Space Biosciences
NASA AMES RESEARCH CENTER

Developing a Genetic Variant Calling Pipeline for Quantifying the Complex Mutagenic Load Accumulated in BioNutrients-1 Production Pack samples

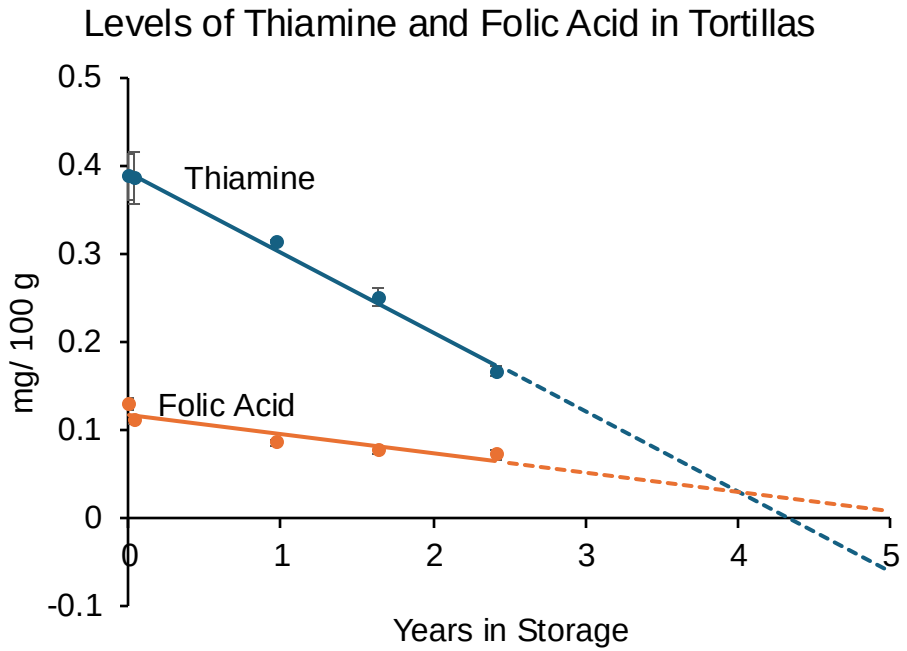
Philip Sweet, Natalie Ball, Barbara S. F. Müller, Sandra Vu, Lisa Anderson, Sadie Downing , Amy Gresser, Aditya Hindupur, John Hogan, Hiromi Kagawa , Aphrodite Kostakis, Matthew Paddock, Hami Ray, Oscar Roque, Sean Sharif, Kevin Sims, Mathangi Soundararajan, Alyssa Villanueva, Junya Zhang, Nicole S. Beisel, Kevin Tyre, Hope L. Hersh, Fang Bai, Frances Donovan, **A. Mark Settles**

Pre-packaged foods in the NASA Food System have a limited shelf life

The NASA food system needs to have a 5-year shelf-life to go to Mars.

Current pre-packaged foods lose palatability and essential nutrients during long-term storage.

Food Stabilization Processes
Dehydrated
Thermostabilized
Extended shelf-life baked goods



Adapted from Zwart et al (2009) *Food Science* 74(7):H209-H217. https://doi.org/10.1111/j.1750-3841.2009.01265.x

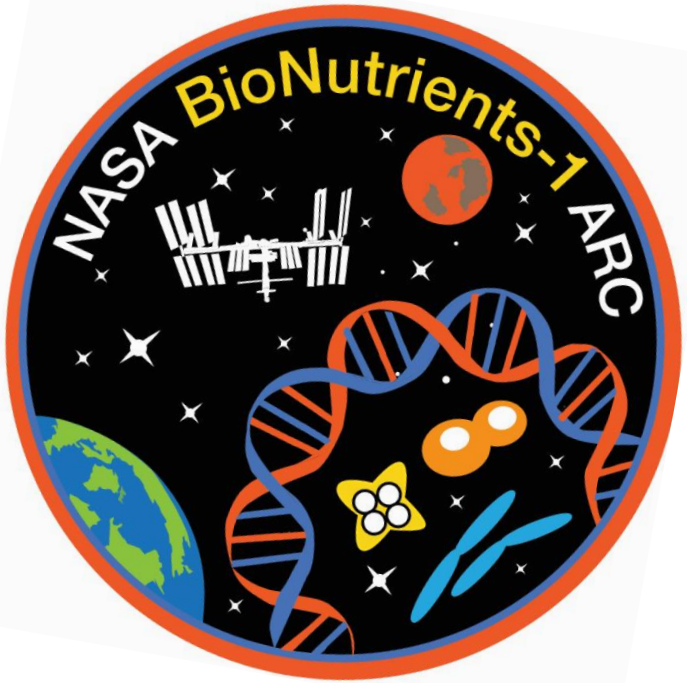
BioNutrients Spaceflight Experiment Series

5-Year ISS Storage-Reactivation Demonstration – NG-11 (04/2019)

BN-1 five-year flight project currently on the International Space Station

- Testing the long-term storage of various microorganisms for the biomanufacturing of space-relevant compounds.
- Validating the performance of the first generation of production packs.

← Focus of this talk



6-Month ISS Reactivation Demonstration – SpX-26 (11/2022)

BN-2 expands the BN-1 flight project scope

- Introducing novel products.
- Broadening the range of microbial food sources.
- Improving upon production pack hardware.

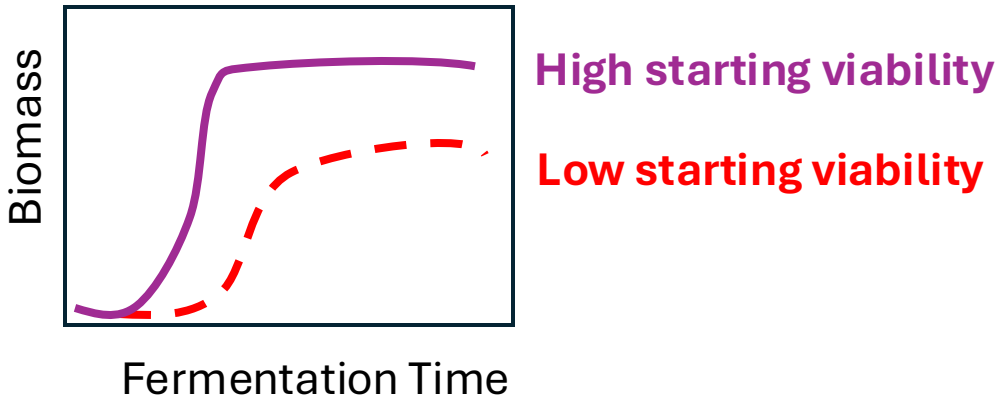
Future flight project – Slated for flight in 2025

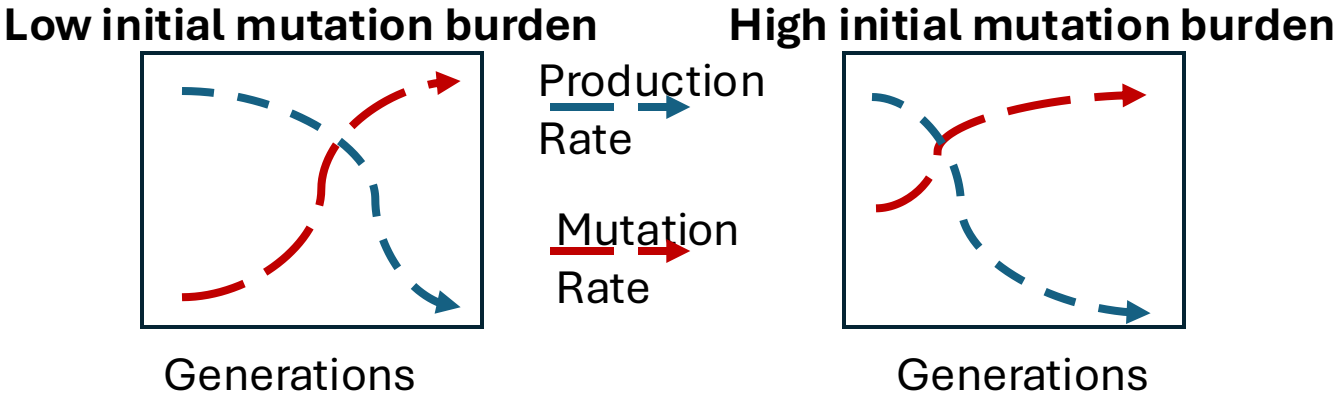
BN-3 expands products and food safety

- Multi-nutrient production in a single bioreactor.
- HACCP plan development and food safety testing.

Reliably producing a nutrient during fermentation requires a predictable starting viability and genomic fidelity

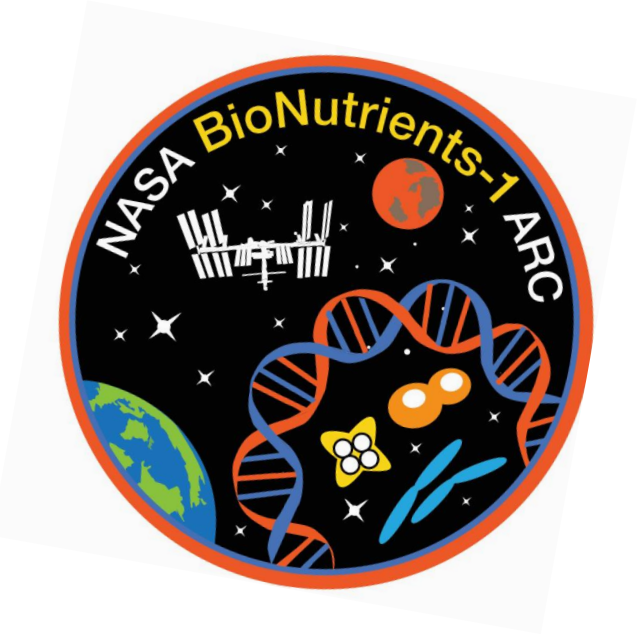
- A decreased starting viability can lengthen the lag-phase and lower the end-state biomass and total nutrient yield.
- An increased mutation rate can disrupt key metabolic genes, lowering the overall fitness of the culture, growth rate and nutrient yield.
- A high initial mutation burden in the production genes can lead to non-producing cells taking over the fermentation.





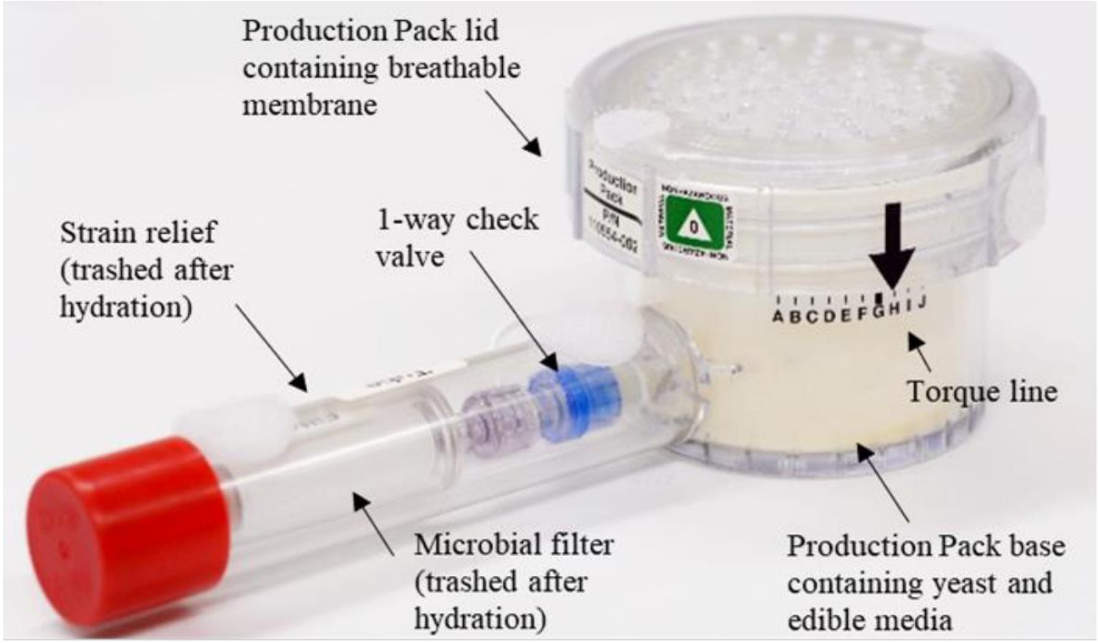
BN-1 seeks to understand the unique risks posed to microbial systems that could be used for long-duration space missions

- How will the viability and genomic integrity of dried cells change over 5 years of storage?
- How will media change or degrade over 5 years of storage and what impact will that have on fermentation?
- At what rate will mutations accumulate in the production organism genome?
- At what rate will mutations accumulate in the genomically engineered pathway?
- What are the impacts of LEO specific factors, such as microgravity, on storage and fermentation?



Simplified bioreactors (Production Packs) were flown to the ISS as part of the BN-1 mission

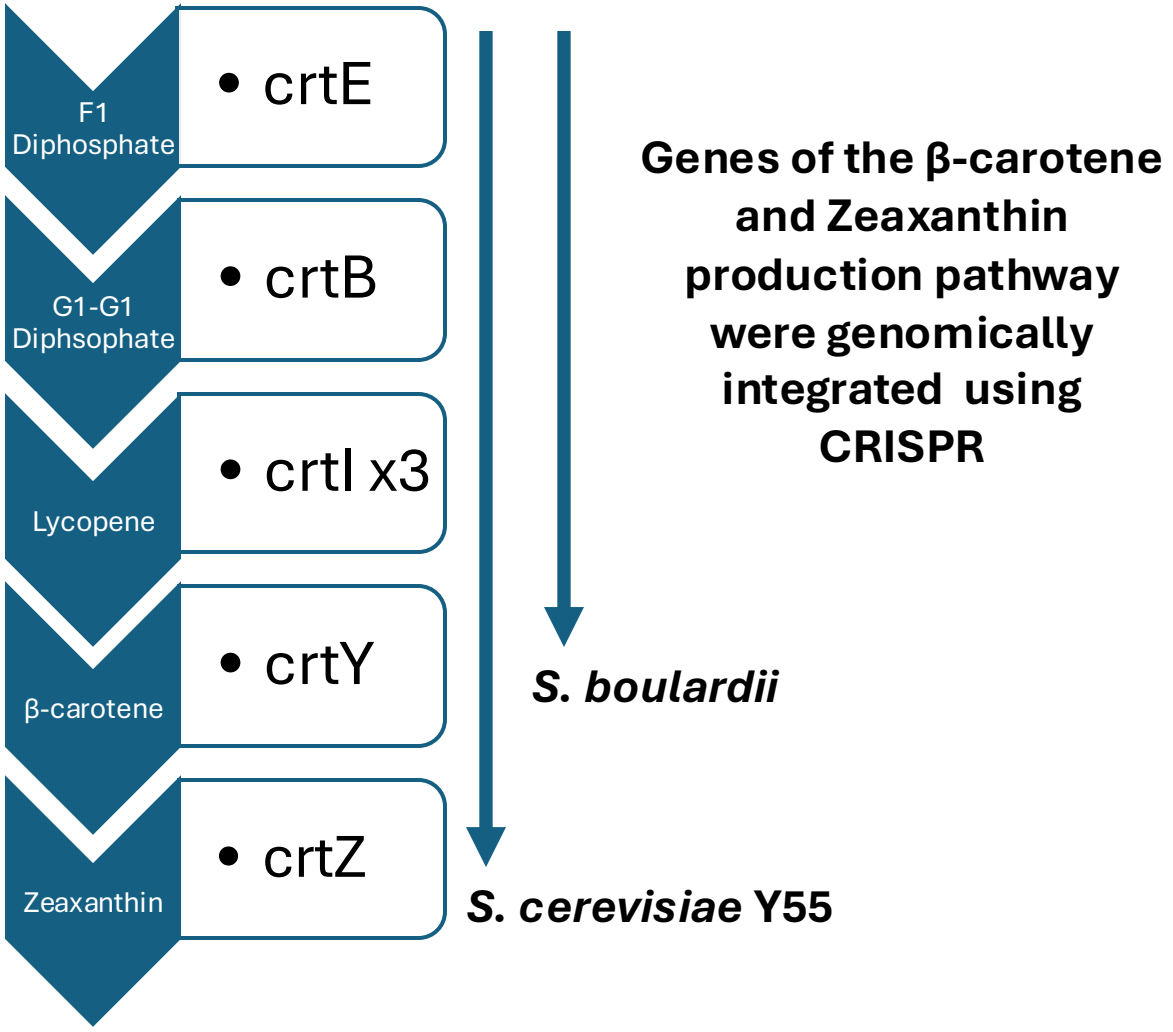
- Production Packs are self-contained fermentation system that include dried media and cell pellets.
- Include either a β-carotene producing strain of *S. c. boulardii* or *a* Zeaxanthin producing strain of *S. cerevisiae* Y55.
- Upon rehydration, Production Packs are shaken and incubated at 30°C for 48h of fermentation.



For BN-1, two strains of *S. cerevisiae* were genetically engineered to produce antioxidants

Strains

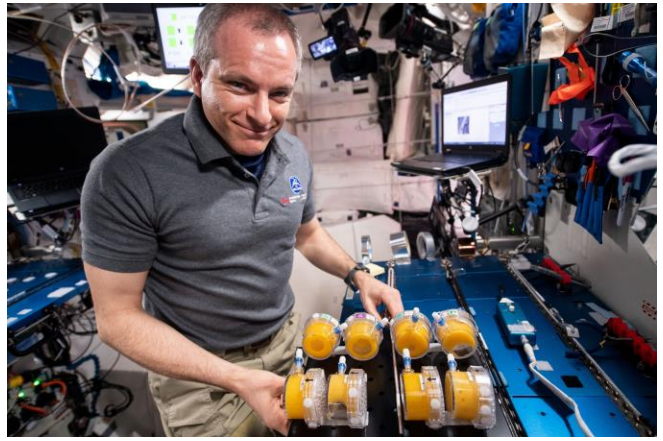
- ***S. c. boulardii***
  - Probiotic, approved for consumption
  - Produces: β-carotene
  - Non-spore-forming
- ***S. cerevisiae* Y55**
  - Produces: Zeaxanthin
  - Spore-forming



The Production Packs experiment captures the effects of storage and fermentation in Low-Earth Orbit (LEO)

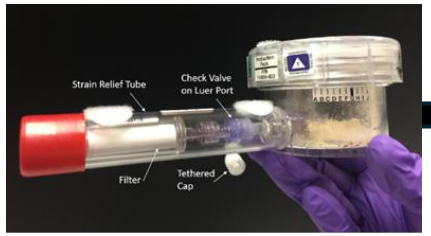
- Six sets of fermentations were conducted over ~5 years of storage on the ISS.
- Paired fermentations were conducted on the ISS and on ground to capture difference due to storage conditions (Ground vs LEO) and fermentation conditions (Ground vs LEO vs Aeration).

Sample	Storage	Growth Condition
IPP	ISS	ISS – SABL incubator
IPPgc	Ground	Stationary, Ground
EPP	ISS	Stationary, Ground
Shaken GC	Ground	Shaken/aerated, Ground



The end-state of the Production Packs fermentations has been characterized using a multi-omics approach

Assemble Production Packs with media and desiccated cells

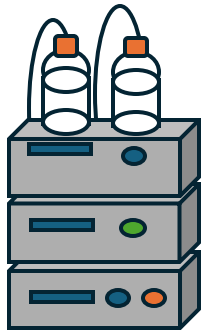


Store on the ISS or on Ground

Rehydrate a batch ~9 months

X4

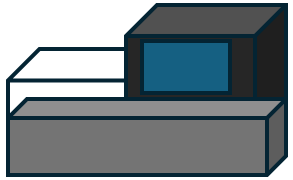
End-State Biomass



Carotenoid Production via HPLC

End-State EtOH

End-State Glucose



Mutation Rates via Whole Genome Sequencing (WGS) of DNA

Differential Expression via RNA sequencing

Whole Genome Sequencing (WGS) will be used to assess the impact of storage and growth in LEO on genomic fidelity of *S. cerevisiae* after fermentation

IPP vs IPPgc
Impact of fermentation in LEO on genomic integrity

IPP vs EPP
Impact of storage in LEO on genomic integrity

T1 vs T3 vs T5 vs T6
Impact of storage length on genomic integrity

***S. boulardii* vs *S. cerevisiae* Y55**
Impact of strain on genomic integrity

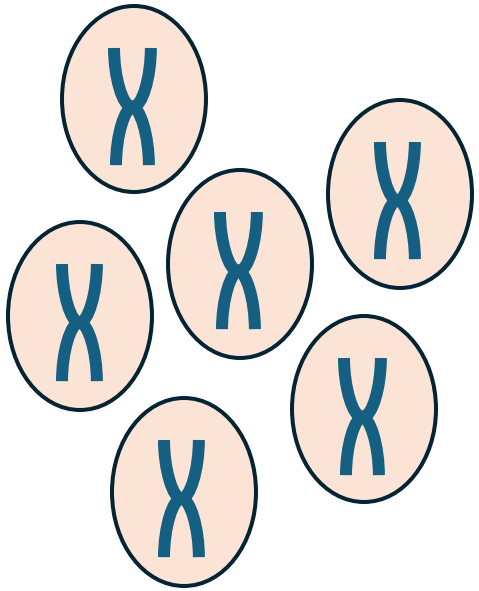
Sample	Storage	Growth Condition
IPP	ISS	ISS – SABL incubator
IPPgc	Ground	Stationary, Ground
EPP	ISS	Stationary, Ground
Shaken GC	Ground	Shaken/aerated, Ground

Sequencing the DNA extracted from the end-state fermentations allows for quantification and characterization of genome mutations

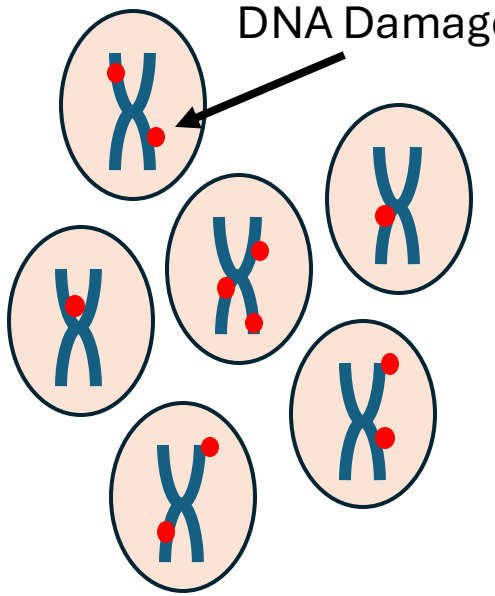
Mutations are generally classified by size and impact on proteins

- **SNP: Single Nucleoid Polymorphism**
  - **Coding:** Located within the coding region of a gene.
    - Synonymous: Doesn't impact the amino acid sequence of the protein
    - Non-synonymous: Introduces an amino acid change
    - Nonsense: Introduces a stop codon, resulting in a truncated protein
  - **Non-coding:** Located in-between coding regions and thus no direct impacton protein coding. Can have indirect effects on regulation and mRNA processing.
- **Indel: Insertion or Deletion**
  - Vary in size.
  - **Frameshift:** Alters the reading frame of a protein
  - If within a protein coding region, can result in a misfolded protein.
  - If within a regulatory region, can results in mis-regulation of nearby genes.

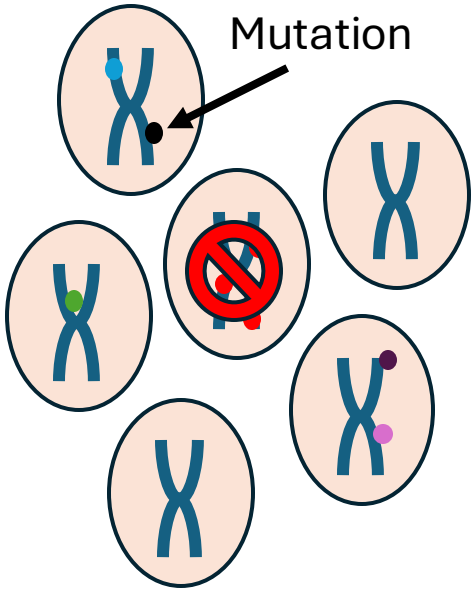
Repairing DNA damage caused by desiccation and prolonged storage is expected to result in a heterogeneous population



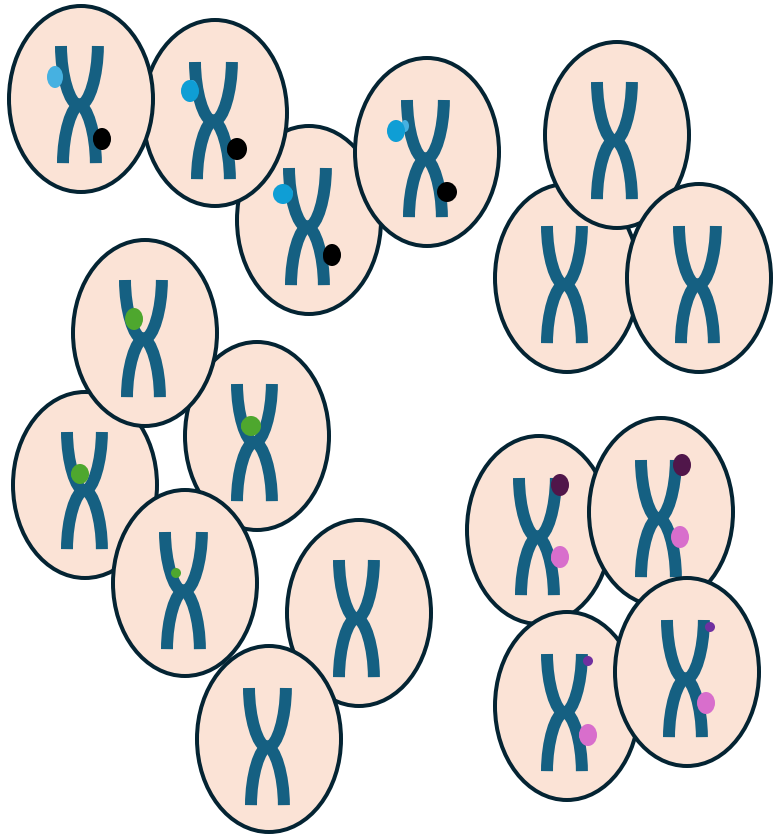
Production Packs start out as genetically homogeneous populations from a common culture



Both the desiccation process and long-term storage result in unique DNA damage events

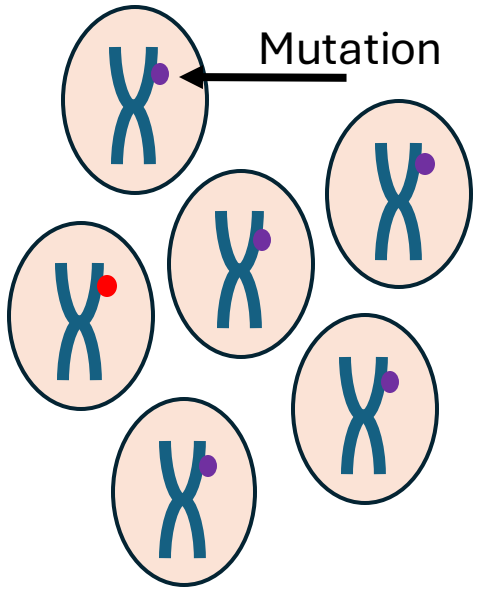


Upon rehydration, cell will attempt to repair DNA damage, sometimes resulting in DNA mutations or cell death



Daughter cells will carry on the unique mutations, resulting in a heterogeneous population

The complex nature of the total DNA extracted from the Production Packs demands a pooled approach to variant calling



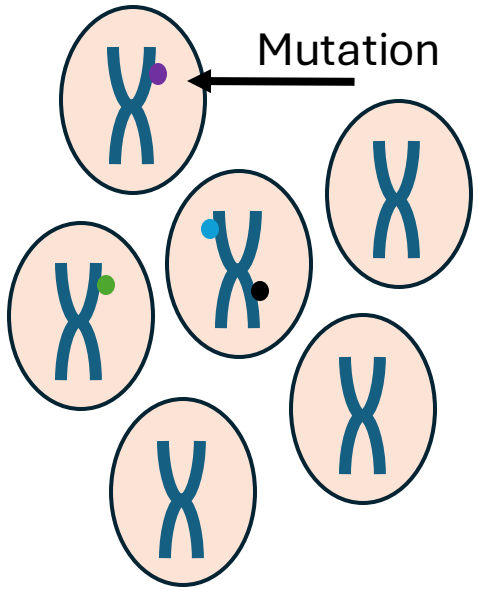
WGS Reads

AGCAGACGT
GGCGTAATCT
AGCAGACGTAAT
GACGTAATCT
GACGTAATCTGA
AGCAGTCGTAATCTGA

Reference Genome

Homogenous Mutations

- Within a homogenous population, such as a liquid culture grown from a single colony, a “True” mutation will be present in the majority of reads (A).
- Low abundance variants are assumed to be sequencing or PCR errors (G)



WGS Reads

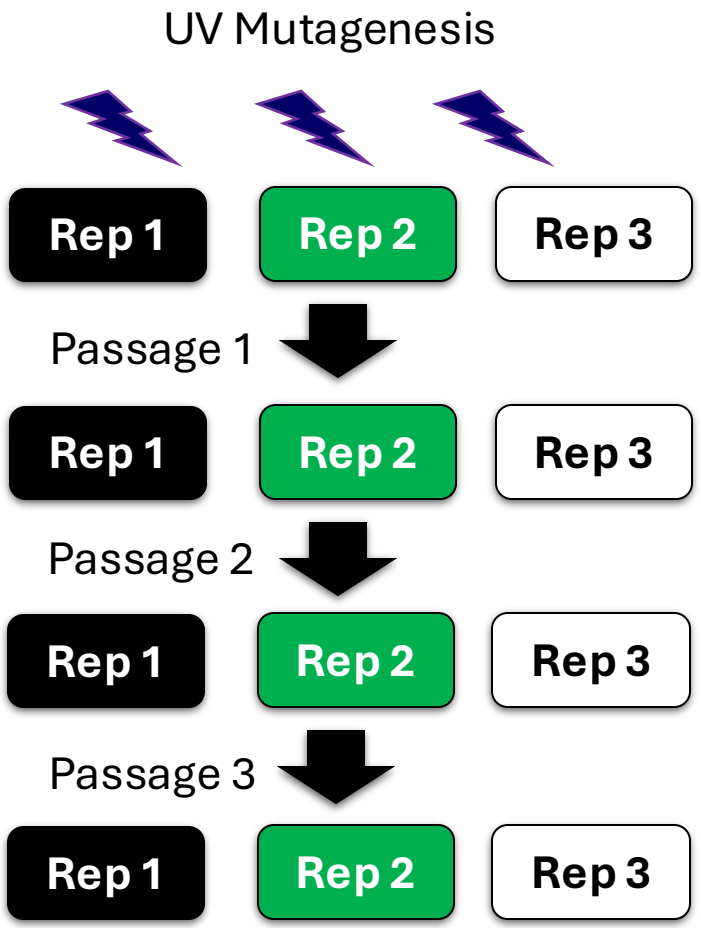
AGCAGACGT *Mutation*
AGCAGACGT
GGCGTAATCT
AGCAGTCGTAAT
GTCGTAATCT
GTCGTAATCTGA
AGCAGTCGTAATCTGA

Reference Genome

Heterogenous Mutations

- Within a heterogenous population, such as the Production Packs samples after rehydration, a “True” mutation (A) would only be present in a subset of reads
- Filtering is required to retain likely mutations without including excess sequencing noise (G)

The Space Algae-1 Mission included a similarly complex genomic sample set

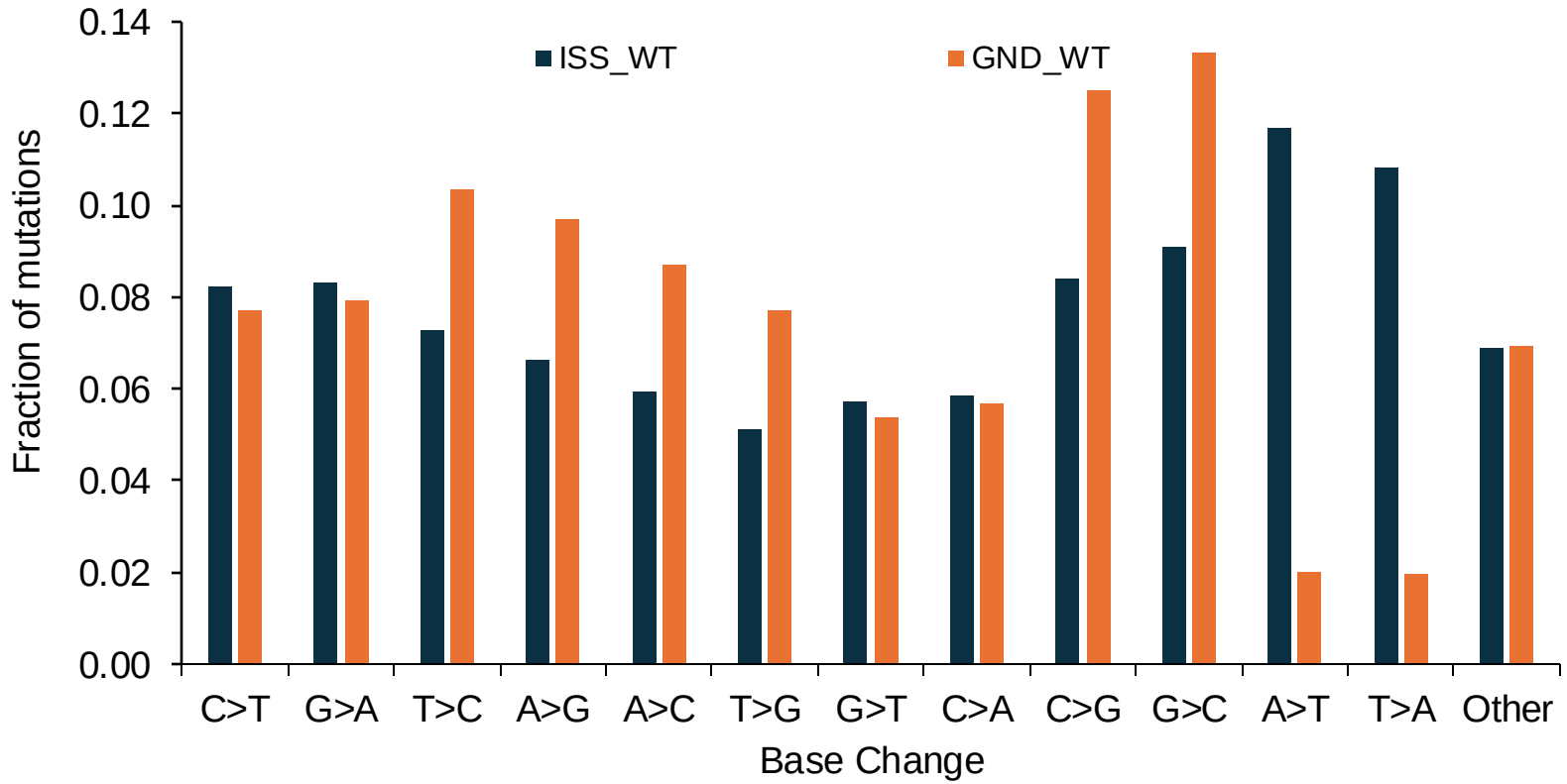


Method

1. Mutagenize with UVA to 5-10% survival.
2. Inoculate ~10^6 viable cells in 100 mL. media.
3. Store in the dark at ambient temperature for 5-7 days for loading, launch, and delivery to ISS.
4. **In space:** Grow 1st culture for 7 days in VEGGIE.
5. Passage 0.5 mL, 3 times to get 4 cultures. ~10 mitotic cell divisions/culture.
6. Store live cultures in dark at ambient temperature.
7. Returned to Ground for processing and comparison with ground controls

The variant caller CRISP was used to process the Space Algae-1 WGS data and identified a space-specific mutation signature

CRISP[1]: Comprehensive **R**ead **A**nalysis for **I**dentification of SNVs (and short indels) from **P**ooled sequencing data



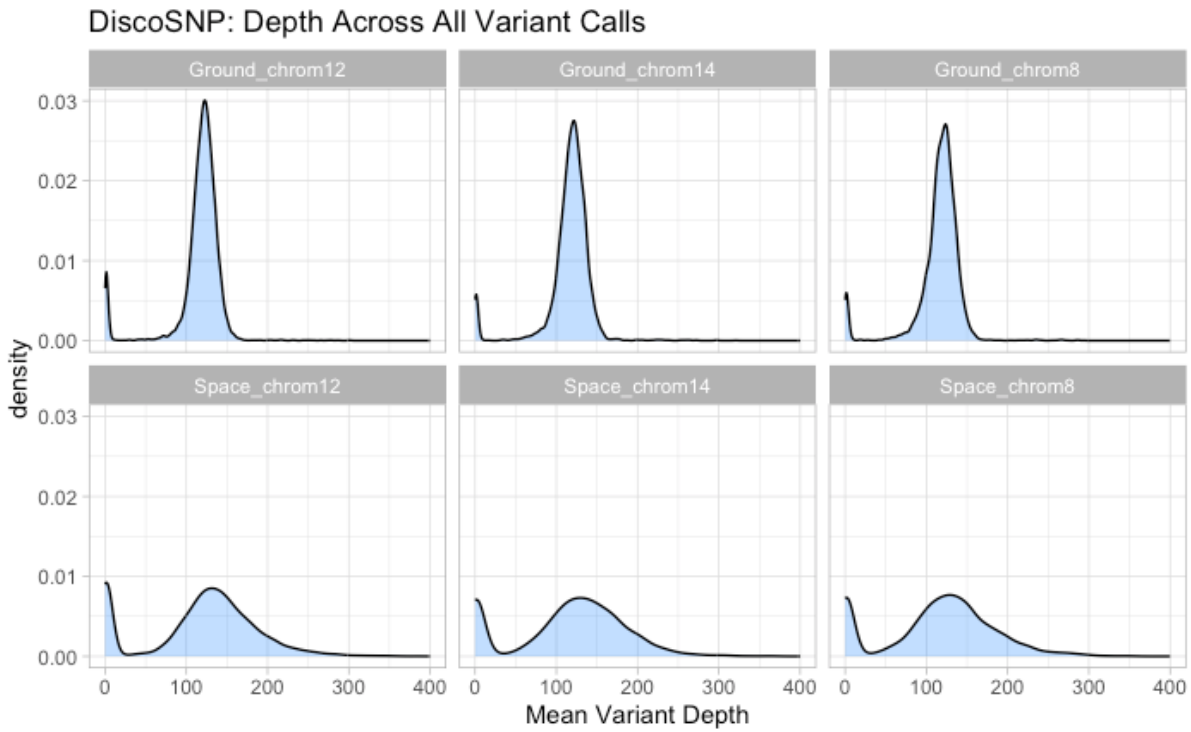
1. Bansal V. A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics. 2010

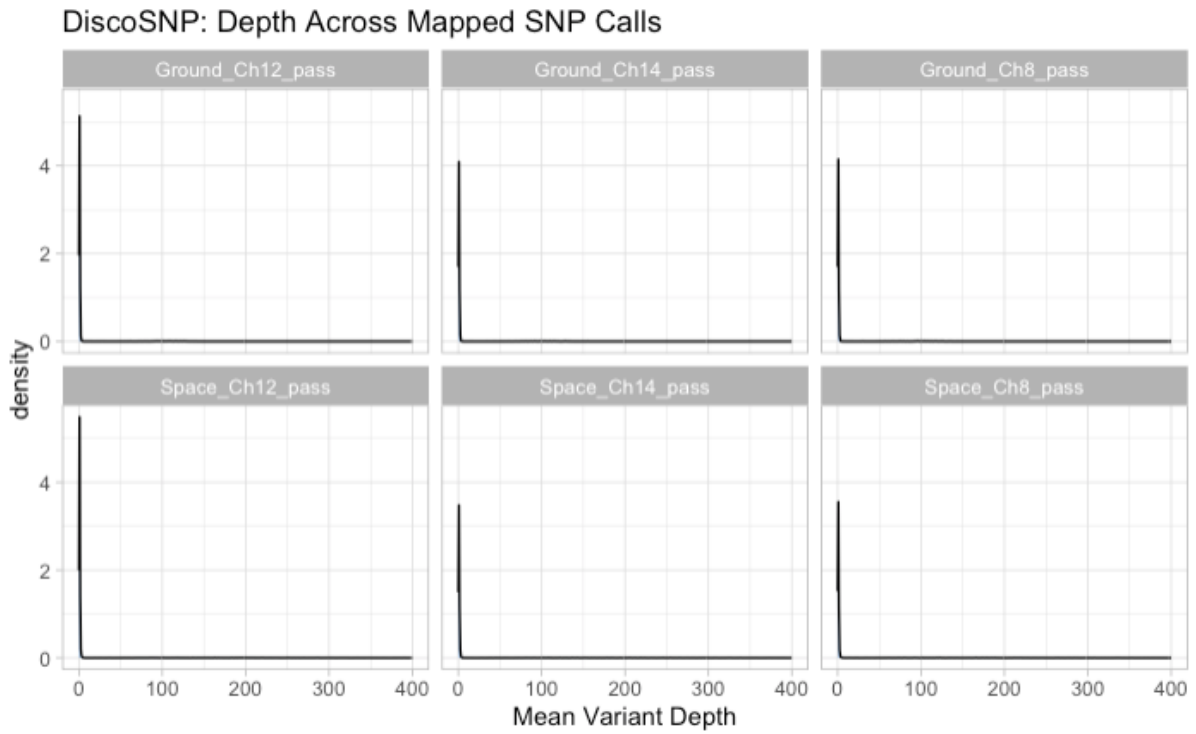
In preparation for the Production Pack WGS analysis, three variant calling tools were benchmarked against CRISP, using the Space Algae-1 dataset

- **GATK for Microbes**[1]
  - An adaption of the **Mutect2** Variant caller designed for bacterial samples where low allele frequencies, varying read depths are expected.
  - Released in 2021
- **DiscoSNP**[2]
  - Reference free approach that detects variants by assembling reads and identifying mismatches.
  - Released in 2017
- **BreSeq**[3]
  - A full-service pipeline, including initial alignments, realignments and variants calling.
  - Utilizes samtools mpileup for variant detect
  - Released in 2014

1. Introducing GATK for microbes. (GATK Team). https://gatk.broadinstitute.org/hc/en-us/articles/360060004292-Introducing-GATK-for-Microbes
2. DiscoSnp++: de novo detection of small variants from raw unassembled read set(s). Pierre Peterlongo, Chloé Riou, Erwan Drezen, Claire Lemaitre. bioRxiv 209965
3. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 2014

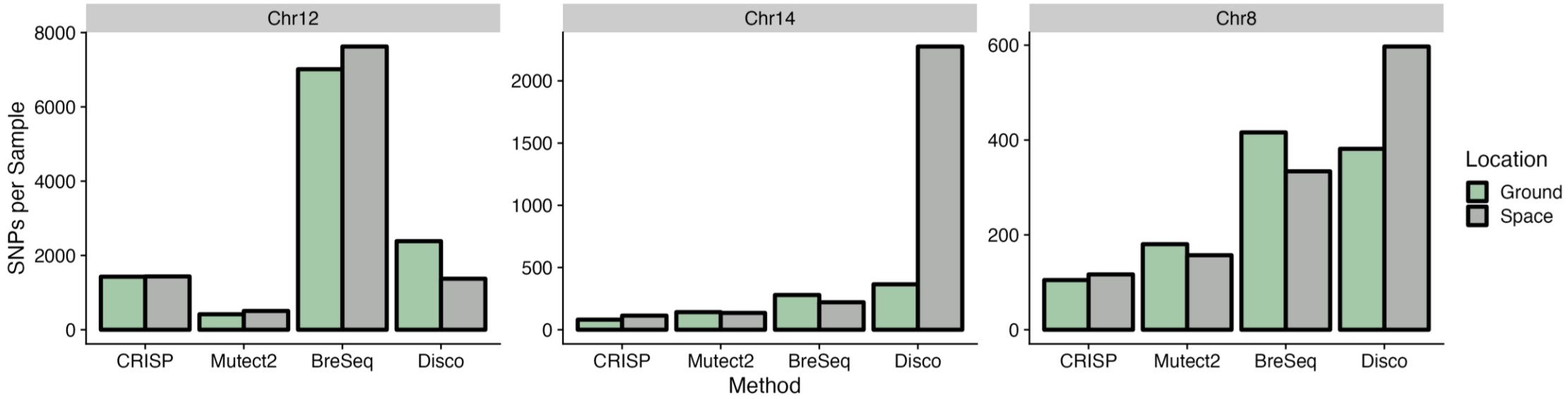
DiscoSNP, the “reference free” variant detection tool, failed to identify mappable variants





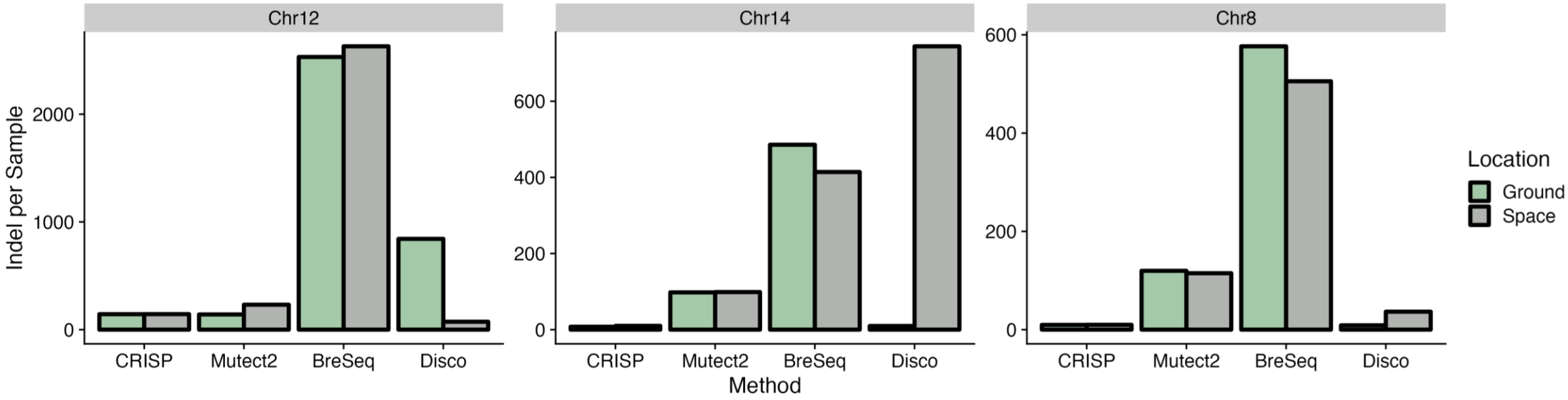
The majority of variants detected by DiscoSNP do not map back to the reference genome

BreSeq and DiscoSNP were less selective than CRIPS and GATK for Microbes in calling SNPs



The GATK for Microbes pipeline utilize the Mutect2 algorithm and detected a similar number of SNPs as CRISP in the Space Algae-1 dataset

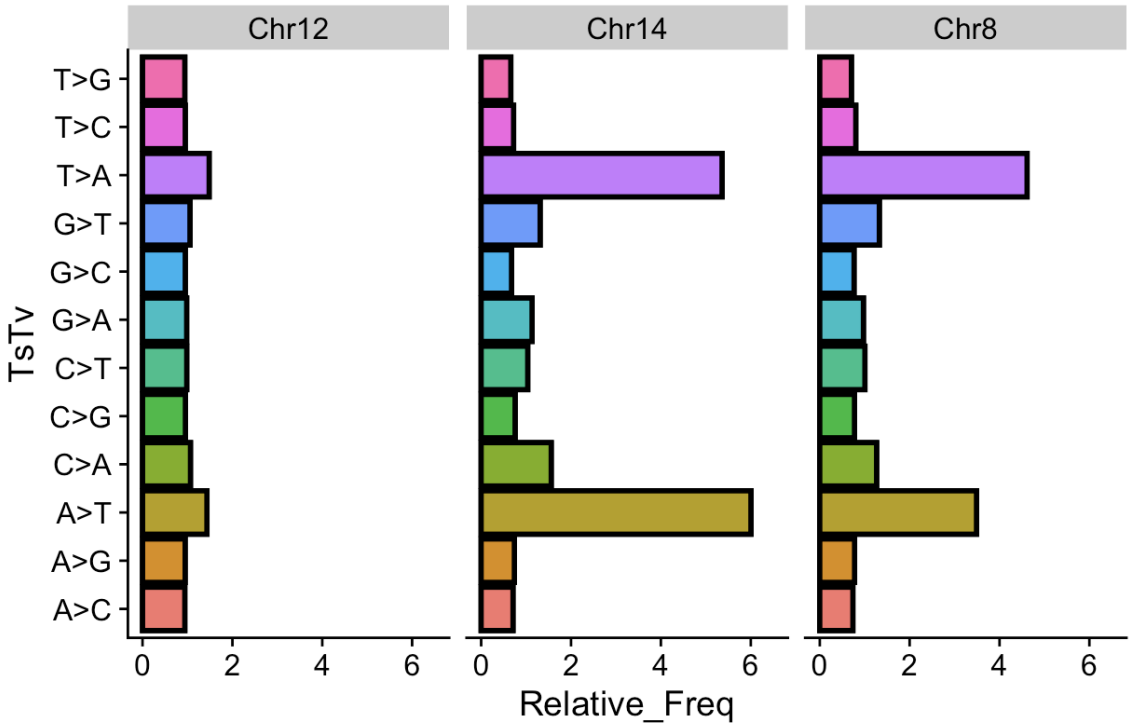
BreSeq and DiscoSNP were less selective than CRIPS and GATK for Microbes in calling Indels



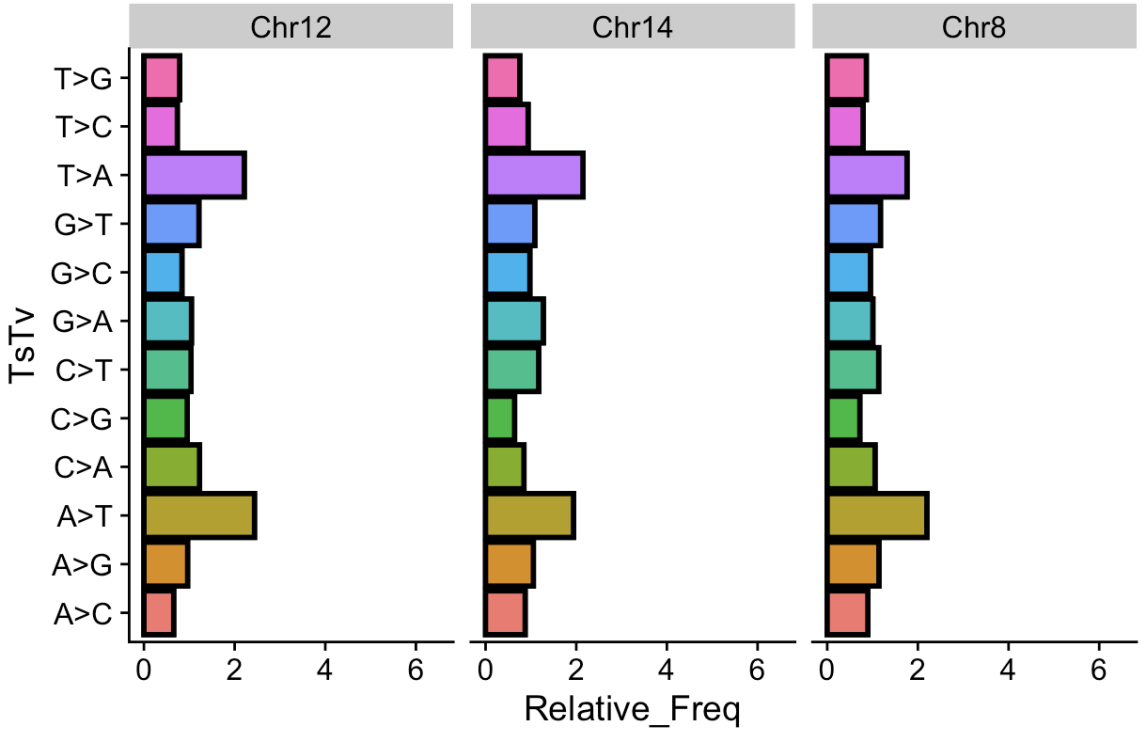
The GATK for Microbes pipeline utilize the Mutect2 algorithm and detected a similar number of Indels as CRISP in the Space Algae-1 dataset

The GATK for Microbes variant caller best recovers the A/T mutation bias detected by CRISP

Ratio of CRISP ”High Quality” SNP



Ratio of GATK for Microbes ”High Quality” SNP



GATK for Microbes best matched the CRISP results and will be utilized for variant calling the Production Pack WGS data

- **DiscoSNP:** Very fast, but the majority of reads accumulated on unmappable sequences and total variant calls are inconsistent with CRISP.
- **BreSeq:** Took the most computational time, detected the greatest number of SNPs and Indels, poor alignment with CRISP results and did not recover the previously observed T/A mutation bias.
- **GATK for Microbes:** Detected a comparable number of SNPs as CRISP as well as slight bias toward A/T mutations in the ratio of TsTv's.

Future Direction

- Classification of Production Pack variants into mutation categories using the SNPeff toolkit (Nonsense, Synonymous, ect)
- Comparison of mutation rates from Production Packs across conditions and timepoints.
- Analysis of trends in which genes are or are not undergoing mutation.
- Changes in nutrient yield, RNA expression, biomass ect will be correlated with mutation rates.
- The Production Pack genome variant calling pipeline can be applied to WGS samples collected for other BioNutrients missions and sample sets.

Planned comparisons

IPP vs IPPgc

Impact of fermentation in LEO on genomic integrity

IPP vs EPP

Impact of storage in LEO on genomic integrity

T1 vs T3 vs T5 vs T6

Impact of storage length on genomic integrity

***S. boulardii* vs *S. cerevisiae* Y55**

Impact of strain on genomic integrity

Synthetic Biology Team


Frances Donovan, PhD Project Manager, PI


Natalie Ball


Hiromi Kagawa, PhD


Sandra Vu

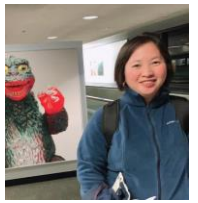

Sadie Downing


Matthew Paddock


Ami Hannon


Hami Ray, PhD, dPM


A. Mark Settles, PhD


Jessica Kong


Philip Sweet, PhD


Candice Tahimic, PhD


Lisa Anderson


Oscar Roque


Alyssa Villanueva


Sean Sharif


Kevin Sims, Payload Manager


Harry Jones, PhD Systems Engineer

Funded By

STMD Game Changing Development

Safety: Daniel Varnum-Lowry
Q/A: Leonard Hee
Logistics at KSC: Satro Narayan

Former team members and students: Aditya Hindupur, Amy Gresser, Aphrodite Kostakis, Asif Rahman, Ava Karanjia, Benjamin Alva, Eliza Zaroff, Eric Litwiller, Jason Samson, Jing Li, John Hogan, Jon Galazka, Julie Levri, Katherine Fisher, Leonard Lee, Matthew Kanan, Marilyn Murakami, Mathangi Soundararajan, Michael Dougherty, Paul Milazzo, William Tyukayev

Image Credit: NASA 3