Real-time detection of emerging outbreaks using large-scale whole genome sequencing

Arthur Pightling, James Pettengill, Tim Muruvanda, Hugh Rand
Biostatistics and Bioinformatics Staff, Office of Analytics and Outreach, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, MD, United States

Plain Language Synopsis

Daily monitoring of the DNA sequence database of foodborne bacteria by the Center for Food Safety and Applied Nutrition's Office of Analytics and Outreach (CFSAN/OAO):

1) provides evidence supporting outbreak investigations,

2) detects emerging outbreaks, and

3) identifies facilities with long-term contamination.

We present two examples of outbreaks involving shrimp and peanut butter that were detected with the tools developed by CFSAN/OAO.

Background

Whole genome sequencing (WGS) can detect single base-pair differences between organisms and can be used to identify sources of foodborne bacterial pathogens.

Investigators use WGS to find matches between contaminated foods, environments, and sick patients during outbreaks and investigations.

Pathogens found during food facility inspections also provide clues about the causes of outbreaks prior to the collection of epidemiologic or traceback evidence and can identify facilities with long-term (resident) contamination.

Bioinformatic tools are necessary to streamline this surveillance process.
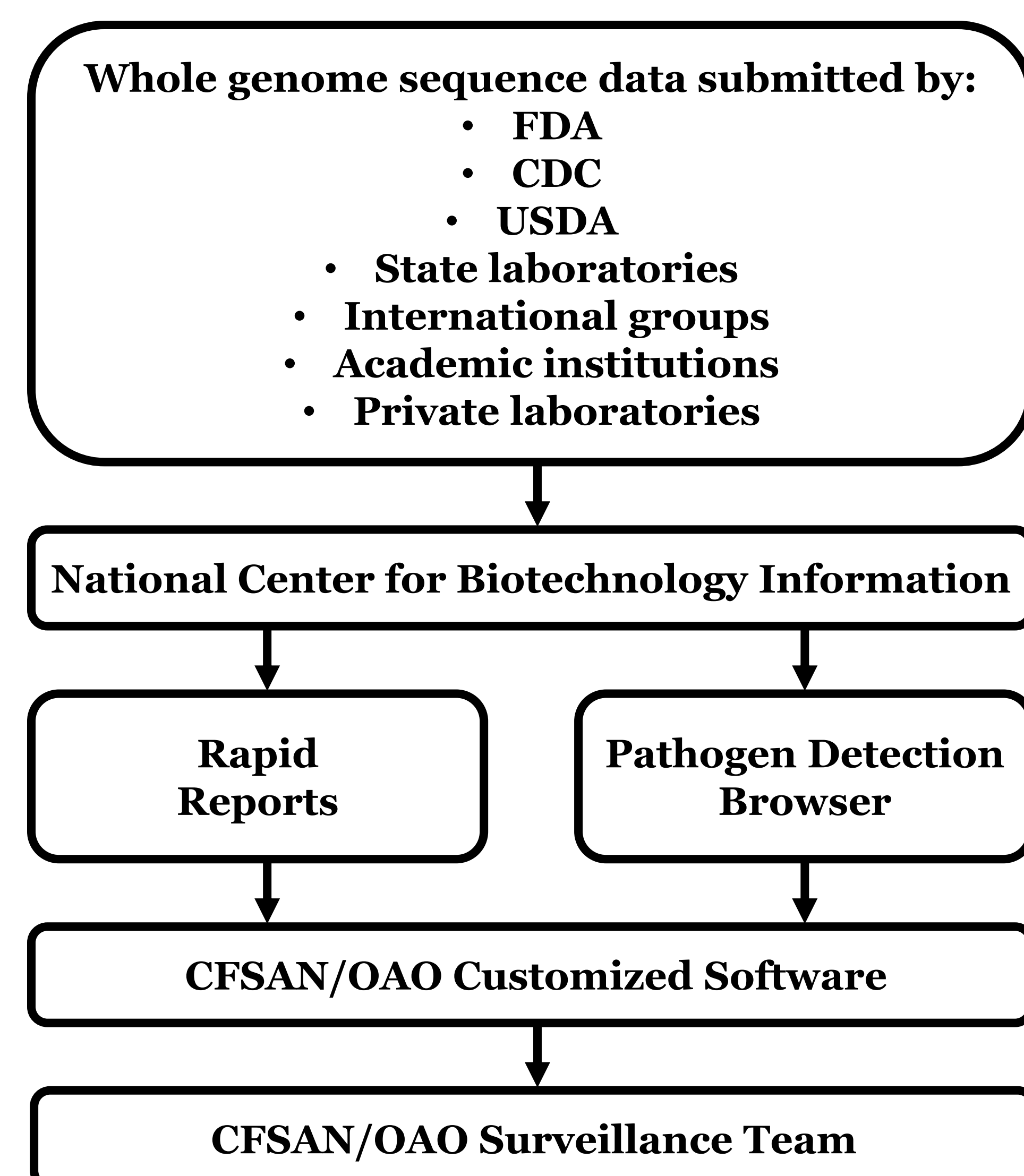
Figure 1. Flowchart illustrating the whole genome sequence data workflow from data generation to signal detection.
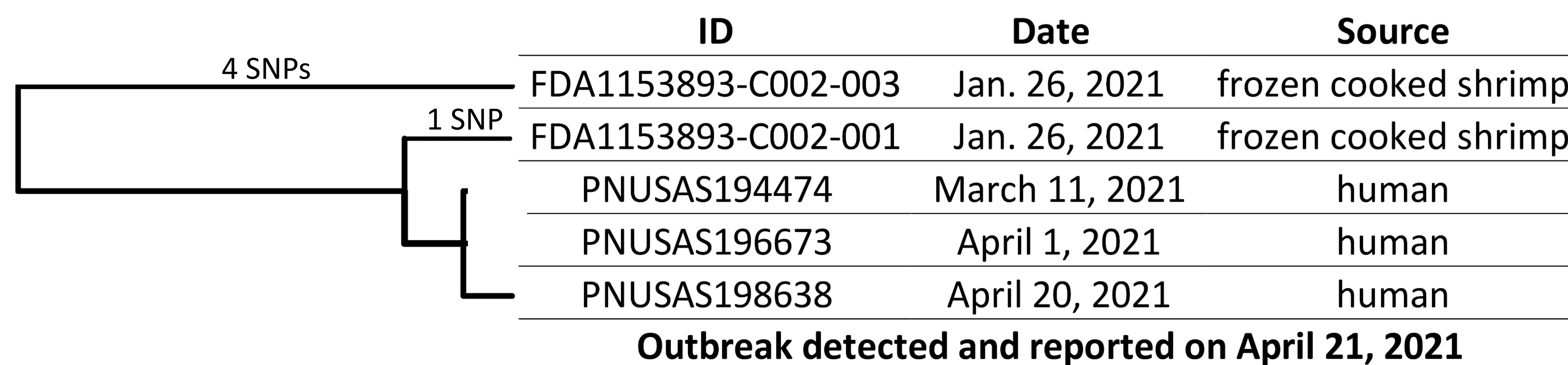
Materials and Methods

The National Center for Biotechnology Information (NCBI) Pathogen Detection (PD) system is an archive of foodborne pathogen WGS data (Figure 1; ncbi.nlm.nih.gov/pathogens/).

The PD Isolates Browser enables users to explore sequence data and identifies groups of closely related sequences and their relationships within that group (ncbi.nlm.nih.gov/pathogens/isolates/).
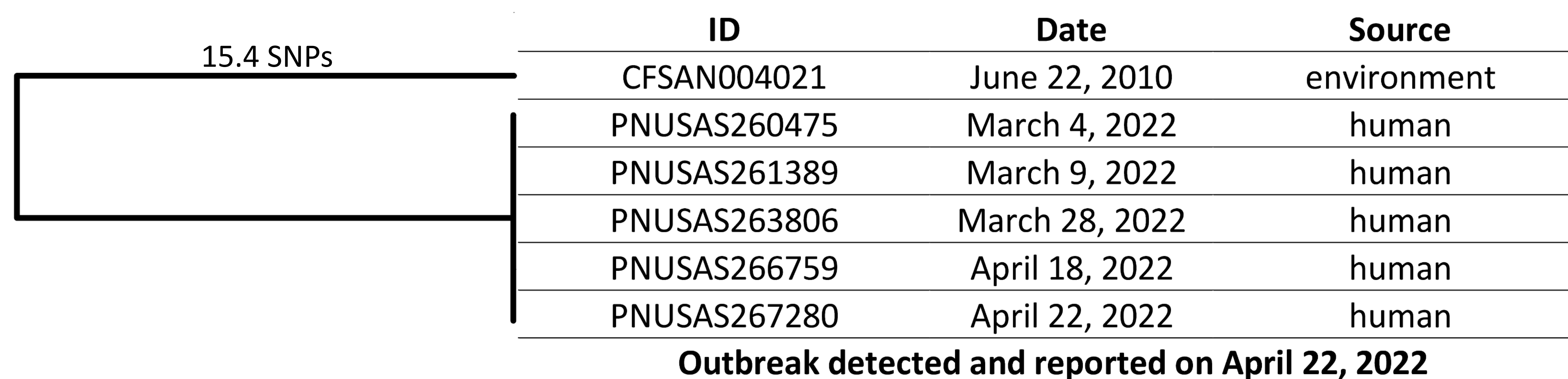
PD Rapid Reports augments the Isolates Browser by providing rapid (< 2hours) results for new data and provides allele distances between newly submitted data and sequences already in the database (ftp.ncbi.nlm.nih.gov/pathogen/Results/Salmonella/Rapid_reports/).

Results are processed with software written by the CFSAN/OAO Bioinformatics Team to filter results that meet match criteria.

Signals are reviewed with CFSAN's Coordinated Outbreak Response and Evaluation Network (CORE) and the Office of Compliance for action.

	ID	Date	Source
4 SNPs	FDA1153893-C002-003	Jan. 26, 2021	frozen cooked shrimp
1 SNP	FDA1153893-C002-001	Jan. 26, 2021	frozen cooked shrimp
	PNUSAS194474	March 11, 2021	human
	PNUSAS196673	April 1, 2021	human
	PNUSAS198638	April 20, 2021	human

Outbreak detected and reported on April 21, 2021

Figure 2. Metadata for five *Salmonella enterica* whole genome sequence datasets associated with an outbreak caused by contaminated shrimp in 2021 with tree that illustrates their genomic relatedness. . The average numbers of nucleotide differences between the clinical and non-clinical isolates are shown along the branches.

	ID	Date	Source
15.4 SNPs	CFSAN004021	June 22, 2010	environment
	PNUSAS260475	March 4, 2022	human
	PNUSAS261389	March 9, 2022	human
	PNUSAS263806	March 28, 2022	human
	PNUSAS266759	April 18, 2022	human
	PNUSAS267280	April 22, 2022	human

Outbreak detected and reported on April 22, 2022

Figure 3. Metadata for six *Salmonella enterica* whole genome sequence datasets associated with an outbreak caused by contaminated peanut butter in 2022 with tree that illustrates their genomic relatedness. The average number of nucleotide differences between the clinical and non-clinical isolates is shown along the branch.

Results and Discussion

Examples of emerging outbreaks detected during routine surveillance:

2021: Three submitted clinical isolates matched previously submitted isolates from imported shrimp. Follow up with CORE resulted in recalls of product and an outbreak investigation (Figure 2; fda.gov/food/outbreaks-foodborne-illness/outbreak-investigation-salmonella-weltevreden-frozen-pre-cooked-shrimp-april-2021).

2022: Submitted clinical isolates matched previous isolates from peanut butter and a food processing facility. Follow up with CORE resulted in recalls of product and an outbreak investigation that confirmed sickened people consumed product from the facility. The data also indicated the likely presence of a resident pathogen (Figure 3; fda.gov/food/outbreaks-foodborne-illness/outbreak-investigation-salmonella-peanut-butter-may-2022).

Conclusion

Surveillance of WGS data is highly effective for preventing illnesses. Tools developed by the CFSAN/OAO Bioinformatics Team and NCBI's PD group database are vital to this effort (Figure 4).
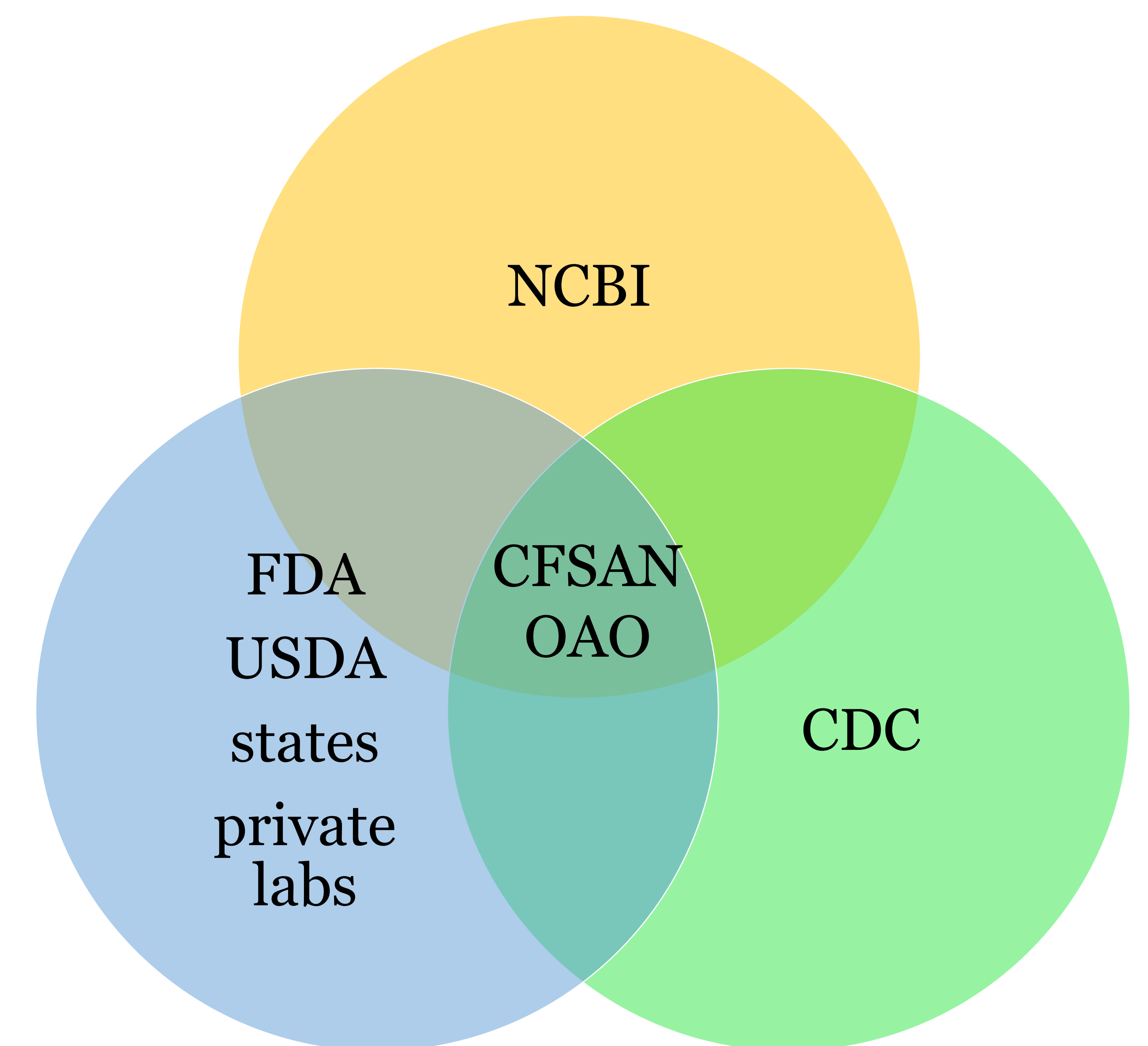
Figure 4. Venn diagram illustrating the relationships between groups that generate whole genome sequence data (from foods, environments [FDA, USDA, states, and private laboratories], and sick people [CDC], the NCBI (which stores and performs preliminary analyses of the data), and CFSAN/OAO (which further analyzes the data, performs daily surveillance, and reports emerging outbreaks for follow-up investigations.

Acknowledgments

We are thankful to the NCBI for their collaboration, expertise, and ongoing efforts to maintain the Pathogen Detection portal.

We are also appreciative of the members of CFSAN's Office of Compliance and the Coordinated Outbreak Response for their expertise and professionalism.

Finally, we are grateful to all of the experts in the FDA's Office of Regulatory Affairs and CFSAN's Office of Regulatory Science that collect samples and generate and manage the sequence data.