FDA's Role In Building the ID NGS Diagnostic Toolkit

Heike Sichtig, MS/PhD

Subject Matter Expert
Principal Investigator

Center for Devices and Radiological Health
Division of Microbiology Devices

NIST-DHS-FDA Standards for Pathogen Detection Workshop
Aug 14-15, 2017

Disclaimer

The information in these materials is not a formal dissemination of information by FDA and does not represent agency position or policy.

Opinions are my own

ID NGS Diagnostic Toolkit Needs

- ID NGS Diagnostic Assay
- Tools to support regulatory review
  - **Reference genomes** for regulatory use to enable sponsor to perform in-silico validation of claims
  - **Reference materials** that sufficiently challenge the entire ID NGS Diagnostic Assay workflow
    - Ideally cell-based
    - Performance for assay's intended use

FDA Current Thinking

NGS Technologies

Targeted (*amplicon, bioinformatics*)

- Scope limited to defined regions that target a specific organism(s), gene(s) or marker(s).
- Targets are selected *a priori* by any lab or bioinformatics method (e.g., amplicon sequencing or a k-mer signature database) based on the diagnostic devices intended use.

Agnostic (*whole genome, shotgun*)

- No *a priori* knowledge of targets.
- Generally can identify all constituents (e.g., organism(s), gene(s) or marker(s), microbiota, human background, and contaminants) in a sample.

Sample Applications

Single Target (Pathogen, Gene, Marker)

Pathogen/Marker Panel

Gene Panel (16S)

Metagenomics

Novel and Emerging Pathogens

FDA Tools for ID NGS Dx

FDA-ARGOS Database

:microbial reference genomes for regulatory use

- ✓ **More flexible regulatory pathway**
  - **Enable In-silico analytical and clinical validation**
  - **Reduce testing burden**
- ✓ **Reference database**

Interagency ID NGS SME Working Group

: team of NGS agency-wide subject matter experts

- ✓ **ID NGS Dx Advisory Board**
- ✓ **Consensus FDA-ARGOS genome vetting**
- ✓ **Keep current on state of the art**
- ✓ **Tackle open questions (i.e. sens/spec)**
- ✓ **NGS Reference Material**

Reference Genomes For Regulatory Use

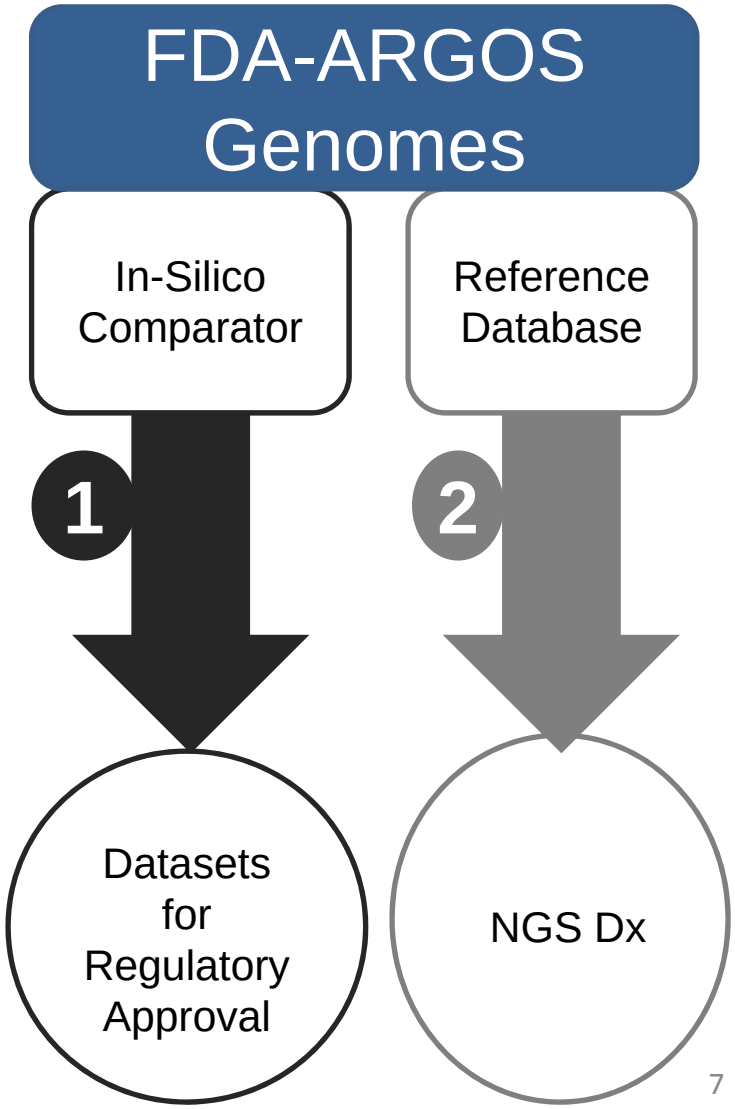
- Support in-silico analytical and clinical validation
  1. Identified by orthogonal reference method
  2. Sequenced and de-novo assembled using 2 sequencing methodologies (preferably long and short read)
  3. High depth of sequencing coverage
  4. 20X over 95 percent of the assembled and polished core genome
  5. Species-specific ANI thresholds that are sufficient for identification

FDA-ARGOS: Goal and Use

- Public Vetted Resource
- Microbial Reference-Grade Genomes for **Regulatory Use**
- US-Initiated
- Medical Countermeasures
- Common clinical
- Near neighbors

- **Coverage for US Needs**
- **Currently not funded to support *Needs for Developing World* and associated *Global Standards***

NCBI Project **PRJNA231221**



In-Silico Comparator Example

DoD/USAMRIID Collaboration

- Sequencing-based diagnostic device
- Generate FDA-ARGOS Reference Genomes
- Datasets for Regulatory Approval

> Enable In-Silico Data Analysis

Endemic African Diseases
Chikungunya virus
Crimean-Congo hemorrhagic fever virus
dengue virus serotype 1
dengue virus serotype 2
dengue virus serotype 3
dengue virus serotype 4
Ebola virus
Lassa virus
Marburg virus (Angola)
Marburg virus (Ci67)
Plasmodium falciparum
Rift Valley fever virus
West Nile virus
Yellow fever virus
Zika virus

FDA-ARGOS Pipeline

FDA-ARGOS microbial genomes are generated in 3 phases:

Phase 1- collection of a previously identified microbe and nucleic acid extraction

Phase 2- sequencing and de novo assembly at UMD

Phase 3- Vetting and data deposit in NCBI databases

FDA-ARGOS Reference Genome Characteristics:

- High depth of base coverage.
- Placed within a pre-established phylogenetic tree.
- Minimum of 20X over 95 percent of the assembled core genome.
- Sample specific metadata, raw reads, assemblies, annotation and details of the bioinformatics pipeline are available.

Bacteria/Fungi/Eukaryote Pipeline

- Collect samples /grow samples
- Extract samples
- Q/C extractions at IGS (>10ug cutoff)
- Batch and library prep/sequence on Illumina –Megablast QC
- Library prep/sequence on PacBio –Megablast QC
- Assemble long/short raw reads with Pilon
- Annotate with in-house pipeline for Q/C
- Data Analytics Q/C Pipeline at FDA
- Register BioSamples and submit raw reads to SRA DB and assemblies to Assembly DB
- NCBI annotates genomes

Bacteria/Fungi/Eukaryote

Genome Coverage (819)

Batch 1: 92 genomes registered	• Rockefeller U, FDA-CFSAN, Children's
Batch 2: 87 genomes registered	• FDA-CFSAN, ATCC, Children's
Batch 3: 85 genomes registered	• USAMRIID, FDA-CFSAN, ATCC, Children's
Batch 4: 76 genomes registered	• USAMRIID, FDA-CBER, FDA-CVM, ATCC, Children's
Batch 5: 93 genomes registered	• USAMRIID, FDA-CBER, ATCC, Children's, BCCDC
Batch 6: 99 genomes processing	• USAMRIID, U o Washington, FDA-NCTR, Children'sl, U of North Carolina
Batch 7: 85 genomes processing	• NYS Wadsworth, U o Louisville, U of North Carolina, Children's, U o Michigan
Batch 8: 68 genomes processing	• Cornell U/Zymo, ATCC, ECBC, Mayo Clinic, University of Ibadan, U o Louisville
Un-batched: 134 orgs collected	• BEI, Cornell U/Zymo, ATCC, U o Louisville, Children's National, Mayo Clinic

Viral Pipeline

Q/C extracted genomic material at IGS (25ng)

Library Prep/sequence on Illumina

- Shotgun
- Amplicon (may require primer set design –Ebola, Zika)
  - Looking into WNV, Dengue –Broad Institute
- RACE

Assemble long/short raw reads with Pilon

Data Analytics Q/C Pipeline at FDA

Register BioSamples and submit raw reads to SRA DB and assemblies to Assembly DB

Virus

Genome Coverage (137)

- 13 Ebola genomes registered (PHAC, UTMB)
- 96 genomes processing
- 27 orgs collected

<u>CBER ZIKA sample prep effort</u>

- Investigate 7 different NGS sample prep methods for 2 samples
  1. PRV/QTX4 -prep 1 (MagNa HP)
  2. PRV/QTX4 -prep 2 (MagNaLV
  3. PRV/QTX4 -prep 3 (QIAamp)
  4. PRV/QTX4 -prep 4 (Trizol 250)
  5. PRV/QTX4 -prep **5 (Trizol 500)**
  6. PRV/QTX4 -prep 6 (Trizol 500 DNAse)
  7. PRV/QTX4 -prep 7 (Trizol 500 Centri-Sep)
- Manuscript in preparation (CBER-led)

Accessing FDA-ARGOS Genomes

Landing page for FDA-ARGOS information directly from GenBank and also linked to from our FDA website.

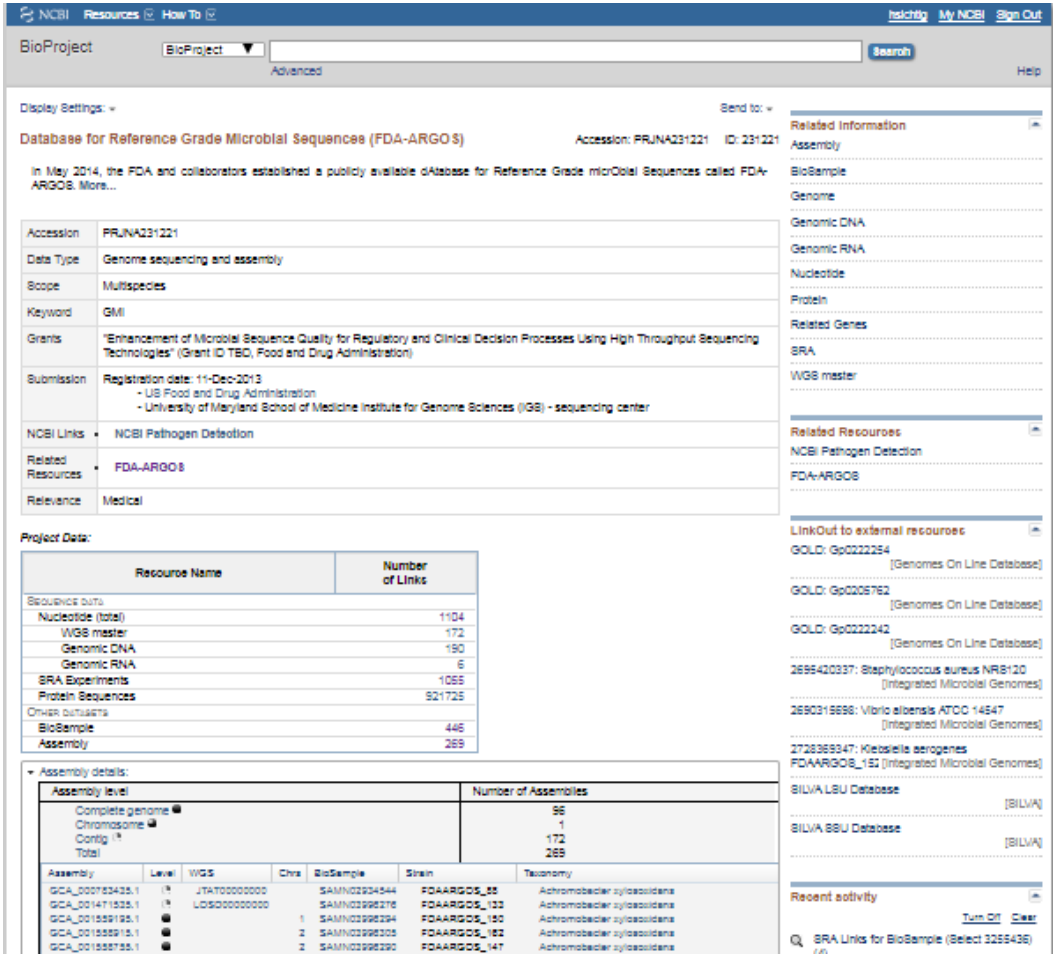
- https://www.ncbi.nlm.nih.gov/bioproject/?term=FDA-ARGOS

>> To get all associated genbank entries, select the Nucleotide database and enter this search term: '231221[BioProject]'

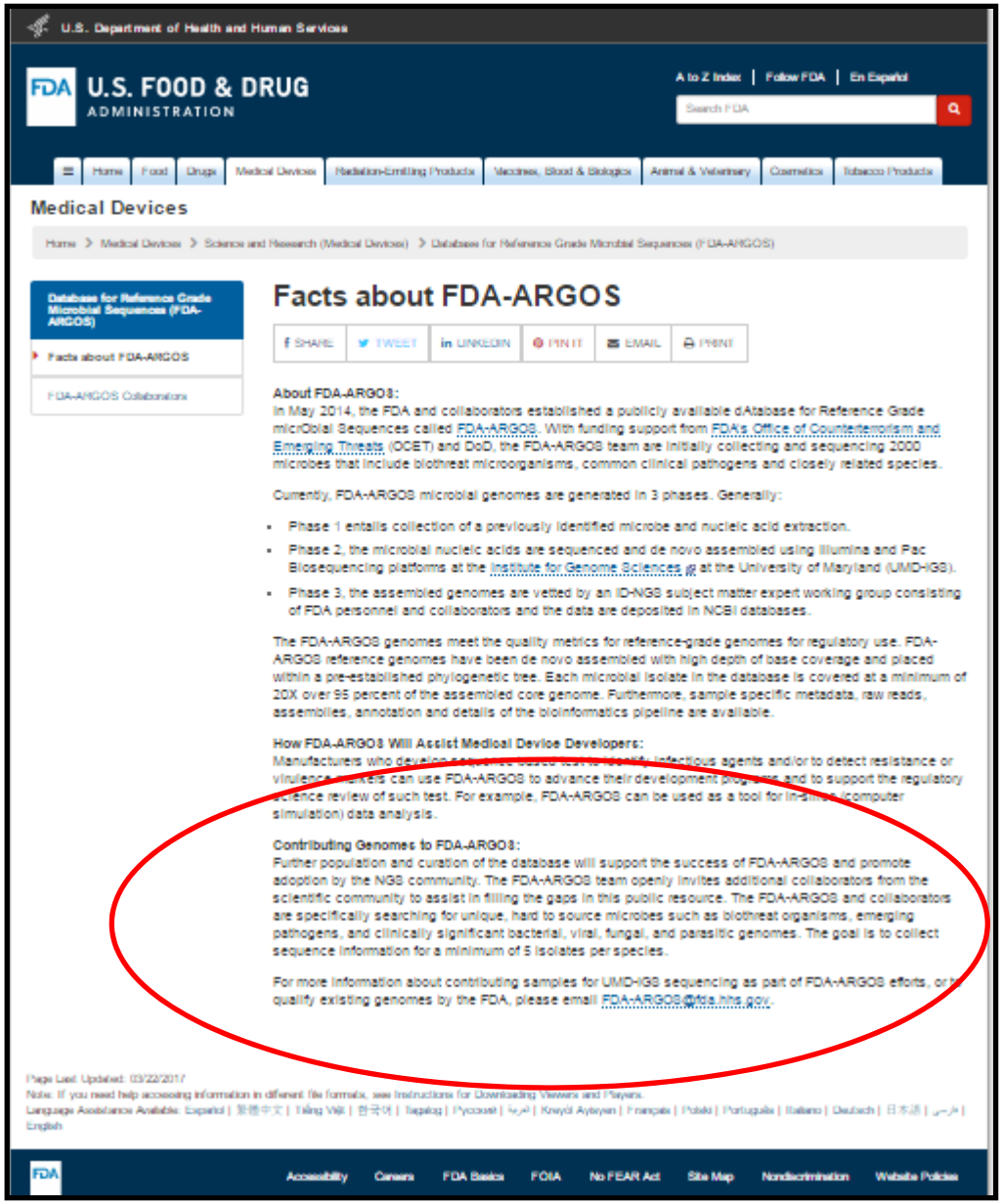
- **GenBank records (annotations, not RefSeq):**
  - https://www.ncbi.nlm.nih.gov/nuccore?term=231221%5BBioProject%5D
- BioSamples:
  - https://www.ncbi.nlm.nih.gov/biosample?Db=biosample&DbFrom=bioproject&Cmd=Link&LinkName=bioproject_biosample&LinkReadableName=BioSample&ordinalpos=1&IdsFromResult=231221
- Assemblies:
  - https://www.ncbi.nlm.nih.gov/assembly?LinkName=bioproject_assembly_all&from_uid=231221
- Raw reads:
  - https://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=231221

NCBI BioProject 231221

:houses FDA-ARGOS genomes generated with the IGS-UMD Sequencing pipeline



https://www.fda.gov/argos



U.S. Department of Health and Human Services

FDA U.S. FOOD & DRUG ADMINISTRATION

A to Z Index | Follow FDA | En Español

Search FDA

Home | Food | Drugs | Medical Devices | Radiation-Emitting Products | Vaccines, Blood & Biologics | Animal & Veterinary | Cosmetics | Tobacco Products

Medical Devices

Home > Medical Devices > Science and Research (Medical Devices) > Database for Reference Grade Microbial Sequences (FDA-ARGOS)

Database for Reference Grade Microbial Sequences (FDA-ARGOS)

- Facts about FDA-ARGOS
- FDA-ARGOS Collaborators

Facts about FDA-ARGOS

SHARE | TWEET | LINKEDIN | PIN IT | EMAIL | PRINT

About FDA-ARGOS:
In May 2014, the FDA and collaborators established a publicly available dAtabase for Reference Grade microbial Sequences called FDA-ARGOS. With funding support from FDA's Office of Counterterrorism and Emerging Threats (OCET) and DoD, the FDA-ARGOS team are initially collecting and sequencing 2000 microbes that include biothreat microorganisms, common clinical pathogens and closely related species.

Currently, FDA-ARGOS microbial genomes are generated in 3 phases. Generally:

- Phase 1 entails collection of a previously identified microbe and nucleic acid extraction.
- Phase 2, the microbial nucleic acids are sequenced and de novo assembled using Illumina and Pac Biosequencing platforms at the Institute for Genome Sciences @ at the University of Maryland (UMD-IGS).
- Phase 3, the assembled genomes are vetted by an IDNGS subject matter expert working group consisting of FDA personnel and collaborators and the data are deposited in NCBI databases.

The FDA-ARGOS genomes meet the quality metrics for reference-grade genomes for regulatory use. FDA-ARGOS reference genomes have been de novo assembled with high depth of base coverage and placed within a pre-established phylogenetic tree. Each microbial isolate in the database is covered at a minimum of 20X over 95 percent of the assembled core genome. Furthermore, sample specific metadata, raw reads, assemblies, annotation and details of the bioinformatics pipeline are available.

How FDA-ARGOS Will Assist Medical Device Developers:
Manufacturers who develop sequence based tests to identify infectious agents and/or to detect resistance or virulence markers can use FDA-ARGOS to advance their development programs and to support the regulatory science review of such test. For example, FDA-ARGOS can be used as a tool for in-silico (computer simulation) data analysis.

Contributing Genomes to FDA-ARGOS:
Further population and curation of the database will support the success of FDA-ARGOS and promote adoption by the NGS community. The FDA-ARGOS team openly invites additional collaborators from the scientific community to assist in filling the gaps in this public resource. The FDA-ARGOS and collaborators are specifically searching for unique, hard to source microbes such as biothreat organisms, emerging pathogens, and clinically significant bacterial, viral, fungal, and parasitic genomes. The goal is to collect sequence information for a minimum of 8 isolates per species.

For more information about contributing samples for UMD-IGS sequencing as part of FDA-ARGOS efforts, or to qualify existing genomes by the FDA, please email FDA-ARGOS@fda.hhs.gov.

Page Last Updated: 03/22/2017

Note: If you need help accessing information in different file formats, see Instructions for Downloading Viewers and Players.

Language Assistance Available: Español | 繁體中文 | Tiếng Việt | 한국어 | Tagalog | Русский | العربية | Kreyòl Ayisyen | Français | Polski | Português | Italiano | Deutsch | 日本語 | فارسی | English

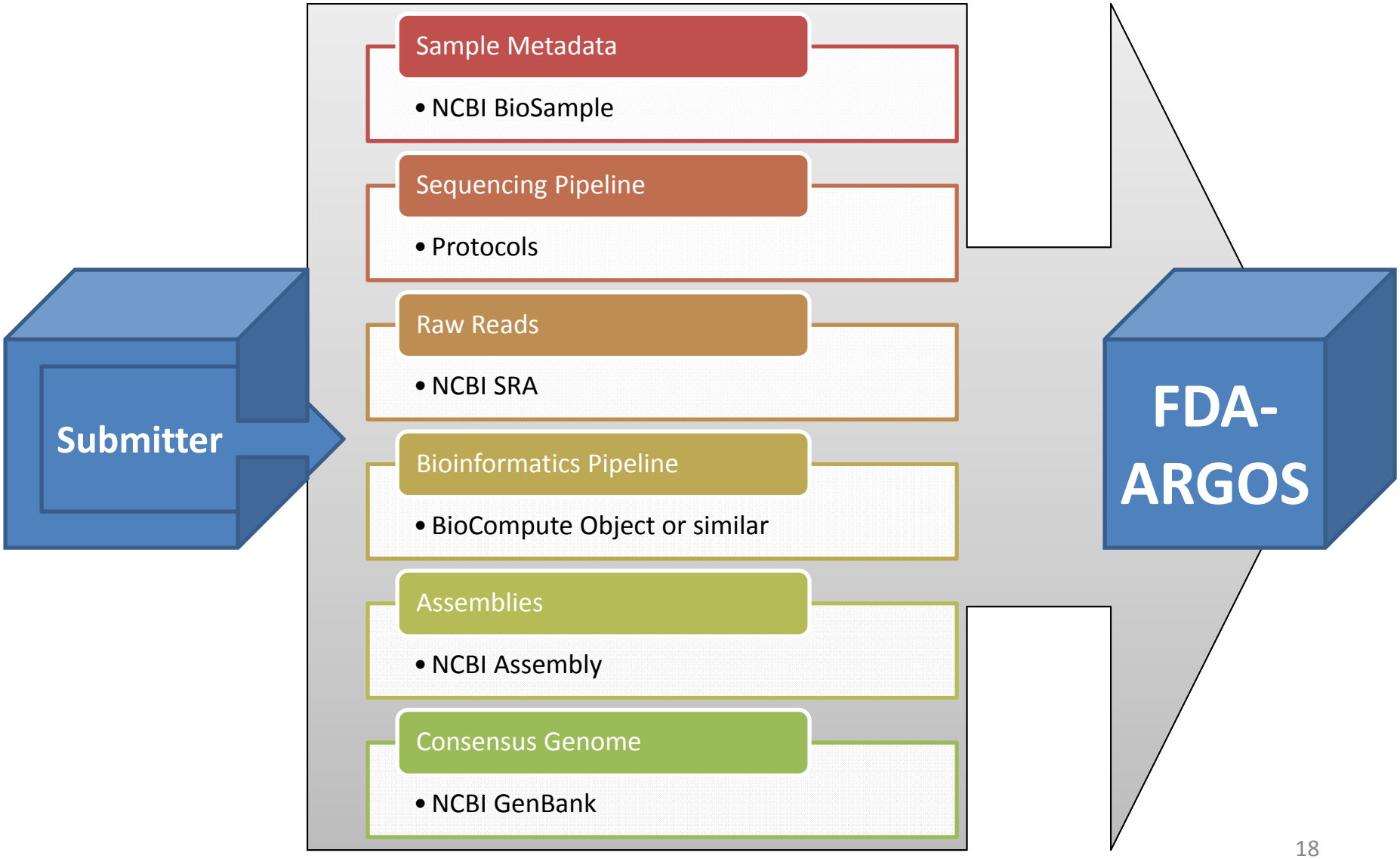
FDA | Accessibility | Careers | FDA Basics | FOIA | No FEAR Act | Site Map | Nondiscrimination | Website Policies

Contributing Genomes to FDA-ARGOS

- Further population and curation of the database will support the success of FDA-ARGOS and promote adoption by the community.
- The FDA-ARGOS team openly invites additional collaborators from the scientific community to assist in filling the gaps in this public resource.
- Specifically searching for unique, hard to source microbes such as biothreat organisms, emerging pathogens, and clinically significant bacterial, viral, fungal, and parasitic genomes.
- The goal is to collect sequence information for a minimum of 5 isolates per species.

- For more information about contributing samples for UMD-IGS sequencing as part of FDA-ARGOS efforts, or to qualify existing genomes by the FDA, please email FDA-ARGOS@fda.hhs.gov.

External Genome Submission



Reference Materials

- Support analytical validation of entire ID NGS Diagnostic assay workflow
  1. Cell-based in clinical matrix (blood, urine, stool) to test from specimen collection to result
  2. Reference material organism panel should sufficiently capture ID NGS assay's claimed target characteristics (intended use)
     - Size of the genome, G/C content, DNA/RNA, Near neighbors, Repetitive content, Commensal, Extremes
- Cross-platform comparison

Micro RM Laundry List

- ❑ High and low G+ C ratio
- ❑ Cell wall
- ❑ Gram neg/gram pos
- ❑ DNA/RNA
- ❑ What kingdom are we dealing with
- ❑ Matrix
- ❑ LOD (high, low samples)
- ❑ Is the goal to quantify
- ❑ Pathogen ratio
- ❑ Interfering substances
- ❑ Stability
- ❑ Danger and legal issues
- ❑ Easy to grow and supply
- ❑ Fragment size (does it make sense to have it as agarose)
- ❑ Quantitation of genome copies (QuBit (more reliable, doesn't work for single stranded RNA), NanoDrop, digital PCR, MassSpec)

Input Material

A. Standard clinical matrix (blood, urine, stool)
B. Standard cells
C. Standard genomic material
D. Standard data sets

Input Material Cont.

- Useful for different steps of the process
  - different standards for every step of the process

1. Pre-analytical
2. Library Prep
3. Sequencing
4. Bioinformatics
5. Metadata standards

Ecosystem of standards, what does the community need

NIST/FDA Reference Material Efforts

- Microbial Genomic DNA Reference Material
  - RM 8375 - Microbial genomic DNA standards for sequencing performance assessment
  - 2 FDA-ARGOS strains/ 2 FDA-CFSAN strains
- Mixed Pathogen DNA Research Material
  - A mixture of genomic DNA from 25 clinically-relevant pathogens plus human genomic DNA.

Build Reference Genomes:

- PacBio/Illumina sequencing of microbial constituents as part of FDA-ARGOS project

Other Reference Material Efforts

- ZymoBIOMICS™ Microbial Community Standards by Zymo Research
  - A mock microbial community consisting of eight bacterial and two fungal strains
- UCSF Control Material

Build Reference Genomes:

- PacBio/Illumina sequencing of microbial constituents as part of FDA-ARGOS project


FDA-ARGOS
Global-ARGOS
COMPREHENSIVE
ACCURACY
EMERGING THREATS
Coverage
Precision
Inclusivity
Rigor
Temporal
VERACITY
Timeliness
Robustness
Specificity
Geography
Sustainability
Spatial

Acknowledgements

FDA-ARGOS team members include representatives from the:

- U.S. Food and Drug Administration
- U.S. Department of Defense
- National Institutes of Health
- Institute for Genome Sciences at University of Maryland

Funding Agencies

FDA's Office of Counterterrorism and Emerging Threats

Joint Program Executive Office for Chemical and Biological Defense (JPEO-CBD)

American Type Culture Collection/ BEI
Bernard Nocht Institute for Tropical Medicine, Germany
British Columbia Centre for Disease Control (BCCDC)
Children's National Medical Center
Defense Threat Reduction Agency (DTRA)
George Washington University
Joint Program Executive Office for Chemical and Biological Defense (JPEO-CBD)
Lawrence Livermore National Lab (LLNL)
Los Alamos National Lab (LANL)
Mayo Clinic
National Biodefense Analysis and Countermeasures Center (NBACC)
National Institutes of Health: National Institute of Allergy and Infectious Diseases (NIH-NIAID)
New York State Wadsworth Laboratories
Public Health Agency Canada (PHAC)
Public Health England (PHE)
Rockefeller University
Rutgers University
Stanford University Medical Center
University of California, San Francisco (UCSF)
University of Colorado Denver
University of Ibadan, Nigeria
University of Louisville
University of Michigan
University of North Carolina at Chapel Hill
University of Texas Medical Branch (UTMB)
University of Washington School of Medicine
U.S. Army Edgewood Chemical Biological Center (ECBC)
U.S. Army Medical Research Institute for Infectious Diseases (USAMRIID)
U.S. Food and Drug Administration
Biodefense and Emerging Infections Research Resources Repository


FDA
U.S. FOOD & DRUG
ADMINISTRATION