gsc genomic STANDARDS consortium

Scott Tighe

Genomics Standards Consortium
Board Member and Secretary

Extreme Microbiome Project (Leader)
International Microbiome Metagenomics Standards Alliance (Steering Committee)

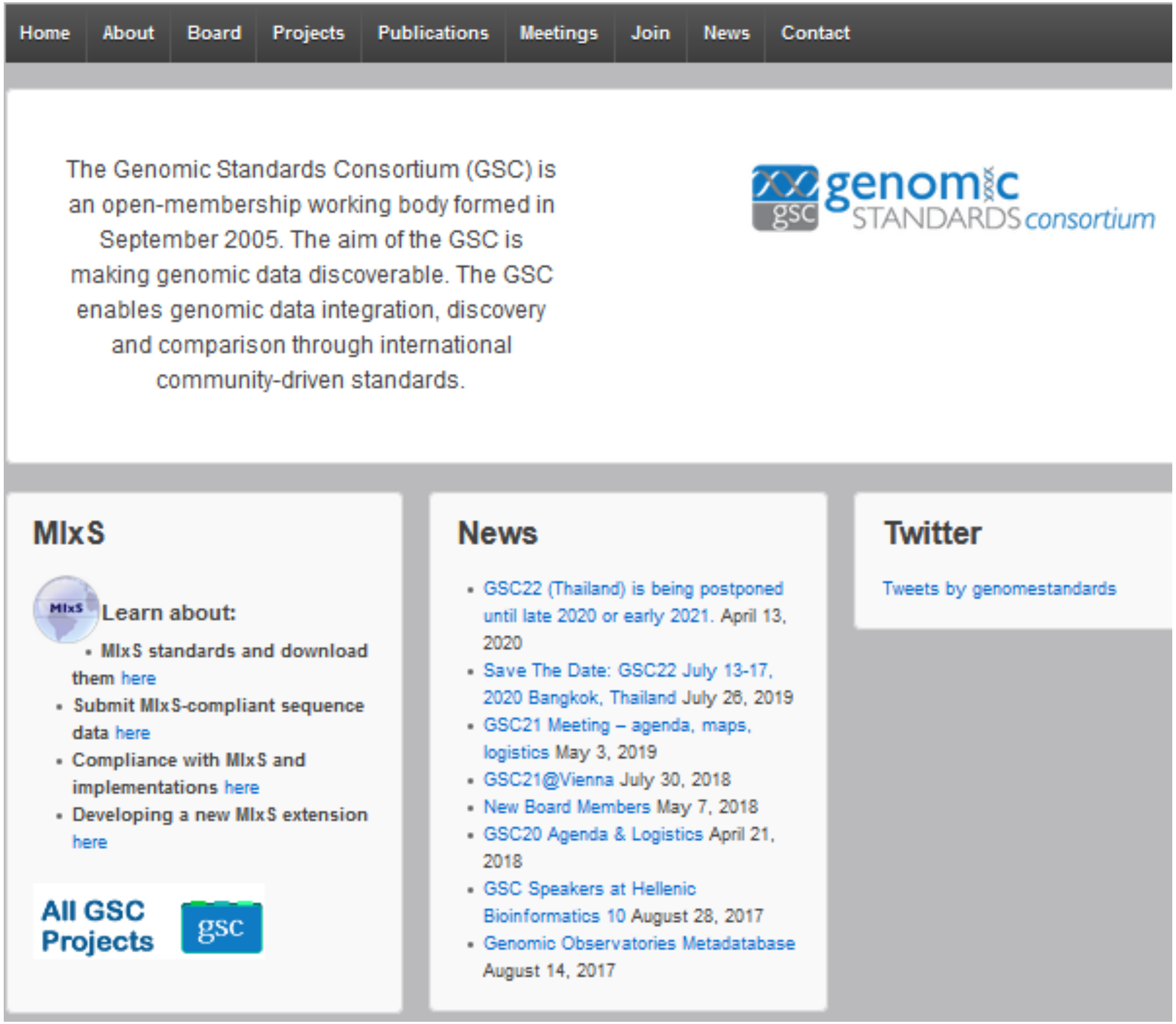
University of Vermont USA
Advanced Genomics Center

What is the GSC

- Open consortium -16 years
- International Collaborative Effort
- Develop Open Metadata Standards for any user and almost all sample types
- Work closely with INSDC

Allows for

- Making genomic data discoverable and usable
- Enabling genomic data integration, discovery and comparison
- International community-driven standards
- Strengths Genomic Databases worldwide
- Development of the MIxS standards



The Genomic Standards Consortium (GSC) is an open-membership working body formed in September 2005. The aim of the GSC is making genomic data discoverable. The GSC enables genomic data integration, discovery and comparison through international community-driven standards.

MIxS

Learn about:

- MIxS standards and download them here
- Submit MIxS-compliant sequence data here
- Compliance with MIxS and implementations here
- Developing a new MIxS extension here

All GSC Projects

News

- GSC22 (Thailand) is being postponed until late 2020 or early 2021. April 13, 2020
- Save The Date: GSC22 July 13-17, 2020 Bangkok, Thailand July 26, 2019
- GSC21 Meeting – agenda, maps, logistics May 3, 2019
- GSC21@Vienna July 30, 2018
- New Board Members May 7, 2018
- GSC20 Agenda & Logistics April 21, 2018
- GSC Speakers at Hellenic Bioinformatics 10 August 28, 2017
- Genomic Observatories Metadatabase August 14, 2017

Twitter

Tweets by genomestandards

GSC Mission

Develop and Integrate MIxS Standards (Minimum Information about any Sequence)

- In coordination with INSDC
- Downloadable MIxS check lists
- Identify the minimum information and categories to be included in a standard metadata checklist

1. Air (download MIxS-air only)
2. Built-environment (download MIxS-built environment only)
3. Host-associated (download MIxS-host associated only)
4. Human-associated (download MIxS-human associated only)
5. Human-gut (download MIxS-human gut only)
6. Human-oral (download MIxS-human oral only)
7. Human-skin (download MIxS-human skin only)
8. Human-vaginal (download MIxS-human vaginal only)
9. Microbial mat/biofilm (download MIxS-microbial mat/biofilm only)
10. Miscellaneous natural or artificial environment (download MIxS-misc. natural or artificial environment only)
11. Plant-associated (download MIxS-plant associated only)
12. Sediment (download MIxS-sediment only)
13. Soil (download MIxS-soil only)
14. Wastewater/sludge (download MIxS-wastewater/sludge only)
15. Water (download MIxS-water only)
16. Hydrocarbon resources-cores (download from GitHub)
17. Hydrocarbon resources-fluids/swabs (download from GitHub)
18. **Minimum Information About a Single Amplified Genome (MISAG) (GitHub)**
19. **Minimum Information About a Metagenome-Assembled Genome (MIMAG) (GitHub)**

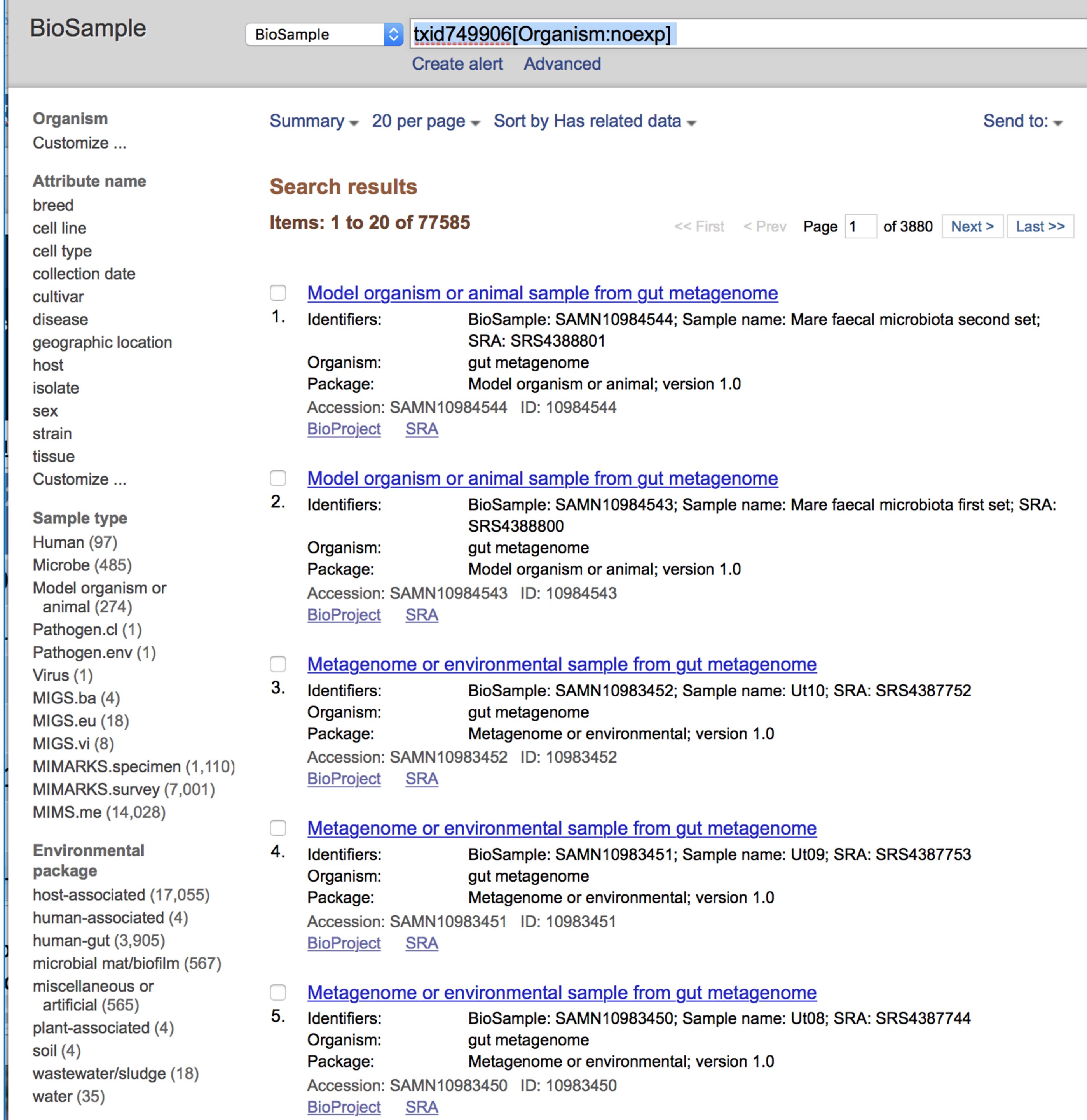
Query and Explore Standardized Metadata

Use attribute list to explore the data

Entrez query ⍰ Attribute & Value

"geo_loc_name=China: Wuhan"[attr]

"Severe acute respiratory syndrome coronavirus 2"[orgn]
NOT "geo_loc_name=China: Wuhan"[attr]



txid749906[Organism:noexp]

Organism
Customize ...

Attribute name
breed
cell line
cell type
collection date
cultivar
disease
geographic location
host
isolate
sex
strain
tissue
Customize ...

Sample type
Human (97)
Microbe (485)
Model organism or animal (274)
Pathogen.cl (1)
Pathogen.env (1)
Virus (1)
MIGS.ba (4)
MIGS.eu (18)
MIGS.vi (8)
MIMARKS.specimen (1,110)
MIMARKS.survey (7,001)
MIMS.me (14,028)

Environmental package
host-associated (17,055)
human-associated (4)
human-gut (3,905)
microbial mat/biofilm (567)
miscellaneous or artificial (565)
plant-associated (4)
soil (4)
wastewater/sludge (18)
water (35)

Search results

Items: 1 to 20 of 77585

☐ 1. Model organism or animal sample from gut metagenome
Identifiers: BioSample: SAMN10984544; Sample name: Mare faecal microbiota second set; SRA: SRS4388801
Organism: gut metagenome
Package: Model organism or animal; version 1.0
Accession: SAMN10984544 ID: 10984544
BioProject SRA

☐ 2. Model organism or animal sample from gut metagenome
Identifiers: BioSample: SAMN10984543; Sample name: Mare faecal microbiota first set; SRA: SRS4388800
Organism: gut metagenome
Package: Model organism or animal; version 1.0
Accession: SAMN10984543 ID: 10984543
BioProject SRA

☐ 3. Metagenome or environmental sample from gut metagenome
Identifiers: BioSample: SAMN10983452; Sample name: Ut10; SRA: SRS4387752
Organism: gut metagenome
Package: Metagenome or environmental; version 1.0
Accession: SAMN10983452 ID: 10983452
BioProject SRA

☐ 4. Metagenome or environmental sample from gut metagenome
Identifiers: BioSample: SAMN10983451; Sample name: Ut09; SRA: SRS4387753
Organism: gut metagenome
Package: Metagenome or environmental; version 1.0
Accession: SAMN10983451 ID: 10983451
BioProject SRA

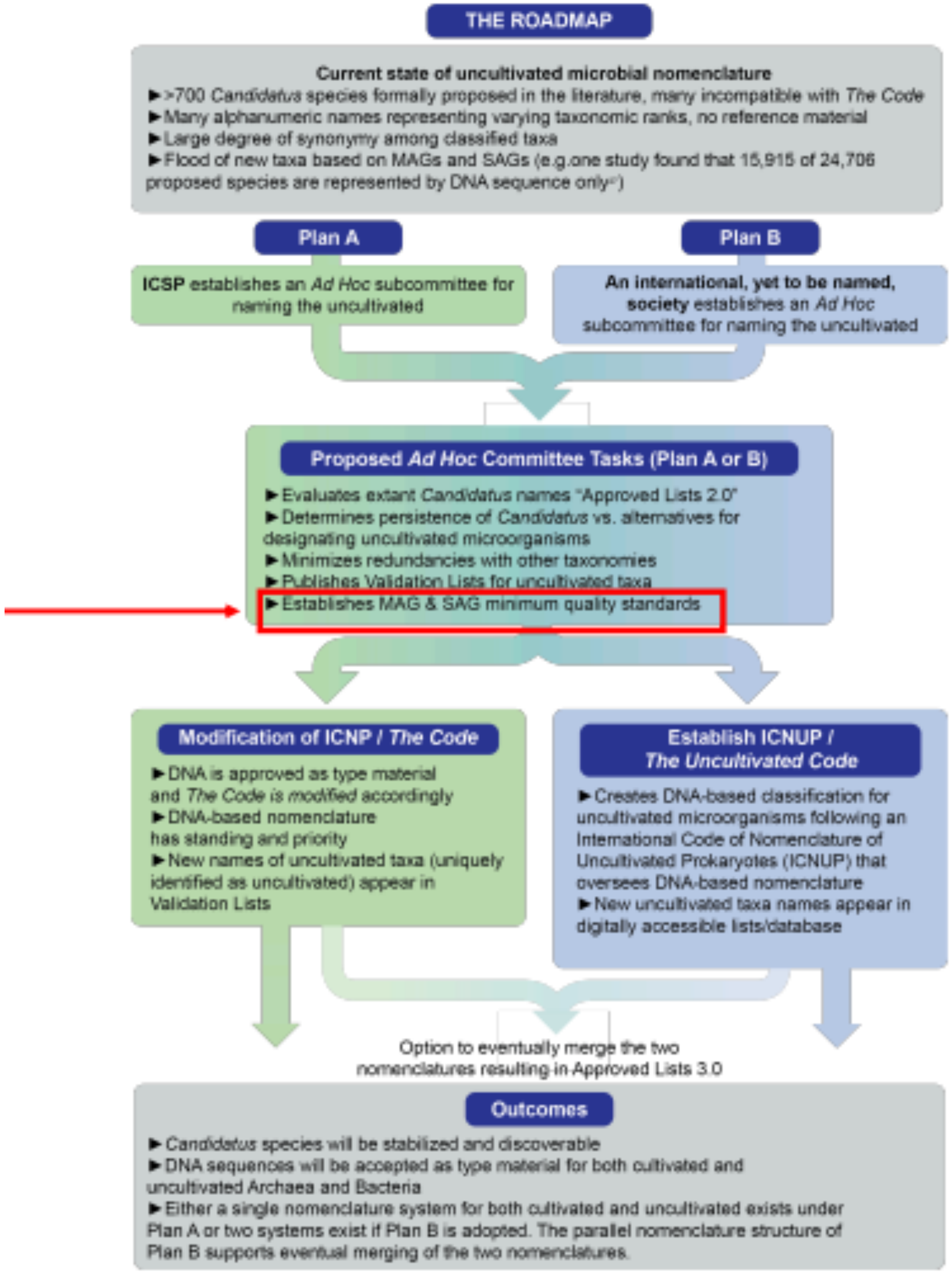
☐ 5. Metagenome or environmental sample from gut metagenome
Identifiers: BioSample: SAMN10983450; Sample name: Ut08; SRA: SRS4387744
Organism: gut metagenome
Package: Metagenome or environmental; version 1.0
Accession: SAMN10983450 ID: 10983450
BioProject SRA

gsc genomic STANDARDS consortium

Leverage GSC to enable proper metadata standards for unculturables

Adopt or modify MIMAG or MISAD standards/checklist as needed to assist SeqCode

Provide community engagement and awareness of SeqCode



THE ROADMAP

Current state of uncultivated microbial nomenclature

- >700 *Candidatus* species formally proposed in the literature, many incompatible with *The Code*
- Many alphanumeric names representing varying taxonomic ranks, no reference material
- Large degree of synonymy among classified taxa
- Flood of new taxa based on MAGs and SAGs (e.g.one study found that 15,915 of 24,706 proposed species are represented by DNA sequence only[illegible])

Plan A

ICSP establishes an *Ad Hoc* subcommittee for naming the uncultivated

Plan B

An international, yet to be named, society establishes an *Ad Hoc* subcommittee for naming the uncultivated

Proposed *Ad Hoc* Committee Tasks (Plan A or B)

- Evaluates extant *Candidatus* names "Approved Lists 2.0"
- Determines persistence of *Candidatus* vs. alternatives for designating uncultivated microorganisms
- Minimizes redundancies with other taxonomies
- Publishes Validation Lists for uncultivated taxa
- Establishes MAG & SAG minimum quality standards

Modification of ICNP / *The Code*

- DNA is approved as type material and *The Code is modified* accordingly
- DNA-based nomenclature has standing and priority
- New names of uncultivated taxa (uniquely identified as uncultivated) appear in Validation Lists

Establish ICNUP / *The Uncultivated Code*

- Creates DNA-based classification for uncultivated microorganisms following an International Code of Nomenclature of Uncultivated Prokaryotes (ICNUP) that oversees DNA-based nomenclature
- New uncultivated taxa names appear in digitally accessible lists/database

Option to eventually merge the two nomenclatures resulting in Approved Lists 3.0

Outcomes

- Candidatus species will be stabilized and discoverable
- DNA sequences will be accepted as type material for both cultivated and uncultivated Archaea and Bacteria
- Either a single nomenclature system for both cultivated and uncultivated exists under Plan A or two systems exist if Plan B is adopted. The parallel nomenclature structure of Plan B supports eventual merging of the two nomenclatures.