



TEXAS
Health and Human
Services

Texas Department of State
Health Services

Genetic Sequencing at DSHS Austin Laboratory: An Introduction

Bonnie Oh, PhD
Maliha Rahman, PhD

2024 ELC Workshop, February 15th-16th, 2024
UT Pickle Research Campus

Genetic Sequencing Branch



TEXAS
Health and Human
Services

- Budded out of Molecular Biology Team in 2016
 - Advanced Molecular Detection Group in 2021
 - Branch in 2023
- 15 members
- Branch Goals:
 - Using next generation sequencing (NGS) for surveillance and identification of public health threats in human, food, and environmental samples.
 - Expand capacity to become a core sequencing lab for Texas





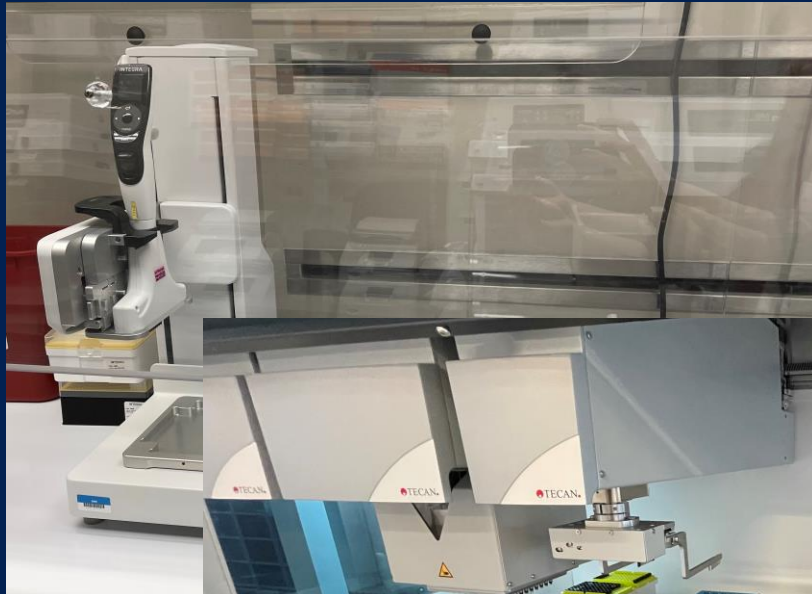
TEXAS
Health and Human
Services

Genetic Sequencing Branch

- Partners with PulseNet (CDC) and GenomeTrakr (FDA)
 - Since 2015, national lab network focusing on clinical, food and environmental surveillance
 - Special projects with CDC (OAMD) /APHL (ISC)
- With epidemiologists, we help identify variants of interest and track outbreaks (e.g. foodborne, hospital acquired)
- Recruit samples from Universities/Laboratories
 - GenomeTrakr project and Wastewater surveillance



New instrumentation to expand capacity



TEXAS

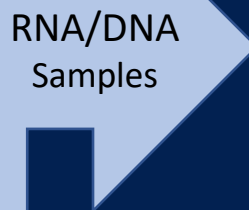
Health and Human Services

Texas Department of State
Health Services

Sample Workflow

Sample Receiving
(Clinical, Food, Environmental)

- Clinical Virology
- Clinical Bacteriology
- Consumer Microbiology
- Molecular Biology
- Parasitology
- Newborn Screening
- University Collaborators



Sample Processing

- Sample queuing (ex: LIMS test ordered)
- cDNA Generation
- DNA Quantification
- Library Preparation
- Quality Control

Sequencing

Resulting

- Submitters
- Epidemiologists
- CDC/FDA
- Collaborators
- Public Data Repositories (ex: NCBI)

Bioinformatic Analyses

- In-house pipelines
- CDC pipelines



Next Generation Sequencing (NGS)

Whole Genome Sequencing (WGS)

- Uses the whole genome of the microorganism to identify serotypes, drug resistance-related genes, toxin genes, cluster analysis, characteristics of interest

Amplicon-Based Sequencing

- Uses primers to target a specific organism or gene region for amplification to identify variants or characteristics of interest

Short-read vs. Long-read sequencing



TEXAS
Health and Human Services

Texas Department of State
Health Services

The Whole Genome Sequencing (WGS) Process

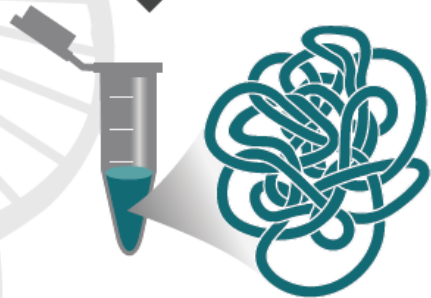
WGS is a laboratory procedure that determines the order of bases in the genome of an organism in one process. WGS provides a very precise DNA fingerprint that can help link cases to one another allowing an outbreak to be detected and solved sooner.

Bacterial Culture



1. DNA Extraction

- 1 Scientists take bacterial cells from an agar plate and treat them with chemicals that break them open, releasing the DNA. The DNA is then purified.



3. DNA Library Preparation

- 3 Scientists make many copies of each DNA fragment using a process called polymerase chain reaction (PCR). The pool of fragments generated in a PCR machine is called a "DNA library."

2. DNA Shearing

- 2 DNA is cut into short fragments of known length, either by using enzymes "molecular scissors" or mechanical disruption.



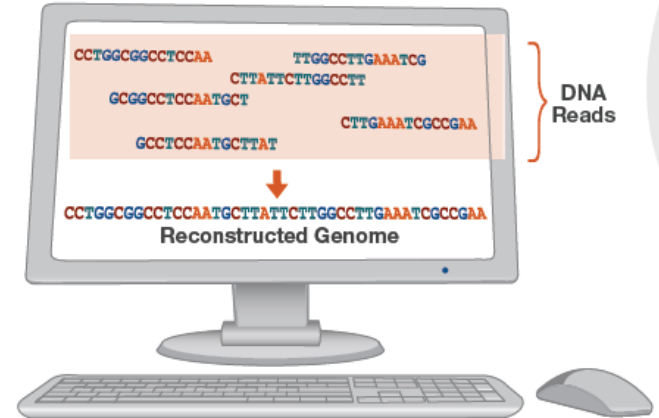
Short Read Sequencing

4. DNA Library Sequencing

- 4 The DNA library is loaded onto a sequencer. The combination of nucleotides (A, T, C, and G) making up each individual fragment of DNA is determined, and each result is called a "DNA read."



5. DNA Sequence Analysis



- 5 The sequencer produces millions of DNA reads and specialized computer programs are used to put them together in the correct order like pieces of a jigsaw puzzle. When completed, the genome sequence containing millions of nucleotides (in one or a few large pieces) is ready for further analysis.

CS264789-A



TEXAS
Health and Human Services

Texas Department of State
Health Services

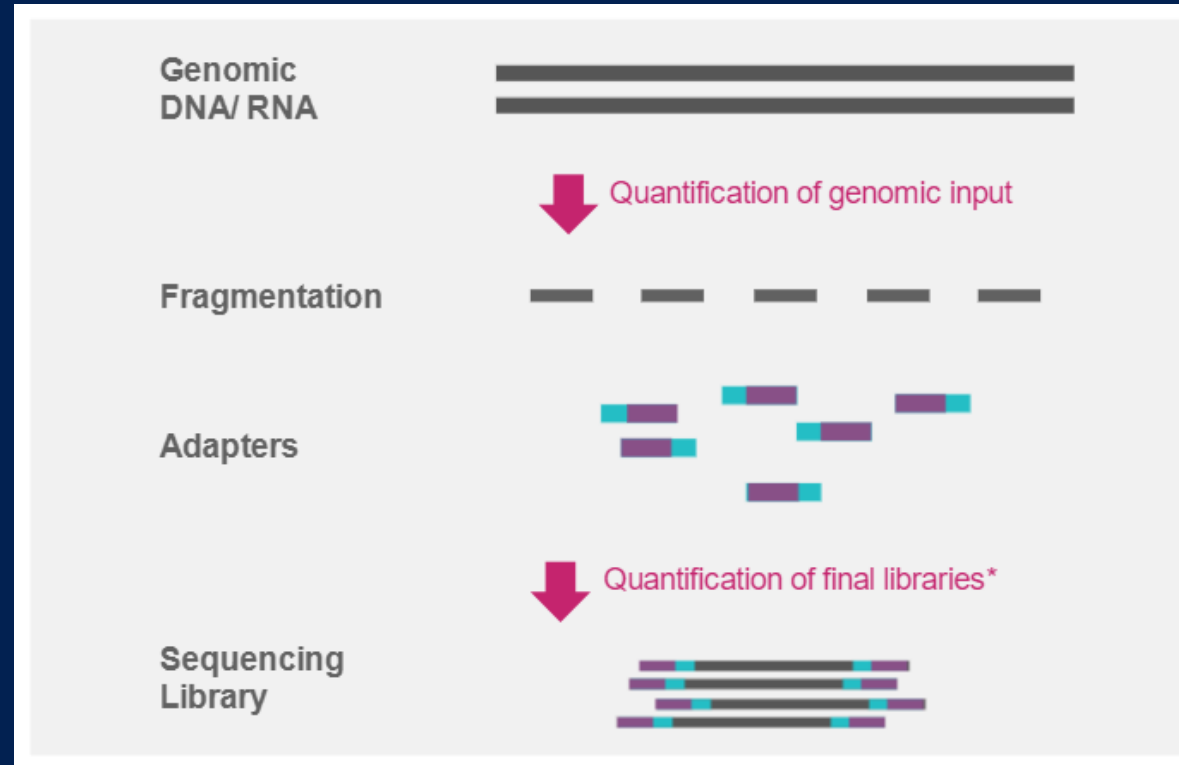
Short Read Sequencing: Illumina WGS Library Preparation



TEXAS

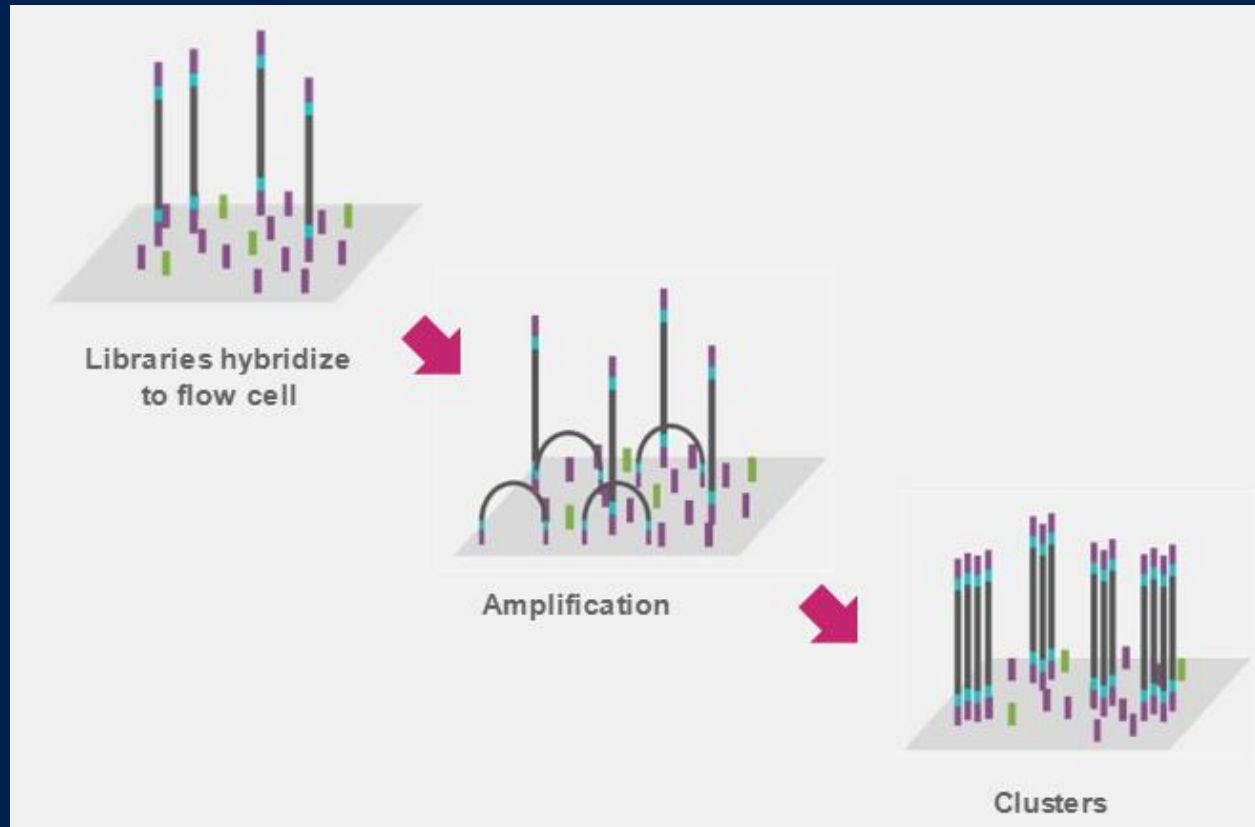
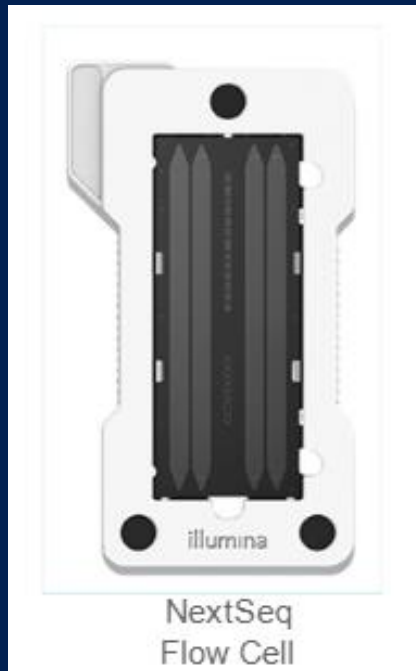
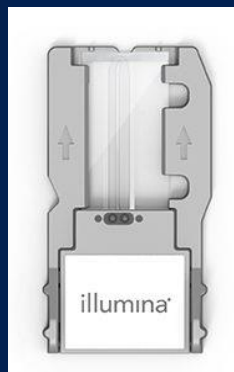
Health and Human Services

Texas Department of State
Health Services



Short Read Sequencing: Sequencing by Synthesis (SBS by Illumina)

Cluster generation occurs on the flow cell



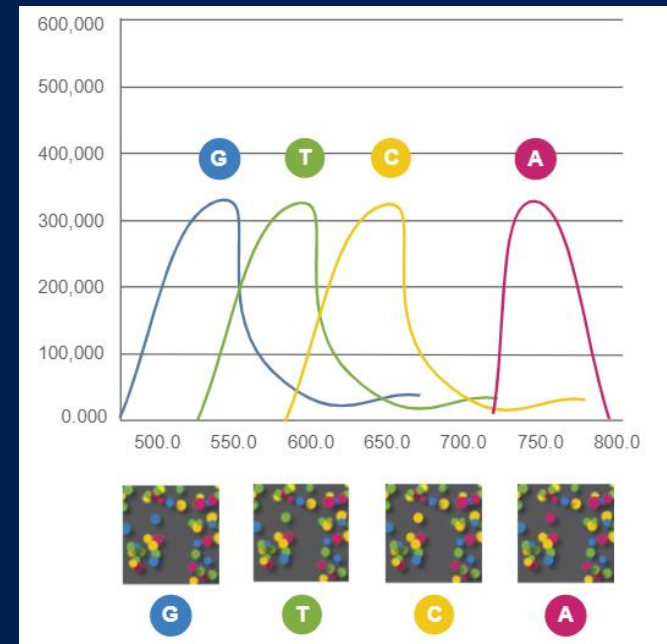
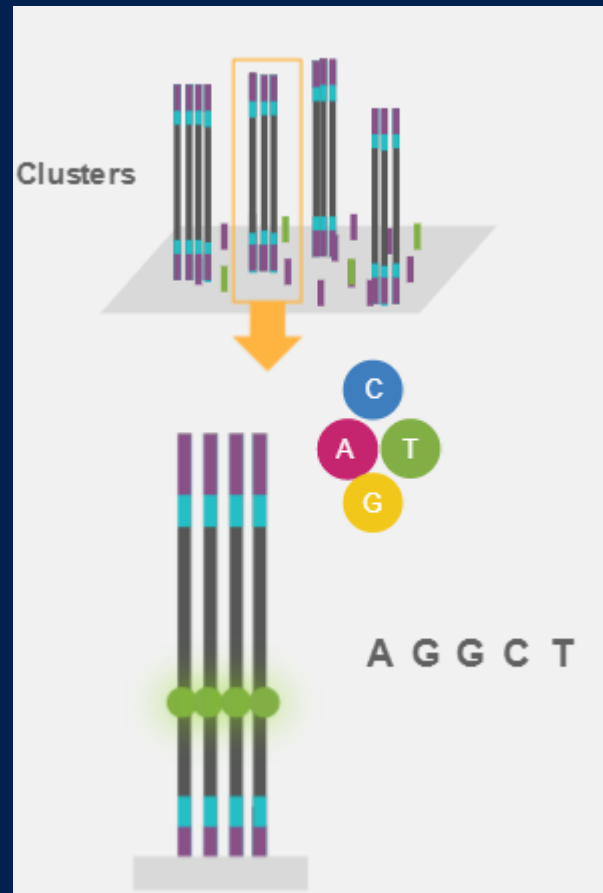
Short Read Sequencing: Sequencing by Synthesis (SBS by Illumina)



TEXAS

Health and Human Services

Texas Department of State
Health Services



- Camera images tiles of clusters
- Base calling
- Generates forward and reverse sequences

Long Read Sequencing: Oxford Nanopore Technologies

- Long-read sequencing has technical advantages for the detection of specific types of genetic variation- preferable for metagenomics studies.
- Flow cell contains protein nanopores where genetic material passes through
- Each nucleotide A,T,C,G has a different electrical charge, which is measured to identify the sequence
- Requires larger initial input of genetic material
- Real time analysis of RNA/DNA



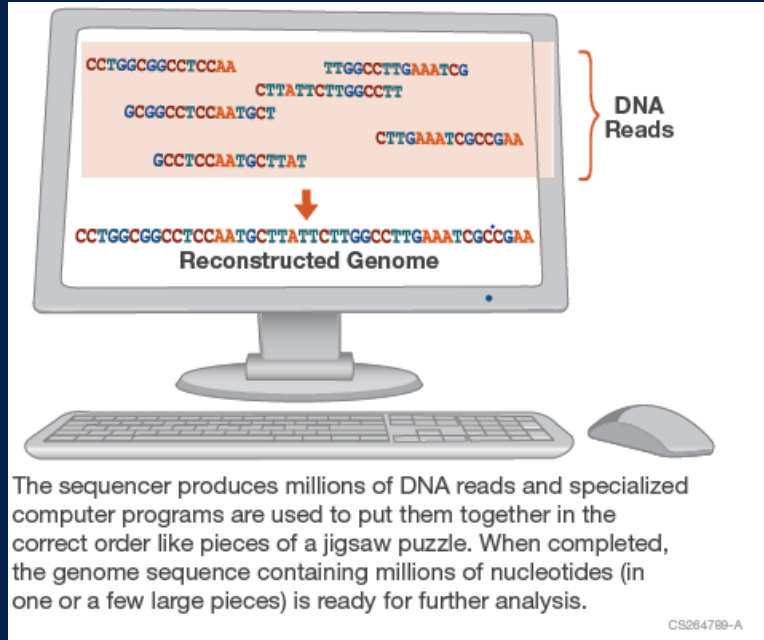
Data Analysis



TEXAS

Health and Human Services

Texas Department of State
Health Services



Find relatedness, build
phylogenetic trees, track
variants

- Align and Assemble fragments
- Forward and Reverse reads (short-read)
- Compare to reference genomes
 - Secret pipeline for Covid
 - BioNumerics for PulseNet/GenomeTrakr
 - Armadillo for antimicrobial resistance
 - Bluebonnet for NBS



TEXAS

Health and Human Services

Texas Department of State
Health Services

Equipment Summary

- Automation:
 - Integra Assist Plus (2)
 - TECAN fluent (2)
 - Clear Dx (1)
- Sequencers:
 - MiSeq (6)
 - NextSeq 2000 (2)
 - NovaSeq 6000 (2)
 - ONT MinION (2)



Our New Sequencing Room



TEXAS

Health and Human Services

Texas Department of State
Health Services



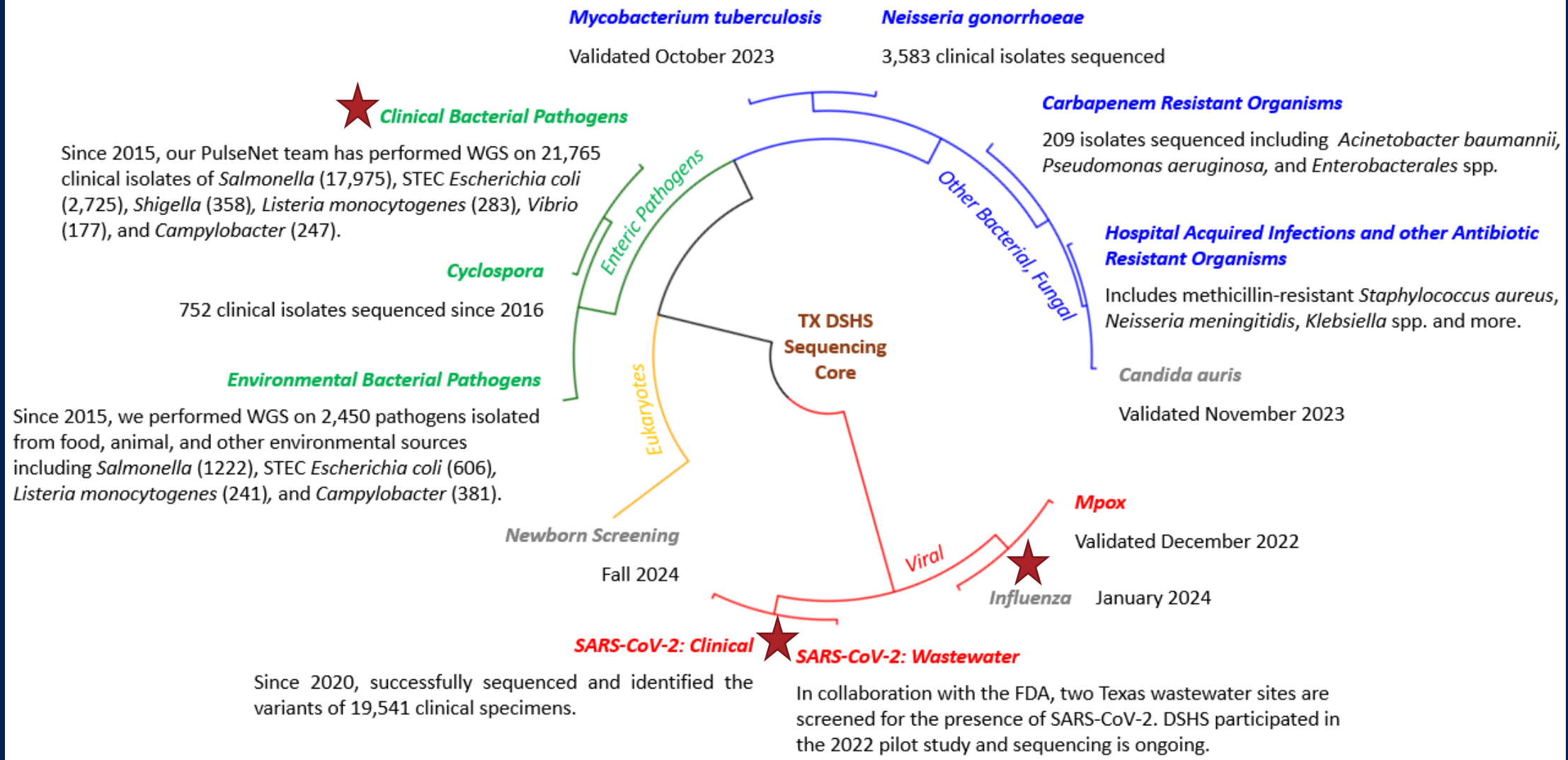
Genetic Sequencing Branch



TEXAS

Health and Human Services

Texas Department of State Health Services



Scope of Testing is Growing

- **Current Testing**

- *Salmonella enterica*
- *Escherichia coli*
- Shigella species
- *Listeria monocytogenes*
- Campylobacter spp.
- Vibrio spp.
- Carbapenem Resistant Organisms
- *Neisseria gonorrhoeae*
- *Mycobacterium tuberculosis*
- Cyclospora
- SARS-CoV-2
- Mpox
- Influenza

- **Future Testing**

- *Candida auris*
- Metagenomics for Unknown Threats (clinical and environmental)
- Lysosomal Storage Disorders for Newborn Screening
- Expanded wastewater surveillance



TEXAS
Health and Human Services

Texas Department of State
Health Services

Genetic Sequencing Branch Testing Numbers

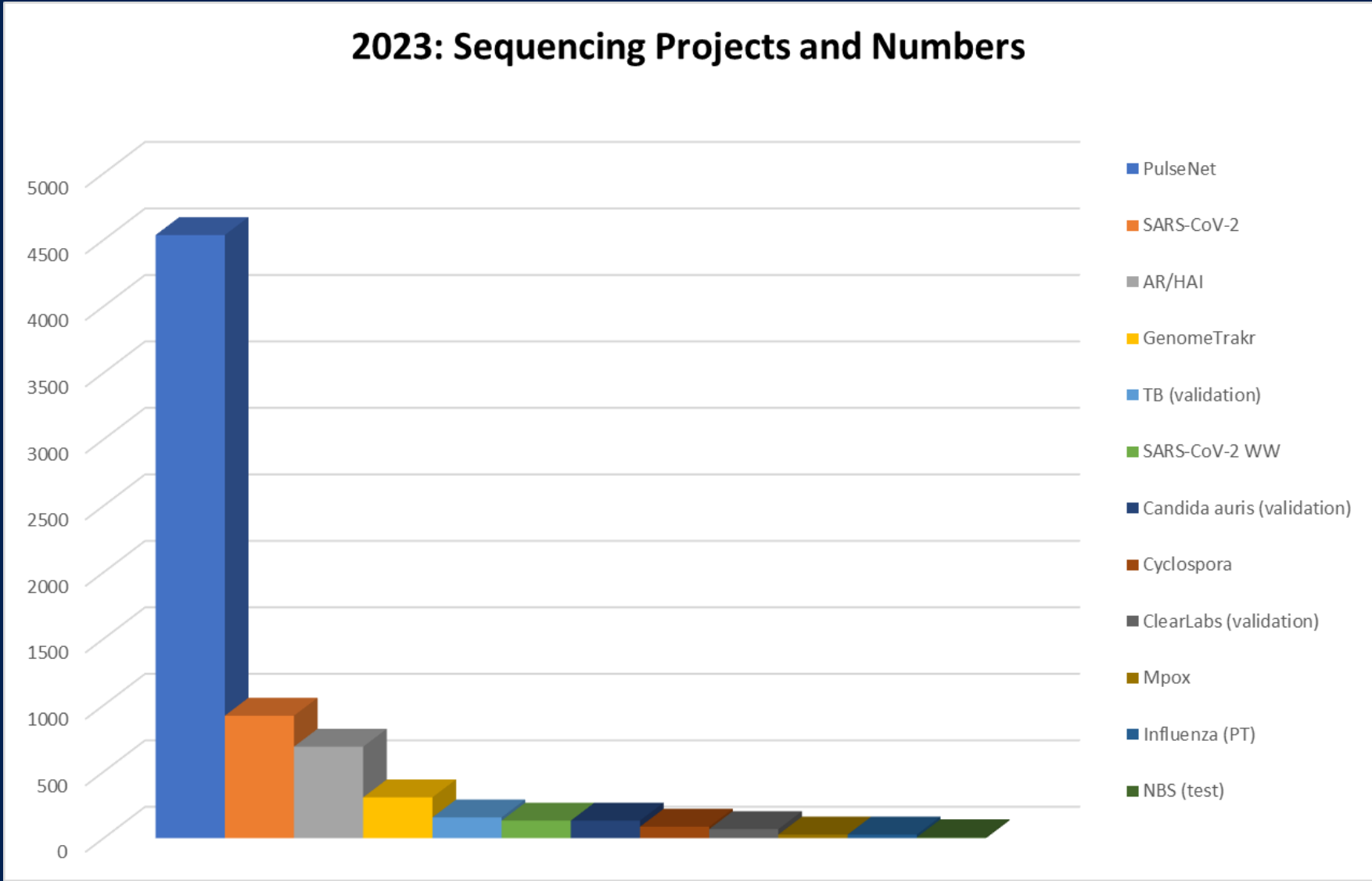


TEXAS

Health and Human Services

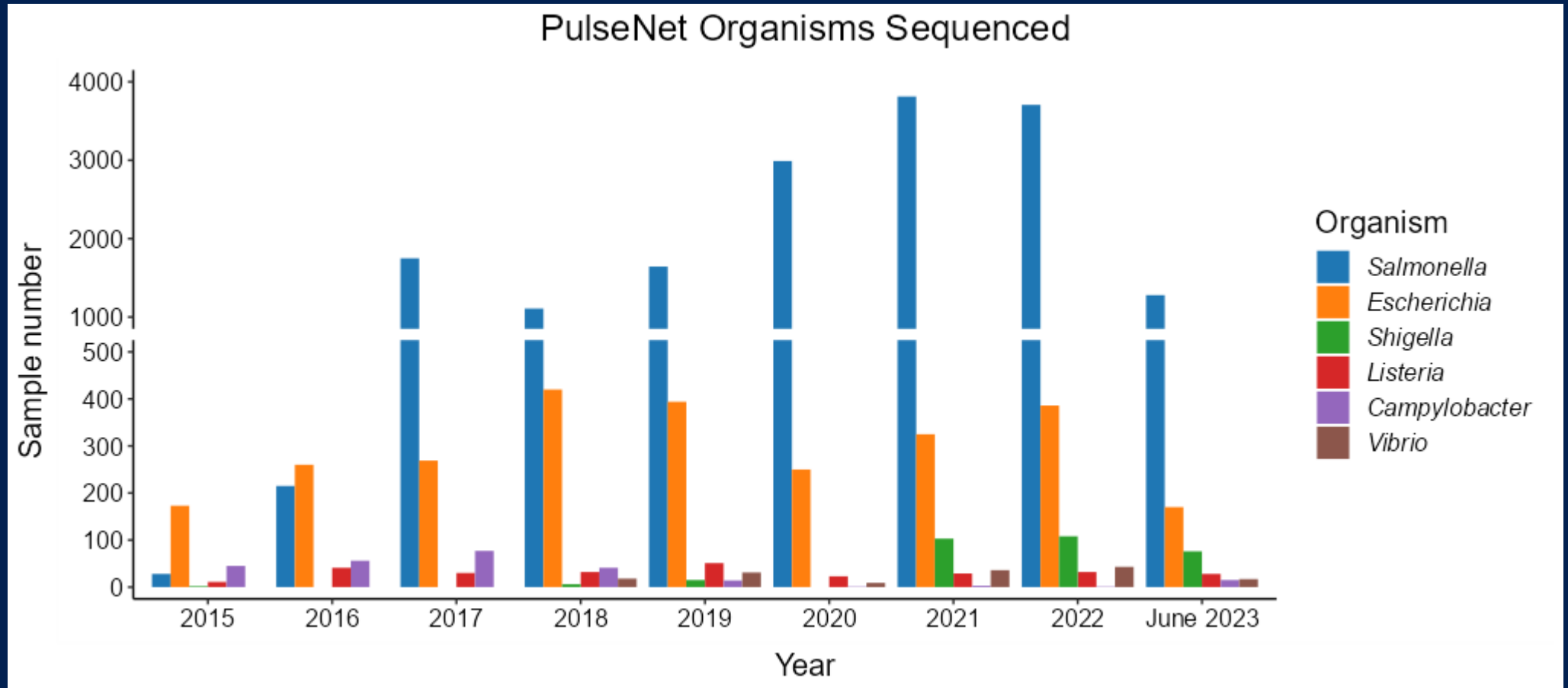
Texas Department of State Health Services

2023: Sequencing Projects and Numbers



Enteric bacterial isolates sequenced at the Texas DSHS Laboratory

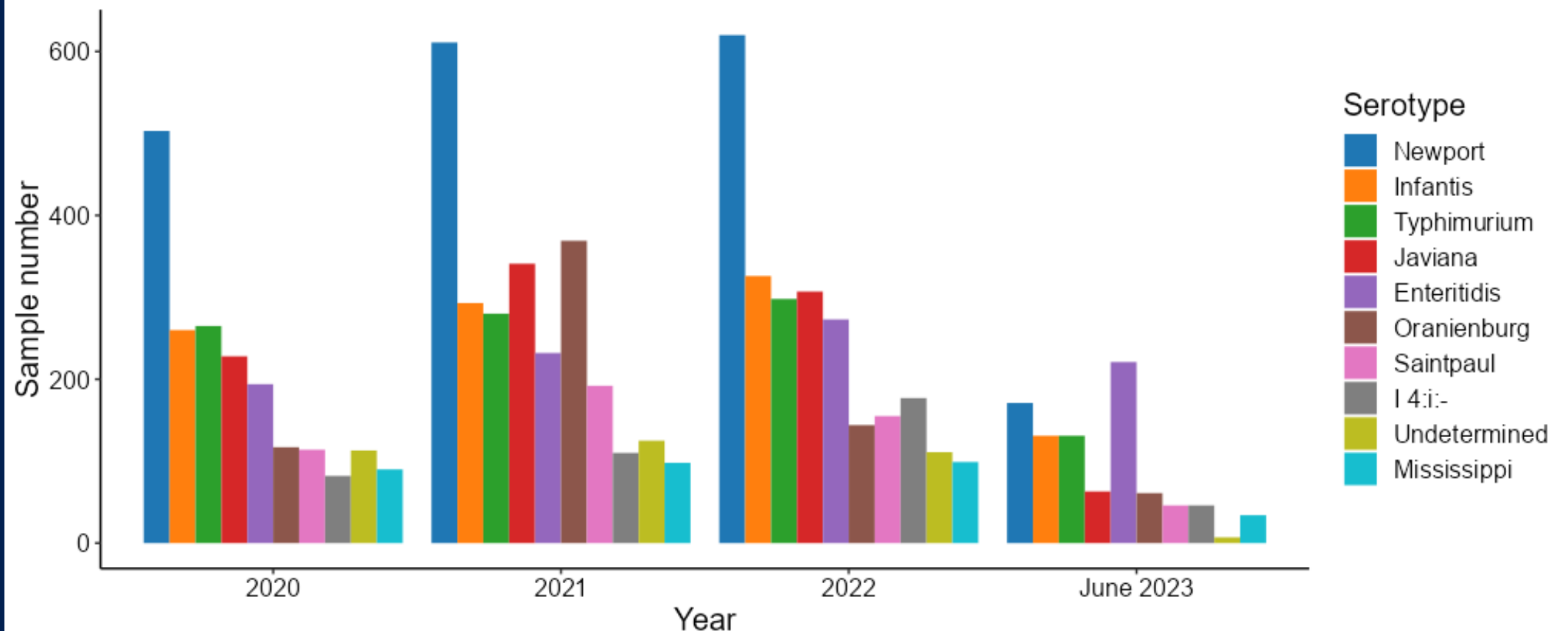
2015 to June 2023



Most common Salmonella enterica serotypes sequenced at the Texas DSHS Laboratory

2020 to June 2023

Top 10 Salmonella Serotypes Sequenced



TEXAS

Health and Human Services

Texas Department of State
Health Services

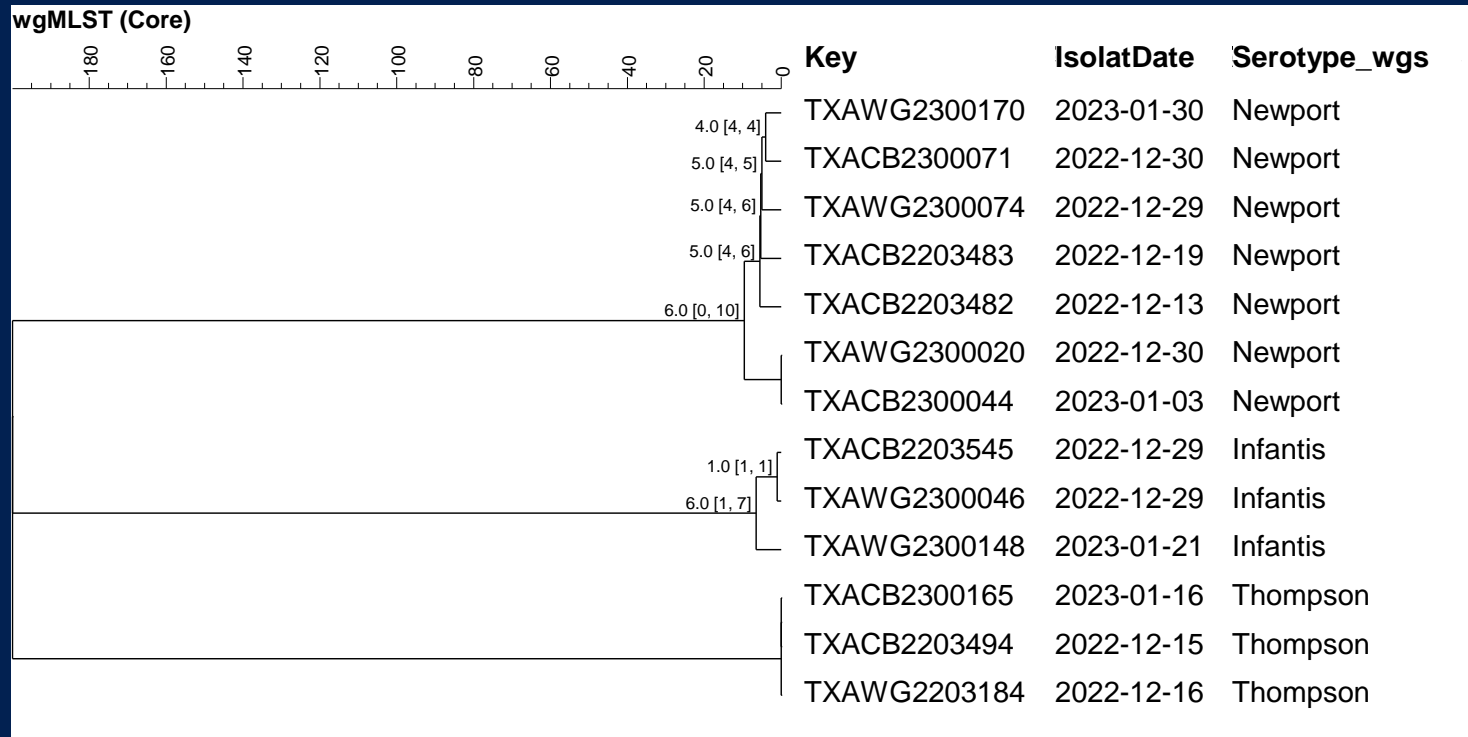
Weekly Cluster Report Example for Enteric Pathogens



TEXAS

Health and Human Services

Texas Department of State Health Services



Salmonella Dendrogram

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
		TXAWG2300170	TXACB2300071	TXAWG2300074	TXACB2203483	TXACB2203482	TXAWG2300020	TXACB2300044	TXACB2203545	TXAWG2300046	TXAWG2300148	TXACB2300165	TXACB2203494	TXAWG2203184	
1		0	4	5	5	5	9	9	2696	2692	2696	2619	2620	2618	
2		4	0	5	5	5	9	9	2698	2694	2698	2622	2623	2621	
3		5	5	0	6	6	10	10	2695	2691	2695	2619	2620	2618	
4		5	5	6	0	6	10	10	2695	2691	2695	2620	2621	2619	
5		5	5	6	6	0	10	10	2700	2696	2700	2621	2622	2620	
6		9	9	10	10	10	0	0	2697	2693	2697	2623	2624	2622	
7		9	9	10	10	10	0	0	2695	2691	2695	2620	2621	2619	
8		2696	2698	2695	2695	2700	2697	2695	0	1	7	2701	2702	2700	
9		2692	2694	2691	2691	2696	2693	2691	1	0	6	2697	2698	2696	
10		2696	2698	2695	2695	2700	2697	2695	7	6	0	2702	2703	2701	
11		2619	2622	2619	2620	2621	2623	2620	2701	2697	2702	0	0	0	
12		2620	2623	2620	2621	2622	2624	2621	2702	2698	2703	0	0	0	
13		2618	2621	2618	2619	2620	2622	2619	2700	2696	2701	0	0	0	

Heatmap

PulseNet Participating Labs Share Outbreak Information



TEXAS

Health and Human Services

Texas Department of State Health Services

Prevention

Symptoms

Diagnosis & Treatment +

Current Outbreaks -

Reporting Timeline

Salmonella Outbreak Linked to Charcuterie Meats -

Where Sick People Lived

When People Got Sick

Investigation Details

Salmonella Outbreak Linked to Cantaloupes +

Salmonella Outbreak Linked to Dry Dog Food +

Salmonella Outbreak Linked to Fresh Diced Onions +

Salmonella Outbreaks Linked to Small Turtles +

Salmonella Outbreak Linked to Ground Beef +

Salmonella Outbreak Linked to Raw Cookie Dough +

Salmonella Outbreak Linked to Flour +

Posted January 18, 2024

This map shows where the 47 people in this *Salmonella* outbreak lived.

This outbreak may not be limited to the states with known illnesses, and the true number of sick people is likely much higher than the number reported. This is because many people recover without medical care and are not tested for *Salmonella*.

Number of Sick People

- 1
- 2 to 3
- 4 to 5
- 6 to 11

Data Table +

[Download Data](#) [XLS - 21 KB]

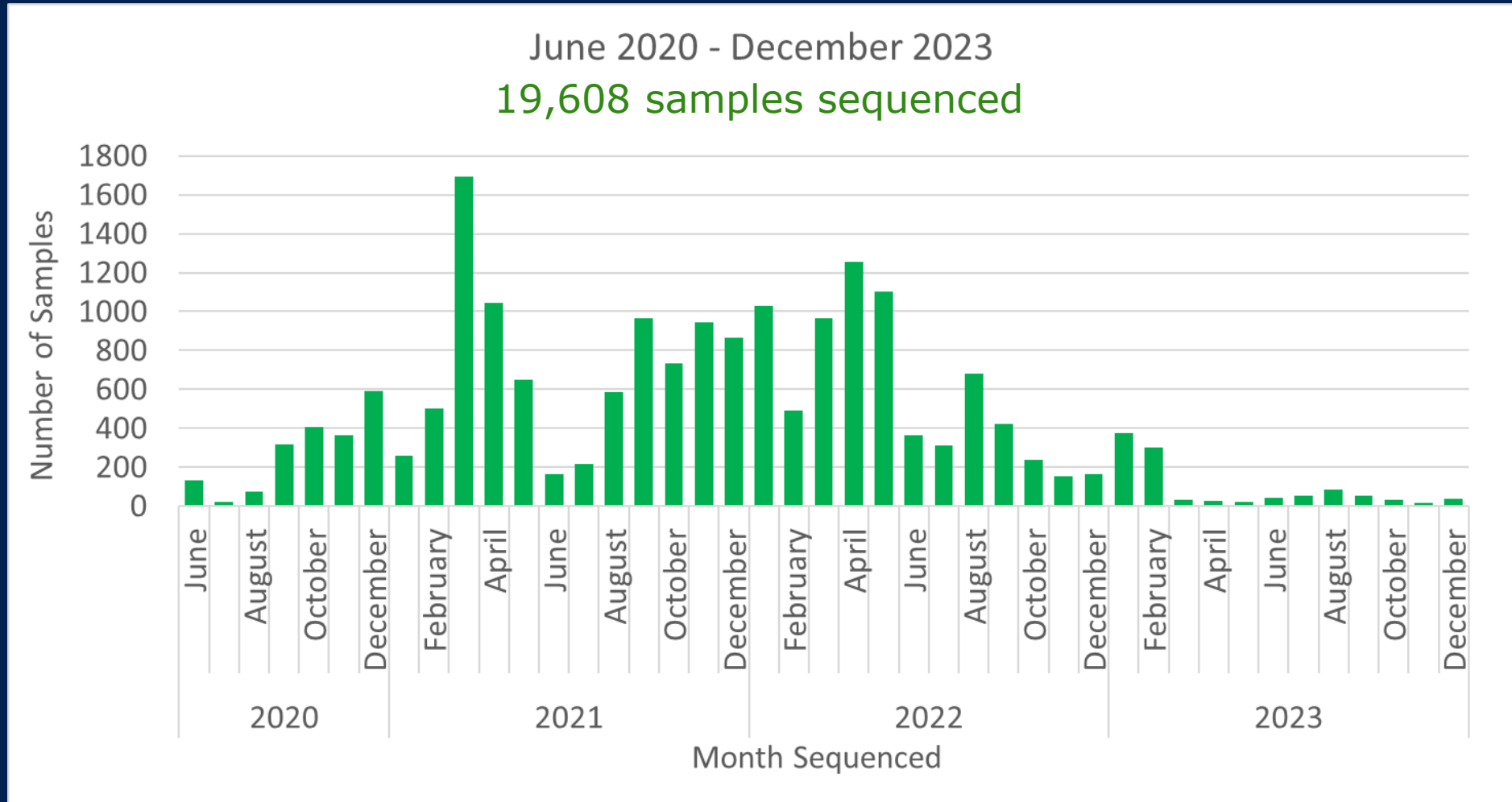
SARS-CoV-2 Samples Sequenced Overtime



TEXAS

Health and Human Services

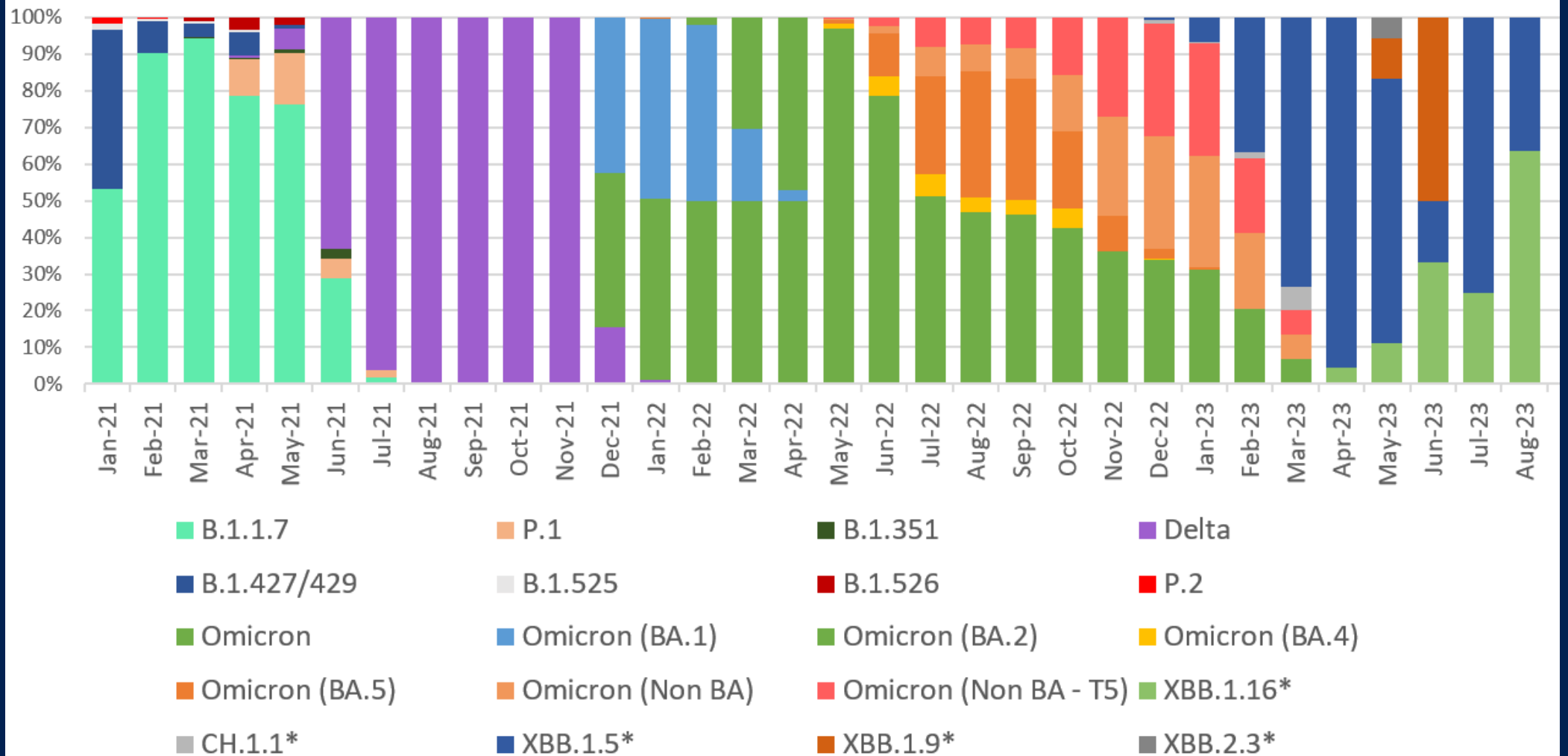
Texas Department of State Health Services



Tracking SARS-CoV-2 Lineages

January 2021-August 2023

GISAID VOC and VOI Lineage Proportions by Collection Date



TEXAS

Health and Human Services

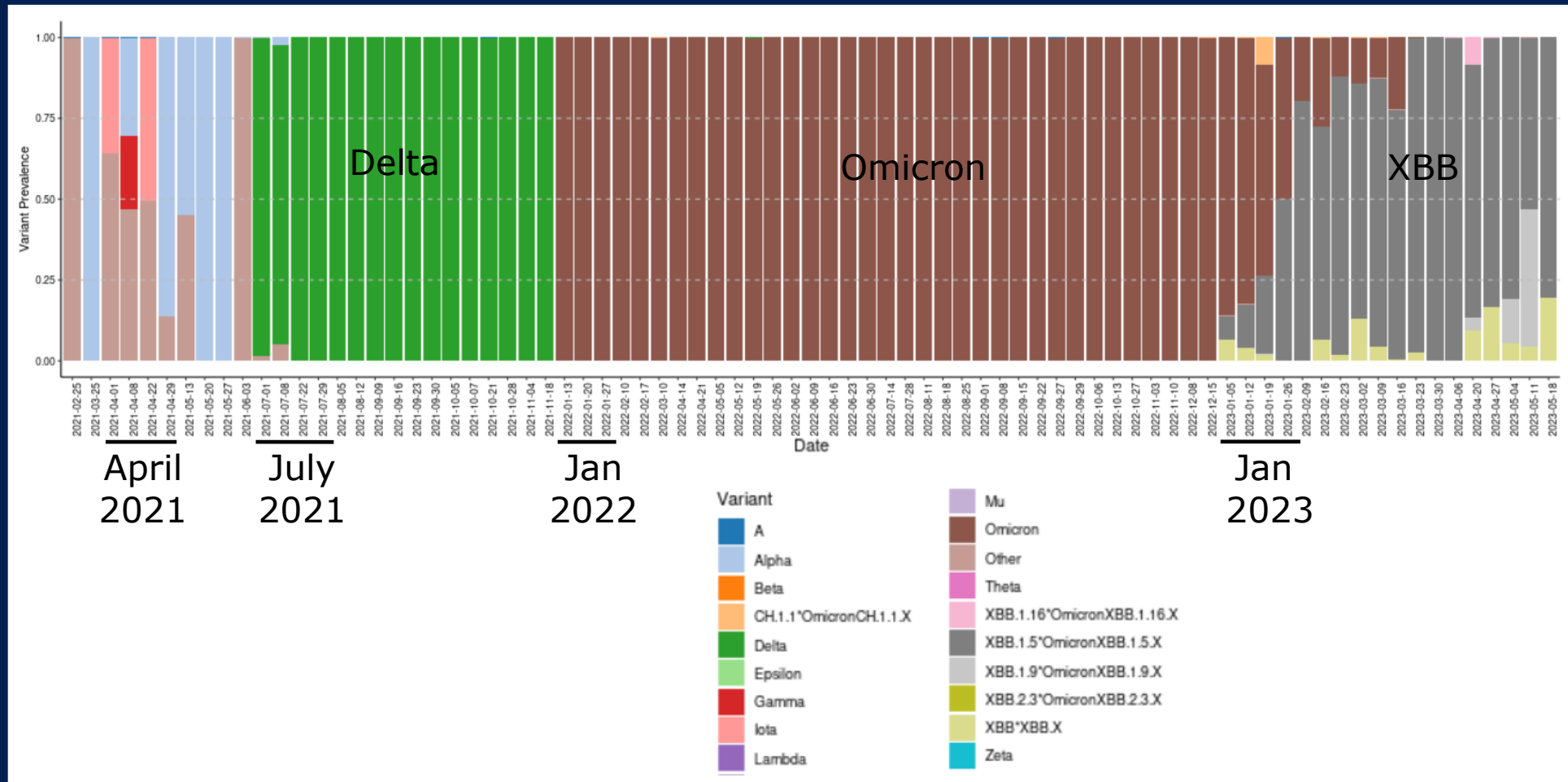
Texas Department of State Health Services

Denton Wastewater SARS-CoV-2 Surveillance

February 2021-May 2023
Variant Prevalence over time

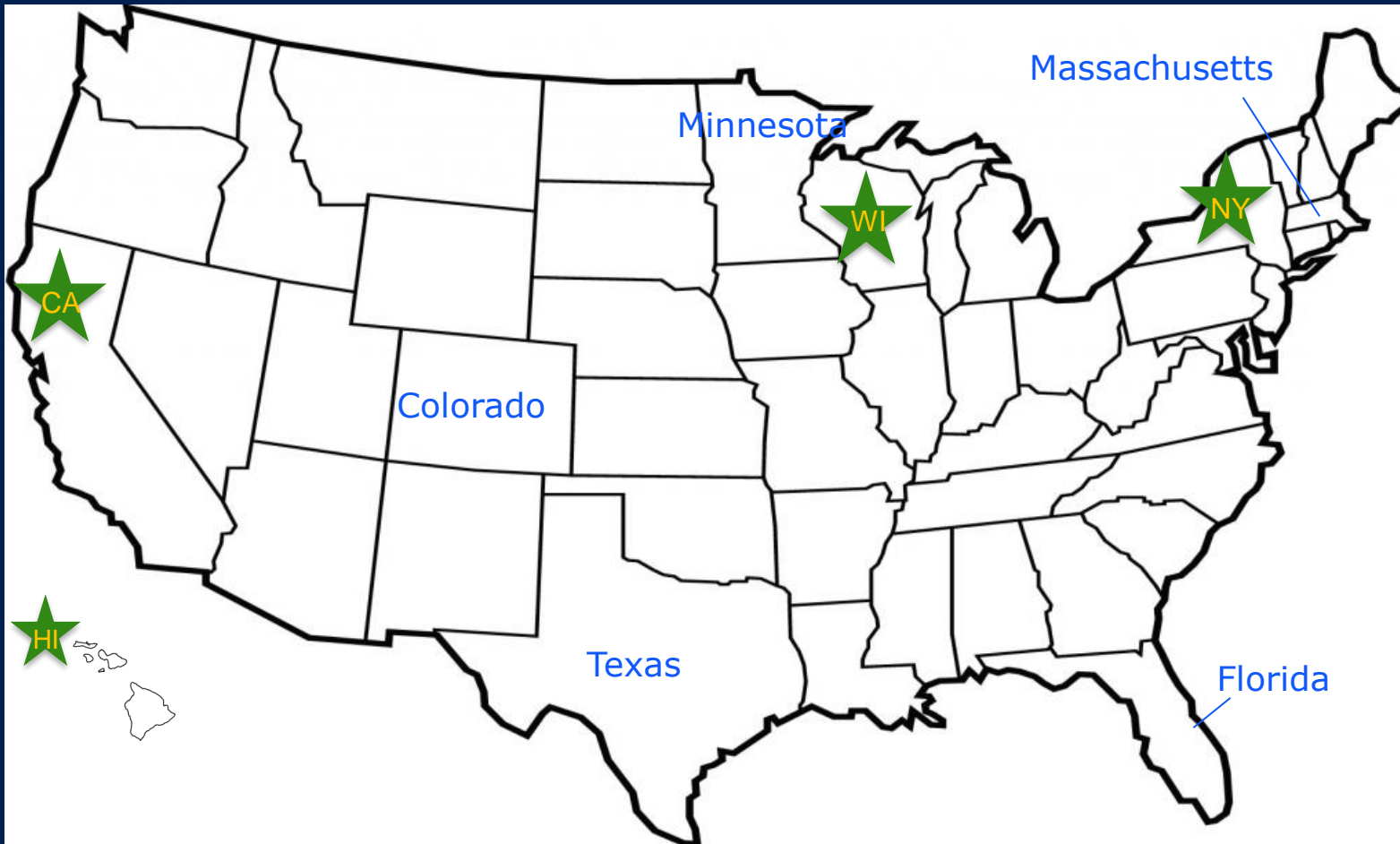


TEXAS
Health and Human Services
Texas Department of State
Health Services





Influenza Sequencing Center



- Real-time NGS data for Flu A/B
- Uses APHL AIMS (APHL Informatics Services) cloud-based environment for data transfer in near real-time and genome assembly analysis
- Project Goal: sequence 500 specimens per year



Influenza Sequencing Center

CDC Yearly Lab Work on Flu Viruses

More than 1 million patient specimens are tested in clinical labs participating in CDC domestic disease surveillance.*

About 100,000 specimens are tested in 93 state/local public health labs.

CDC conducts full genetic sequencing on about 7,000 flu viruses each year.

CDC tests about 2,000 flu viruses to determine their immune properties.

CDC prepares as many as 50 viruses for possible use in vaccine production.

*Influenza data current as of 2020-2021, as reported by CDC's Influenza Division, National Center for Immunization and Respiratory Diseases (NCIRD)

- Increased recruitment will improve surveillance
- New ISC locations will expand sequencing capabilities
- Improve vaccine development

<https://www.cdc.gov/flu/resource-center/freeresources/graphics/infographic-lab-work.htm>



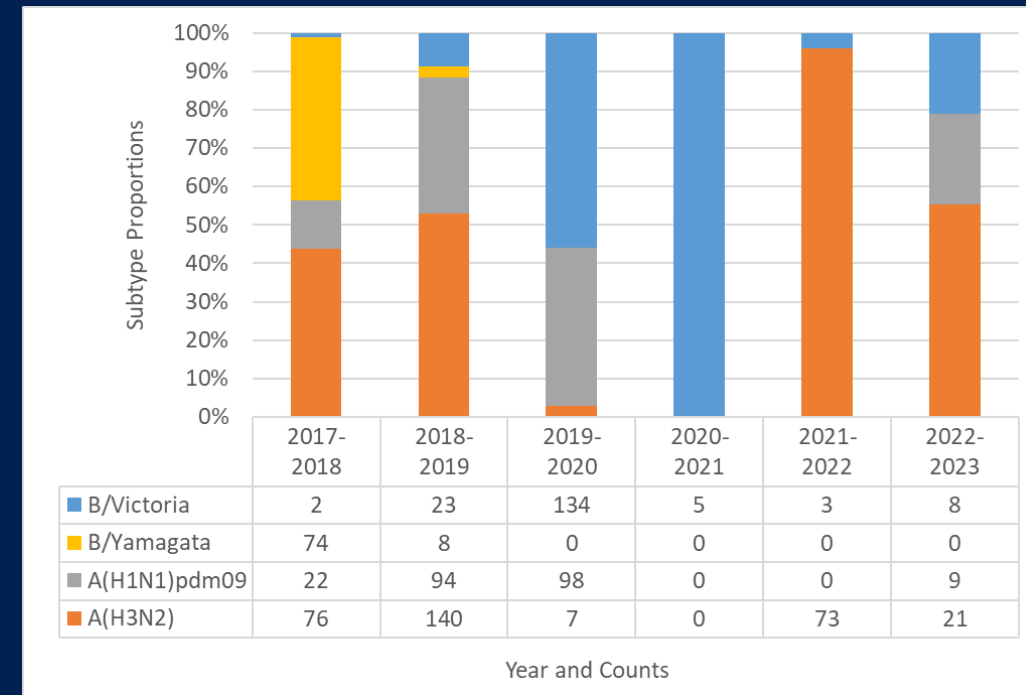
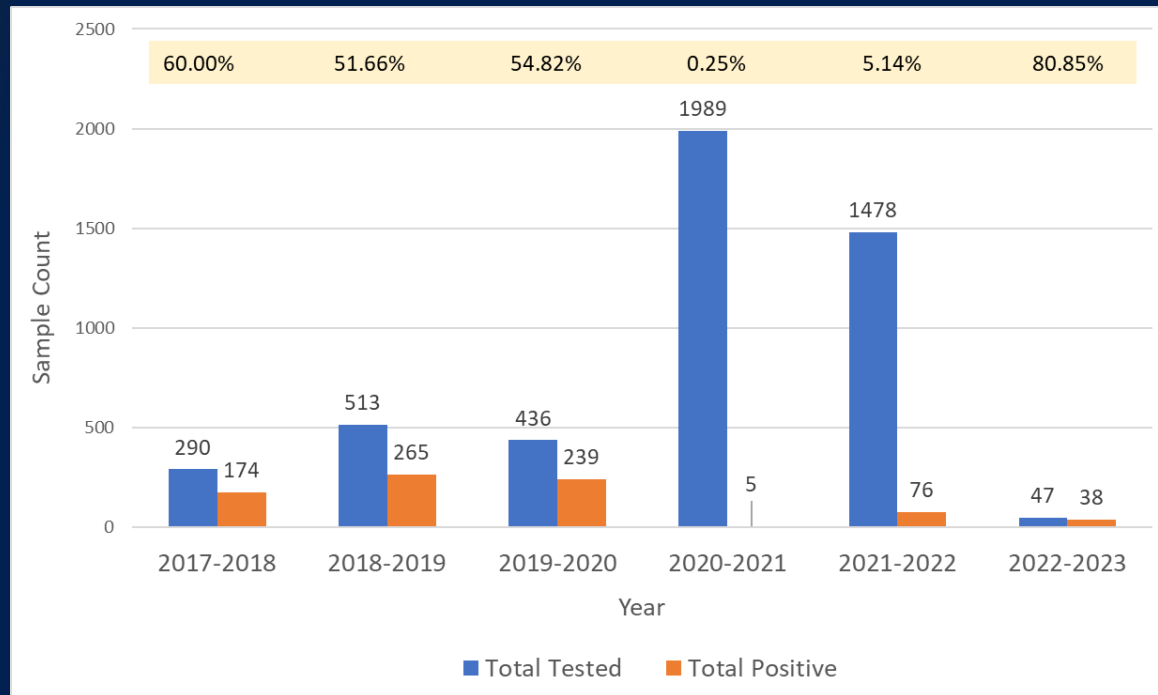
DSHS Influenza Positivity Rates and Breakdown



TEXAS

Health and Human Services

Texas Department of State Health Services



We need to increase recruitment to better represent Texas Flu data

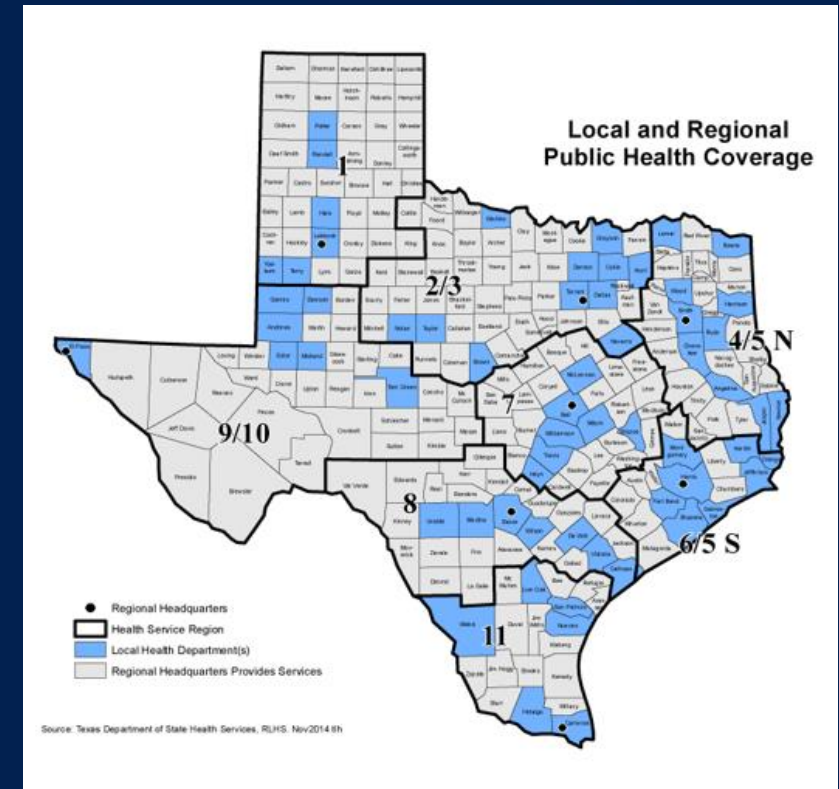
Working together...



TEXAS
Health and Human
Services

Texas Department of State
Health Services

- We are here to improve Texas public health surveillance
- More samples improves surveillance coverage
- Please help increase recruitment for:
 - Foodborne organisms like E. coli, Listeria, Campy
 - Cyclospora
 - SARS-CoV-2
 - Influenza samples
 - Mpox samples



Genetic Sequencing Branch



TEXAS
Health and Human
Services

Texas Department of State
Health Services





TEXAS
Health and Human
Services

**Texas Department of State
Health Services**

Thank you

Bonnie.Oh@dshs.texas.gov

Maliha.Rahman@dshs.texas.gov