




GSC: <http://gensc.org/> @genomestandards

**THE UNIVERSITY OF ARIZONA**  
**BIO5 Institute**



## GSC24: Challenges of Reproducibility in Genomics

- **Dates:** August 5-9, 2024
- **Location:** Bio5 Institute, University of Arizona, Tucson, AZ, USA
- **Sessions:**  
Standards in Genomic Reproducibility, Virome Standards, Omic Standards, GSC MixS Standards, IMMSA: A year long discussion on data reuse



# SPONSORS



## Agenda Schedule



### Monday, August 5 GSC board meeting

- o 1:00–4:00 pm: Board Meeting; 5:30 pm Board Dinner

### Tuesday, August 6 GSC main meeting

#### Session 1: Reproducibility Challenges Session Chair: Nikos Kyrpides

- 8:45–9:00 **Check In**  
**Meeting Opening:** GSC President: Lynn Schriml, Local Hosts: Bonnie Hurwitz & Ramona Walls
- 9:00–9:30 Keynote: [Greg Caporaso](#) (Northern Arizona University, The Translational Genomics Research Institute): **Reproducing, replicating and replaying bioinformatics workflows**
- 9:30–9:50 [Chris Hunter](#) (GigaScience): **The reproducibility challenges in publishing**
- 9:50–10:10 [Susanna-Assunta Sansone](#) (OICR, University of Oxford, UK): **Reproducibility across contexts: spotlight on TIER2**
- 10:10–10:20 [Shawn Levy](#) (Chief Scientific Officer, [Element Biosciences](#)): **Avidity Sequencing on the AVITI System - Sequencing Reimagined**
- 10:20–11:00 **Coffee Break**
- 11:00–11:20 [João Carlos Setubal](#) (University of São Paulo): **An Anecdote regarding Metagenome Data Access, and the lessons we can learn from it**
- 11:20–11:40 [Bonnie Hurwitz](#) (University of Arizona, Bio5 Institute): **Planet Microbe: a platform that integrates microbiological, genomic, ecological, oceanographic, and computational methods through FAIR data.**
- 11:40–12:00 [Taukondjo Shikongo](#) ([Convention on Biological Diversity](#)): **Nagoya Protocol and Digital Sequence Information Rights (Zoom)**
- 12:00–1:30 **Lunch**
- 1:30–2:00 **Dawn Field Award**  
[Savvas Paragkamian](#) (Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC)): **Data integration and reproducibility, the case of the Island Sampling Day in Crete.**
- 2:00–3:00 **Panel:** Reproducibility Challenges panel discussion
- 3:00–3:30 **Coffee Break & Group Photo**
- 3:30–4:45 Afternoon Working Groups
  - CIG - MIXS standards
  - GA4GH Experimental Metadata Hackathon
- 4:45pm Reception transportation pickup, departing from Bio5
- 5:30–7:30 pm Opening Reception at [Arizona Sonora Desert Museum](#): Transportation provided

## Wednesday, August 7 GSC main meeting

### Session 2: Virome Genomic Standards Session Chair: Emiley Eloë-Fadrosh

- 8:45-9:00 Day 2 Announcements
- 9:00-9:20 [Stacy Carrington-Lawrence](#) (NIH/NIA Division of Aging Biology): **NIH Human Virome Program; Opportunities and Challenges for Cross Consortium Standardization**
- 9:20-9:40 [Arvind Varsani](#) (Arizona State University): **The diverse world of eukaryote-infecting viruses + genomic standards = huge challenge**
- 9:40-10:00 [Simon Roux](#) (DOE Joint Genome Institute): **5 years of viromics standard: success, challenges, and future of the MIUViG checklist**
- 10:00-10:15 [Yi Xu](#) (Zymo): **One for All: RNA-Seq's Growing Role in Unraveling Microbial Transcriptomes**
- 10:15-11:00 Coffee Break

### Session 3: Omic Standards Session Chair: Chris Hunter

- 11:00-11:20 [Ramona Walls](#) (Critical Path Institute, Bio5 Institute): **The Research, Technology, Standards Triangle in 'Omics**
- 11:20-11:40 [Ruth Timme](#) (FDA, [Center for Food Safety and Applied Nutrition](#)): **GenomeTrakr's Efforts to Improve Pathogen Data Standards: Advancing Interoperability for Public and Private Repositories**
- 11:40-12:00 [Tim Van Den Bossche](#) (Ghent University): **Challenges and solutions for metadata management in metaproteomics**
- 12:00-1:00pm Lunch
- 1:00-1:20 [Peter Woollard](#) (EMBL-EBI): **eDNAqua-Plan: Exploring Standards in Aquatic eDNA, Presented by Peter Woollard, EMBL-EBI**
- 1:20-1:40 [Neil Davies](#) (Gump South Pacific Research Station): **Sampling Nature and Omic Observatories in an AI World**
- 1:40-2:00 [Briana Benton](#) (ATCC): **Setting the Standard in Microbial Genomics: Using Reference Assemblies from the ATCC® Genome Portal as Ground Truth**
- 2:00-2:20 [Tianyi Wang](#) (George Washington University): **BioCompute Objects: A Standard for Transparent and Reproducible Omic Data Analysis Communication**
- 2:30-3:00 Coffee Break
- 3:00-5:00 Afternoon Working Groups
- TWG - CIG interactions
  - [NMDC](#) workshop: **MIxS implementation in the NMDC and improving adoption**

## Thursday, August 8 GSC main meeting

### Session 4: Challenges of 'Omic data reuse - IMMSA ([International Microbiome and Multi'Omic Standards Alliance](#)). Session Chairs: Ishi Keenum and Jason Kralj

- 8:45-9:00 Day 3 Announcements
- 9:00-9:10 [Scott Jackson](#) (NIST - National Institute of Standards and Technology): **History of IMMSA**
- 9:10-9:30 [Ben Callahan](#) (NC State University, College of Veterinary Medicine): **Training on one study and predicting on another using publicly available microbiome datasets**
- 9:30-9:50 [Julie Dunning Hotopp](#) (Institute for Genome Sciences, University of Maryland School of Medicine): **A "Research Parasite's" Perspectives on the Challenges of Data Reuse**
- 9:50-10:30 **Coffee Break**
- 10:30-11:45 IMMSA **Panel** Discussion Panel: Data reuse challenges identified in the last year.
- 11:45-1:00 **Lunch**

### Session 5: Genomics and Standards challenges Session Chair: Bonnie Hurwitz

- 1:00-1:20 [Melanie Culver](#) ([University of Arizona](#)): **Genomic Challenges in Non-Model and Endangered Species**
- 1:20-1:40 [Dario Copetti](#) (Arizona Genomics Institute): **Saguaro genomics across technologies and (new) standards**
- 1:40-2:00 [Ryan Bartelme](#) (Pivot Bio): **Challenges with 'omics operations and standards in agriculture technology**
- 2:00-2:30 **Keynote:** [Jack Gilbert](#) (UCSD): **Precision microbiome medicine and standards**
- 2:30-3:30 Coffee Break
- 3:30-5:00 Afternoon Working Groups:
  - iSamples, OmicBON

## Friday, August 9 GSC main meeting

8:45 Handoff GSC26

### Session 6: Discussion of Reproducibility Challenges to Solutions

Session Chair: Lynn Schriml

- 9:00-9:30 **Keynote:** [Elisha Wood-Charlson](#) (Lawrence Berkeley National Laboratory, KBase): **Going beyond FAIR to COPE with complex biological data**
- 9:30-10:00 **Coffee Break**
- 10:00-12:00 **Panel:** Discussion of Challenges to Solutions
- 12:00 **Lunch**



# One for All

Universal rRNA Depletion  
for **ALL** Organisms



Unlock a new world of RNA-Seq with  
Zymo-Seq RiboFree® Total RNA Library Kit

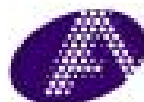
Scan to learn more!





Access and download thousands of  
reference-quality genomes with a supporting  
membership

# ATCC GENOME PORTAL



**ATCC**

**Credible leads to Incredible™**

# AVITI 24™

**E** Element  
Biosciences

## Sequencing Reimagined



### Core Sequencing

Flexible, accurate, and affordable sequencing for any application

#### ACCURATE

Reach Q50 data with Cloudbreak UltraQ™ kit

#### FLEXIBLE

Dual independent flow cell with two lanes

#### COMPATIBLE

Direct loading of linear libraries with Cloudbreak Freestyle™

#### SIMPLIFIED WORKFLOW

On-flow cell enrichment with Trinity™

#### AFFORDABLE

20–70% more reads per run at no extra cost with Expert Mode HD

### Teton CytoProfiling

Targeted, multiomic CytoProfiling in <24 hours

#### MULTIOMIC

DNA, RNA, proteins, phosphorylated proteins, and morphology

#### SUBCELLULAR RESOLUTION

200nm resolution: multi-feature cell segmentation

#### HIGH SENSITIVITY

With avidity base chemistry (ABC) sequencing

#### HIGH THROUGHPUT

Up to 1 million cells per run with 10cm<sup>2</sup> usable area per side in 24h

#### VERSATILE

On-flow cell cultured cells or fixed cell suspensions

# Never miss a microbial target

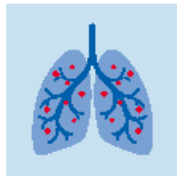


## QIAseq® xHYB Viral and Bacterial Panels

### Detect pathogens by enrichment with hybrid capture technology

Transform targeted bacterial and viral whole-genome sequencing using our high-sensitivity, probe-based enrichment panels.

- cDNA synthesis and enrichment in one kit
- Multiplex with ease: 384 Unique Dual Indices for high-throughput multiplexing
- Simple data analysis: Complimentary access to GeneGlobe® for data analysis



**Respiratory:** Characterize 89 separate viral targets, including SARS-CoV-2, influenza, rhinovirus, enterovirus and more



**MPXV:** Complete set of reagents necessary for the hybrid capture enrichment of the whole genome of all MPXV clades, including ITRs



**Antimicrobial resistance:** Sequence and detect the presence of 2786 antimicrobial resistance genes (6200+ AMR markers) in bacteria



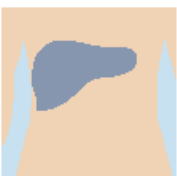
**Mycobacterium tuberculosis:** Comprehensive whole genome sequencing of tuberculosis without the need for cell culture



**Viral sexually transmitted infections:** Sequence and detect common STI viruses: HBV, HIV-1 and 19 types of high-risk HPV



**Adventitious Agent:** Detect 132 separate viral targets that can be problematic contaminants in bioreactors for biotherapeutics manufacturing



**HepC:** Enrich all hepatitis C virus genotypes, allowing deep characterization of the full viral diversity



Learn more at [www.qiagen.com/qiaseq-xhyb-viral-and-bacterial](https://www.qiagen.com/qiaseq-xhyb-viral-and-bacterial)



For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at [www.qiagen.com](https://www.qiagen.com) or can be requested from QIAGEN Technical Services or your local distributor.

Trademarks: QIAGEN®, Sample to Insight®, QIAseq®, GeneGlobe® (QIAGEN Group).  
GPRO-7741 1130120 07/24 © 2024 QIAGEN, all rights reserved.

Ordering [www.qiagen.com/shop](https://www.qiagen.com/shop) | Technical Support [www.support.qiagen.com](https://www.support.qiagen.com) | Website [www.qiagen.com](https://www.qiagen.com)



## **Workshop Descriptions:**

**CIG - MIxS standards:** The Compliance and Interoperability working group will meet to work on in-progress standards (e.g., eDNA, ancientDNA), MIxS checklists and extensions.

### **GA4GH: Experimental Metadata Hackathon:**

The purpose of this hackathon is to review and improve the core genomics metadata that the GA4GH Experiments Metadata Standard Group has already generated. We built the current terms from existing knowledge, but are aware of much interest and other knowledge out there. There will be a short introduction, we will work as several teams reviewing and improving the core metadata. A secondary objective is to share awareness of what we are doing and ideally have time for even occasional experts of particular experiment type domain experts. The GA4GH Experiments Metadata Standard Group is focusing on defining a minimal set of metadata properties for genomic datasets, to make it usable in a variety of contexts, particularly for discovery and analysis. These are intrinsic to the FAIR (Findability, Accessibility, Interoperability, and Reuse) principles. We focus on the properties of the methodology and equipment used in a genomic experiment, and more precisely on library preparation and instrument run. It provides context around the preparation of biological samples into libraries for a given laboratory experiment run, and the execution context for that run.

**TWG - CIG interactions:** Technical working group/CIG workflow, preparing for the next release.

**NMDC Workshop:** MIxS implementation in the [NMDC](#) and improving adoption

Presenting on NMDC, NMDC using [GSCs MIxS environmental extensions](#) (the app and [submission portal](#)), and then hosting a discussion around implementing standards & ways to improve. We will first introduce the NMDC [Submission Portal](#); this NMDC tool lowers the barrier to capturing valid metadata that conforms to the MIxS standard. We will highlight the features we've incorporated into the UI to make metadata completion and validation of MIxS terms user-friendly. Following this demo, we will present a tool currently under development, the [NMDC Field Notes App](#), designed to capture metadata using a mobile device. By capturing metadata while in the field, this application will help researchers reliably collect metadata in real time and reduce errors related to delayed transcription.

## iSamples/OmicBON: Sampling Nature and Omic Observatories in an AI World

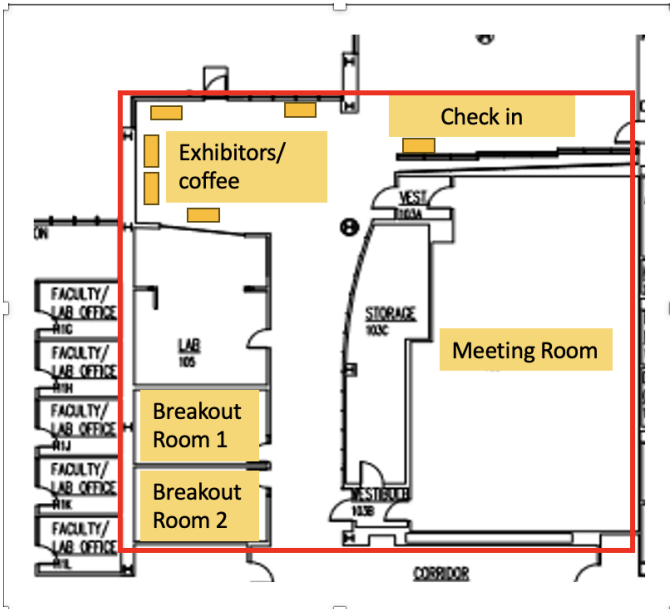
Omic observations are inevitably tied to material samples, so tracking the (meta)data linked to those samples is critical to data discovery, quality assurance, and (re)use. In May 2024, the GSC (Davies, Schriml, Buttigieg) participated in a “Sampling Nature” workshop at the Smithsonian in Washington, DC. Hosted by the National Museum of Natural History (Meyer), the workshop was co-organized by the Sampling Nature Research Coordination Network, the iSamples project, and Omic BON. It explored how to reuse (meta)data and samples to improve biodiversity observations, combine them with data from other domains (e.g., geoscience, archeology), and contribute to global digital interoperability frameworks such as UNESCO-coordinated Ocean Data and Information System (ODIS). Use cases discussed at the workshop emphasized the need to build functional and cross-validated data flows across natural history value chains, engaging field sites, labs, repositories, and museums to create an integrated corpus of material sample (meta)data. For example, participants identified the need for a new GSC ‘Life on Rocks’ extension of MIxS - using more modern, linked open data approaches - to describe living materials (e.g., lichens) that are found on geological specimens in research collections (e.g., the Polar Rock Repository). The workshop participants also drafted a collective definition of “AI-ready samples”. The crux of this definition is that the (meta)data representing samples must be 1) serialized using modern, globally adopted, portable, and cross-domain standards, and 2) marked up with machine-actionable and widely used semantic resources. Further, AI-ready sample metadata must adequately describe the provenance of the sample, which includes not only scientific metadata, enabling (re)use across domains, but also legal and social metadata, which are vital for ethical reuse across a range of potential purposes. Through organizations like Omic BON, the GSC should contribute to guidance for the ethical sampling of nature, supporting sample management systems (like iSamples) that operationalize FAIR and CARE data principles. While machines will play an evermore important role, humans must remain in the loop, setting the goals, defining the rules, and deciding how to steward samples and their derived data in an AI-world for the common good. The progress reported above will be further developed during this year's GSC workshop, and help shape the workshop's other themes.

**University of Arizona:**

<https://maps.arizona.edu/Files/VisitorMap.pdf>



**Bio5 Building: Main meeting site, coffee breaks, vendors**



# BSRL: Bioscience Research Laboratories

