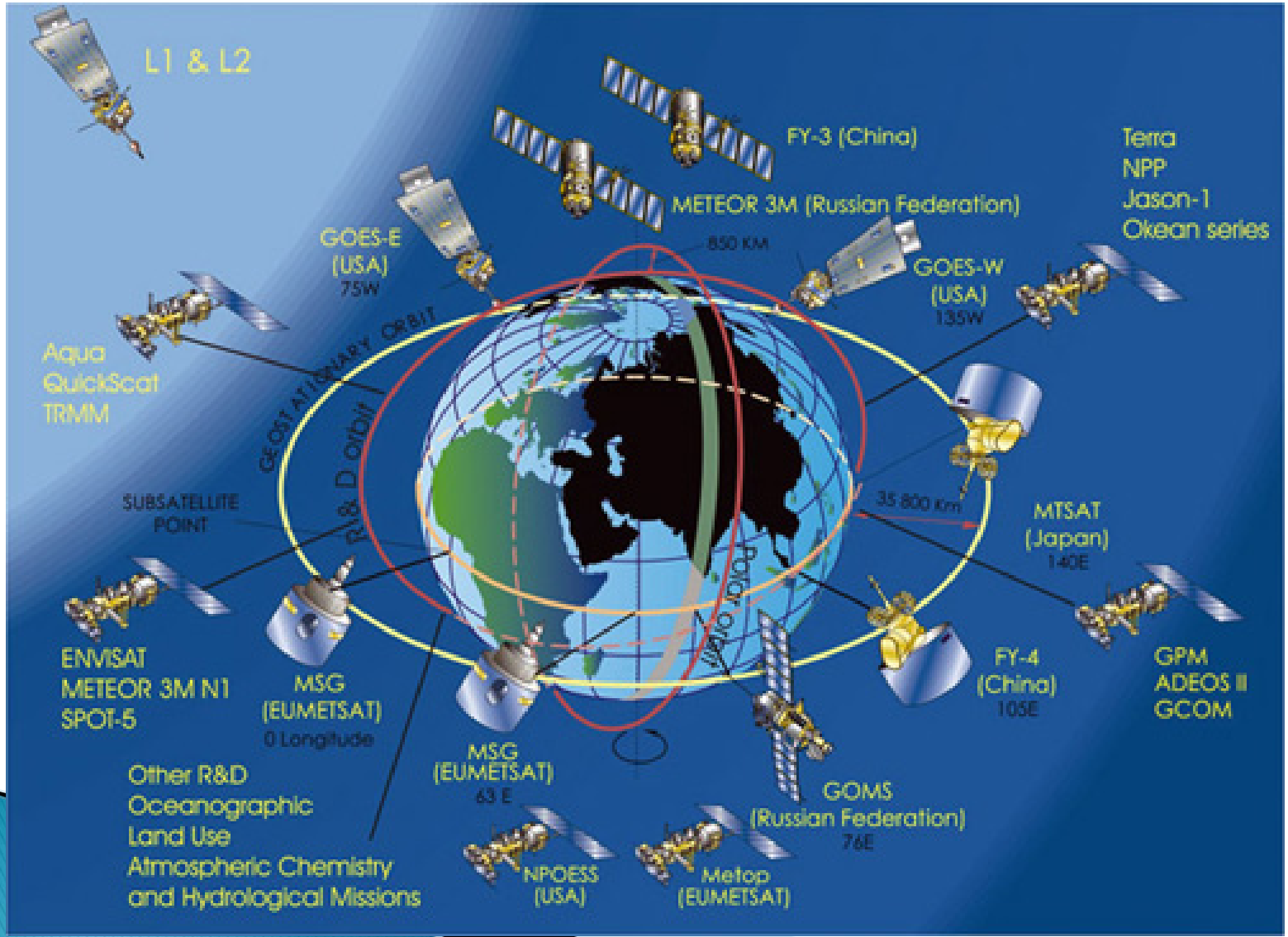
The GMI Global Repository WG

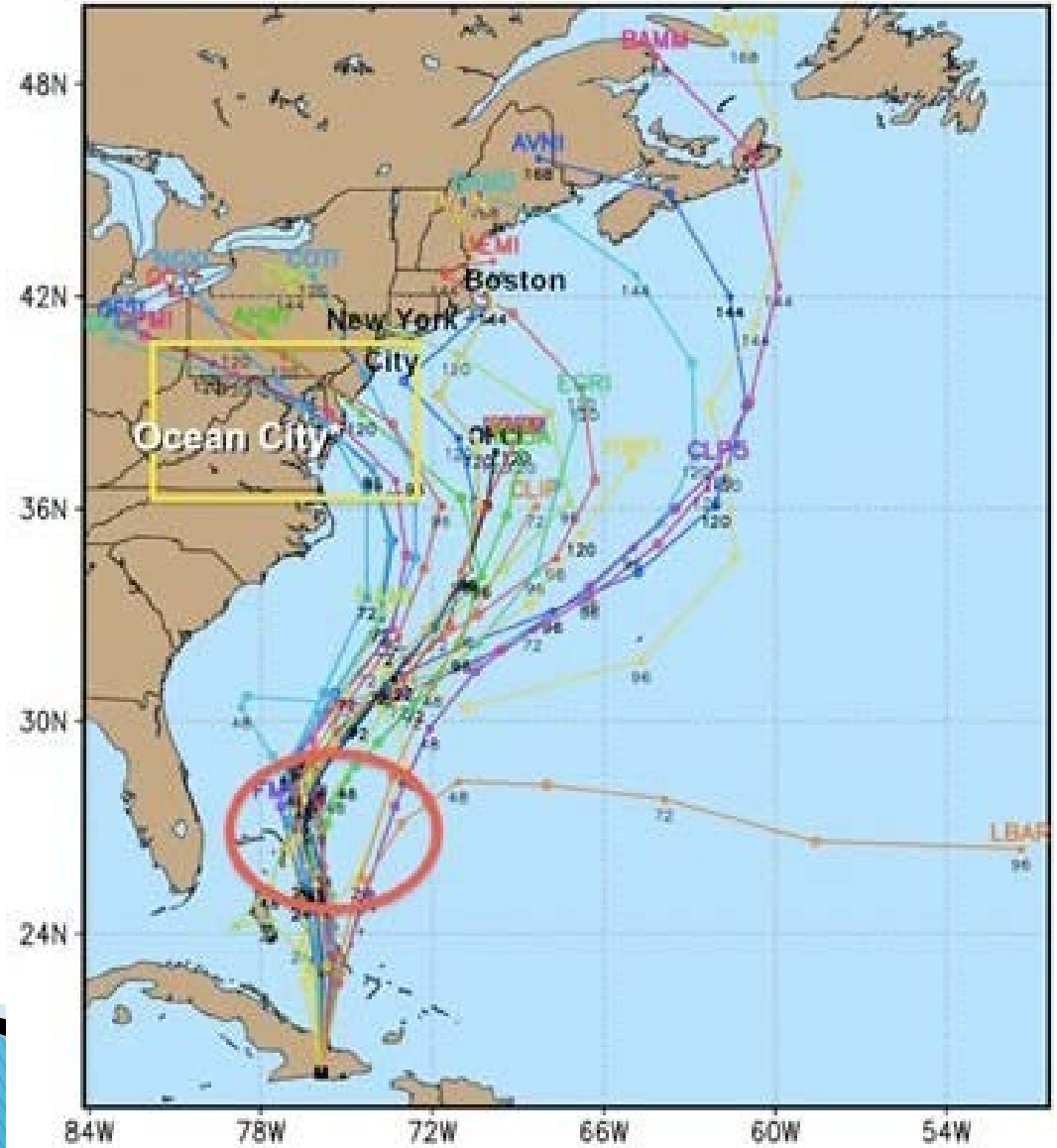
James Ostell

National Center for Biotechnology Information
US National Institutes of Health, Bethesda, MD, 20894-6513

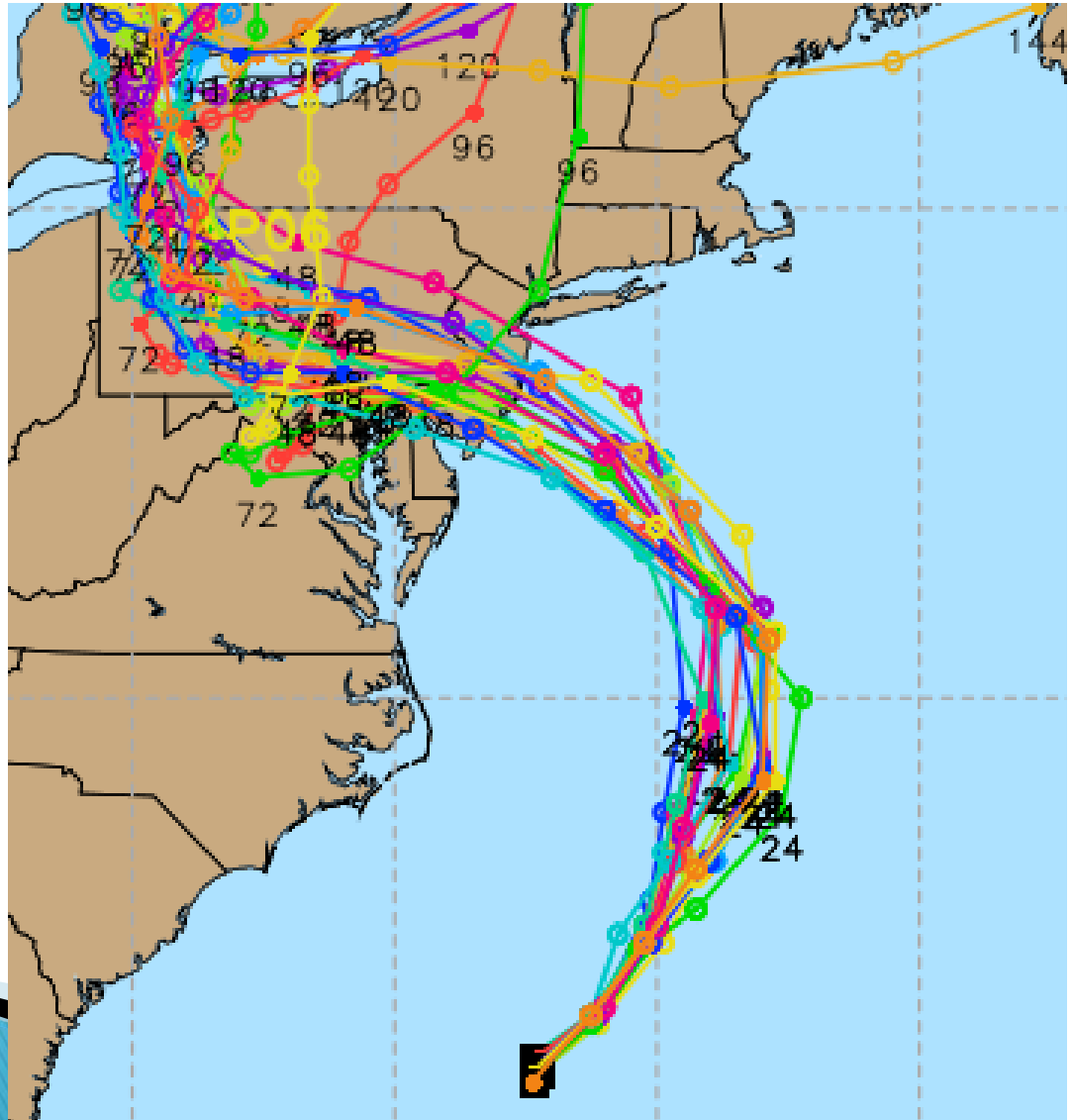
Weather as a Modern Global Infrastructure



Many Models on the Data, Fully Public, On the News



Citizens and Government All See the Power and Limits of Science



DNA Forensics…

Genome sequence is agnostic. One biological assay could work on all pathogen species

To be immediately useful all we need is the genome and a little metadata.

Functional prediction can be developed and refined more slowly from this base.

A Pathogen Genome Is The Fingerprint

The Well-Traveled Salad.

Do You Know Where Your Food Has Been?

As consumers, many of us fail to recognize that even our domestic and local food supplies are part of a global network. The daily activity of consuming food directly links our health as humans to the health of crops and produce, food animals, and the environments in which they are produced.

LETTUCE
Canada, Chile, Dominican Republic, Mexico, Peru, USA

CUCUMBERS
Canada, Honduras, India, Mexico, Spain, USA

FETA CHEESE
Canada, Denmark, Egypt, Germany, Greece, Israel, Italy, Turkey, UK, USA

VINAIGRETTE
Argentina, Brazil, Canada, Chile, China, France, Germany, Greece, India, Indonesia, Italy, Mexico, Morocco, Peru, Portugal, Spain, Thailand, Tunisia, Turkey, USA, Vietnam

OLIVES
Greece, Israel, Mexico, Spain, USA

SPROUTS
Argentina, Australia, Bangladesh, Canada, China, Egypt, France, India, Morocco, Nepal, Pakistan, South Africa, Spain, Turkey, USA

MANDARIN ORANGES
Israel, Mexico, Morocco, South Africa, Spain

CROUTONS
Argentina, Australia, Brazil, Canada, China, France, India, Mexico, Netherlands, Poland, Russia, Switzerland, Uruguay, USA, Vietnam

TOMATOES
Canada, Dominican Republic, Holland, Israel, Italy, Mexico, USA

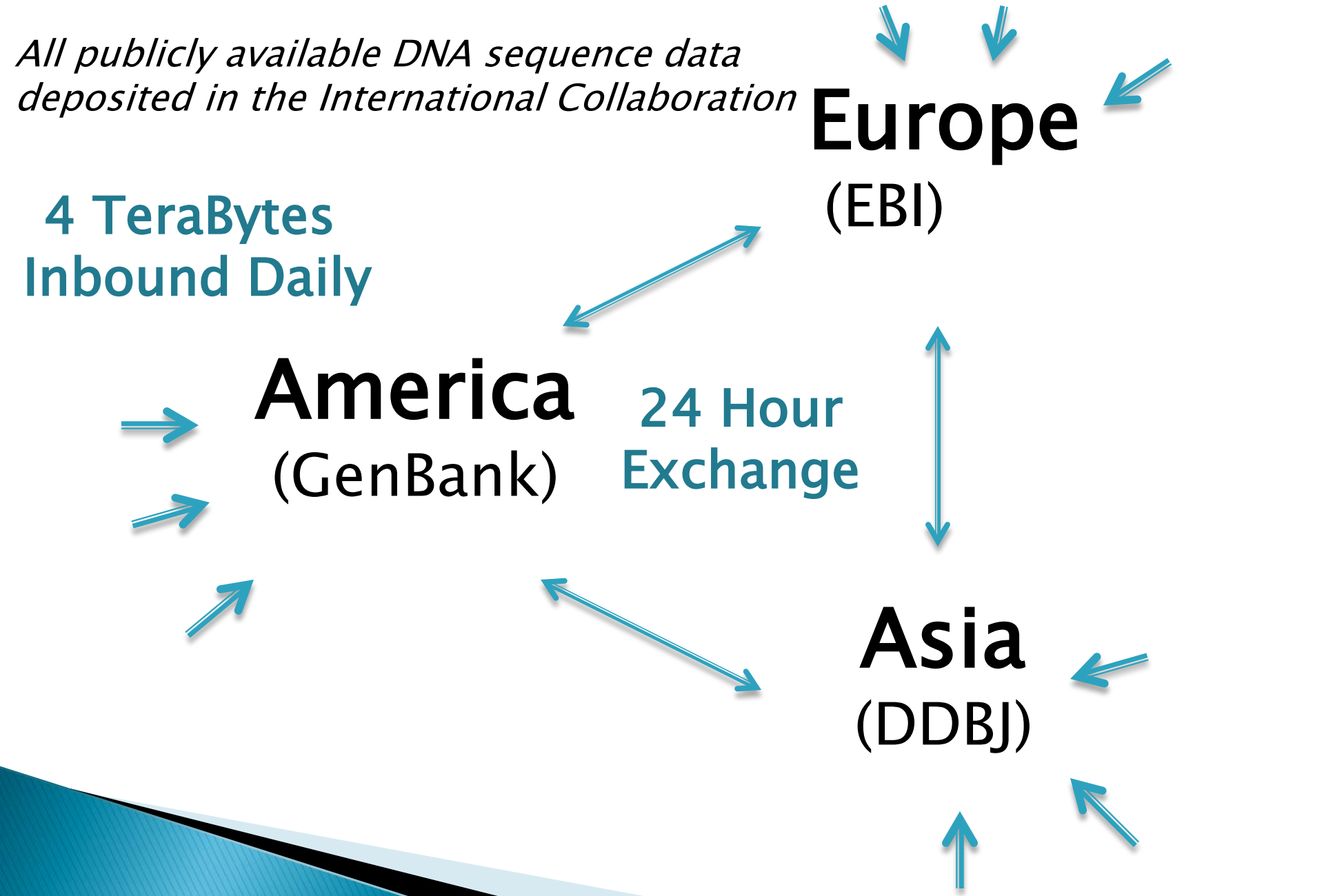
ONIONS
Canada, China, Germany, India, USA

A "One Health" approach to food safety—bringing together expertise and resources from the clinical, veterinary, wildlife health, and ecology communities—has the potential to reveal the sources, pathways, and factors driving the outbreaks of foodborne illness and possibly prevent them from occurring in the first place.

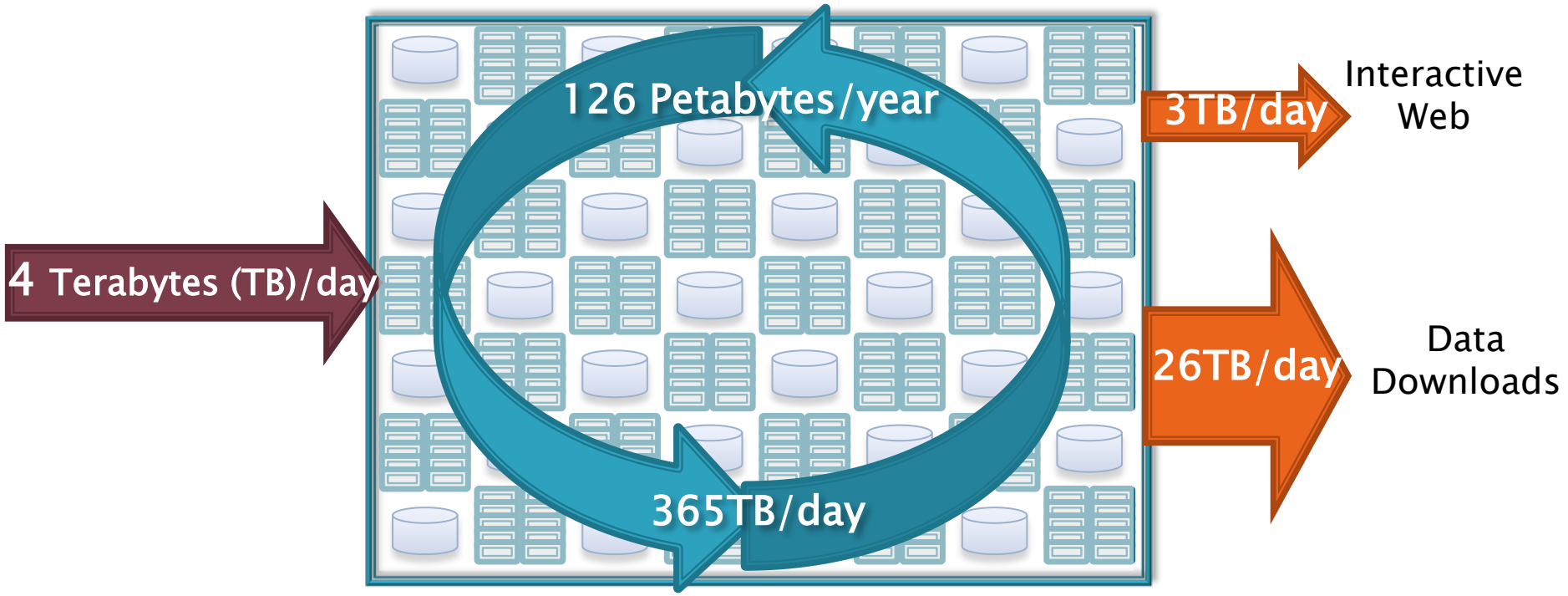
NOTE: Countries are listed in alphabetical order and not by volume of export.

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
www.iom.edu

All publicly available DNA sequence data deposited in the International Collaboration



Daily Data Processing at NCBI



MINIMAL PATHOGEN METADATA (FOODBORNE OUTBREAKS)

What

sample_name
organism
strain/isolate

Category (attribute_package)
1a) **Clinical/Host-associated**
1a1) **specific_host**
1a2) **isolation_source**
1a3) host-disease
OR
1b) **Environmental/Food/Other**
1b1) **isolation_source**

When

collection_date

Where

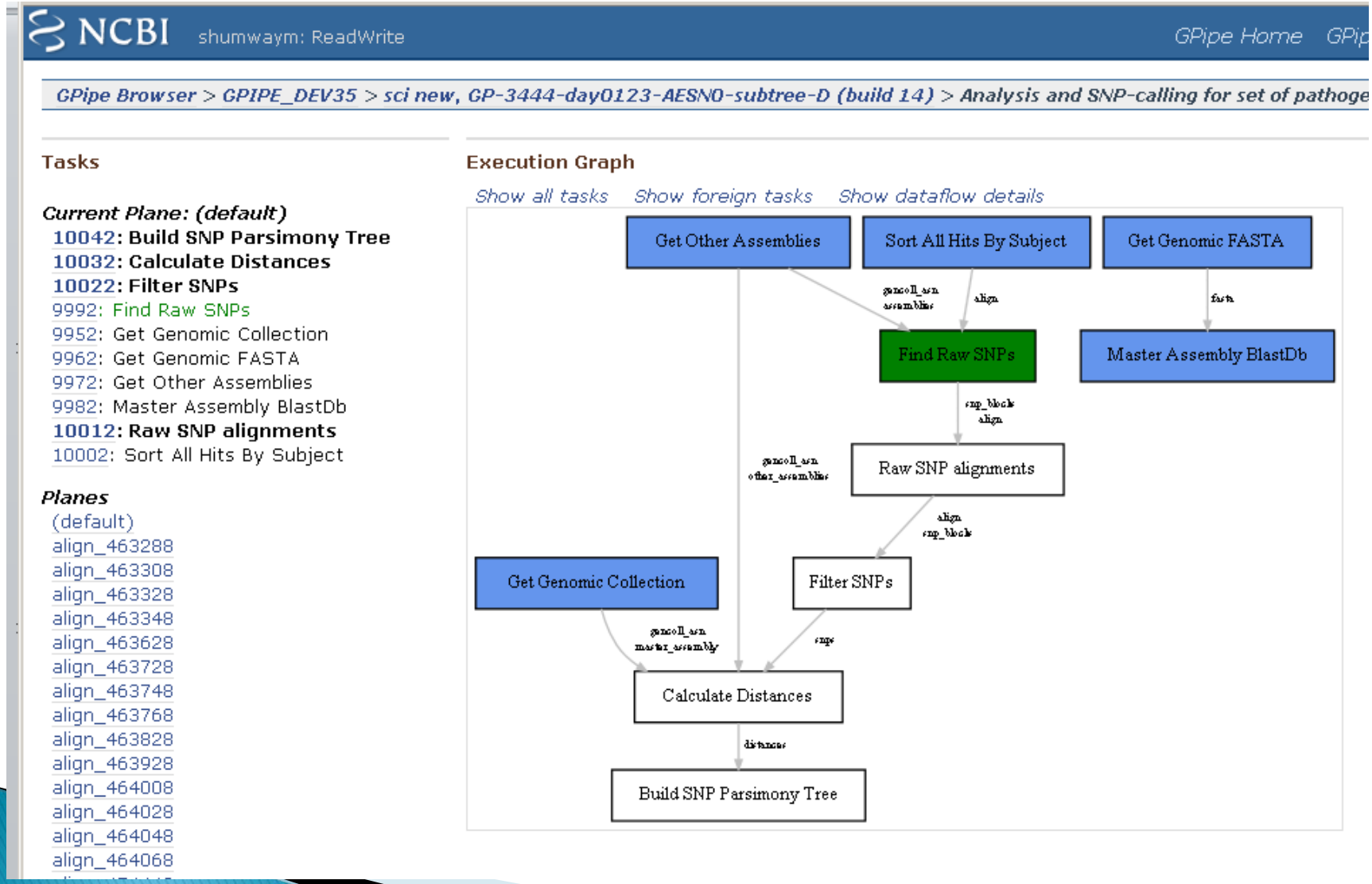
Geographic location

6a) geo_loc_name
OR
6b) lat_lon

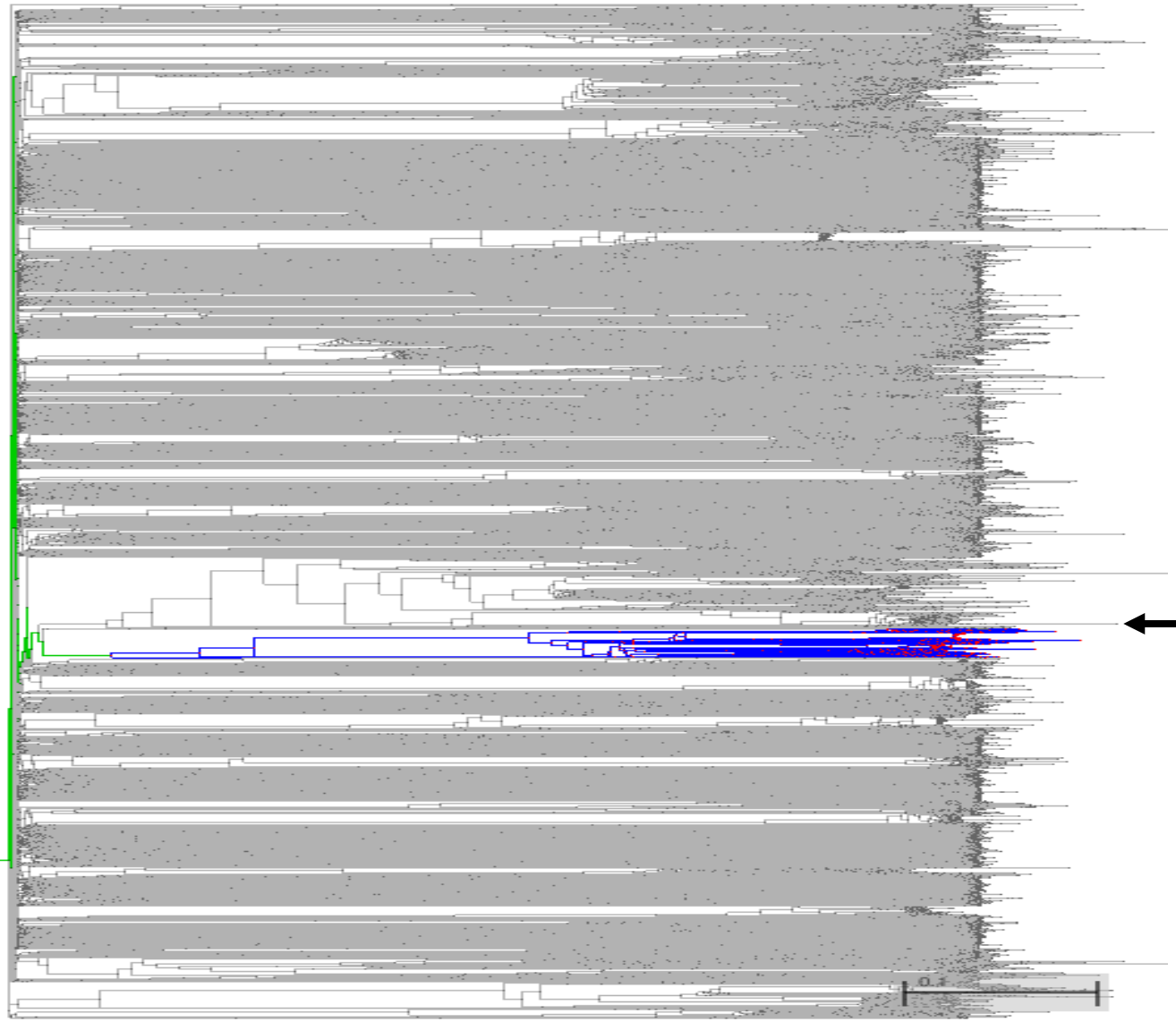
Who

collected by

Production Pipelines

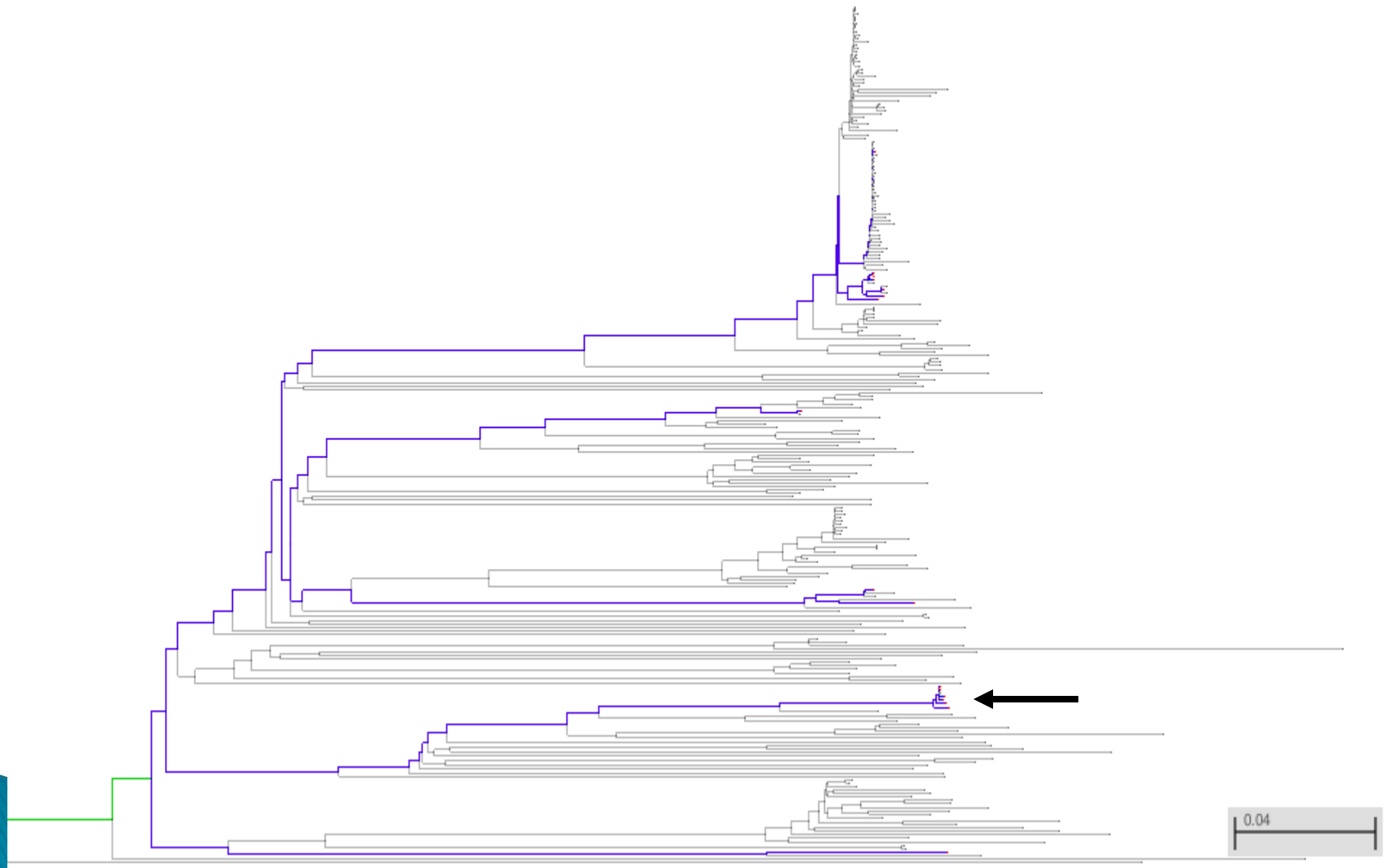


Example Surveillance Workflow

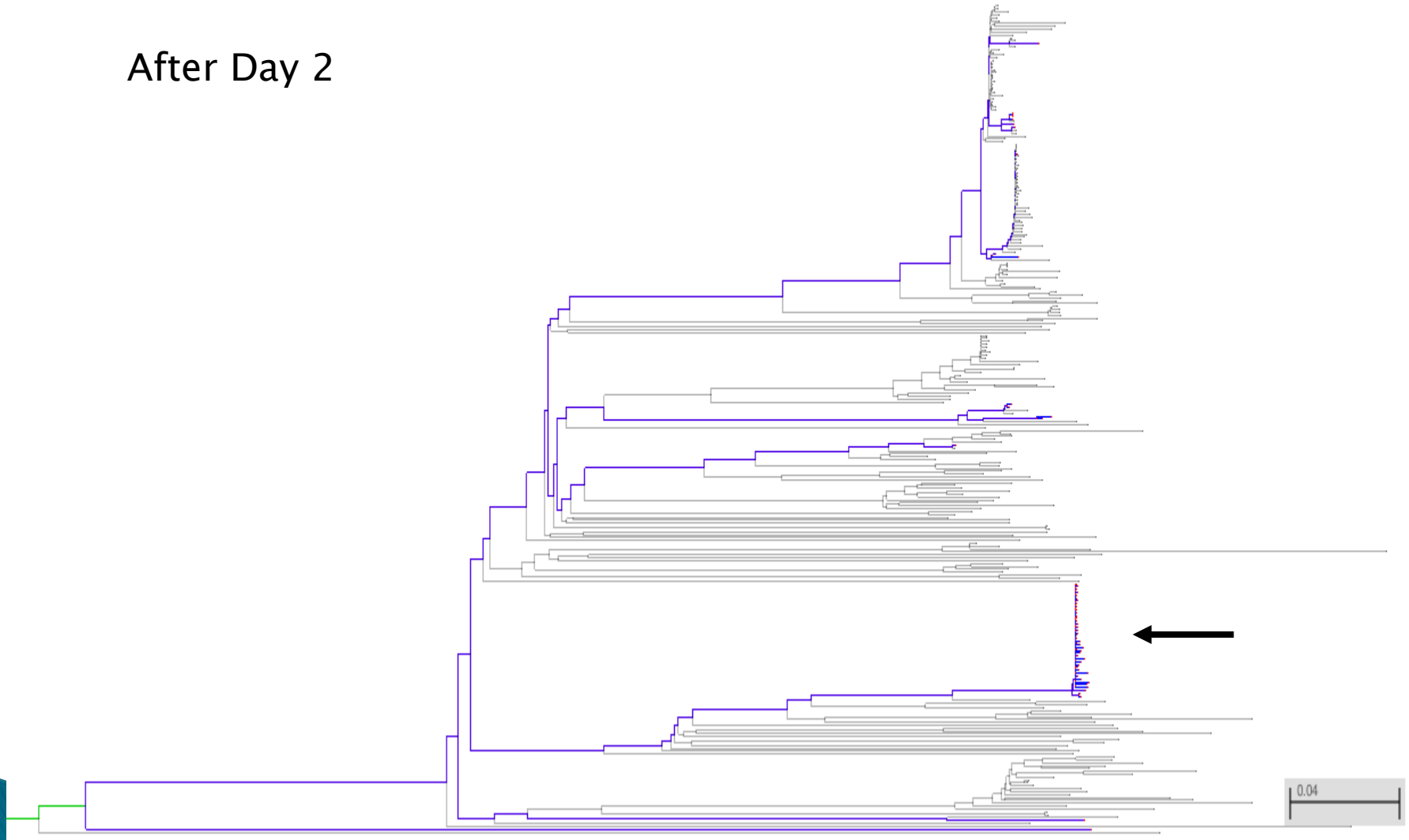


Existing *Salmonella* clade in combined *Eubacteria* kmer tree

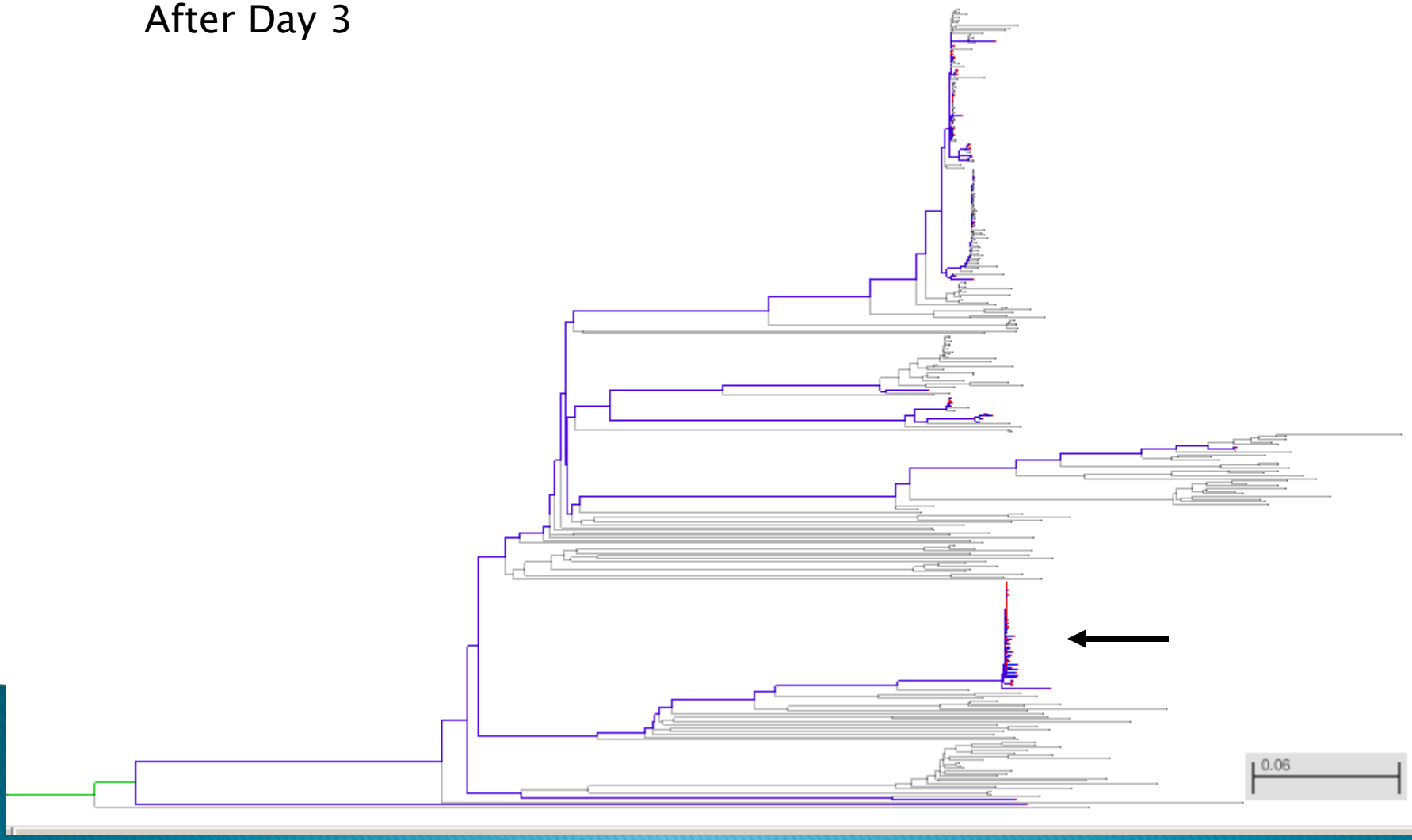
After Day 1



After Day 2

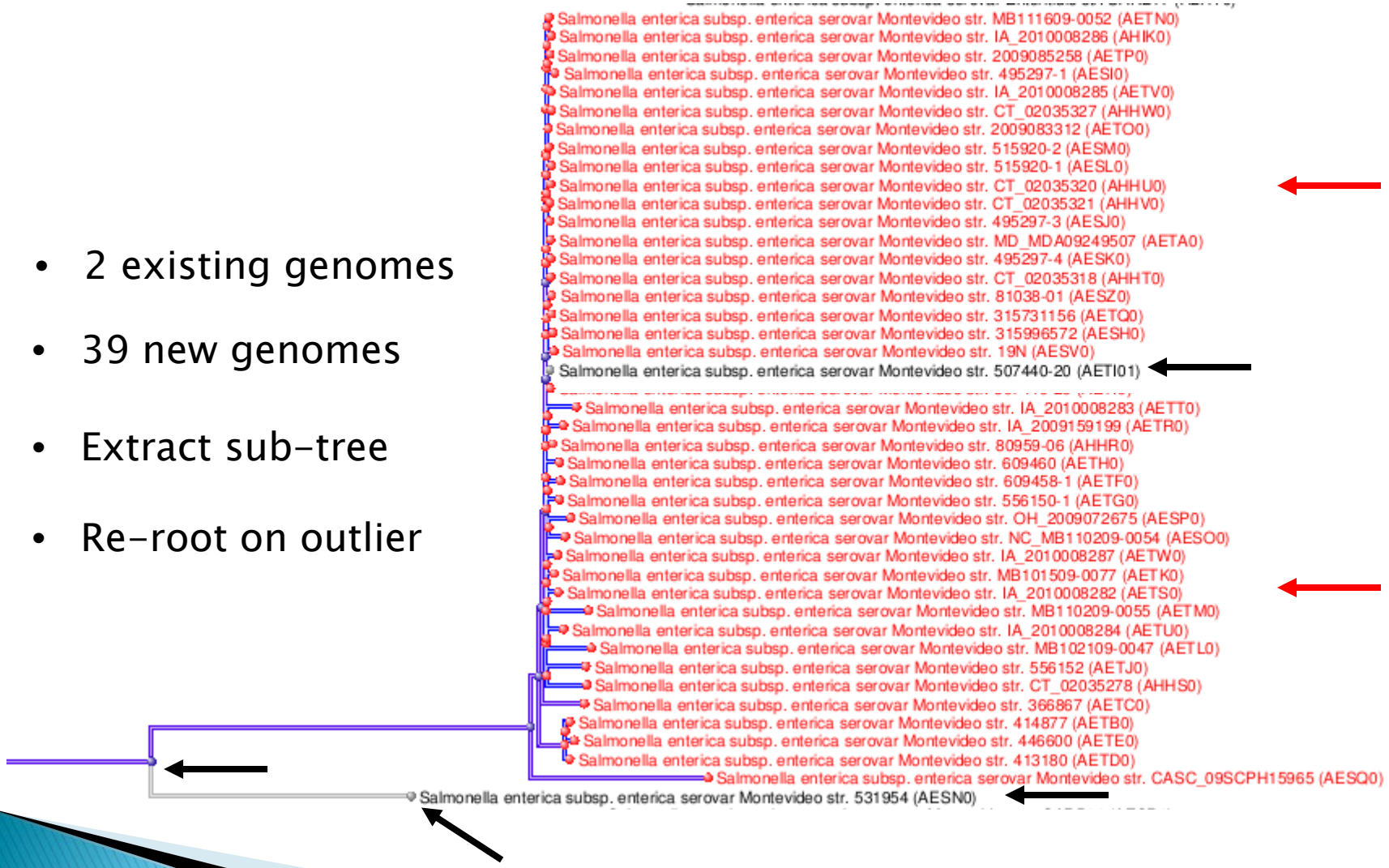


After Day 3



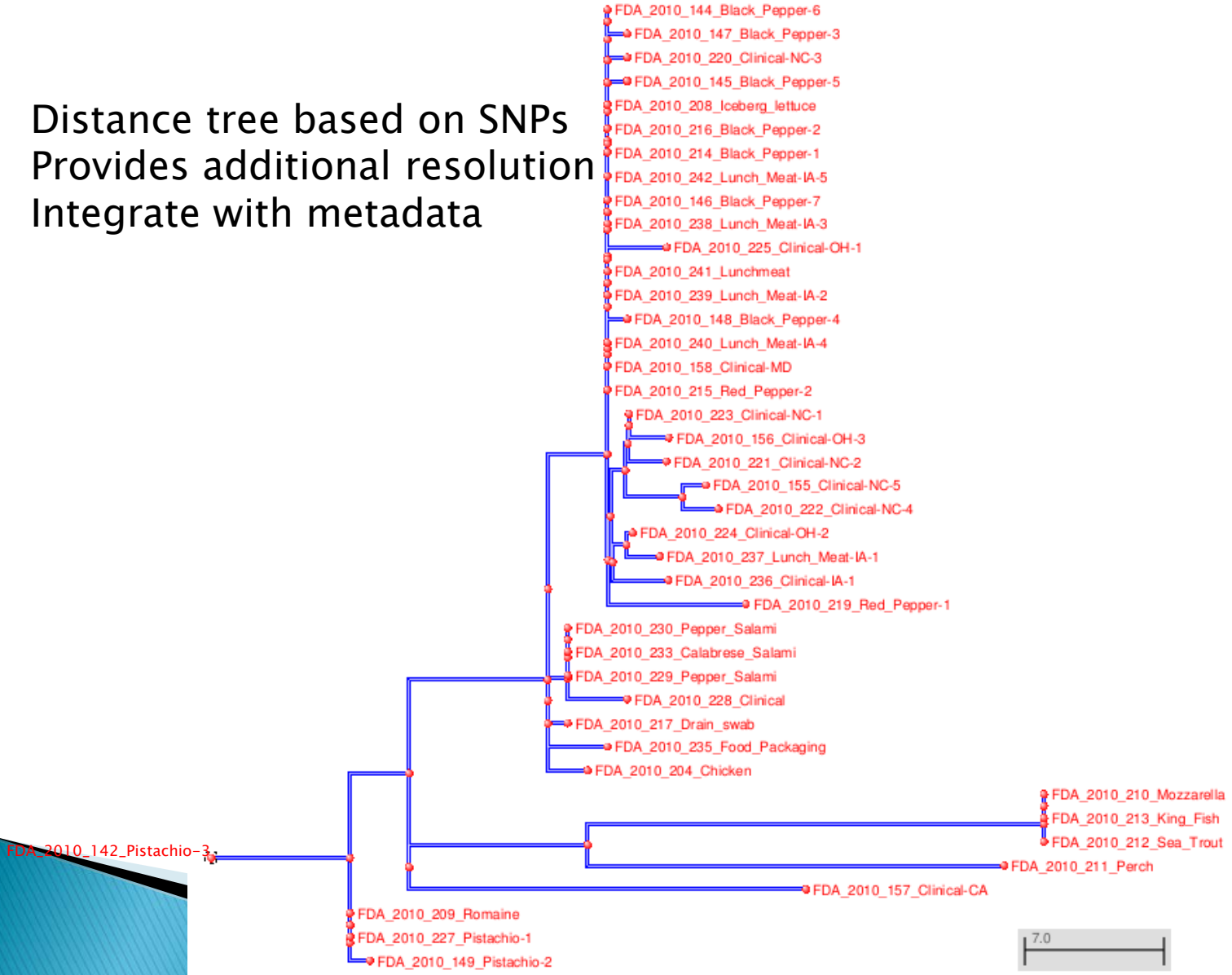
Extract cluster, Montevideo serovar

- 2 existing genomes
- 39 new genomes
- Extract sub-tree
- Re-root on outlier

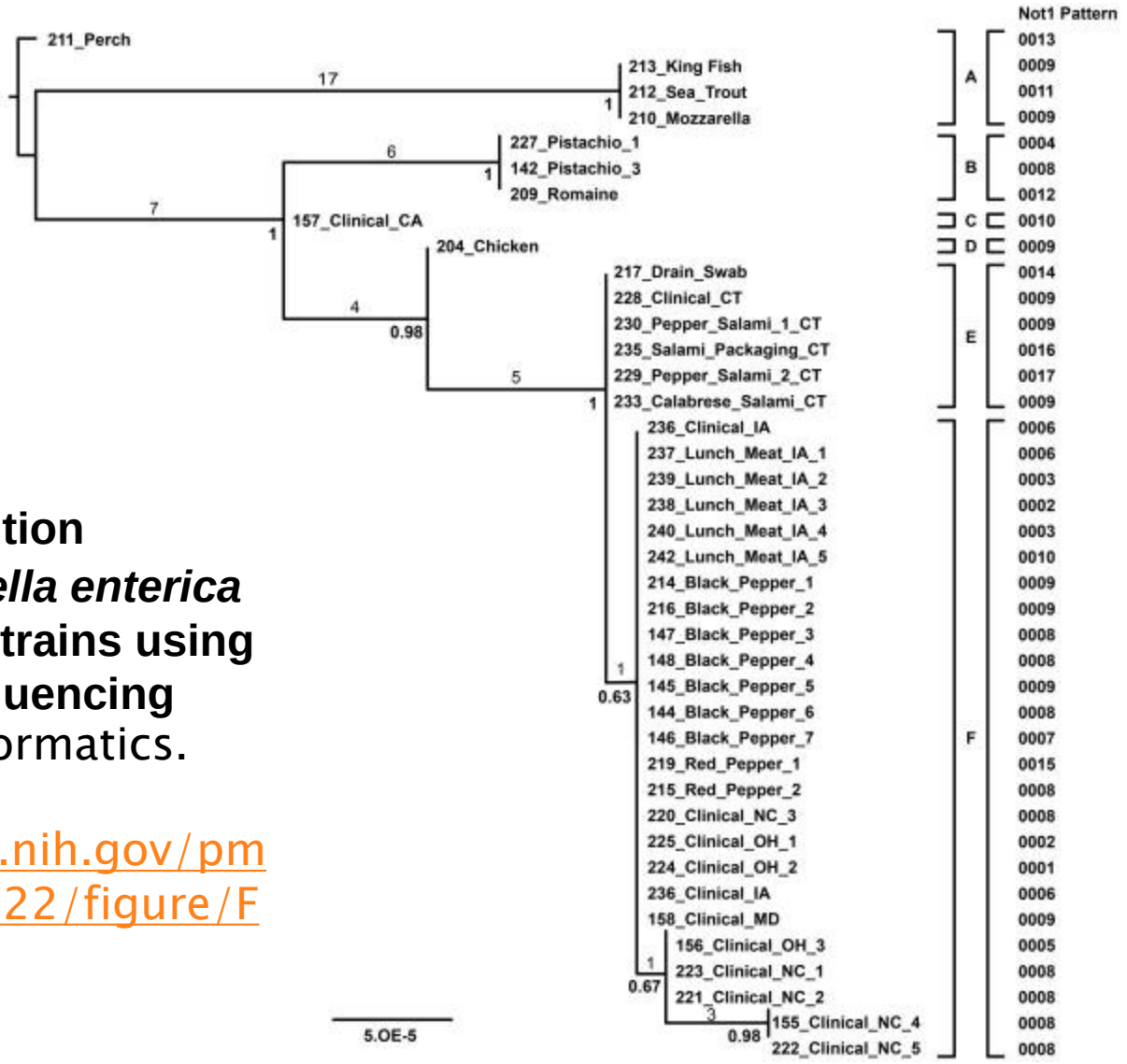


Compute new subtree

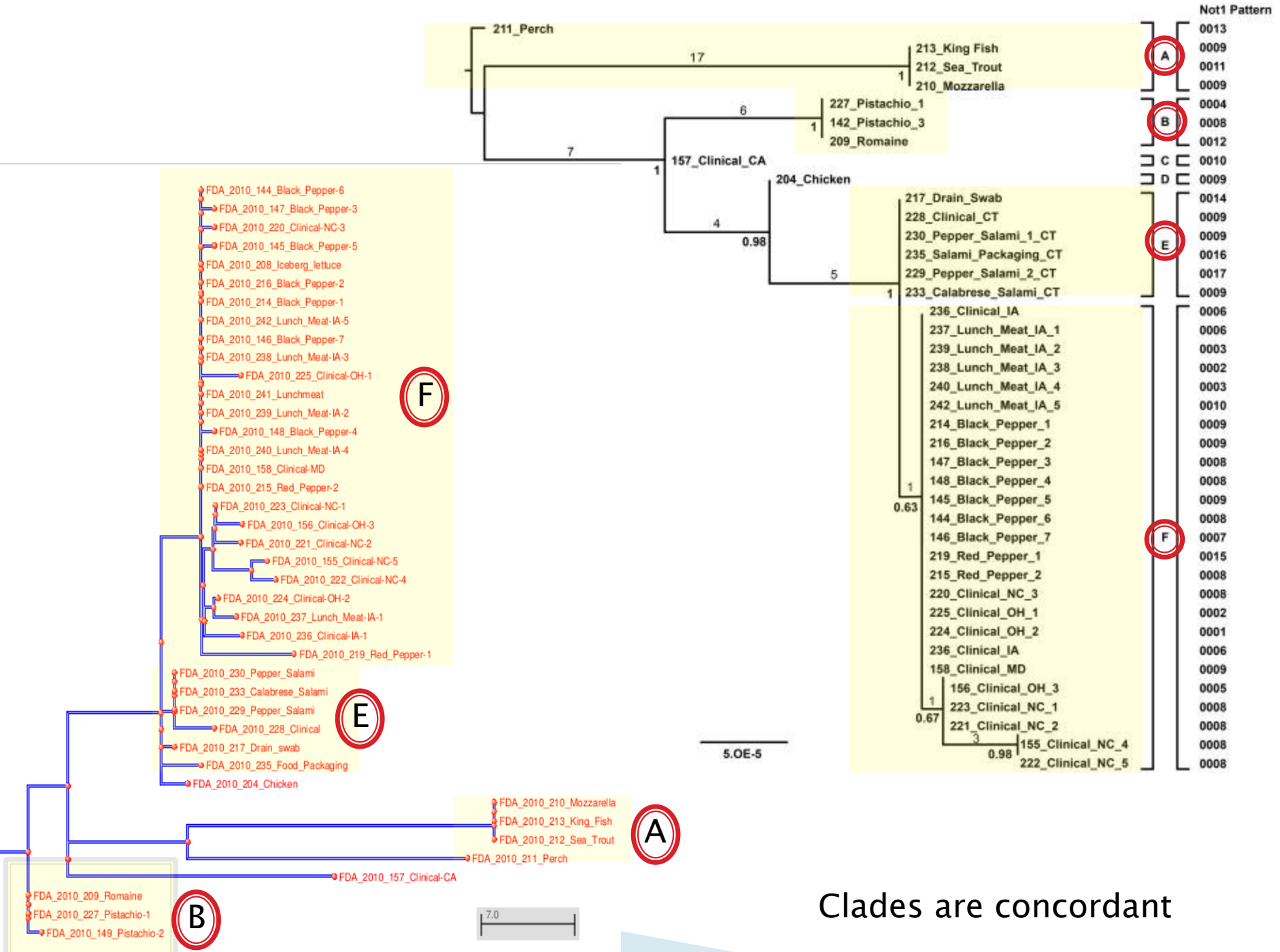
- Distance tree based on SNPs
- Provides additional resolution
- Integrate with metadata



Compare with published dataset



Luo et al. **High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach** BMC Bioinformatics. 2012; 13: 32. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3368722/figure/F3/

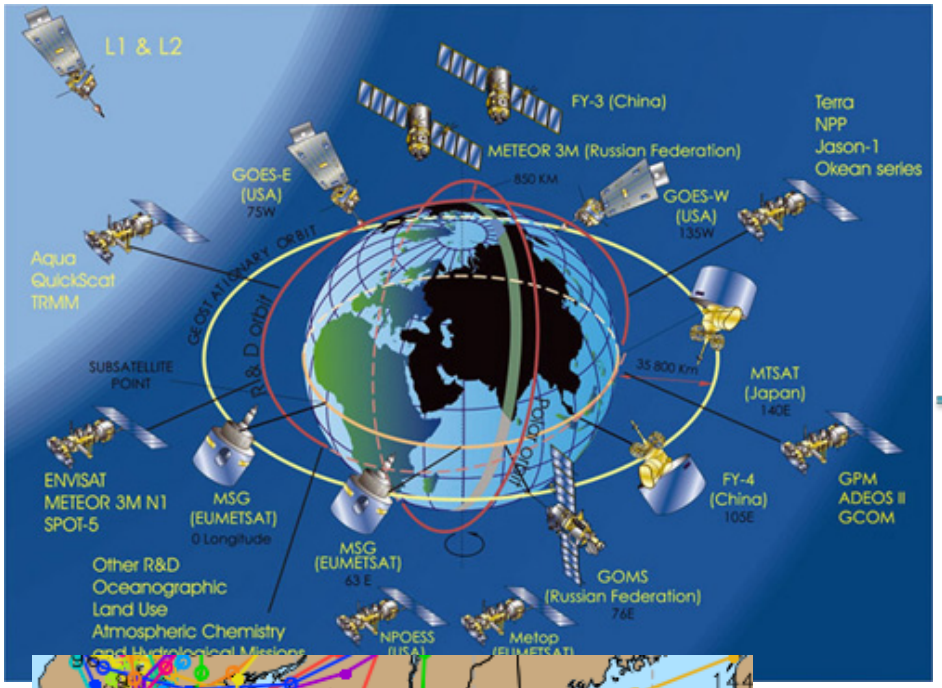


Clades are concordant

Pipeline Performance Today for 42 Salmonella Genomes

- Place Reads in Taxonomic Tree to Clade Level by Kmers
- Assemble Reads by Both de novo and Reference Guided Methods and Integrate
- Annotate All CDS and Structural RNAs
- Call SNPs Against Closely Related Genomes and Compute a New SubTree

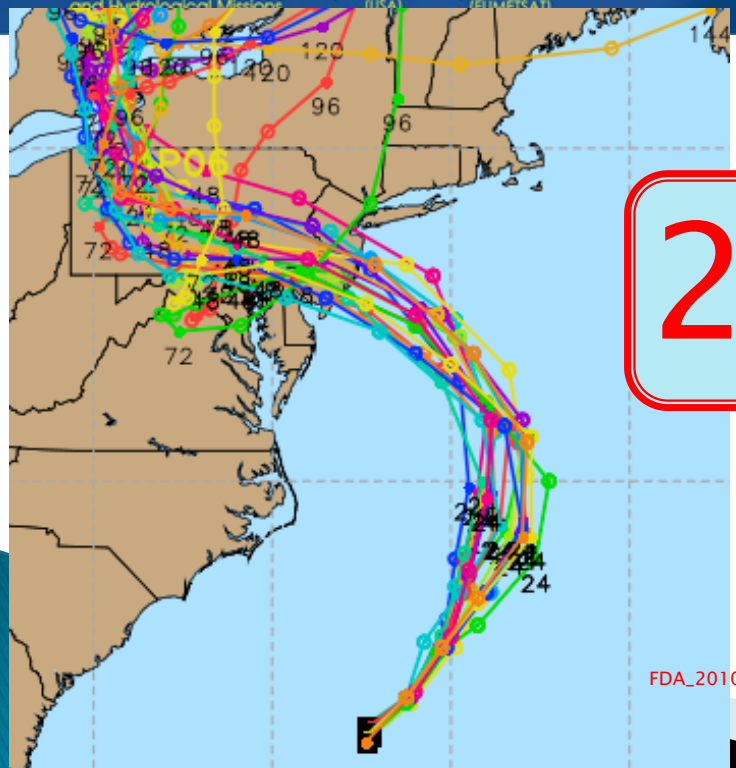
- Wall Clock Time = 4 hours



Europe
(EBI)

America
(GenBank)

Asia
(DDBJ)



2014