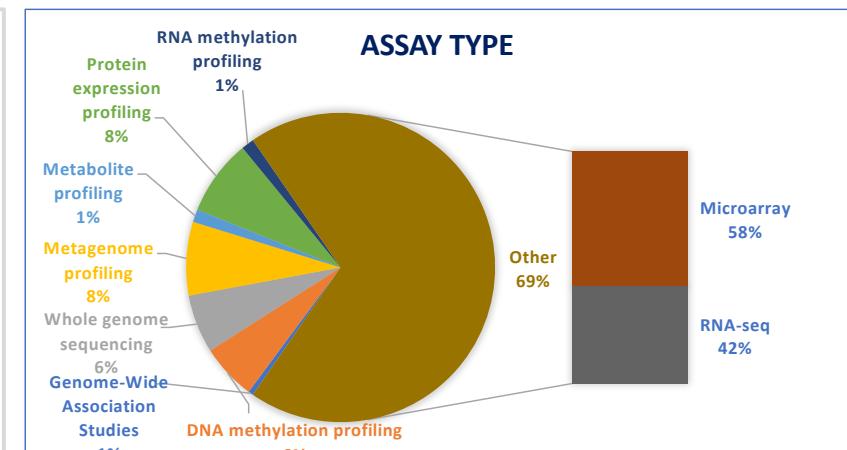
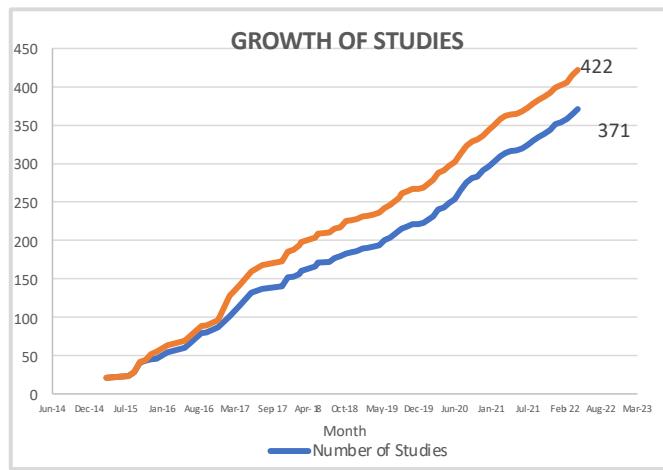
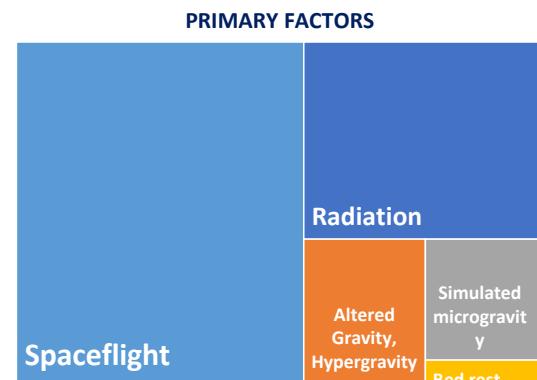
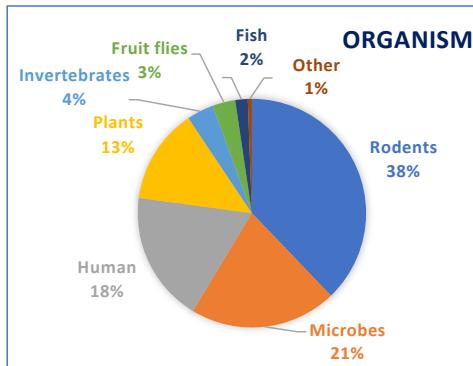
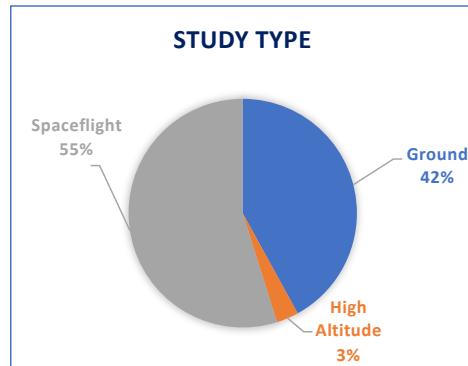
Species

>10

Assays

>135TB

Data



GeneLab: Environmental Data (<https://genelab.nasa.gov>)



Open Science for Life in Space

[Home](#)

About ▾

Data & Tools ▾

Research & Resources ▾

Working Groups ▾

Help ▾

Keywords



Welcome to NASA GeneLab - the first comprehensive space-related omics database; users can upload, download, share, store, and analyze spaceflight and spaceflight-relevant data from experiments using model organisms.



Data Repository

Search and upload spaceflight datasets



Analyze Data

Perform large-scale analysis of biological
omics data



Environmental Data

Radiation data collected during experiments
conducted in space



Collaborative Workspace

Share, organize and store files



Submit Data

Have space-relevant data to submit
to GeneLab?



Visualize Data

Interact with GeneLab processed data

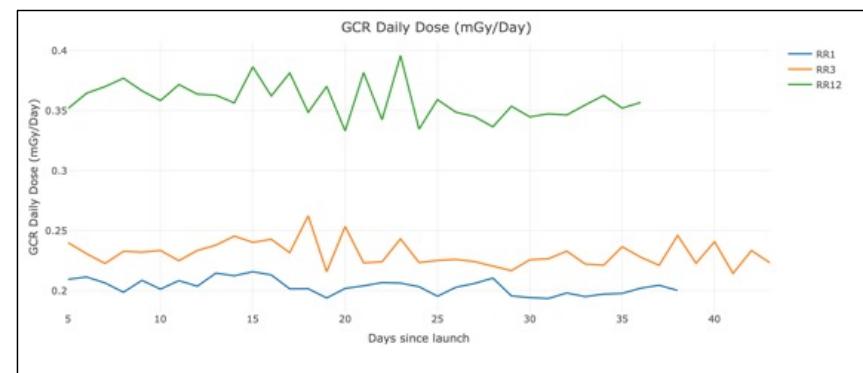
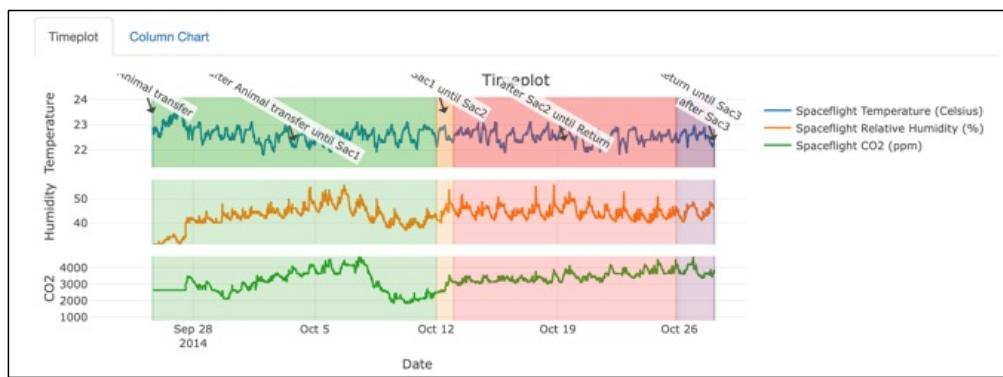
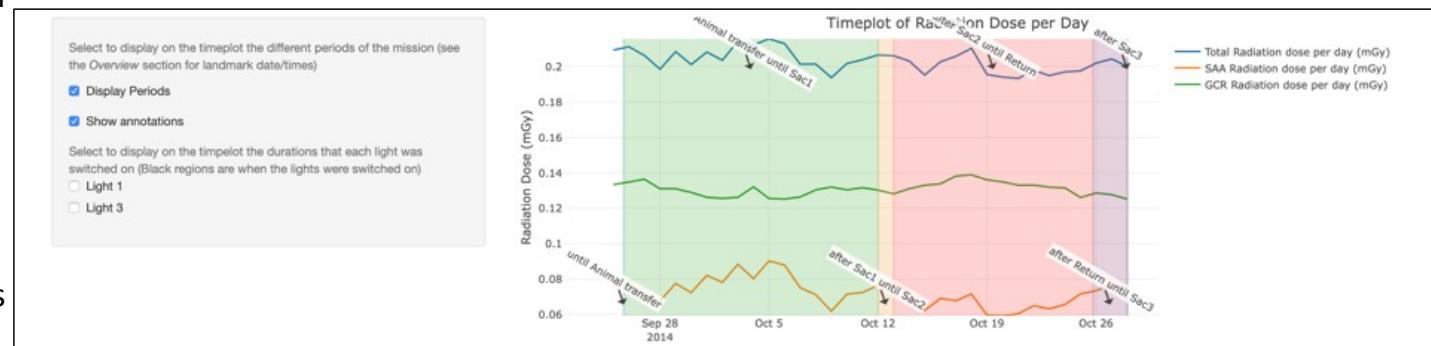
GeneLab: Environmental Data



Users can view environmental data relevant to a dataset.

This includes temperature, relative humidity, CO₂, radiation exposure.

Users can compare conditions of different datasets.



GeneLab: Data Visualization Portal (<https://genelab.nasa.gov>)



Open Science for Life in Space

[Home](#)

About ▾

Data & Tools ▾

Research & Resources ▾

Working Groups ▾

Help ▾

Keywords



Welcome to NASA GeneLab - the first comprehensive space-related omics database; users can upload, download, share, store, and analyze spaceflight and spaceflight-relevant data from experiments using model organisms.



Data Repository

Search and upload spaceflight datasets



Analyze Data

Perform large-scale analysis of biological
omics data



Environmental Data

Radiation data collected during experiments
conducted in space



Collaborative Workspace

Share, organize and store files



Submit Data

Have space-relevant data to submit
to GeneLab?



Visualize Data

Interact with GeneLab processed data



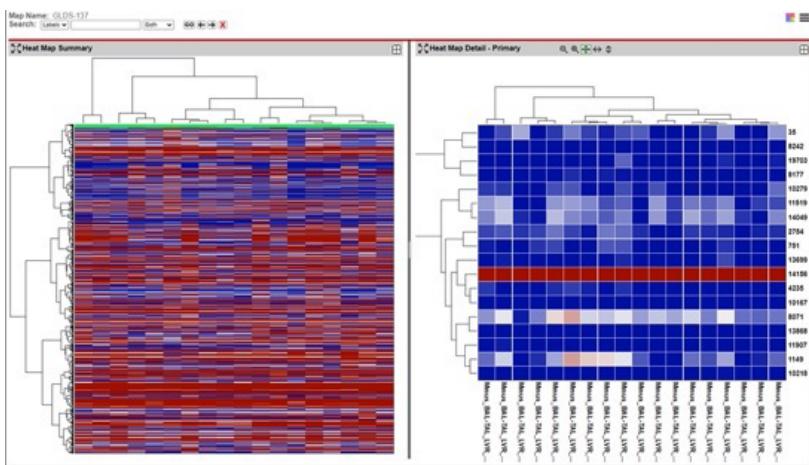
GeneLab: Data Visualization Portal



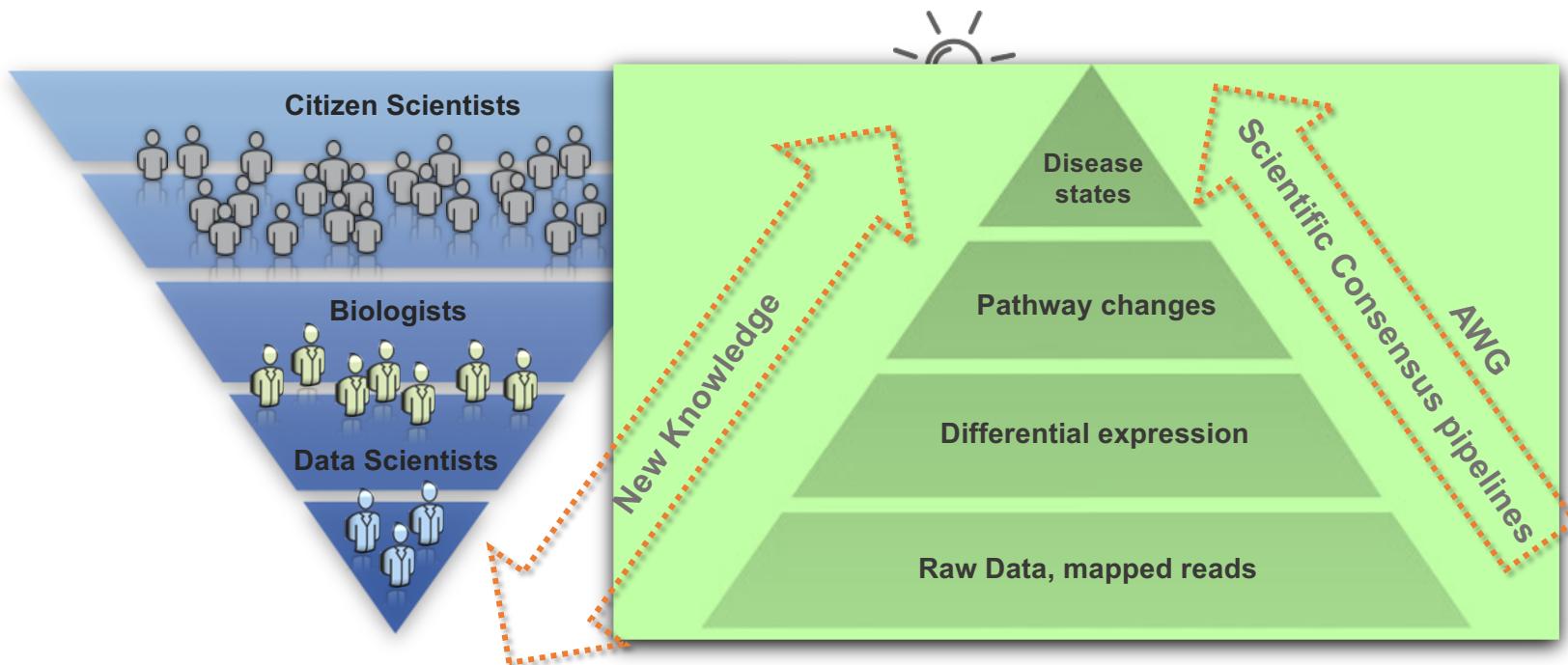
GLDS-137 DGE

ENSEMBL	Symbol	LOG2FC	PVAL	ADJP
Pvalb	Pvalb	-6.29773327428316		
Actn3	Actn3	-5.97815189279096		
Casq1	Casq1	-3.95963363476749		
Tcap	Tcap	-3.25435847812658		
Eef1a2	Eef1a2	-4.65720447979742		
Trnc2	Trnc2	-6.08619947458785		
Jph2	Jph2	-0.8961587251770401		
Trdn	Trdn	-1.2468444937389904		
Ddit4	Ddit4	-0.9511855044660951		
Pgam2	Pgam2	-5.63776597708189		

Showing 1 to 10 of 22,228 entries



GeneLab Data Democratization



Re-use of Data and Enabling New Discoveries



40 enabled publications (10 publications produced by the AWGs) using data available in GeneLab.

Mammalian and Invertebrate Models as Complementary Tools for Gaining Mechanistic Insight on Muscle Responses to Spaceflight

by Thomas Cahill¹, Henry Cope², Joseph J. Bass³, Elijah G. Overbey⁴, Rachel Gilbert^{5,6}, Willian Abraham da Silveira^{1,7}, Amber M. Paul^{5,8,9}, Tejaswini Mishra¹⁰, Raúl Herranz¹¹, Sigrid S. Reinsch⁵, Sylvain V. Costes⁵, Gary Hardiman^{1,12}, Nathaniel J. Szewczyk^{3,13} and Candice G. T. Tahimic^{5,14,*}



An Integrative Network Science and Artificial Intelligence Drug Repurposing Approach for Muscle Atrophy in Spaceflight Microgravity

Vidya Manian^a, Jairo Orozco-Sandoval and Victor Diaz-Martinez^b

Article

Knowledge Network Embedding of Transcriptomic Data from Spaceflown Mice Uncovers Signs and Symptoms Associated with Terrestrial Diseases

Charlotte A. Nelson¹, Ana Uriarte Acuna^{2,3}, Amber M. Paul^{2,4}, Ryan T. Scott^{2,3}, Atul J. Butte^{5,6}, Egla Cekanaviciute², Sergio E. Baranzini^{1,5,7,*} and Sylvain V. Costes^{2,8}

Cell: The biology of spaceflight package

A coordinated package of 29 scientific papers published in five *Cell Press* journals featuring 9 papers utilizing data or resources in GeneLab. *High impact research highlights:*

- [Comparative Transcriptomics Identifies Neuronal and Metabolic Adaptations to Hyper gravity and Microgravity in *Caenorhabditis elegans*, *iScience*](#)
- [Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact, *Cell*](#)
- [NASA GeneLab RNA-seq consensus pipeline: standardized processing of short-read RNA-seq data, *iScience*](#)



Computational and Structural Biotechnology Journal

Volume 19, 2021, Pages 2223-2235



Reanalysis of the Mars500 experiment reveals common gut microbiome alterations in astronauts induced by long-duration confinement

N.J.B. Brereton^a, F.E. Pitre^a, E. Gonzalez^{b,c}

Rad-Bio-App: a discovery environment for biologists to explore spaceflight-related radiation exposures

Richard Barker^a, Sylvain V. Costes, Jack Miller, Samrawit G. Gebre, Jonathan Lombardino & Simon Gilroy^a

npj Microgravity 7, Article number: 15 (2021) | [Cite this article](#)



GeneLab Power Users



GeneLab Analysis Working Groups (AWGs) consist of 250+ scientists from multiple space agencies, international institutions, and industry. Scientists meet monthly with each group to analyze data in the GeneLab repository. We invite you to join - <https://genelab.nasa.gov/awg/join>!

ANIMAL

68 members

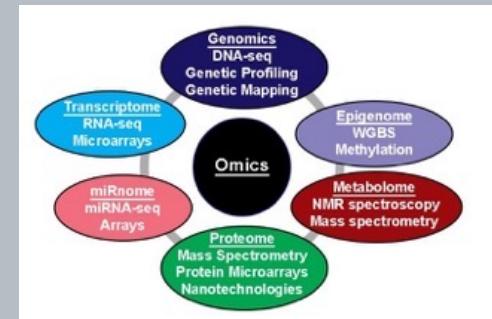
Facilitates the use of omics in understanding basic mechanisms by which animals and constituent tissues and cells adapt to the spaceflight environment.



MULTI-OMICS

128 members

Interactions between the different omics to provide complete understanding of the entire system begin studied.



PLANTS

58 members

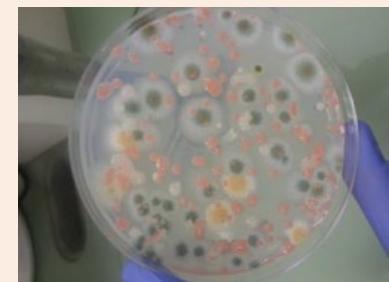
Share and discuss the latest developments in **AstroBotany** – the discipline of botany concerned with interactions between plant biology and space environment.



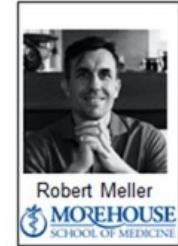
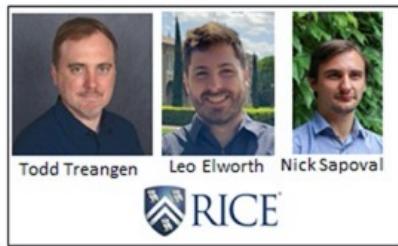
MICROBES

50 members

Focuses on analyzing microbial datasets within GeneLab that includes gene-expression, proteomic, metabolomic and environmental metagenomic datasets.



Analysis Working Groups (AWGs) are Driving Data Reuse



Accelerating the pace of scientific discovery through open data



- GeneLab has established the AWGs with the goal of maximizing the generation of new knowledge from these rare and complex spaceflight datasets. With over 250+ scientists and experts worldwide the four working groups focus Animal, Plants, Microbe and Multi-omics strategies.
- Since 2018, the AWGs have published 10 peer-reviewed publication re-using data in GeneLab:
 - [Mammalian and Invertebrate Models as Complementary Tools for Gaining Mechanistic Insight on Muscle Responses to Spaceflight, International Journal of Molecular Sciences](#)
 - [Immunological and hematological outcomes following protracted low dose/low dose rate ionizing radiation and simulated microgravity, Scientific Reports](#)
 - [NASA GeneLab RNA-seq consensus pipeline: standardized processing of short-read RNA-seq data, iScience](#)
 - [NASA GeneLab RNA-seq consensus pipeline: standardized processing of short-read RNA-seq data, Cell System](#)
 - [Comparative Transcriptomics Identifies Neuronal and Metabolic Adaptations to Hypergravity and Microgravity in *Caenorhabditis elegans*, iScience](#)
 - [Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact, Cell](#)
 - [Fundamental Biological Features of Spaceflight: Advancing the Field to Enable Deep-Space Exploration, Cell](#)
 - [Multi-omics analysis of multiple missions to space reveal a theme of lipid dysregulation in mouse liver, Science Reports](#)
 - [NASA GeneLab Platform Utilized for Biological Response to Space Radiation in Animal Models, Cancers](#)
 - [Mice Exposed to Combined Chronic Low-Dose Irradiation and Modeled Microgravity Develop Long-Term Neurological Sequelae, International Journal of Molecular Sciences](#)

GeneLab Enabled Publications – 40+



Article

Detection of Genes in *Arabidopsis thaliana* L. Responding to DNA Damage from Radiation and Other Stressors in Spaceflight

Vidya Manian^{1,2,*}, Jairo Orozco-Sandoval¹ and Victor Diaz-Martinez³



International Journal of
Molecular Sciences



Article

Mammalian and Invertebrate Models as Complementary Tools for Gaining Mechanistic Insight on Muscle Responses to Spaceflight

Thomas Cahill¹, Henry Cope², Joseph J. Bass³, Elijah G. Overbey⁴, Rachel Gilbert^{5,6}, William Abraham da Silveira^{1,7}, Amber M. Paul^{5,8,9}, Tejaswini Mishra¹⁰, Raúl Herranz¹¹, Sigrid S. Reinsch⁵, Sylvain V. Costes⁵, Gary Hardiman^{1,12}, Nathaniel J. Szewczyk^{3,13} and Candice G. T. Tahimic^{5,14,*}



Article

Comparative Transcriptomics Identifies Neuronal and Metabolic Adaptations to Hypergravity and Microgravity in *Caenorhabditis elegans*

Craig R.G. Willis,¹ Nathaniel J. Szewczyk,^{2,3} Sylvain V. Costes,⁴ Ingrid A. Udranszky,⁵ Sigrid S. Reinsch,⁴ Timothy Etheridge,^{7,8,*} and Catharine A. Conley⁶



www.nature.com/npjgrav

ARTICLE OPEN

Comparison of *Bacillus subtilis* transcriptome profiles from two separate missions to the International Space Station

Michael D. Morrison¹, Patricia Fajardo-Cavazos¹ and Wayne L. Nicholson¹

OPEN

Multi-omics analysis of multiple missions to space reveal a theme of lipid dysregulation in mouse liver

Afshin Beheshti^{1,8*}, Kaushik Chakravarty^{2,8}, Homer Fogle¹, Hossein Fazelinia³, Willian A. da Silveira⁴, Valery Boyko⁵, San-Huei Lai Polo⁶, Amanda M. Saravia-Butler⁵, Gary Hardiman⁵, Deanne Taylor⁵, Jonathan M. Galazka⁷ and Sylvain V. Costes^{5,7*}



www.nature.com/npjgrav

COMMENT OPEN

Rad-Bio-App: a discovery environment for biologists to explore spaceflight-related radiation exposures

Richard Barker^{1,52}, Sylvain V. Costes², Jack Miller^{3,4}, Samrawit G. Gebre⁴, Jonathan Lombardino⁵ and Simon Gilroy^{1,52}

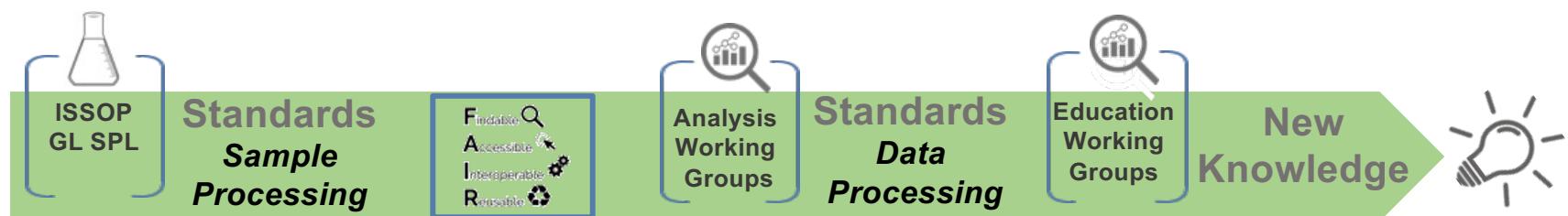
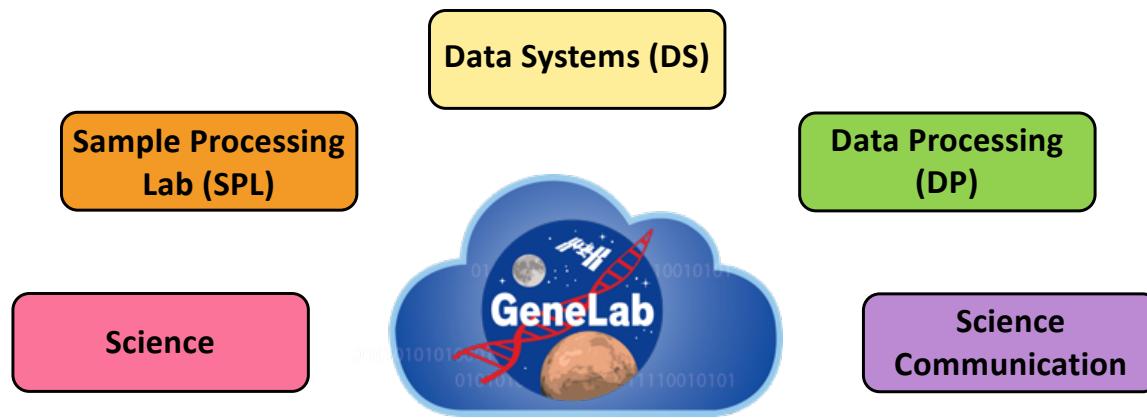


Article

Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact

Willian A. da Silveira,^{1,23} Hossein Fazelinia,^{2,23} Sara Brin Rosenthal,^{1,23} Evangelia C. Laiakis,^{4,23} Man S. Kim,^{2,23}

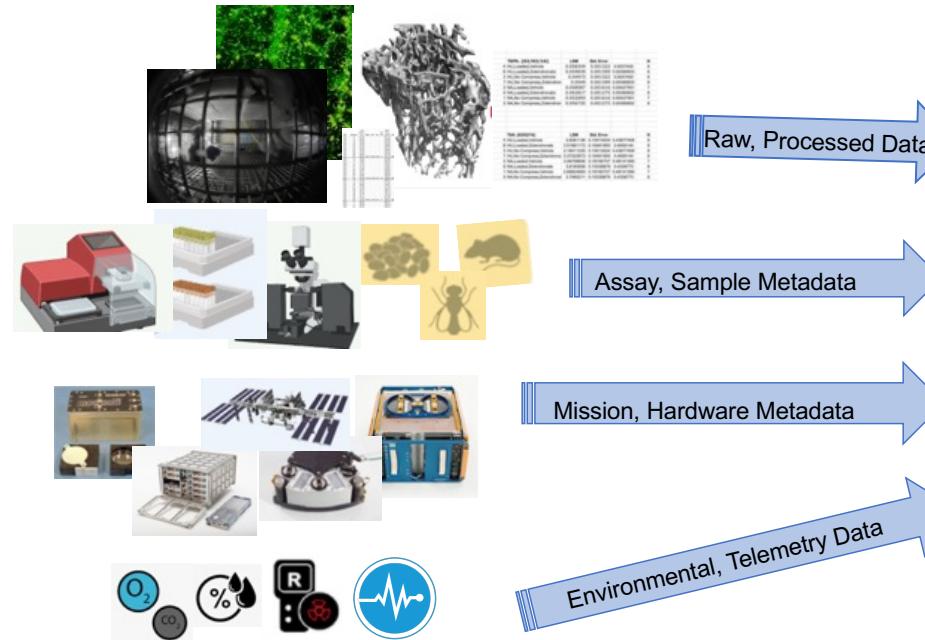
GeneLab Teams



Integrating data beyond *omics*



In 2021, the Ames Life Sciences Data Archive (ALSDA) transitioned to use the GeneLab Data Systems. The goal of this integration was to develop a multi-project submission portal to reduce the burden on PIs submitting data and enabling discovery of both omics and phenotypic data.



Open Science Data Repositories

- Phenotypic-Physiological-Behavioral-Imaging-Level Data
- Mission Telemetry data from spacecraft and hardware associated to experimental studies
- Federated with GeneLab (Molecular Omics-Level Data)

(Images from Ronca et al., 2019 <https://doi.org/10.1038/s41598-019-40789-y>; Cells, Bone, RR Habitat NASA-ARC)

ALSDA Science Community-Driven Curation Standards



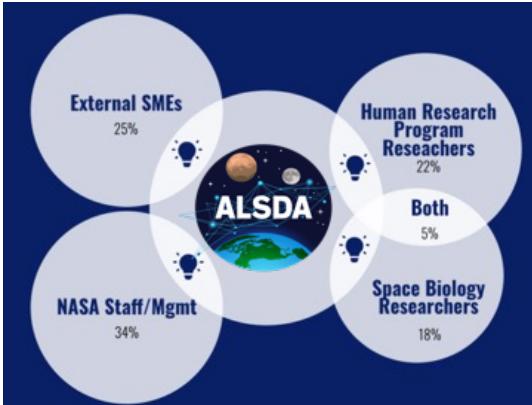
- Last year, ALSDA transitioned their legacy database to a new infrastructure to comply to **FAIR and Open data**.
 - **AWG** is a community of PIs and other SMEs, who contribute their expertise to establish a consensus on standards for the assay, sample, and study design metadata which must be collected from PIs to **ensure data reusability**.

ALSDA Analysis Working

Group:

~100 members

Hosted 5 monthly meetings



Email to learn more and join!
arc-dl-alsda@mail.nasa.gov

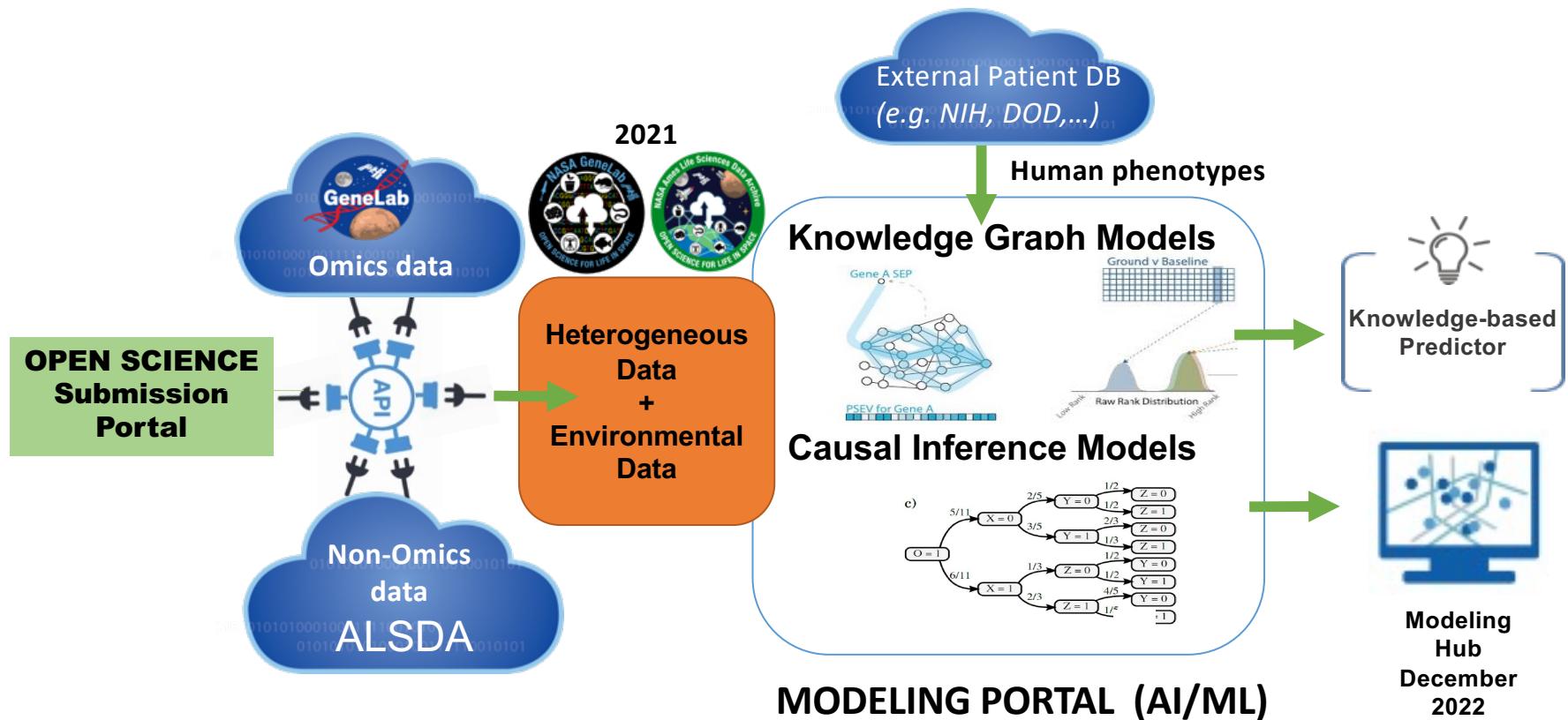
(Credit: Amazing ALSDA AWG Members)

Assay Metadata Configurations	Drafted	AWG Covered	Portal	Dataset Creation
Bone Microstructure (MicroCT)	✓	09/08/21	✓	✓
Behavior (Novel Object Recognition)	✓	09/08/21	✓	✓
Molecular/Cellular Imaging (L ₁ /Fluor Microscopy)	✓	09/08/21	✓	✓
Flow Cytometry (Flow Cytometry)	✓	09/08/21	✓	✓
Behavior (Elevated Plus Maze)	✓	12/08/21	✓	✓
Behavior (Open Field)	✓	12/08/21		
Behavior (Gait)	✓	12/08/21		
Protein Analysis (Western Blot/Immunoblot)	✓	12/08/21	✓	✓
Calcium Uptake (Spectrofluorometry)	✓	12/08/21	✓	✓
Behavior (Barnes Maze)	✓	12/08/21		
Osteocellular (Histomorphometry)	✓	02/16/22		
Mechanical Test (3pt-Bend)	✓	02/16/22		
Intra-Ocular Pressure (Rebound Tonometry)	✓			
Blood Analysis (Serum Metabolites)				
Echocardiogram (Ultrasoundography)	✓			
Protein Analysis (ELISA)	✓	02/16/22		
Behavior (Radial Arm; Water Maze)				
Behavior (Morris; Water Maze)				
Muscle Physiology (Muscle Fiber Typing)				
Mineral Density and Composition (DXA)				
Neurophysiology (Electrophysiological Recordings)				
Physiology (Blood Pressure)				
Physiology (Pulse Oximetry)				
Metabolites (HPLC with ECD)				
Cardiovascular Imaging (MRI)				
Neuroimaging (MRI)				
Eye Microstructure (microCT)				
Grip Force (Force Transducer)				
Electron Microscopy (Scanning)				
Electron Microscopy (Transmission)				
Behavior (Spatial Novelty Y Maze)				
Behavior (Tail Suspension Test)				
Behavior (Startle Test)				
Behavior (Prepulse Inhibition)				
Behavior (Contextual Fear Conditioning)				
Behavior (Rotarod)				
Behavior (Water Maze)				
Behavior (Wire Hang)				
Behavior (Escher Staircase Track)				

With the support of ALSDA AWG:
Completed 10 New Assay metadata standards for the Submission Portal

ALDSA Assay Metadata, Image, Outcome Measure Configuration for microCT		Micro-Computed Tomography (microCT; µCT)	
File Folders will contain:		Western Blot / Immunoblot	
Assay .log file (txt)		Western Blot / Immunoblot	
Supporting Library (e.g., Protocols, Data, Metadata files, etc.)		Western Blot / Immunoblot	
ALDSA Assay Metadata, Setup, Outcome Measure Configuration for Barnes Maze		Barnes Maze	
File Folders will Optionally Contain: Study Metadata File, Processed Data Outputs (e.g., Images, Text), Configuration for Gait, Video		Measurement: Behavior Gait	
NOTE: Data from Barnes Maze is not included in the Supporting Library.		Configuration for Gait	
Supporting UN DOI: 10.33234/Champlain-PAID-793154		PDX ALDSA Assay Metadata, Setup, Outcome Measure Configuration for Open Field	
Scanner ID: PAXID: 10.33234/Champlain-PAID-793154		Open Field	
Volume of Subjects		ALDSA Assay Metadata, Image Configuration for Microscopy	
Brittleness, Rad %		Microscopy	
Volume of Scale		ALDSA Assay Metadata, Setup, Outcome Measure Configuration for Open Field	
Contract Size		Elevated Plus Maze (EPM)	
Version		ALDSA Assay Metadata, Setup, Outcome Measure Configuration for Open Field	
Killed		Measurement: Behavior Ontology	
X-ray Info		https://ncit.nci.nih.gov/ncitbrowser/pages/concept_details.jsp?conceptId=NO_Thesaurus&code=C16326	
Integration Time		Supporting UN DOI: 10.33793/TOBIA	
Frame rate		Supporting UN DOI: 10.33793/TOBIA	
Other Behaviors after the Barnes Maze		Supporting UN DOI: 10.33793/TOBIA	
Filtration		Wall, NatProtocols, 2007; https://doi.org/10.1089/natprot.2007.44	
Segmentation		Supporting UN DOI: 10.33793/TOBIA	
Segmentation		Supporting UN DOI: 10.33793/TOBIA	
Intervals Between Vectr Speed & Interval		Parameter	
Interval before Prior Hallucinations (min)		Example Value	
Prior Hallucinations		ALDSA ASSAY Metadata for EPM	
Gait Repetition		COGNITIVE ASSAY ARRANGEMENT	
Barnes Maze Access & Duration		Tracking and Analysis system (name/version)	
Other Behaviors after the Barnes Maze		e.g. Ethovision XT 12.0	
Number		Other behavioral assays conducted prior or concurrent to EPM (free text)	
Number		e.g. Open Field, Elevated Plus Maze, NOR	
Number		Vector of the sequence of all behavioral tests performed in order, and the position of EPM (free text)	
Subject Count		e.g. 1. Open Field, 2. Elevated Plus Maze, 3. NOR, 4. Rotarod	
Subject Count		EPM-Assay Repeated (unit: YES/NO)	
Subject Count		e.g. No	
Subject Count		Number of times EPM-Assay Repeated (unit: numerical)	
Subject Count		e.g. 0	
Subject Handling Frequency and Technique (free text)		PHYSICAL ENVIRONMENT	

From multiple databases to a knowledge-based system



Introduction to Planetary Protection

What is JPL

- -Founded in 1930s, by a few Caltech doctoral students who were building alcohol fueled rocket motors with professor Theodore Von Karman
- -Federally funded, governed by Caltech
- -5000 employees, over 1000 students.
- -10+ active missions: Juno, Cassini, MRO, MSL, Mars 2020, Dawn, Europa Clipper, InSight.



What is Planetary Protection?

Planetary Protection (PP) aims to:

- Protect other worlds from microorganisms found on Earth.
- Prevent forward and backward contamination.



Why?

- How can we look for life on other planets if we bring our own?
- Meet NASA Planetary Protection requirements
- Comply with international treaty obligations

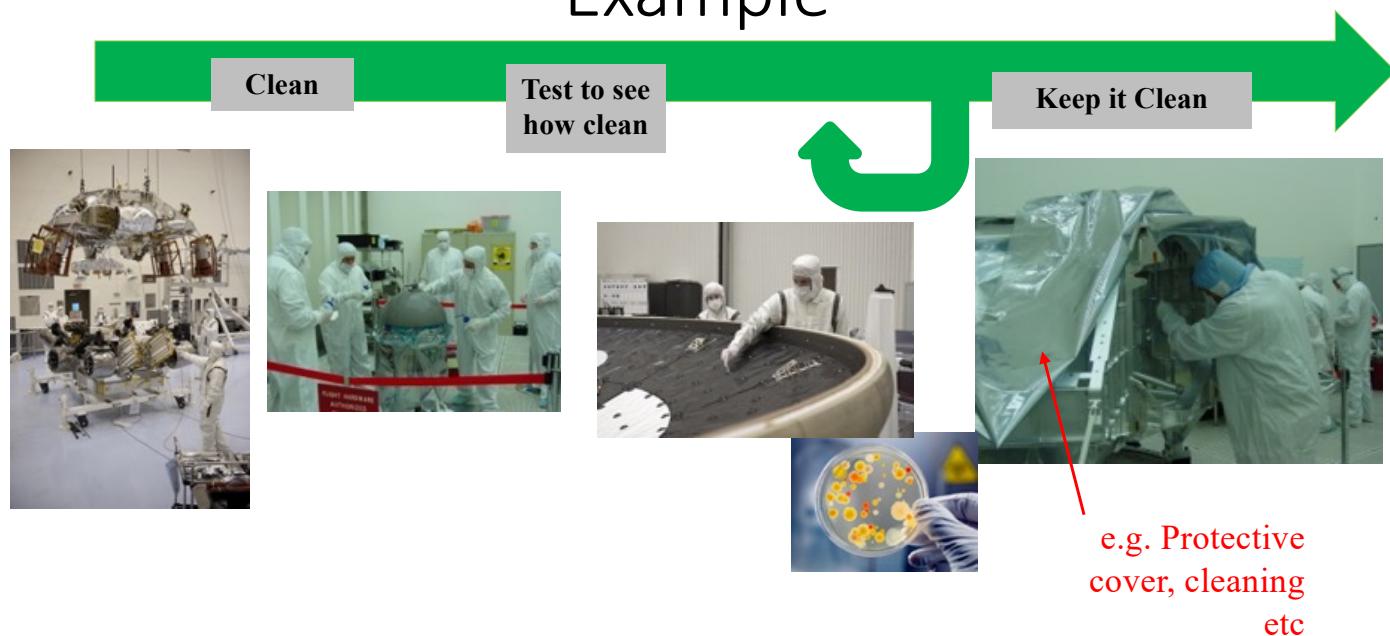


Jet Propulsion Laboratory
California Institute of Technology

Bioburden Requirements for a Mars-Bound Biologically Sensitive Mission

- PP requirements as per NASA Procedural Requirements 8020.12D
 - $\leq 5 \times 10^5$ total spores at launch
 - $\leq 3 \times 10^5$ total spores on planned landing hardware
 - ≤ 300 spores/m² for exposed surfaces
- For a biologically sensitive mission, each mission develops a biological contamination control campaign to include: hardware design, material selection, material processing, manufacturing, assembly, testing, launch operations, space flight operations, and surface operations.
- Microbial bioburden reduction and monitoring is conducted throughout the mission life cycle.
- Monitoring and/or verification of biological cleanliness is conducted using direct wipe and swab samples extracted and processed using a cultivation based technique (NASA standard assay) which measures heat-tolerant organisms.

Planetary Protection Implementation Process Example



Planetary Protection Biological Cleanliness Verification Sampling



Clean hardware (and table or bag)
with an Iso-Propyl alcohol wipe prior
to sample or install



Assay Swab or Wipe
(water is used as solvent)



Count Plates – 24h, 48h, and 72h



Process swabs and wipes,
~3 hours required post-assay

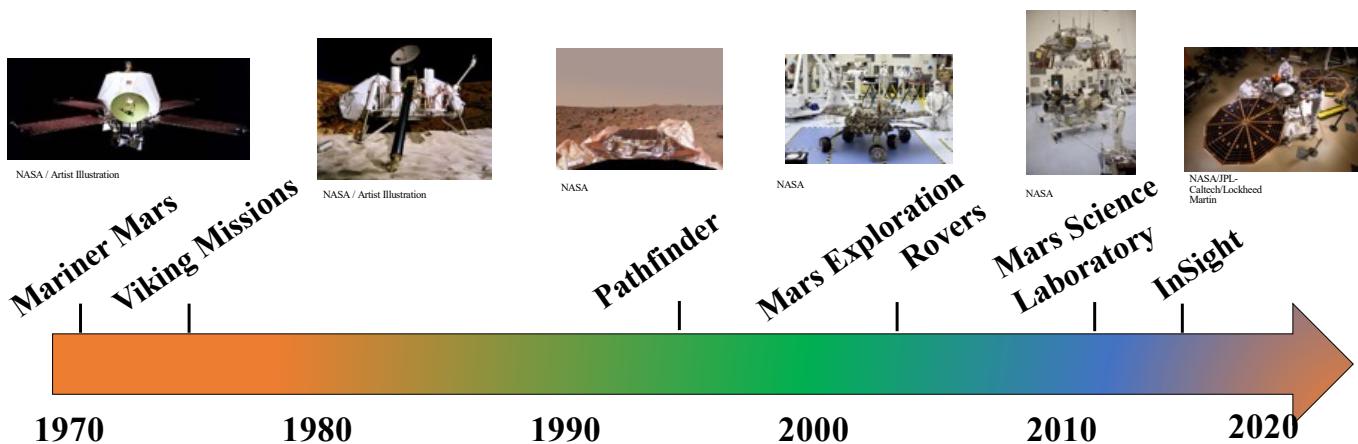
NASA Standard Spore Assay



- Does detect:
 - spores
 - viable, cultivable, aerobic
 - 32°C, 3 days
- Does not detect:
 - non-cultivable, vegetative cells grown under various conditions
 - NASA standard assay underestimates the total microbial burden.



Bioburden Accounting for Mars Landed Spacecraft Raw Spore Counts to Calculated Bioburden



Sum of the Means Approach

The total effective area sampled was represented by:

$$A_s = \sum_{i=1}^{n_s} a_{si} f_{si} e_{si}$$

The total number of spores counted was:

$$N_{tot} = \sum_{i=1}^{n_s} N_{si}$$

The bioburden density, B, was given by:

$$B = N_{tot} / A_s$$

The estimate of the total bioburden, N, was given by:

$$N = B A_0$$

Variable	Definition
A_0	the total area represented by a group or sample set, m ²
n_s, n_w	the total number of swabs or wipes
n_{tot}	the total number of samples
a_{si}, a_{wj}	is the area sampled by the i th swab and the j th wipe, (m ²)
f_{si}, f_{wj}	the pour fractions for swabs and wipes
e_s, e_w	the recovery efficiencies for swabs and wipes
A_s	the total effective area, (m ²)
N_{si}	the number of CFU counted in the i th swab sample
N_{wj}	the number of CFU counted in the j th wipe sample
N_{tot}	the total number of CFU in a group, (spores/m ²)
B	

Poisson and Gaussian Statistics

Examples equations utilized:

$$\sigma = 1 \div (A_0 A_s)$$

$$B_{max} = 1/A_s + 3\sigma$$

$$\sigma = 1 \div A_s$$

$$B_{max} = 1/A_s + 3\sigma$$

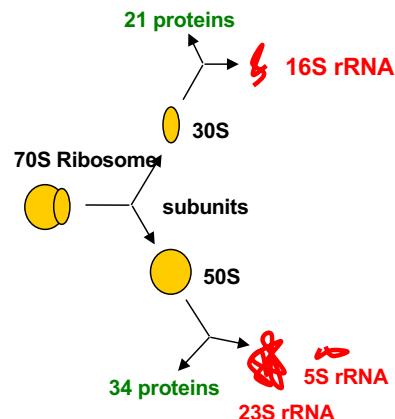
$$\sigma = \sqrt{(N/A_s)}$$

$$B_{max} = N/A_s + 3\sigma$$

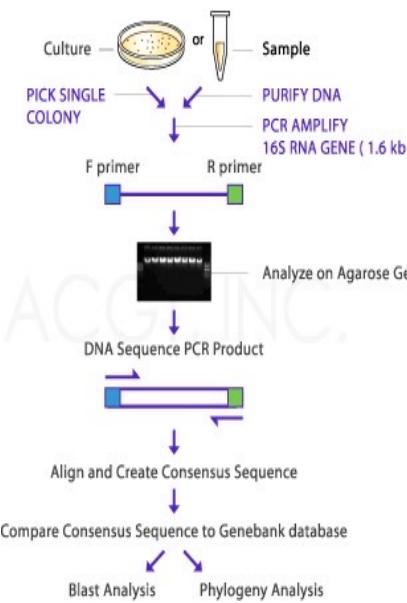


Jet Propulsion Laboratory
California Institute of Technology

Bacterial Identification



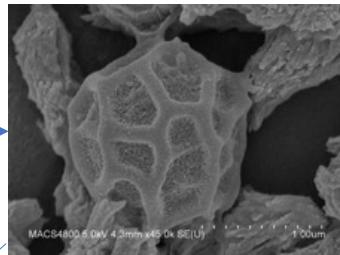
- 16s rRNA is evolutionarily a highly conserved gene.
- It is used for phylogenetic classification of bacterial isolates at species level.
- 97.5% sequence similarity yields an accurate species level identification
- 95% yields an accurate genus level identification



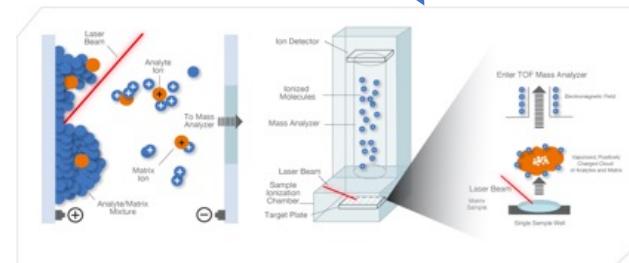
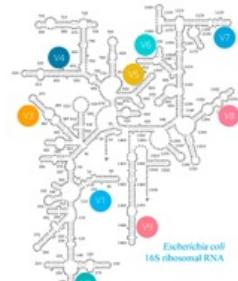
Overall Workflow



Surface samples of spacecraft hardware and associated surfaces are taken using sterile wipes and swabs



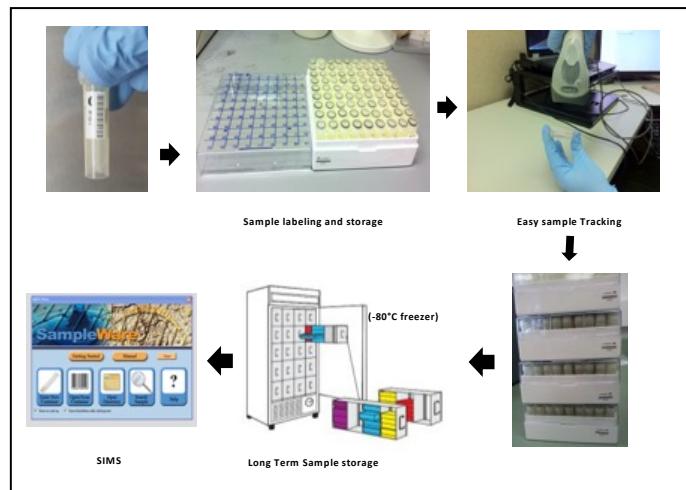
Hardy microorganisms are isolated via NASA standard spore assay (heat shock at 80°C for 15 minutes)



Microbial Identification via MALDI-TOF MS and 16S rRNA Sequencing

JPL Microbial Culture Collection

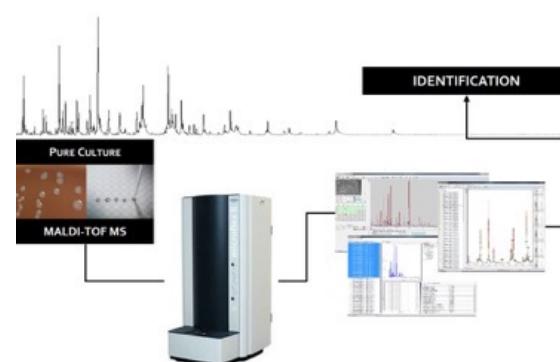
- Viking (1975): 1454 strains
- Mars Pathfinder (1996): 62 strains
- Mars Odyssey (2001): 97 strains
- Mars Exploration Rovers (2003): 870 strains
- Phoenix (2007): 824 strains
- Mars Science Laboratory (2012): 1437 strains
- InSight (2018): 1000 strains
- M2020 (2020): 3685 strains



- In total, JPL microbial archive includes 9,429 **spacecraft related cultures** that span NASA's robotic exploration of Mars ranging from 1975-2017.

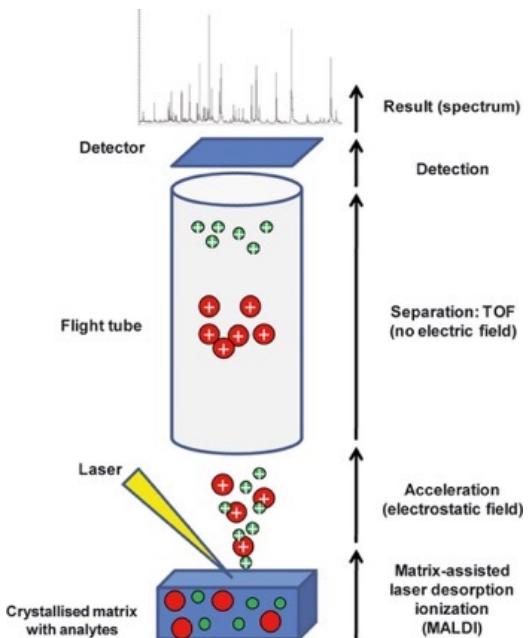
What is MALDI-TOF MS

- Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry
- Generates ions from large molecules while minimizing fragmentation
- Identification based on a protein profile



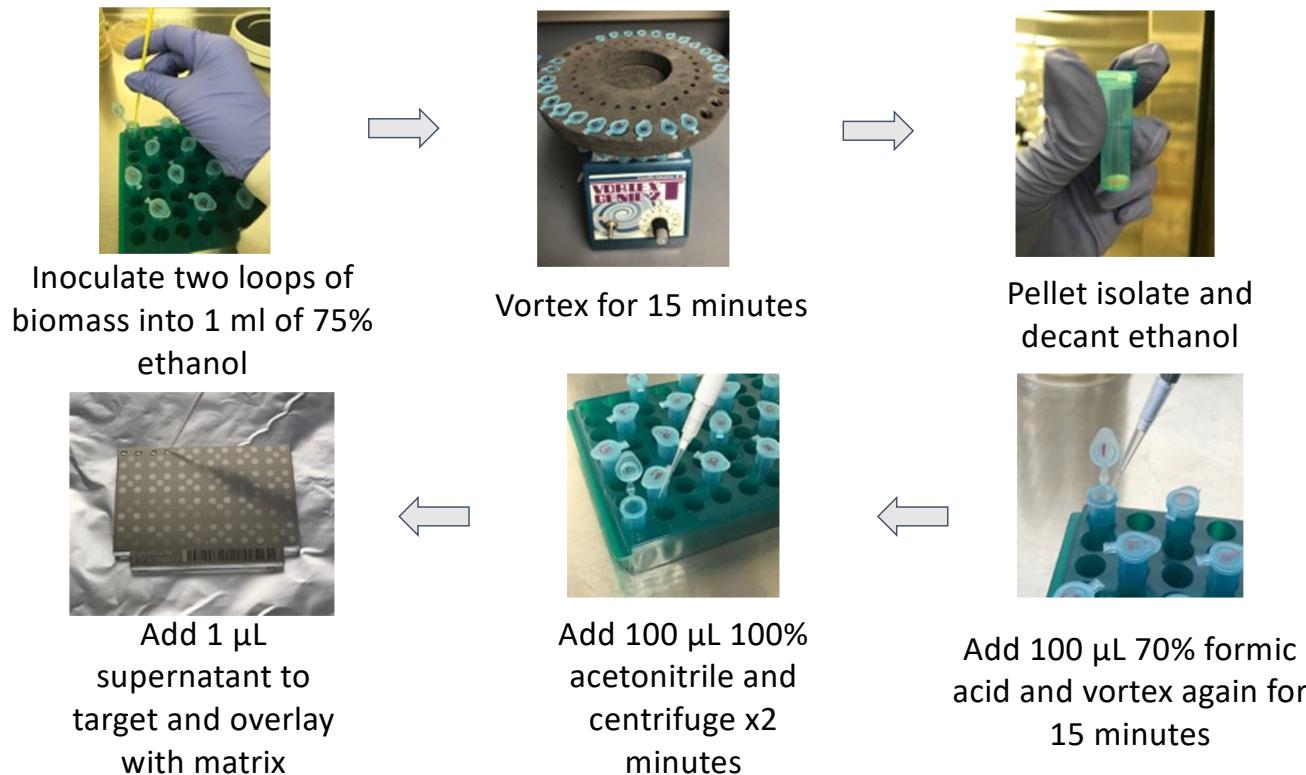
5/19/22

For required markings, please visit <https://mh.jpl.nasa.gov>



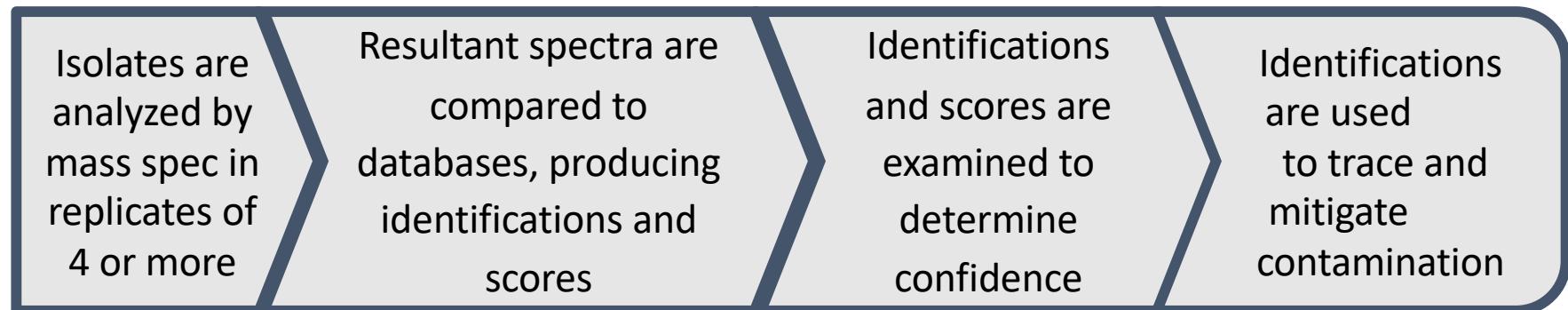
Identification Techniques (MALDI-TOF MS)

Formic Acid Tube Extraction



MALDI-TOF Background

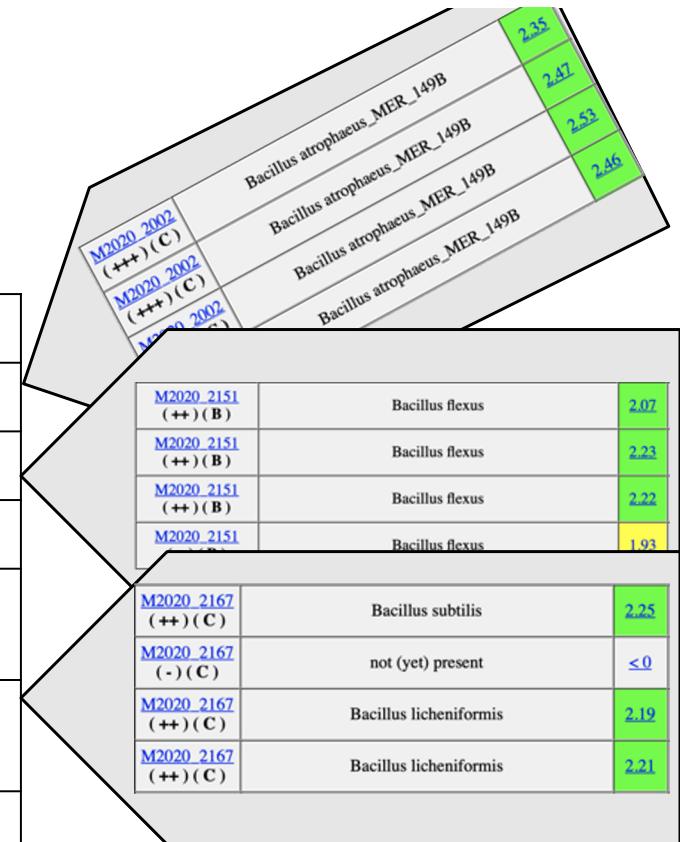
Goal : Rapid species- level identification of bacterial colonies



Slide provided by: Akemi Hinzer

MALDI-TOF Scoring Strategy

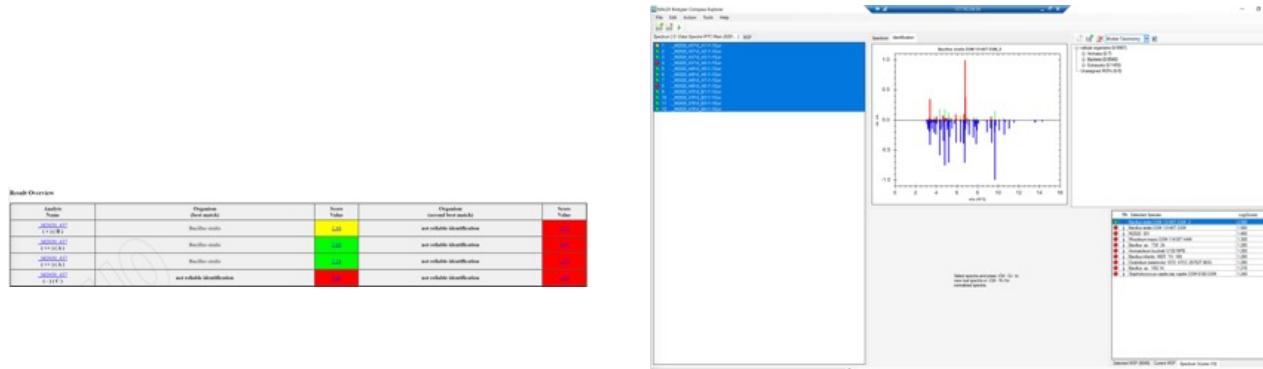
Match Type	Score Met	≥ 3 Scores	Interpretation
Perfect ID	2.2	✓	Confident in identification
Redo Species	2.2		Identified without confidence
Genus Match	2.0	✓	Confident in genus identification
Redo Genus	2.0		Genus identified without confidence
Taxonomic Group	2.0	✓	Genus identified with mismatched species
Not Reliable	-		All scores were below 2.0
No Peaks			Experiment produced no peaks to match



Slide provided by: Akemi Hinzer

Data Analysis

- Mass spectra of unknown isolates are compared against
 - The Bruker reference database containing:
 - 8540 bacterial strains
 - The JPL in-house curated database containing:
 - 509 bacterial strains

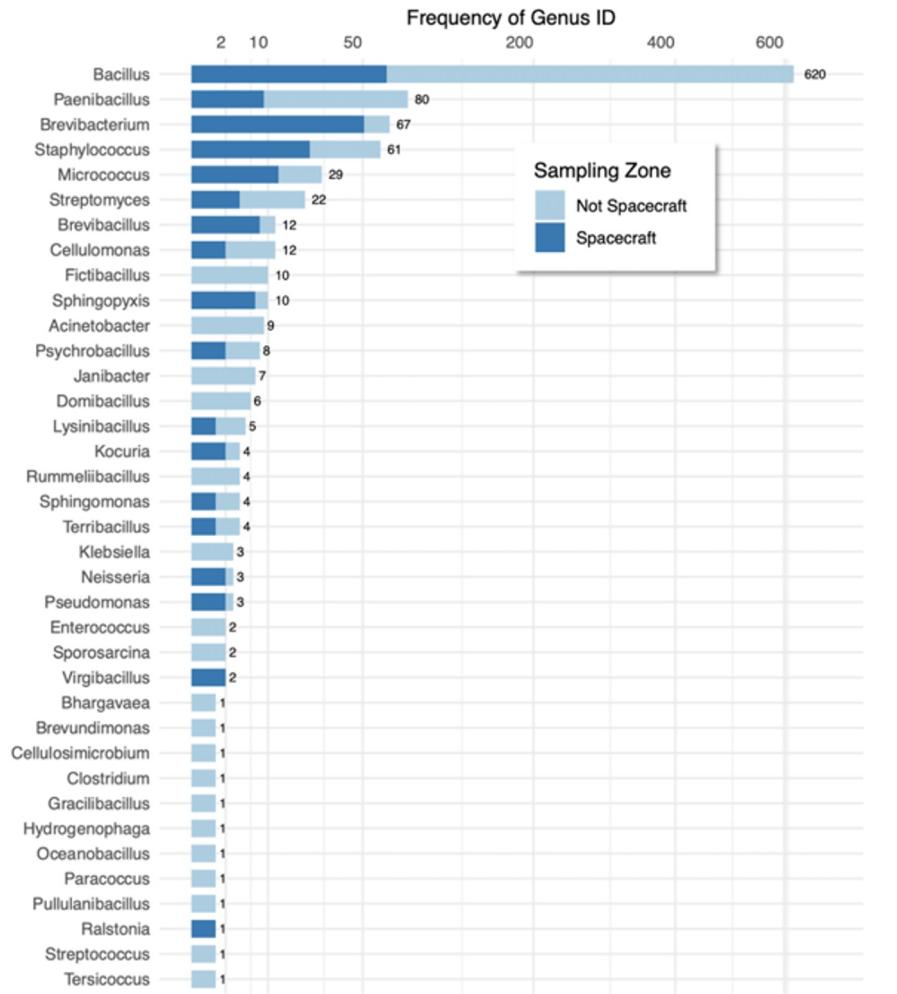
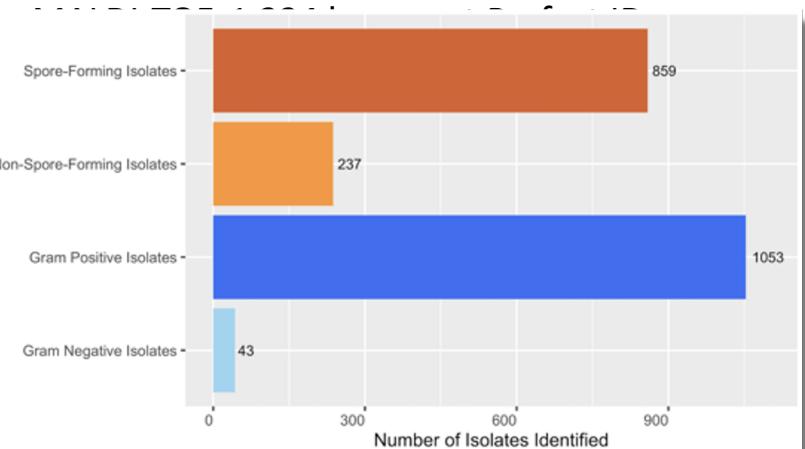


Preliminary Findings

Of the 16,839 samples collected, 1,686 plates produced colonies

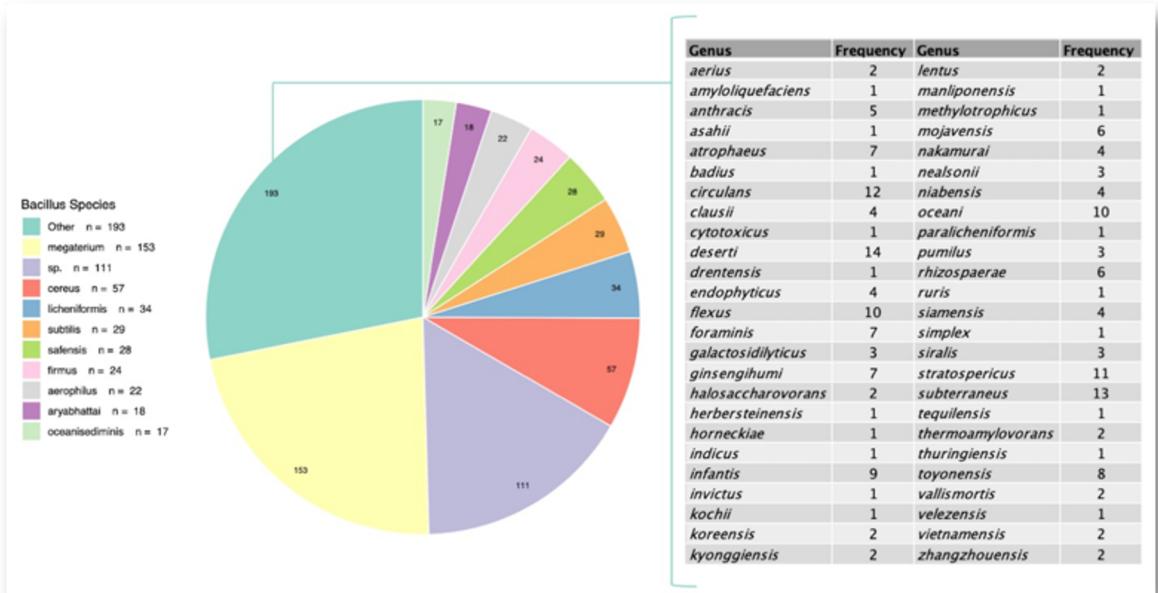
From these positive plates, 4,232 colonies were subcultured and archived

Of the 3,101 isolates analyzed so far by

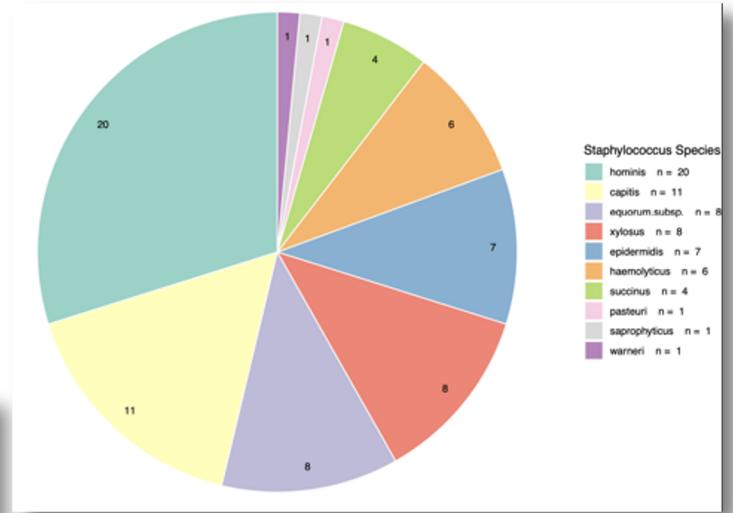


Slide provided by: Akemi Hinzer

Preliminary Findings



Genus	Frequency	Genus	Frequency
<i>aerius</i>	2	<i>lentus</i>	2
<i>amylolyquefaciens</i>	1	<i>manilponensis</i>	1
<i>anthracis</i>	5	<i>methylotrophicus</i>	1
<i>asahii</i>	1	<i>mojavensis</i>	6
<i>atrophaeus</i>	7	<i>nakamurae</i>	4
<i>badius</i>	1	<i>nealsonii</i>	3
<i>circulans</i>	12	<i>niabensis</i>	4
<i>clausii</i>	4	<i>oceani</i>	10
<i>cytotoxicus</i>	1	<i>paralicheniformis</i>	1
<i>deserti</i>	14	<i>pumilus</i>	3
<i>drentensis</i>	1	<i>rhizospaerae</i>	6
<i>endophyticus</i>	4	<i>ruris</i>	1
<i>flexus</i>	10	<i>siamensis</i>	4
<i>foraminis</i>	7	<i>simplex</i>	1
<i>galactosidilyticus</i>	3	<i>siralis</i>	3
<i>ginsengihumi</i>	7	<i>stratospericus</i>	11
<i>halosaccharovorans</i>	2	<i>subterraneus</i>	13
<i>herbersteiniensis</i>	1	<i>tequilensis</i>	1
<i>horneckiae</i>	1	<i>thermoamylovorans</i>	2
<i>indicus</i>	1	<i>thuringiensis</i>	1
<i>infantis</i>	9	<i>toyonensis</i>	8
<i>invictus</i>	1	<i>vallisimortis</i>	2
<i>kochii</i>	1	<i>velezensis</i>	1
<i>koreensis</i>	2	<i>vietnamensis</i>	2
<i>kyonggiensis</i>	2	<i>zhangzhouensis</i>	2



Slide provided by: Akemi Hinzer

What if there are no matches to known sequences

Novel Species

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY Volume 70, Issue 1

Research Article Free

Bacillus glennii sp. nov. and **Bacillus saganii** sp. nov., isolated from the vehicle assembly building at Kennedy Space Center where the Viking spacecraft were assembled

Arman Seylemezian¹, Logan Ott¹, Sarah Wolf¹, Jomel Fragante¹, Oscar Yip¹, Rüdiger Pukall², Peter Schumann², Parag Vaishampayan³

View Affiliations

First Published: 03 February 2020 | <https://doi.org/10.1099/ijsem.0.003714>

Info Sections

Side by side view PDF Tools Share

ABSTRACT

Two Gram-stain-positive, motile, endospore-forming, aerobic strains, isolated from environmental air sampling at the vehicle assembly building at Cape Canaveral, grew optimally at pH 7–9 (optimum, pH 9) for s 23a^T. Both strains displayed growth in 0–5% NaCl with an optimum at 1% NaCl. Strains V44-8^T and V47-23a^T grew optimally at 32 °C (15–32 °C) and 21 °C. Strains contained meso-diaminopimelic acid as the diagnostic diamino acid, phosphatidylethanolamine and diphosphatidylglycerol. The predominant and iso-C₁₅:0. Strain V47.23a^T shared its highest 16S rRNA sequence similarity values between strains V44-8^T and DSM-103964^T at 96.6%. Based on position inferred from 16S rRNA gene sequence analyses, the isolates represent two novel species in the genus *Bacillus* for which the names *Bacillus* sp. nov. (=DSM 105192^T) and *Bacillus* saganii sp. nov. (V47-23a^T) (nA

Keywords

planetary protection, *Bacillus*, Viking, Kennedy Space Center, cleanroom

> Int J Syst Evol Microbiol. 2010 May;60(Pt 5):1031–1037. doi: 10.1099/ijsm.0.008979-0.

Pubmed 2009 Aug 7.

Bacillus horneckiae sp. nov., isolated from a spacecraft-assembly clean room

Parag Vaishampayan¹, Alexander Probst¹, Srinivasan Krishnamurthy², Sudeshna Ghosh¹, Shariff Osman¹, Alasdair McDowell³, Arunachalam Ruckmani², Shanmugam Mayilraj², Kasthuri Venkateswaran¹

Affiliations + expand

PMID: 19666815 DOI: 10.1099/ijsm.0.008979-0

Abstract

Five Gram-stain-positive, motile, aerobic strains were isolated from a clean room of the Kennedy Space Center where the Phoenix spacecraft was assembled. All strains are rod-shaped, spore-forming bacteria, whose spores were resistant to UV radiation up to 1000 J m⁻². The spores were subterminally positioned and produced an external layer. A polyphasic taxonomic study including traditional biochemical tests, fatty acid analysis, cell-wall typing, lipid analyses, 16S rRNA gene sequencing and DNA-DNA hybridization studies was performed to characterize these novel strains. 16S rRNA gene sequencing and lipid analyses convincingly grouped these novel strains within the genus *Bacillus* as a cluster separate from already described species. The similarity of 16S rRNA gene sequences among the novel strains was >99%, but the similarity was only about 97% with their nearest neighbours *Bacillus pocheonensis*, *Bacillus firmus* and *Bacillus bataviensis*. DNA-DNA hybridization dissociation values were <24% to the closest related type strains. The novel strains had a G+C content 35.6–37.0 mol% and could liquefy gelatin but did not utilize or produce acids from any of the carbon substrates tested. The major fatty acids were iso-C(15 : 0) and anteiso-C(15 : 0) and the cell-wall diamino acid was meso-diaminopimelic acid. Based on phylogenetic and phenotypic results, it is concluded that these strains represent a novel species of the genus *Bacillus*, for which the name *Bacillus horneckiae* sp. nov. is proposed. The type strain is 1P01SC(T) (=NRRL B-59162(T) =MTCC 9535(T)).

Paenibacillus xerothermodurans sp. nov., an extremely dry heat resistant spore forming bacterium isolated from the soil of Cape Canaveral, Florida

Navjot Kaur^{1,2}, Arman Seylemezian^{2,3}, Prashant P. Patil³, Prabhu Patil³, Srinivasan Krishnamurthy¹, Joseph Varelas⁴, David J. Smith⁵, Shanmugam Mayilraj³, Parag Vaishampayan²

View Affiliations

First Published: 21 August 2018 | <https://doi.org/10.1099/ijsem.0.002967>

Info Sections

Side by side view PDF Tools Share

ABSTRACT

i.e. endospore-producing, facultative anaerobic bacterial strain, designated ATCC 27380^T, was isolated from the soil of Cape Canaveral, Florida, USA. Growth was observed at 20–42 °C (optimum, 37 °C), at 0 °C and in the presence of 0.5–3% NaCl (optimum, 0.5%). The cell wall contained meso-diaminopimelic acid and the isoprenoid quinone was MK-7. The polar lipids present were dihydroethanolamine, diphosphatidylglycerol and one unknown phospholipid. The main fatty acids were C₁₆:0. Phylogenetic analysis based on 16S rRNA gene sequencing affiliated strain ATCC 27380^T with *Paenibacillus rigui* JCM 16352^T (97.0%). Type strains exhibited 16S rRNA gene sequence similarity values below 95.9%. The draft genome size of 4,361,187 bases, with a G+C content of 51.0%. The average nucleotide identity and similarity values between strain ATCC 27380^T and *P. rigui* JCM 16352^T were 72.5% and 18.5%, below the threshold suggested for species differentiation (96% and 70%, respectively). The average nucleotide identity between strain ATCC 27380^T and *P. rigui* JCM 16352^T was 68.72%, which was above the threshold of 65%. Based on phenotypic, genotypic and chemotaxonomic data, strain ATCC 27380^T represents a new species in the genus *Paenibacillus*, for which the name *Paenibacillus xerothermodurans* sp. nov. is proposed.

Novel Genera

Description of *Tersicoccus phoenicis* gen. nov., sp. nov. isolated from spacecraft assembly clean room environments

Parag Vaishampayan ¹, Christine Moissl-Eichinger ², Rüdiger Pukall ³, Peter Schumann ³, Cathrin Spröer ³, Angela Augustus ⁴, Anne Hayden Roberts ⁴, Greg Namba ⁴, Jessica Cisneros ⁴, Tina Salmassi ⁴, Kasthuri Venkateswaran ¹

Affiliations + expand

PMID: 23223813 DOI: [10.1099/ijns.0.047134-0](https://doi.org/10.1099/ijns.0.047134-0)

Abstract

Two strains of aerobic, non-motile, Gram-reaction-positive cocci were independently isolated from geographically distinct spacecraft assembly clean room facilities (Kennedy Space Center, Florida, USA and Centre Spatial Guyanais, Kourou, French Guiana). A polyphasic study was carried out to delineate the taxonomic identity of these two isolates (1P05MA(T) and KO_PS43). The 16S rRNA gene sequences exhibited a high similarity when compared to each other (100 %) and lower than 96.7 % relatedness with *Arthrobacter crystallopileus* ATCC 15481(T), *Arthrobacter luteolus* ATCC BAA-272(T), *Arthrobacter tumbae* DSM 16406(T) and *Arthrobacter subterraneus* DSM 17585(T). In contrast with previously described *Arthrobacter* species, the novel isolates maintained their coccoid morphology throughout their growth and did not exhibit the rod-coccus life cycle typically observed in nearly all *Arthrobacter* species, except *A. agilis*. The distinct taxonomic identity of the novel isolates was confirmed based on their unique cell-wall peptidoglycan type (A:11:20; Lys-Ser-Ala2) and polar lipid profile (presence of phosphatidylglycerol, diphosphatidylglycerol, phosphatidylinositol, an unknown phospholipid and two unknown glycolipids). The G+C content of the genomic DNA was 70.6 mol%. The novel strains revealed MK-9(H2) and MK-8(H2) as dominant menaquinones and exhibited fatty acid profiles consisting of major amounts of anteiso-C15 : 0 and anteiso-C17 : 0 and moderate amounts of iso-C15 : 0 discriminating them again from closely related *Arthrobacter* species. Based on these observations, the authors propose that strains 1P05MA(T) and KO_PS43 be assigned into a separate genus *Tersicoccus* gen. nov. For this new taxon, comprising strains 1P05MA(T) and KO_PS43, we propose the name *Tersicoccus phoenicis* gen. nov., sp. nov. (the type species of *Tersicoccus*), represented by the type strain *Tersicoccus phoenicis* 1P05MA(T) (= NRRL B-59547(T) = DSM 30849(T)).



Introduction to the Command Line

What is a Terminal?



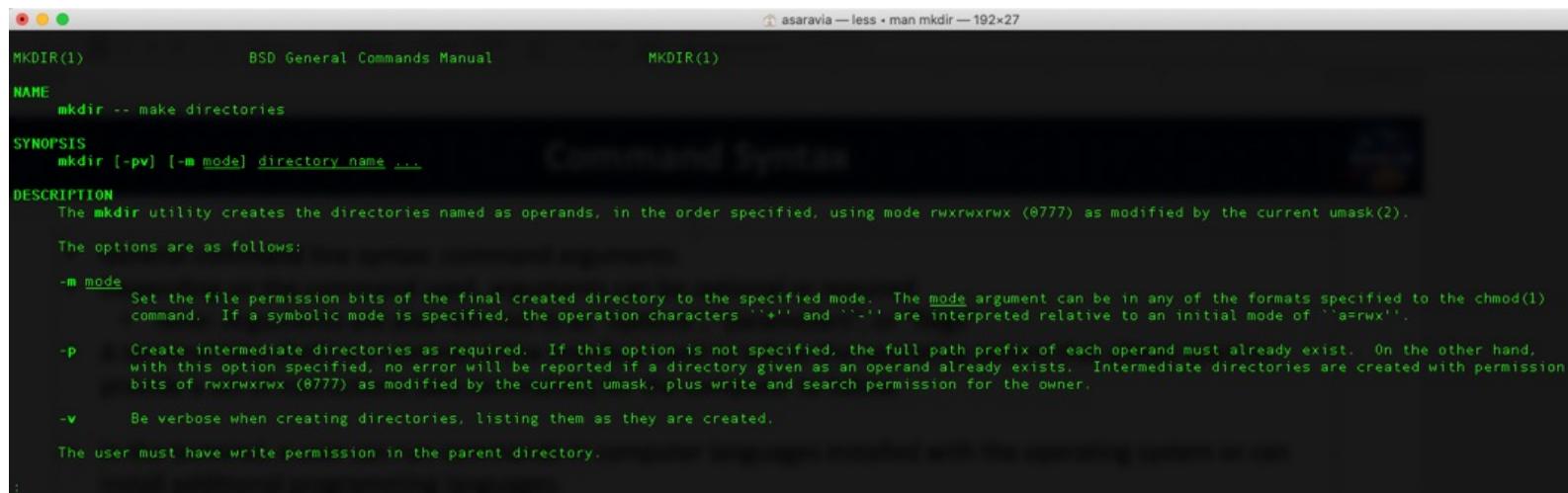
- A terminal window is a command-line interface used to interact with your computer (without graphics)
- The terminal window displays a blank text window, or “command line”, where you can type commands to communicate with your computer
- A terminal uses a specific language interpreter, or “shell”, to interpret commands typed into the command line
- Unix is a computer operating system written in the C programming language
 - By default, terminals on Linux and Mac computers run a Unix shell
 - Unix shells come in a variety of flavors including zsh, csh, tcsh, and bash, each having slightly different features and functionality
 - We will be using a Unix shell simulator, bash, during this bootcamp

A screenshot of a Mac OS X terminal window titled "asaravia — -zsh — 149x28". The window shows a standard terminal prompt and a multi-line text message about government information system usage. The message states that users are accessing a U.S. Government information system and outlines rules for use, including monitoring and potential legal consequences for unauthorized access. The message ends with the user's name and session ID: "(base) asaravia@ARLAL0121011036 ~ %".

Command Syntax



- General command line syntax: command arguments
- Depending on the command used, arguments can be optional or required
 - Note: Arguments are also referred to as “options”, “parameters”, or “flags”
- All available arguments for a command can be found by looking at the command’s manual or help page, which can be opened in the terminal.
 - To view a command’s manual, we use the `man` function in Unix followed by the command of interest
 - Example: man <command_of_interest>
 - To view a command’s help page, we use the `--help` or `-h` flag for the command of interest
 - Example: <command_of_interest> --help
- Below is the manual of a common Unix command, mkdir, which was displayed in a terminal by running the command: man mkdir



A screenshot of a terminal window titled "asaravia — less + man mkdir — 192x27". The window displays the man page for the "mkdir" command. The text is in green on a black background. The man page includes sections for NAME, SYNOPSIS, and DESCRIPTION, along with detailed explanations of options like -m, -p, and -v.

```
MKDIR(1)           BSD General Commands Manual          MKDIR(1)

NAME
    mkdir -- make directories

SYNOPSIS
    mkdir [-pv] [-m mode] directory name ...

DESCRIPTION
    The mkdir utility creates the directories named as operands, in the order specified, using mode rwxrwxrwx (0777) as modified by the current umask(2).

    The options are as follows:

    -m mode
        Set the file permission bits of the final created directory to the specified mode. The mode argument can be in any of the formats specified to the chmod(1)
        command. If a symbolic mode is specified, the operation characters '+' and '-' are interpreted relative to an initial mode of ``a=rwx''.

    -p
        Create intermediate directories as required. If this option is not specified, the full path prefix of each operand must already exist. On the other hand,
        with this option specified, no error will be reported if a directory given as an operand already exists. Intermediate directories are created with permission
        bits of rwxrwxrwx (0777) as modified by the current umask, plus write and search permission for the owner.

    -v
        Be verbose when creating directories, listing them as they are created.

    The user must have write permission in the parent directory.

:
```

Running Unix Commands



- The SYNOPSIS section of the mkdir manual shows how the command should be run:

```
SYNOPSIS
  mkdir [-pv] [-m mode] directory_name ...
```

- For the mkdir command, `mkdir` is written first, followed by the flags indicated in the brackets, and ending with the directory name.
 - Note: Anything in brackets [] is optional and any part of the command not in brackets is required.
 - For the mkdir command, which parts are optional and which are required?
- Do you notice how the different parts of the command are separated? **With spaces!**
- On the command line, spaces are special
 - Spaces are used to know how to properly separate distinct parts of a command
 - For this reason, it's not ideal for file or directory (aka folder) names to contain spaces. Thus, it's better to use dashes (-) or underscores (_) instead of spaces when naming files and directories.
 - Example: “example_test1.txt” is preferred over “example test1.txt”

Running Unix Commands

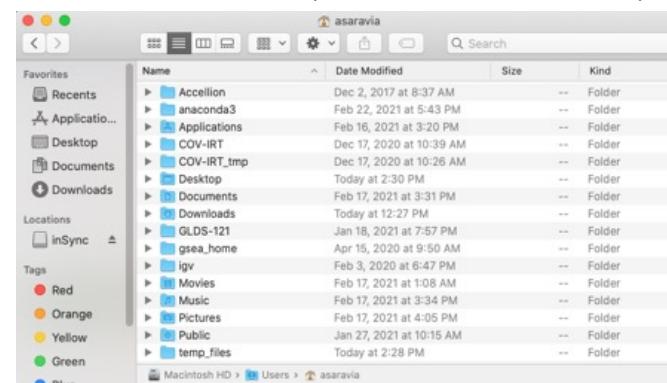


- Before we run the `mkdir` command to make a new directory, we first need to decide where to make the new directory.
- On computers, files are organized into directories (aka folders).
- When you open a terminal, it opens to a particular directory on your system, referred to as your “working (aka current) directory”.
- You can see what your working directory is by executing the `pwd` command, which stands for “print working directory”
- In the example below, you’ll see the output of the `pwd` command:

```
[(base) asaravia@ARLAL0121011036 ~ % pwd  
/Users/asaravia
```

- What is the current (aka working) directory?
- You can list what’s in your working directory with the `ls` command as shown on the left below (note: `ls` stands for “list”):

```
[(base) asaravia@ARLAL0121011036 ~ % ls  
Accellion    Documents      Music          igv  
Applications  Downloads      Pictures       temp_files  
COV-IRT      GLDS-121      Public  
COV-IRT_tmp   Library       anaconda3  
Desktop      Movies         gsea_home
```



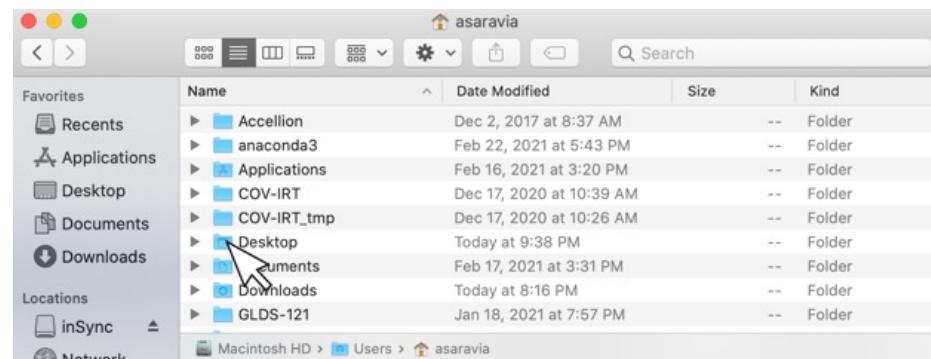
- We can also navigate to the `/Users/asaravia` directory using the Finder window as shown on the right above.
 - Do you notice any similarities between the files shown in `/Users/asaravia` on the command line versus the Finder window?

Running Unix Commands



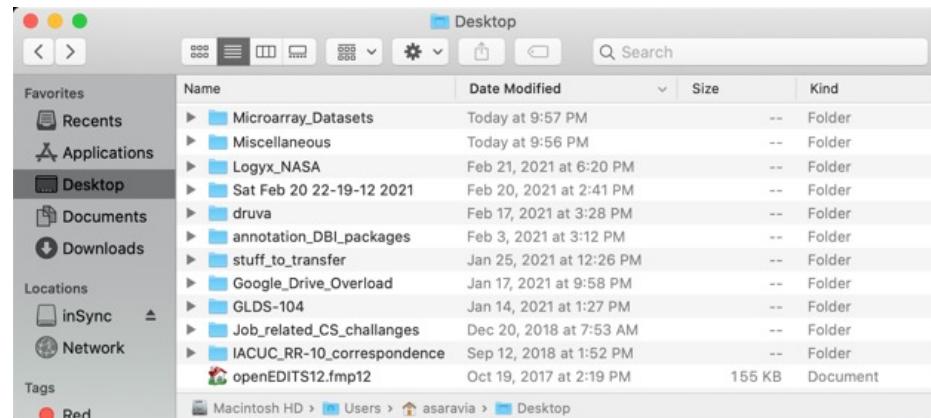
- Similar to how you can navigate to a different directory (aka folder) within your current directory by double clicking on the folder icon in the Finder window, you can use the `cd` command to change directories in your terminal (note: `cd` stands for “change directory”).
- Below, we will navigate to the Desktop using the command line (left) and in Finder (right):

```
[(base) asaravia@ARLAL0121011036 ~ % cd Desktop  
[(base) asaravia@ARLAL0121011036 Desktop % pwd  
/Users/asaravia/Desktop
```



- What is the working directory now?
- How are directories distinguished on the command line?
- Let's see what's on the Desktop using the `ls` command in the terminal below:

```
[(base) asaravia@ARLAL0121011036 Desktop % ls  
GLDS-104           Miscellaneous  
Google_Drive_Overload Sat Feb 20 22-19-12 2021  
IACUC_RR-10_correspondence annotation_DB1_packages  
Job_related_CS_challanges druva  
Logyx_NASA          openEDITS12.fmp12  
Microarray_Datasets stuff_to_transfer
```



- Compare the content on the Desktop in the terminal with what is shown in Finder



Running Unix Commands



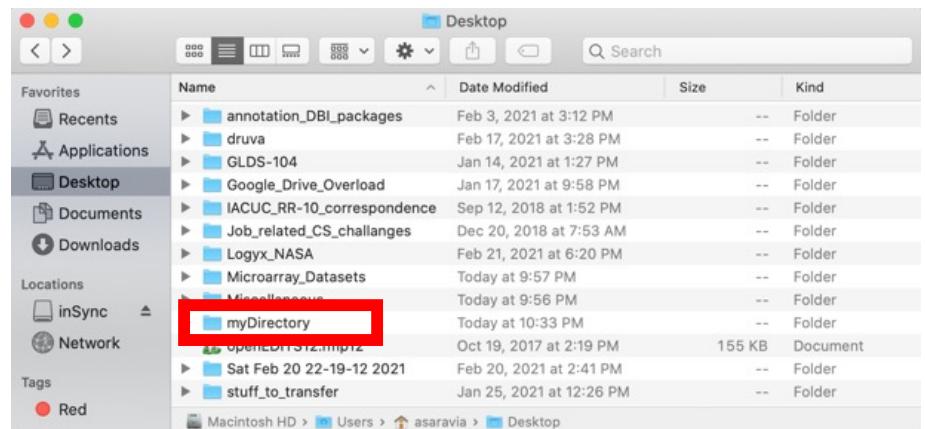
- Now that we've set the Desktop as our working directory, let's use the `mkdir` command to create a new folder on the Desktop.
- Recall the SYNOPSIS section of the mkdir manual, which shows how the command should be run:

```
SYNOPSIS
  mkdir [-pv] [-m mode] directory name ...
```

- In the example below, we'll make a directory titled "myDirectory" on the Desktop using the command line then view what's on the Desktop using both the command line and Finder:

```
[(base) asaravia@ARLAL0121011036 Desktop % mkdir myDirectory

[(base) asaravia@ARLAL0121011036 Desktop % ls
GLDS-104                               Sat Feb 20 22-19-12 2021
Google_Drive_Overload                   annotation_DB1_packages
IACUC_RR-10_correspondence            driva
Job_related_CS_challanges              GLDS-104
Logyx_NASA                             Google_Drive_Overload
Microarray_Datasets                     IACUC_RR-10_correspondence
Miscellaneous                          Job_related_CS_challanges
myDirectory                            Logyx_NASA
stuff_to_transfer                      Microarray_Datasets
                                         Miscellaneous
                                         openED1TS12.1mp12
                                         stuff_to_transfer
```



Note: The `mkdir` command is the equivalent of creating a new folder in Finder.

Creating Variables



- There will be times when we need to set our working directory to a folder that is deep within a directory structure. Rather than navigating one directory at a time, like you would in Finder, you can navigate directly to your folder of interest with a single `cd` command in the terminal by writing the explicit path to that directory as shown:

```
[(base) asaravia@ARLAL0121011036 ~ % pwd  
/Users/asaravia  
[(base) asaravia@ARLAL0121011036 ~ % cd /Users/asaravia/Desktop/myDirectory  
[(base) asaravia@ARLAL0121011036 myDirectory % pwd  
/Users/asaravia/Desktop/myDirectory
```

- If there is a location that we need to visit often, rather than writing out the explicit path every time, we can create a variable containing the explicit path. In Unix, you can do this using an equal sign to assign the path to a variable name as shown:

```
[(base) asaravia@ARLAL0121011036 ~ % pwd  
/Users/asaravia  
[(base) asaravia@ARLAL0121011036 ~ % myDir=/Users/asaravia/Desktop/myDirectory  
[(base) asaravia@ARLAL0121011036 ~ % echo $myDir  
/Users/asaravia/Desktop/myDirectory  
[(base) asaravia@ARLAL0121011036 ~ % cd $myDir  
[(base) asaravia@ARLAL0121011036 myDirectory % pwd  
/Users/asaravia/Desktop/myDirectory
```

- What command was introduced in the example above?
- What was added to the variable name when using the variable in Unix?

Creating Variable Names



- When creating variable names there are a few import things to keep in mind:
 - Do not name your variable after a common function or a variable built-in to the coding language being used
 - Avoid using spaces and any characters other than letters, numbers, dashes, and underscores – Why?
 - Choose informative, memorable, and short variable names
- Example: You want to create a variable to store gene expression data from Drosophila melanogaster (aka fruit flies) that were grown at 10°C, 20°C, or 30°C. Which of the names below would be good variable names to use?
 1. GeneExpressionData_Drosophilamelanogaster_DifferentTemperatures
 2. Dmel_GE_dT
 3. FruitFlies_GE_dT
 4. FruitFlies_GE_10°C_20°C_30°C
 5. echo
 6. myData
 7. Dmel\$GE\$dT
 8. Dmelanogaster GE Data

Jupyter Notebook



- Throughout this bootcamp, we will be running several of the commands we just discussed. However, since working directly on the command line takes time to get used to, we will be using Jupyter Notebooks to run our commands.
- Jupyter Notebook (JN) was created as part of Project Jupyter (<https://jupyter.org/>) and is a web application that allows you to create documents that have both narrative text and live code; meaning you can run commands through a web browser (instead of a terminal).
- Additional advantages of JNs:
 - Facilitates sharing code and results between groups
 - Ease of publishing your code
 - Allows for reproducibility

4c. Volcano Plot

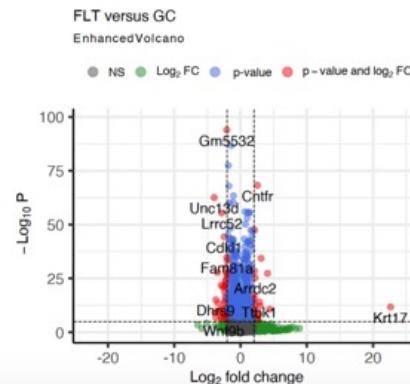
Finally, let's make a volcano plot to identify a few interesting genes. A volcano plot is a scatterplot which shows the relationship of the adjusted p-value significant by adjusted p-value are labeled.

First, we'll use the default settings from the `EnhancedVolcano()` function: log₂ Fold Change cutoff > |2|, and the adjusted p-value cutoff is < 10e-6.

You can read about `EnhancedVolcano()` and see some examples by clicking [here](#)

```
# Volcano plot showing genes differentially expressed in FLT vs GC
EnhancedVolcano(DGE_output_table,
  lab = DGE_output_table$SYMBOL,
  x = 'Log2fc_(FLT)v(GC)',
  y = 'Adj.p.value_(FLT)v(GC)',
  title = 'FLT versus GC',
  pCutoff = 10e-6,
  FCcutoff = 2,
  pointSize = 3.0,
  labSize = 6.0,
  colAlpha=0.5)

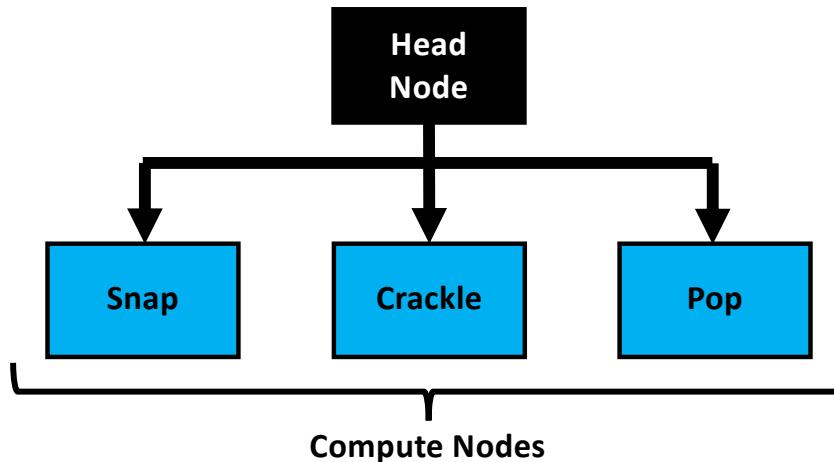
## Save your volcano plot
ggsave(file.path(DGE_plots,'GLDS-104_volcano_DGE.png'), width = 6, height = 8.5, dpi = 300)
```



Compute Cluster

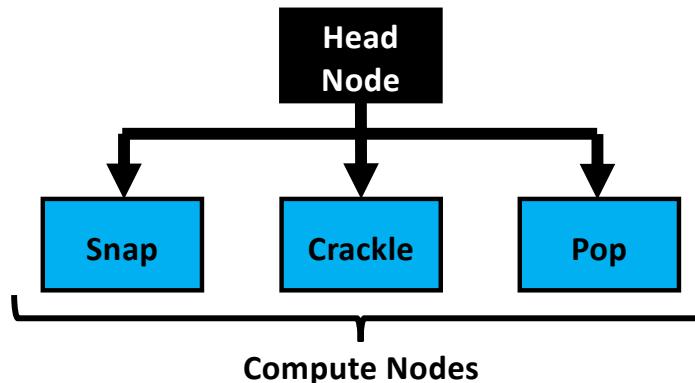


- When we execute commands in our JNs, they will be run on a simulated compute cluster.
- A compute cluster is a set of servers, which are like computers, that are all connected, and each server is used for a specific purpose. Below is an example of a relatively simple cluster containing 4 servers (aka nodes) – 1 head node and 3 compute nodes:



- **Head node:** The server that allows you to connect to a cluster (aka the login server). You should not run any commands that require a lot of computational resources on the head node. Why do you think that is?
- **Compute nodes:** Servers that are used to run commands, programs, and scripts (aka jobs). Compute nodes usually contain more computational resources than the head node.

Compute Cluster



- There are often multiple people logged into a cluster at the same time, and they all want to run jobs. This could be a problem for the servers in the cluster. Why?
 - Hint: Think about what happens when you try to run a bunch of applications on your personal computer. Now imagine multiple people each running several applications on a server...
- To avoid multiple people running jobs all at the same time and risking freezing the servers, compute clusters are equipped with a job scheduler.
- Rather than executing jobs directly, users submit their jobs to the job scheduler, which will check each compute node to see if there are enough resources available to run the job. If there are, the job will begin running on one of the compute nodes; if not, the job will get queued and will not start until enough resources are available.
- During this bootcamp, we will be “submitting” all computationally expensive jobs to simulated cluster job scheduler, SLURM.

SLURM Array Script



```
#!/bin/bash
#
#SBATCH --job-name="GLDS-104_raw_fastqc"
#SBATCH --output=raw_fastqc_out_logs/GLDS-104_raw_fastqc_%A_%a.out
#SBATCH --partition=priority
#SBATCH --mem=6000
#SBATCH --array=1-12
#
#SBATCH --mail-user=user@nasa.gov
#SBATCH --mail-type=END

. ./profile

echo "GLDS-104_raw_fastqc"
echo ""

start=$(date +%s)
echo "start time: $start"
echo $HOSTNAME
echo "My SLURM_ARRAY_TASK_ID: " ${SLURM_ARRAY_TASK_ID}

source activate /miniconda38_admin/envs/fastqc_09-2021

in_dir=/GLDS-104/00-RawData/Fastq
out_dir=/GLDS-104/00-RawData/FastQC_Reports

sample=$(cat samples.txt | sed -n ${SLURM_ARRAY_TASK_ID}p)

echo ""
echo "SAMPLE: ${sample}"

echo ""
echo "FastQC version: "
fastqc --version
echo ""

call="fastqc -o $out_dir ${in_dir}/${sample}_R*_raw.fastq.gz"

echo $call
echo ""
eval $call

echo ""
end=$(date +%s)
echo "end time: $end"
runtime_s=$(echo $(( end - start )))
echo "total run time(s): $runtime_s"
sec_per_min=60
sec_per_hr=3600
runtime_m=$(echo "scale=2; $runtime_s / $sec_per_min;" | bc)
echo "total run time(m): $runtime_m"
runtime_h=$(echo "scale=2; $runtime_s / $sec_per_hr;" | bc)
```

Now, let's view an example of the standard output from this command:

https://processing-00.readthedocs.io/en/latest/processing_00.html

SLURM Array Script – SLURM commands



```
#!/bin/bash ← Specifies that the Unix terminal is running bash
#
#SBATCH --job-name="GLDS-104_raw_fastqc" ← Specifies the job name
#SBATCH --output=raw_fastqc_out_logs/GLDS-104_raw_fastqc_%A_%a.out ← Instructs SLURM where to write standard output
#SBATCH --partition=priority ← Indicates which queue to add the job to
#SBATCH --mem=6000 ← Specifies the required RAM in Mb
#SBATCH --array=1-12 ← Indicates the job numbers to run
#
#SBATCH --mail-user=user@nasa.gov ← Specifies the e-mail address to send SLURM notifications
#SBATCH --mail-type=END ← Instructions SLURM to send an e-mail when jobs complete
```

SLURM Array Script – nonSLURM commands



```
. ~/profile ← Instructions bash to use the user profile settings
echo "GLDS-104_raw_fastqc" } Prints the job name followed by blank line to the standard output
echo ""

start=$(date +%) } Records then prints the job start date in seconds to the standard output
echo "start time: $start"
echo $HOSTNAME ← Prints the name of the server running the job to standard out
echo "My SLURM_ARRAY_TASK_ID: " $SLURM_ARRAY_TASK_ID ← Prints the SLURM job ID number to standard out

source activate /miniconda38_admin/envs/fastqc_09-2021 ← Activates the required conda environment

in_dir=/GLDS-104/00-RawData/Fastq
out_dir=/GLDS-104/00-RawData/FastQC_Reports } Specifies the directory holding the input files and where to print the output files

sample=$(cat samples.txt | sed -n ${SLURM_ARRAY_TASK_ID}p) ← Links each array job number to a sample name
echo ""
echo "SAMPLE: ${sample}" } Prints a blank line followed by the sample name to standard out

echo ""
echo "FastQC version: " } Prints a blank line followed by the FastQC tool version number
fastqc --version
echo "" and then another blank line to standard out
```

SLURM Array Script – nonSLURM commands



```
call="fastqc -o $out_dir $in_dir/${sample}_R*_raw.fastq.gz" ← Stores the fastqc command as a variable
```

```
echo $call  
echo "" } Prints the exact fastqc command to standard out followed by a  
eval $call blank line then executes the fastqc command
```

```
echo ""  
end=$(date +%s)  
echo "end time: $end"  
runtime_s=$(( end - start ))  
echo "total run time(s): $runtime_s"  
sec_per_min=60  
sec_per_hr=3600  
runtime_m=$(echo "scale=2; $runtime_s / $sec_per_min;" | bc)  
echo "total run time(m): $runtime_m"  
runtime_h=$(echo "scale=2; $runtime_s / $sec_per_hr;" | bc)
```

} Prints a blank line followed by the job end date in seconds then calculates and prints the job run time in seconds, minutes, and hours to standard out



Questions?