Update from TC34/SC9/WG25

Genomic sequencing of foodborne microorganisms — General requirements and guidance for bacterial genomes

Errol Strain, Ph.D.
Director, Biostatistics and Bioinformatics Staff
FDA Center for Food Safety and Applied Nutrition

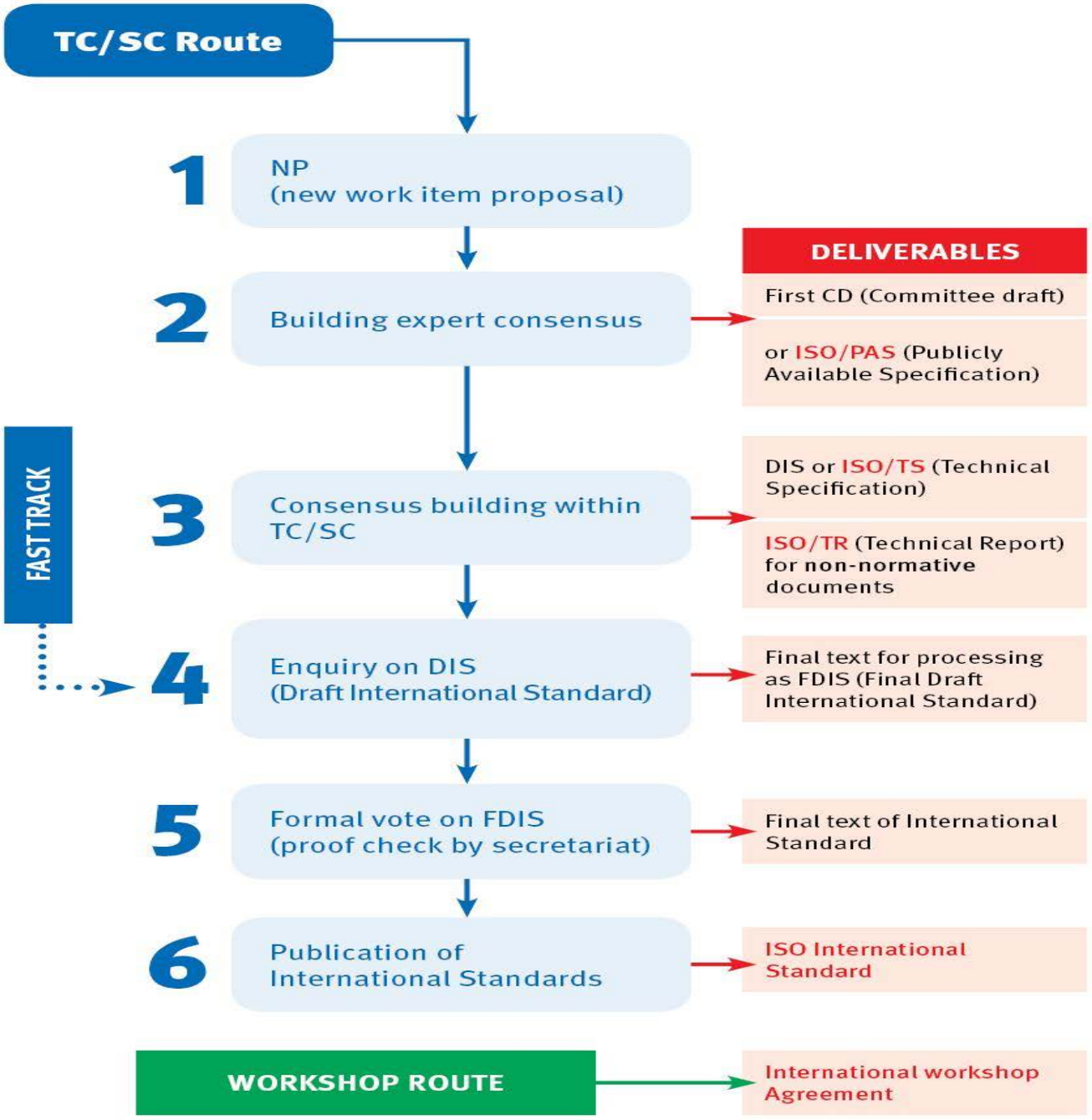
April 5th, 2018
EURL-AR Workshop 2018

US Proposal for Working Group for Novel Genomic Reference Standards

"This ISO/CEN proposed workgroup for genomic methods for 2014 defines the general principle and the technical protocol for the **validation of molecular genomic methods** (i.e., **whole-genome sequencing (WGS)** and DNA microarray technology) in the field of microbiological analysis of food, animal feeding stuffs, environmental, and veterinary samples for the validation of fast-approaching and highly sensitive genomic methods **as alternative tools to conventional identification, differentiation, and traceback approaches**."

Caveats

- Working group draft has been submitted to SC9 as a New Work Item Proposal (NWIP)
  - Feedback expected in early May
- If members affirm the need for the standard it will proceed to a committee draft
- Content may change
- Not intended for viruses and/or metagenomics


TC/SC Route
1
NP
(new work item proposal)
2
Building expert consensus
DELIVERABLES
First CD (Committee draft)
or ISO/PAS (Publicly Available Specification)
FAST TRACK
3
Consensus building within TC/SC
DIS or ISO/TS (Technical Specification)
ISO/TR (Technical Report) for non-normative documents
4
Enquiry on DIS
(Draft International Standard)
Final text for processing as FDIS (Final Draft International Standard)
5
Formal vote on FDIS
(proof check by secretariat)
Final text of International Standard
6
Publication of International Standards
ISO International Standard
WORKSHOP ROUTE
International workshop Agreement

Overview

Drafted in 3 sections

1. Laboratory Operations
   - DNA extraction & isolate to sequence data
2. Bioinformatic Pipelines
   - Functional and/or cluster predictions
3. Metadata
   - Enabling traceback, surveillance, data sharing

Scope of the Draft

1. Handling of bacterial cultures;
2. Genomic DNA isolation;
3. Sequencing library preparation, sequencing, and assessment of raw DNA sequence read quality and storage;
4. Bioinformatic analysis, including methods such as high quality single nucleotide polymorphism (hqSNP) analysis, core genome and whole genome multi-locus genotyping (cgMLST, wgMLST), and bioinformatic pipeline validation; and
5. Metadata capture and sequence repository deposition.

Principle

- Applicable to any organization that handles samples, performs sequencing, or performs bioinformatic analyses for WGS analysis of foodborne bacteria
- Includes guidance for laboratory and bioinformatic workflows
- Platform agnostic – independent of sequencing chemistry and bioinformatic methods

Challenges

WGS is a non-targeted assay

- We're not evaluating performance at a particular SNP/Allele

Fit-for-Purpose

- There is no recipe
- Details are left user
  - Quality Thresholds, Coverage, Sensitivity/Specificity
  - Test/Validation Data Sets

Laboratory and Bioinformatics workflows are not independent

- Some methods may be more/less robust to changes in quality in laboratory sequencing and/or bioinformatic analysis

Laboratory Operations

- General Laboratory Requirements
  - Elements of GLP: personnel, SOPs, proficiency testing
  - Minimize cross-contamination & sample mix-ups
    - Unlike PFGE & MLST errors may not be identified until bioinformatic analysis stage
  - Steps: DNA Isolation, library preparation, multiplexing, assessment of read quality
  - Quality assessments specific to sequencing platform
    - Not defined in standard

Bioinformatics Pipelines

1. Predictions on an individual isolate
    1. Is the toxin present or absent?
    2. Was the strain type correctly predicted?
2. Predictions about a group/cluster of isolates
    1. Accuracy of phylogenetic trees\dendrograms
    2. Estimates of genetic or allelic distance
3. Focus on fit for purpose
    – SNPs and/or wgMLST methods are okay if they work

Bioinformatics Pipelines

Implementation of validated commercial/public pipelines vs in-house

- If using a validated pipeline it may only be necessary to ensure that it is installed correctly

If using a custom pipeline then additional steps are required

- Development of validation data sets (real and simulated)
- Establishment of performance metrics, different species or strains may be different levels of genetic variation

* In general labs should not develop their own pipelines

Bioinformatics cont.

Assess quality of sequence data

- Read qualities/scores
- Filter/trim if necessary
- Check for contamination

Assemblies (N50, # of contigs, …)

SNPs (hqSNPs)

- Reference sequence, coverage, etc.

cgMLST and wgMLST

- Criteria for allele determinations

Phylogenetic or dendrogram construction

Interpretation and Reporting

- "Results from bioinformatic pipelines should be interpreted in the context of information regarding metadata about the origins of isolates and epidemiology (i.e. traceback information)."
- Comparison to historical gold-standards may be a challenge
  - WGS is much higher resolution. Isolates that were previously indistinguishable may not have measurable differences

Metadata

Integration with other public and private food safety resources requires some standardization of metadata

- Labs may or may not share data, but they will likely compare internal WGS data to external databases

List suggested and required/minimal metadata fields

- Minimal Data for Matching (MDM)

Ontologies and controlled vocabularies

- Improve data quality and usability

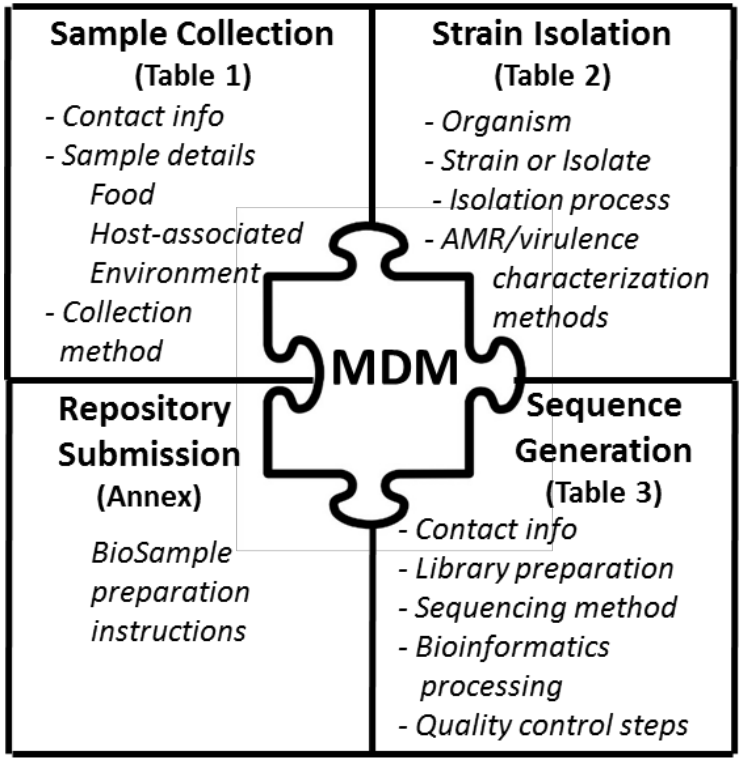
Metadata

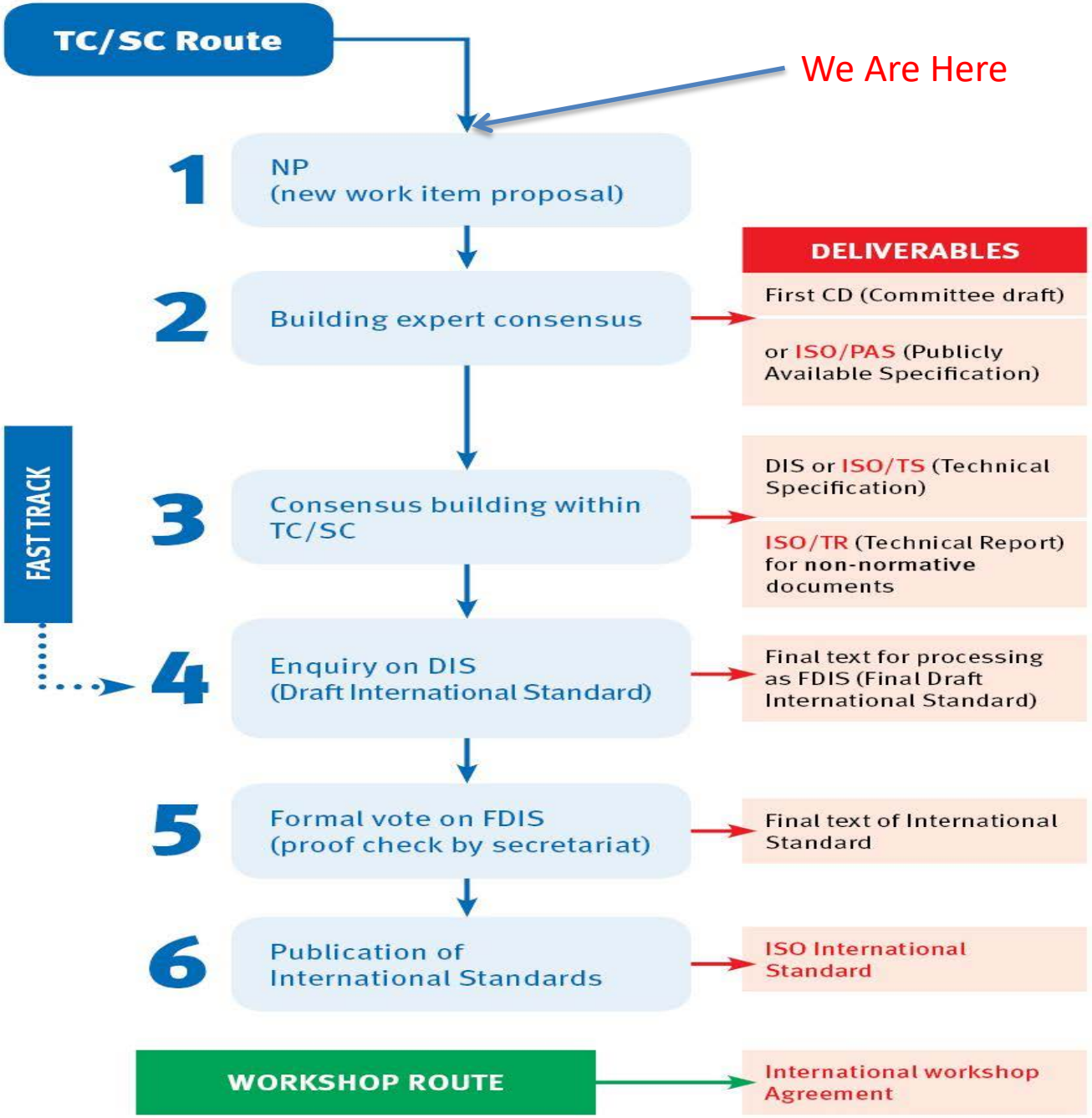
What is good quality contextual information?

- ***Fit-for-purpose***
  - *Descriptive, detail*
- ***Easily understood***
  - *Clear, minimal ambiguity*
- ***Data integrity***
  - *Free of errors*
- ***Auditable***
  - *Traceable, attribution, chain-of-custody, IP*
- ***Interoperable***
  - *Same structure, cross domains*

Metadata

MDM – Minimal Data for Matching




TC/SC Route
We Are Here
1
NP
(new work item proposal)
2
Building expert consensus
DELIVERABLES
First CD (Committee draft)
or ISO/PAS (Publicly Available Specification)
FAST TRACK
3
Consensus building within TC/SC
DIS or ISO/TS (Technical Specification)
ISO/TR (Technical Report) for non-normative documents
4
Enquiry on DIS
(Draft International Standard)
Final text for processing as FDIS (Final Draft International Standard)
5
Formal vote on FDIS
(proof check by secretariat)
Final text of International Standard
6
Publication of International Standards
ISO International Standard
WORKSHOP ROUTE
International workshop Agreement

FDA U.S. FOOD & DRUG ADMINISTRATION

Questions?