JGI JOINT GENOME INSTITUTE
A DOE OFFICE OF SCIENCE USER FACILITY

Metagenome analysis of low-biomass samples: A JGI User Facility Perspective

Emiley Eloe-Fadrosh
Metagenome Program Head

NASA Planetary Protection Workshop
November 21, 2024

The DOE Joint Genome Institute at a glance

JGI MISSION:
Provide advanced genomic capabilities, large-scale data, and professional expertise to support the global research community in studies of complex biological and environmental systems. We optimize our service to the community through responsibly managing our people and resources.

A U.S. Department of Energy Office of Science User Facility

- JGI established in 1997, User facility from 2004
- IGB in Berkeley, CA opened in Dec 2019
- ~260 staff; ~$92M annual funding

Mission Areas, Programs & Infrastructure

Mission Areas

Bioenergy

Carbon Cycling

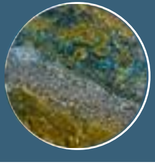
Biogeochemistry

Programs

Plants

Fungi

Metagenomes



Microbes

DNA Synthesis Science

Infrastructure

DNA Sequencing

- Illumina platforms: NovaSeq, HiSeq, MiSeq, Next Seq
- PacBio platforms: RSII, Sequel
- Oxford Nanopore platforms: MinION, PromethION

Advanced Genomic Technologies

- Single-cell Seq
- Metabolomics
- ChIP-Seq, Tn-Seq, DAP-Seq
- Epigenomics
- Transcriptomics

Computational Analysis

- Assembly/ Annotation/Analysis
- High-Performance Computing
- Big Data Integration & Analysis Systems

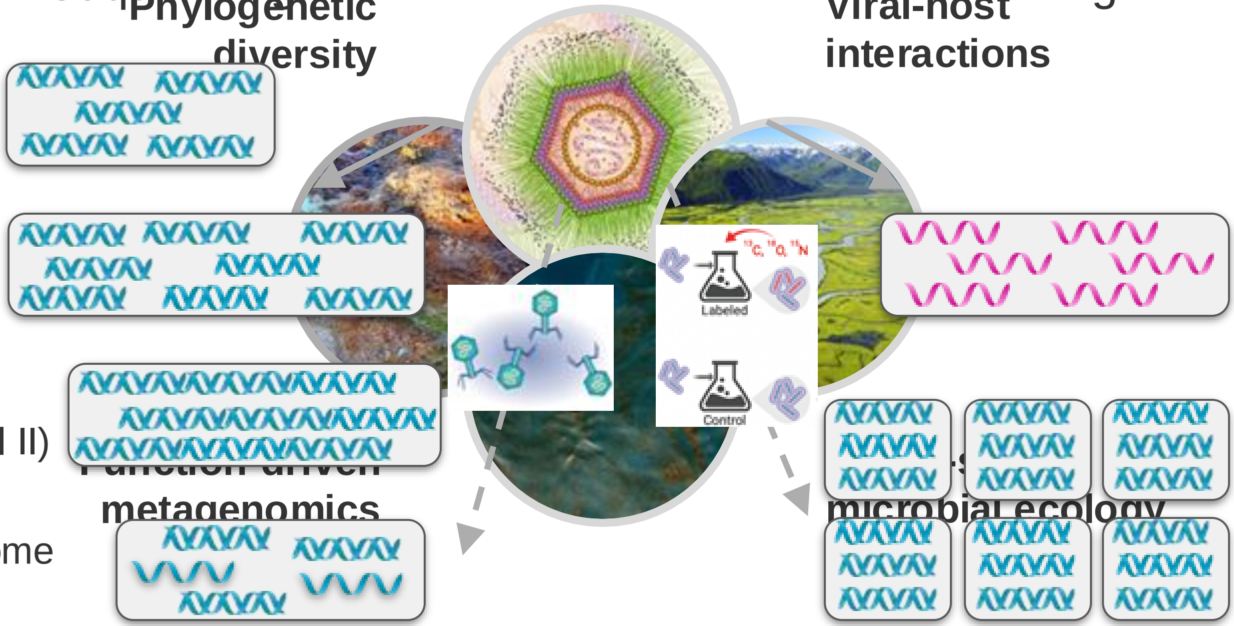
DNA Synthesis

- Design & Pathway Assembly
- Host Engineering

JGI Metagenome Program - Products

Metagenome

shotgun DNA sequencing



Phylogenetic diversity

Viral-host interactions

Function driven metagenomics

microbial ecology

Minimal draft (~35M reads)

Standard draft (~150M reads)

Improved draft (PacBio Sequel II)

Viral metagenome (ss & dsDNA)

Metatranscriptome

shotgun RNA sequencing
RNA-Seq

Ribo-depletion
or
PolyA selection

SIP metagenomes (density fractions enrichment analysis)

CSP Annual Call: 10Tb cap (up to 50Tb for multi-PI projects)

FICUS / CSP New Investigator: 3Tb cap

JGI Metagenome Program – Standard output

DNA/RNA sample

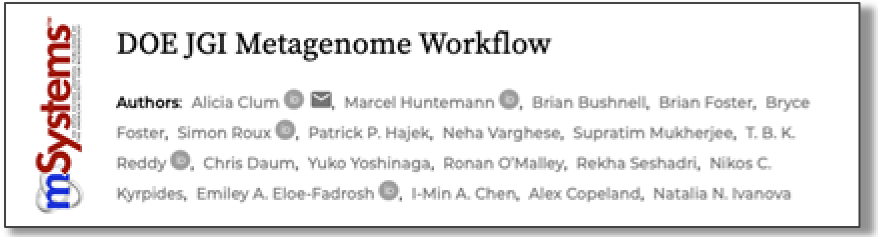
- Sample QC
- Library preparation
- Sequencing & QC

Filtered reads

- Assembly (metaSPAdes for metagenomes)
- Read mapping
- Genome binning (Metabat2) and QC
- Functional annotation (IMG pipeline)

Annotated assembly

- Data release through JGI Data Portal (download) and IMG (exploration/analysis) plus IMG/VR (Virus genomes)



mSystems

DOE JGI Metagenome Workflow

Authors: Alicia Clum, Marcel Huntemann, Brian Bushnell, Brian Foster, Bryce Foster, Simon Roux, Patrick P. Hajek, Neha Varghese, Supratim Mukherjee, T. B. K. Reddy, Chris Daum, Yuko Yoshinaga, Ronan O'Malley, Rekha Seshadri, Nikos C. Kyrpides, Emiley A. Eloe-Fadrosh, I-Min A. Chen, Alex Copeland, Natalia N. Ivanova

JGI Data Portal

JGI IMG/M
INTEGRATED MICROBIAL GENOMES & MICROBIOMES

JGI IMG/VR
INTEGRATED MICROBIAL GENOMES / VIRUS

Library prep for metagenomes

Sample Registration

Sample QC:

volume concentration quality

Two paths for library creation

384-well automation: fixed PCR cycles

- Illumina PCR-free WGS libraries
- Illumina low input (1 ng) WGS libraries (9 cycles of PCR)
- Illumina RNA-seq (1 ug input 8 cycles of PCR, 100 ng input with 14 cycles of PCR)

8-striptube manual libraries:

Ultra-low input (undetectable to <1 ng) - flexible PCR cycles (>15) &:

- Illumina ss/dsDNA-seq (1 ng) with flexible PCR cycles (>12)
- PacBio PCR-free long reads WGS (1 ug)
- PacBio low input (20 ng) long reads WGS with flexible PCR cycles (>12)

Example of ultra-low input samples: ice core viruses

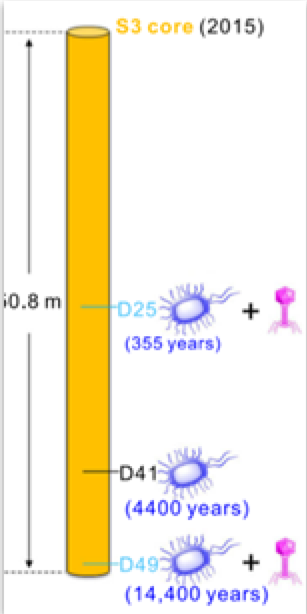
- **Viromes from ice-cores**
    - Low input, contamination-prone
    - First view of viruses “trapped in ice”
- **High level of genetic novelty**
    - Novel phages linked to dominant ice-adapted bacteria
    - Remote similarity to plant-associated, soil, freshwater sequences

User Programs

CSP# 503428


Zhiping Zhong


Matthew Sullivan



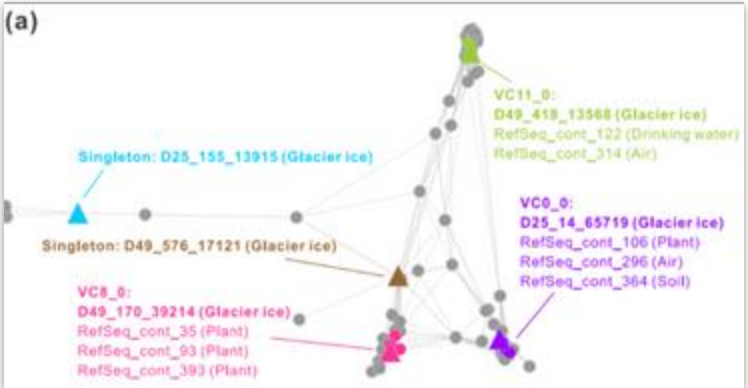

Zhong et al. Microbiome (2021) 9:160
https://doi.org/10.1186/s40168-021-01106-w

Microbiome

RESEARCH Open Access

Glacier ice archives nearly 15,000-year-old microbes and phages

Zhi-Ping Zhong[1,2,3], Funing Tian[2,3], Simon Roux[4], M. Consuelo Gazitúa[2], Natalie E. Solonenko[2,3], Yueh-Fen Li[2,3], Mary E. Davis[1], James L. Van Etten[5], Ellen Mosley-Thompson[1,3,6], Virginia I. Rich[1,2,3], Matthew B. Sullivan[1,2,3,7*] and Lonnie G. Thompson[1,3,8*]




Smithsonian MAGAZINE

SMARTNEWS Keeping you current

Scientists Extract 15,000-Year-Old Viruses From Tibetan Glacier

NEWS

15,000-year-old viruses never seen by humans discovered in glacier ice


WORLD

Scientists discover more than 30 viruses frozen in ice, most never seen before

Jordan Mendoza USA TODAY

Published 2:49 p.m. ET Jul. 21, 2021 | Updated 6:01 p.m. ET Jul. 21, 2021

Primary template-directed amplification: addressing coverage bias


Charles Gawad

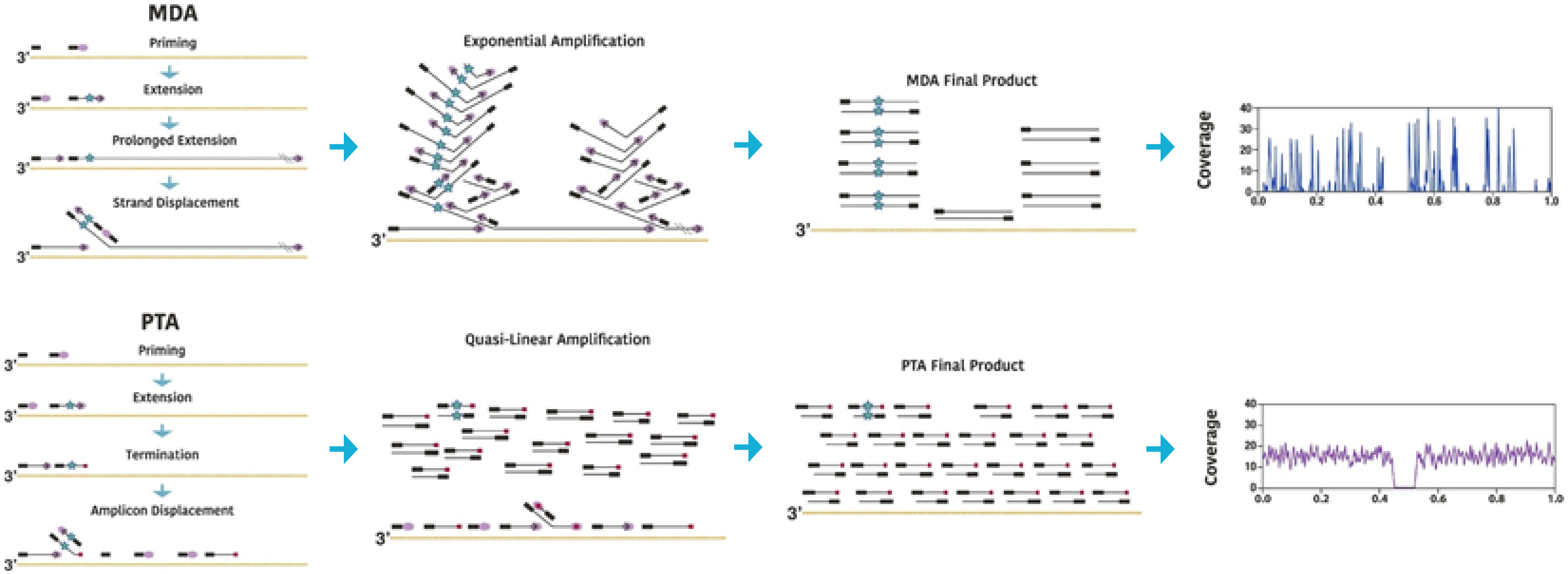

Veronica Gonzalez-Pena

Stanford
MEDICINE
School of Medicine

Comparison of PTA to MDA on human cells

Addition of terminators limits over-amplification

Quasi-linear amplification improves coverage uniformity



Gonzalez-Pena et al. 2021 PNAS

Can PTA help improve the quality of microbial single cells?


Charles Gawad


Veronica Gonzalez-Pena

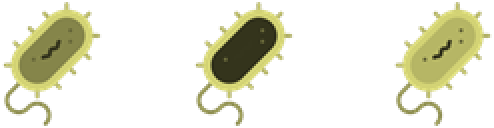
Stanford MEDICINE
School of Medicine

Microbial isolates
(n = 34)

E. coli
(n = 12)

P. putida
(n = 12)

B. subtilis
(n = 10)



PTA | MDA | WGA-X

Environmental microbes
(n = 185)


Mountain View Sough
(n = 97)


LBNL soil
(n = 27)

PTA | MDA | WGA-X

ISME

ISME Communications, 2024, 4(1), ycae085

https://doi.org/10.1093/ismeco/ycae085

Advance access publication: 12 July 2024

Brief Communication

scMicrobe PTA: near complete genomes from single bacterial cells

Robert M. Bowers[1,‡], Veronica Gonzalez-Pena[2,‡], Kartika Wardhani[2], Danielle Goudeau[1], Matthew James Blow[1], Daniel Udwary[1], David Klein[2], Albert C. Vill[3], Ilana L. Brito[3], Tanja Woyke[1], Rex R. Malmstrom[1,*,‡], Charles Gawad[2,3,*,‡]


Bob
Bowers


Danielle
Goudeau


Tanja
Woyke


Rex
Malmstrom

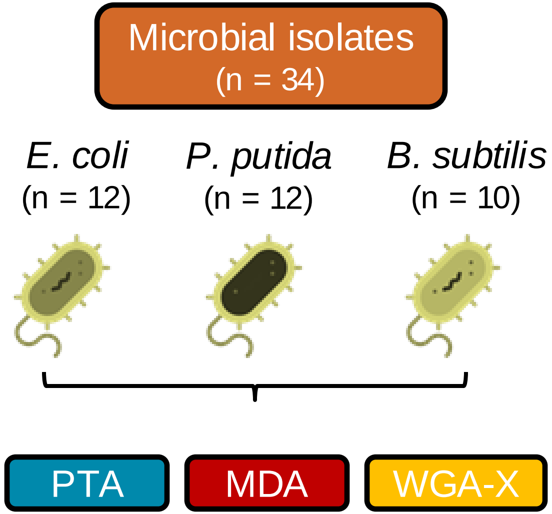
Can PTA help improve the quality of microbial single cells?

Charles Gawad

Veronica Gonzalez-Pena

Stanford MEDICINE

School of Medicine



Microbial isolates (n = 34)

E. coli (n = 12)

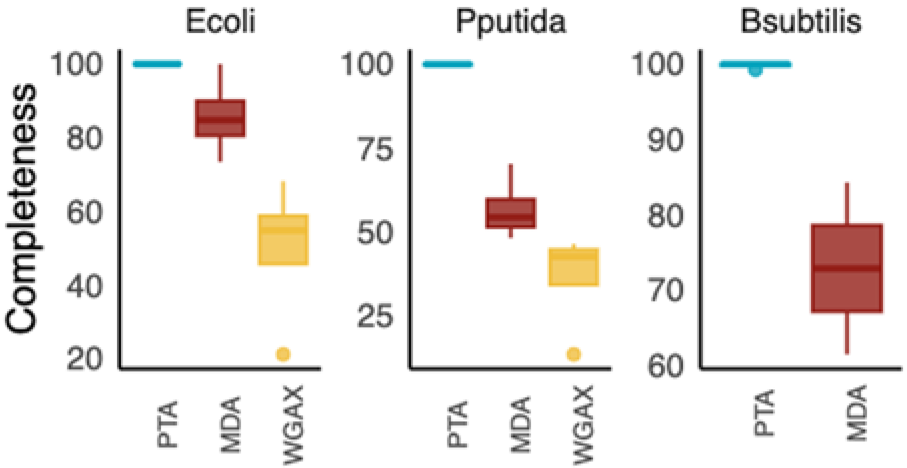
P. putida (n = 12)

B. subtilis (n = 10)

PTA

MDA

WGA-X



Analyses

- Genome coverage uniformity
- Genome quality statistics

Environmental benchmarking: Overall quality improves

Environmental microbes
(n = 185)

Mountain View Sough
(n = 97)

LBNL soil
(n = 27)

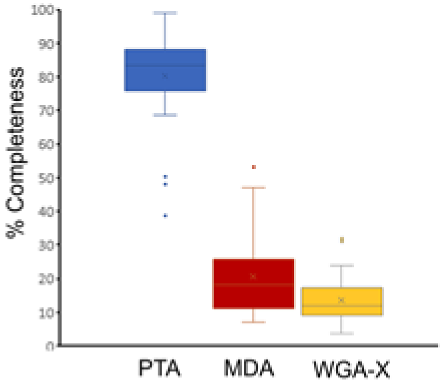
PTA

MDA

WGA-X

Analyses

- Genome quality statistics
- Taxonomic composition



- Significant improvement in SAG completeness
- Moderate contamination in PTA’d SAGs
- Increase in medium- and high-quality SAGs