INTERNATIONAL FRESH PRODUCE ASSOCIATION

Challenges and Opportunities in the Fresh Produce Supply Chain

Eric L. Stevens, PhD

Vice President – Food Safety and Regulatory Affairs

Opportunities

- FDA/CDC had many conversations with industry on the use of WGS pre 2019
- Industry accepts that WGS is one part of evidence (alongside epidemiological and traceback)
- Industry does not routinely upload their own isolates they have sequenced (used typically for resident vs transient when the need arises) but does see the benefit of NCBI's Pathogen detection
- Opportunity = industry getting insight into ongoing/emerging outbreaks

What is the Global Pathogen Surveillance and Screening Pipeline? (Scraper)

The Scraper is an automated Python pipeline that monitors NCBI's public genomic databases daily, tracking pathogen isolates submitted by federal surveillance labs and other uploaders. It answers one question for fresh produce industry stakeholders: "Does this touch my business?“

We monitor:

- *Salmonella enterica*: All serovars
- *Escherichia coli*: Pathogenic strains (e.g. STEC)
- *Listeria monocytogenes*: All strains

Sources:

- CDC PulseNet — clinical human isolates (national foodborne illness network)
- FDA GenomeTrakr — environmental & food isolates (state labs, academic partners)
- FDA NARMS — antimicrobial resistance monitoring (retail & clinical)

How the scraper works

Daily automatic runs. Each run follows the same sequence:

1 Fetch
Retrieves isolates published to NCBI in the last 48 hours

2 Normalize
Standardizes submitter names, organisms, and geographic fields

3 History
Appends new isolates to the permanent 177,000+ row archive

4 Enrich
Adds SNP cluster assignments from NCBI Pathogen Detection

5 Analyze
Analyzes data based on statistical track or source quality signal track to assign context tier labels

6 Export
Writes timestamped Excel bundle: raw isolates, daily & weekly tabs

Weekly report generates every Thursday. Biweekly reports run manually on demand.

How Signals Become Tier Labels

Escalate+	Statistical spike + clinical isolates + (≥2 submitters OR containing Produce link). Highest confidence signal.
Escalate	Clinical human cases detected, with a statistical spike.
Monitor+	Produce link in isolation OR environmental isolates OR ≥2 submitters for clinical isolates without statistical spike.
Monitor	Trending upward for 2+ weeks without crossing the spike threshold. Worth watching. not yet escalated.
Background	Within normal range. No actionable signals this week.

Output

1 Raw Isolates

- ❑ Every isolate fetched in the report window
- ❑ Full metadata: organism, serovar, source, location, submitter
- ❑ SNP cluster assignments (PDG#, PDS#, PDT#)

2 Daily Summary

- ❑ Per-day counts by organism, submitter, source type, and geography
- ❑ Useful for spotting submission bursts or lab-specific patterns

3 Weekly Summary

- ❑ Per-serovar weekly counts with rolling stats and z-scores
- ❑ Context tier labels (Escalate / Monitor / Background)
- ❑ Week-over-week deltas and coverage notes

4 Cluster Score

- ❑ On-demand genomic second pass for flagged signals
- ❑ Concentration flag: concentrated / moderate / dispersed
- ❑ Which criteria fired and what PDS# subcluster dominates

What the GPSSP Cannot Tell You

The scraper is a surveillance signal tool, not an outbreak investigation tool. These limitations apply to all outputs.

Genetics ≠ Causation	SNP clusters reflect genetic similarity only. They are not proof of a common source without epidemiological/traceback evidence.
Tiers are signals	"Escalate" does not mean outbreak confirmed. It means the signal meets defined thresholds and warrants attention.
Publication lag	Isolates may be submitted to NCBI weeks or months before being published. Or uploaded weeks to months to years after isolation.
Incomplete metadata	Geographic data may be country-only. Serovar is missing for 53,000+ isolates — this is a real NCBI data gap, not a GPSSP limitation.
Cluster indexing lag	New isolates take 2–4 days to appear in Pathogen Detection. Fresh spikes (< 2 weeks old) will have incomplete cluster data.

Challenges

- WGS (by itself) can never establish causation. Datasets are only as good as the data going into them.
- Epidemiological data for suspected vehicle/source exposure is not released in its raw form (e.g. what questions were used, what the answers were)
- Other Metadata important to investigations is also not released in its raw form (e.g. age, sex, illness onset, location*)
- Industry has no way to get insight into this critical data that could help explain distribution patterns, commodities, source, etc in real time (like WGS data*)
- Challenge = industry not getting insight into the whole picture

September 3, 2026

Genetically identical isolates spread across continents over decades

This Salmonella Typhimurium phylogenetic tree shows isolates that :

- Essentially the same genotype (0 SNP difference)
- Span 17 years and multiple countries on 4 continents
- Sources: Clinical, Environmental, Food.
- If we were to based our investigations only on genetics, we'd be tempted to call this one connected event.
- Genomic identity here reflects that Typhimurium's core genome doesn't churn out enough SNPs per year to distinguish "same clone circulating for two decades" from "same outbreak this month.
- Are ALSO commonly used in research settings

environmental/other, 2025-09-13, China: Shanghai, duck, SJTUF15790, PDT002921737.1
environmental/other, 2025-09-13, China: Shanghai, duck, SJTUF15787, PDT002921740.1
environmental/other, 2025-09-13, China: Shanghai, beef, SJTUF15780, PDT002921748.1
clinical, 2025-09-23, China: Shandong, HNSDD99, PDT002937198.1
clinical, 2025-10-03, clinical, ATCC 14028, PDT002945014.1
environmental/other, 2025-11-10, Brazil: Sao Paulo, Sao Paulo, oat, SL72_25, PDT002982294.1
NULL, 2025-11-26, Mozambique, milk, PDT003000924.1
NULL, 2025-11-26, Mozambique, water, PDT003000925.1
environmental/other, 2025-11-25, China: Gansu, Jujube, PDT002997537.1
environmental/other, 2025-11-25, China: Gansu, Hairtail fish, PDT002997536.1
environmental/other, 2025-11-25, China: Gansu, patient, PDT002997241.1
clinical, 2013-10-18, CDC_2009K1283, PDT000018238.2
clinical, 2013-10-18, CDC_2009K1158, PDT000017238.1
environmental/other, 2018-12-05, chicken tissues, ATCC 14028, PDT000414969.1
environmental/other, 2018-05-25, not provided, broiler chicken, NC_STy8876448, PDT000319714.1
clinical, 2026-02-10, USA, PNUSAS563728, PDT003089574.1
NULL, 2026-03-10, Indonesia: Banyumas, Central Java, Traditional-Local Fermented Food (Tempe Dage), TD1, PDT003138375
environmental/other, 2018-05-25, Brazil, swine slaughterhouse, NC_STy9727578, PDT000317606.2
clinical, 2018-06-16, United Kingdom: United Kingdom, human, 535267, PDT000330647.1
clinical, 2018-07-18, United Kingdom: United Kingdom, human, 277853, PDT000351211.1
clinical, 2017-06-01, United Kingdom: United Kingdom, human, 340262, PDT000214158.2
environmental/other, 2026-05-26, Germany, Poultry, 20-SA03022-0, PDT003225206.1
environmental/other, 2026-05-26, Germany, Poultry, 20-SA03021-0, PDT003225207.1
environmental/other, 2026-05-29, Germany, Feed, plant-based, 22-SA01546-0, PDT003228119.1
NULL, 2026-06-15, PDT003253192.1
NULL, 2026-05-26, Germany, 21-SA00363-0, PDT003225802.1
environmental/other, 2025-07-07, China:Hangzhou, cattle, SAL20-275, PDT002825885.1
environmental/other, 2019-04-06, E40, PDT000486872.1
NULL, 2019-01-23, MA.0711R8927, PDT000450300.1
NULL, 2019-01-23, MA.0911S25602, PDT000450792.1
NULL, 2019-05-15, PDT000501241.1
NULL, 2019-05-15, PDT000501248.1
NULL, 2021-01-18, Poland, 1081_13, PDT000937583.1
NULL, 2025-03-14, Ireland, PDT002668469.1
NULL, 2025-03-07, Ireland, PDT002658531.1
NULL, 2025-03-07, Ireland, PDT002657959.1
NULL, 2025-03-07, Ireland, PDT002658535.1
NULL, 2025-06-05, not provided, PDT002795191.1
NULL, 2025-06-05, not provided, PDT002796219.1
clinical, 2025-03-22, Portugal, feces, PDT002680690.1
NULL, 2026-06-05, Ireland, PDT003241375.1
NULL, 2026-06-05, Ireland, PDT003241376.1
NULL, 2022-03-16, South Africa: KwaZulu-Natal, Food, Other, CF00227001, PDT001267786.1
clinical, 2022-04-20, South Africa: Eastern Cape, Human, Stool, YA00298168, PDT001291437.1
NULL, 2020-05-14, 80 - Somme, PDT000735365.1
NULL, 2020-02-14, PDT000683864.1
NULL, 2020-02-14, PDT000683866.1
NULL, 2020-12-12, South Africa, Plant, Food, Cabbage, CFSAN095761-2, PDT000912353.1
NULL, 2020-12-12, Laboratory, ATCC14028 UPretoria, PDT000912354.1
NULL, 2020-12-12, South Africa, Plant, Food, Spinach, CFSAN095764-2, PDT000912356.1
NULL, 2020-12-12, South Africa, Plant, Food, Cabbage, CFSAN095755-2, PDT000912357.1
NULL, 2020-12-12, South Africa, Plant, Food, Carrot, CFSAN095762-2, PDT000912358.1
NULL, 2020-12-12, South Africa, Plant, Food, Carrot, CFSAN095763-2, PDT000912359.1
NULL, 2020-12-12, South Africa, Plant, Food, Cabbage, CFSAN095756-2, PDT000912360.1
NULL, 2020-12-12, South Africa, Plant, Food, Cabbage, CFSAN095757-2, PDT000912361.1
NULL, 2020-12-12, South Africa, Plant, Food, Carrot, CFSAN095760-2, PDT000912363.1
NULL, 2020-12-12, South Africa, Plant, Food, Lettuce, CFSAN095759-2, PDT000912364.1

Conclusions

- Industry is following public WGS data closely (and it is extremely helpful)
- Data sharing needs to go both ways
- Industry knows supply chains/distribution patterns
- Challenge: how can we move to the next stage?

Acknowledgements

IFPA Food Safety Team

- Max I. Teplitski, PhD. Chief Science Officer
- Christina Ferreira, PhD. Vice President Food Safety and Quality
- Eric L. Stevens, PhD. Vice President Food Safety and Regulatory Affairs
- Jorge Quintanilla-Portillo, PhD. Food Safety Manager (**GPSSP Lead developer**)
- Alison P. Decker-Saltzmann. Science and Technology Coordinator