

bettercallsal

Application of bettercallsal pipeline on environmental samples

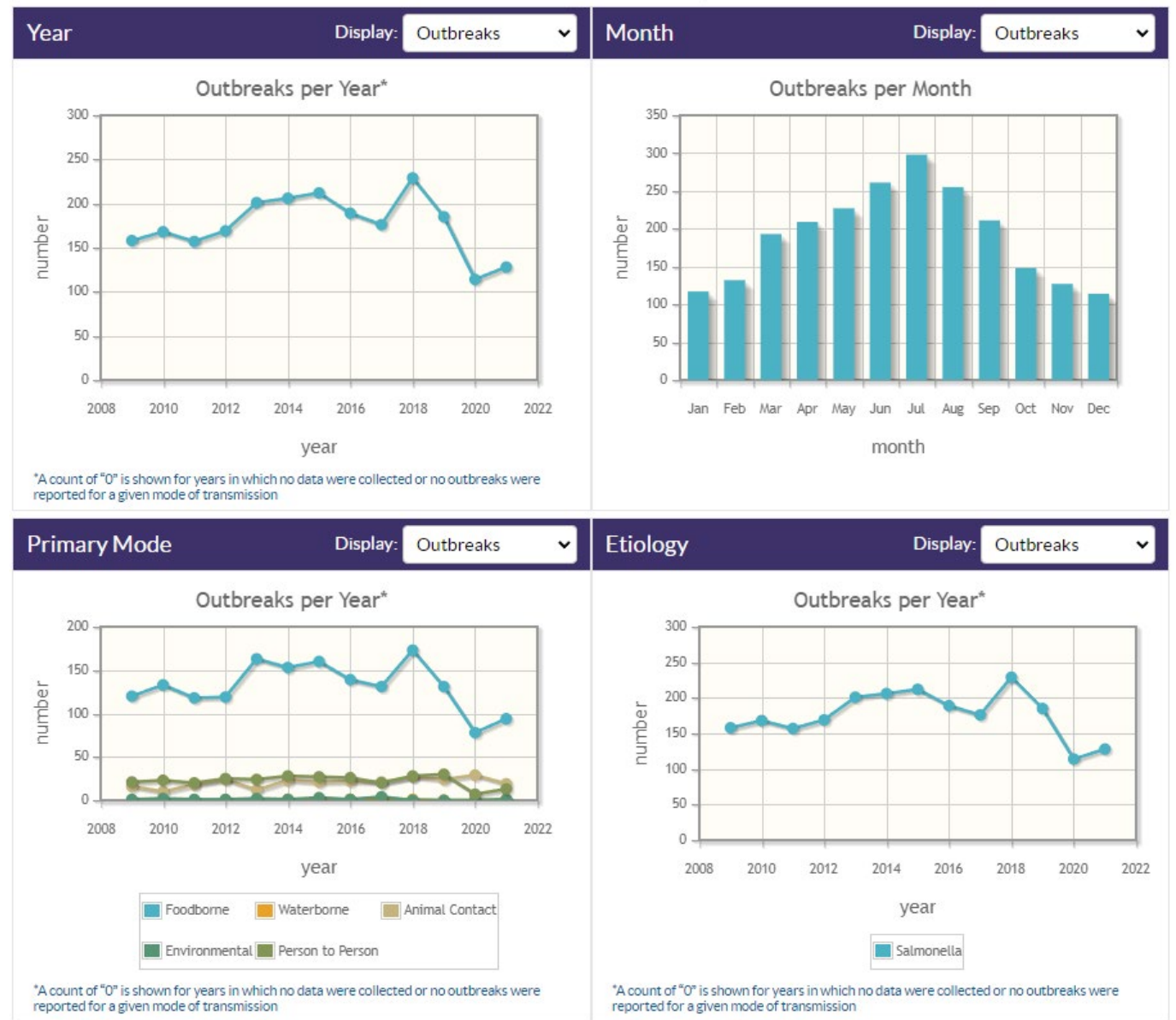
Padmini Ramachandran
FDA/CFSAN

Outbreaks- *Salmonella*

Salmonella outbreak- one among the leading cause.

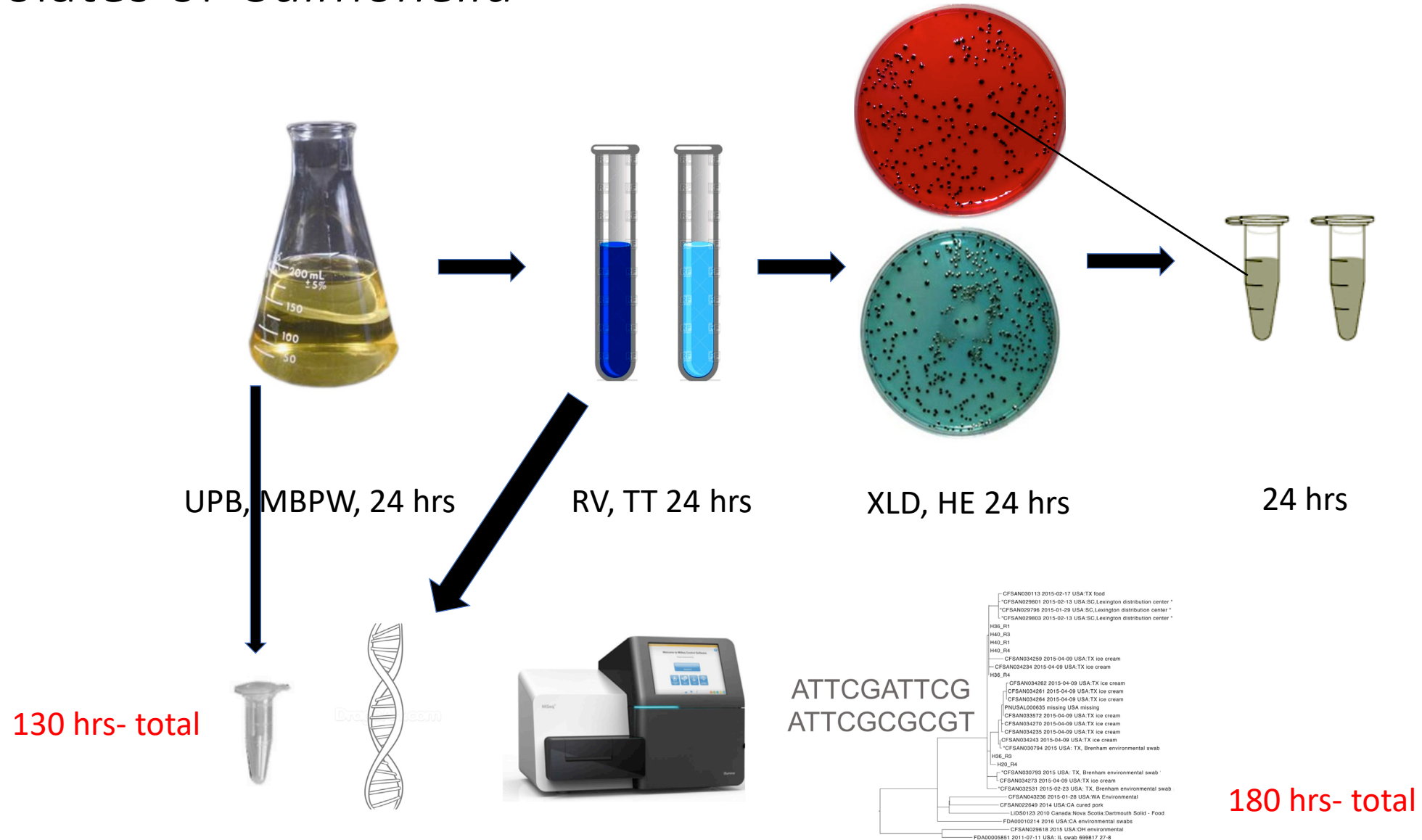
NORS – CDC – Mostly foodborne

~3% of *Salmonella* outbreaks are multi-serovar



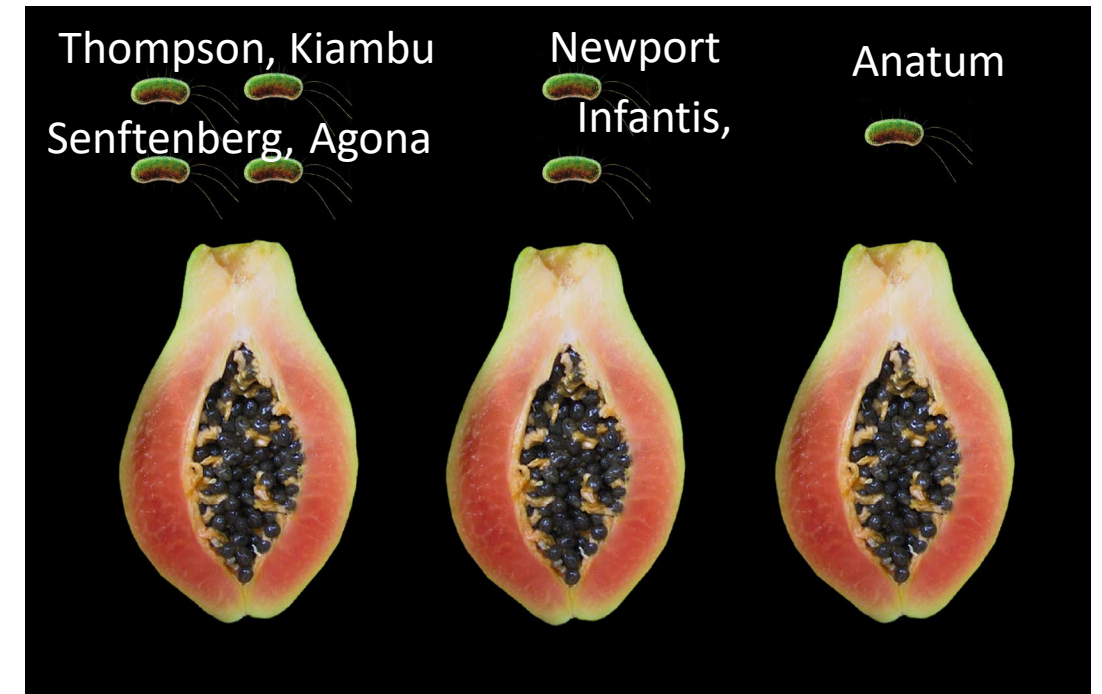
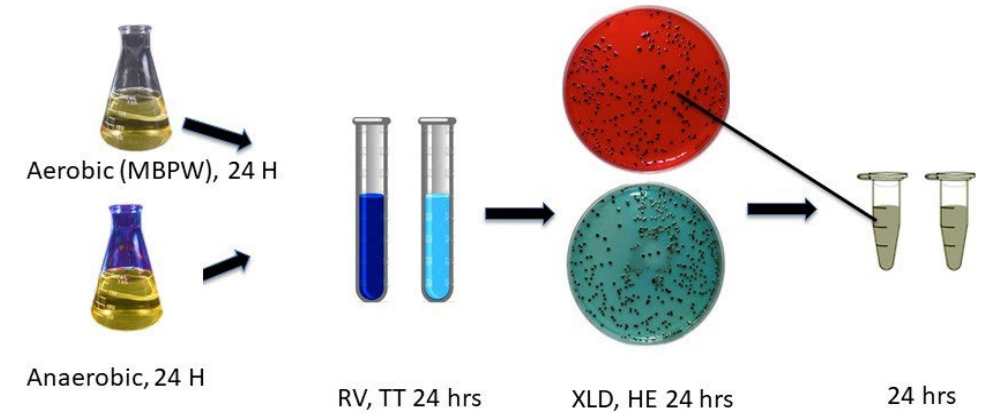
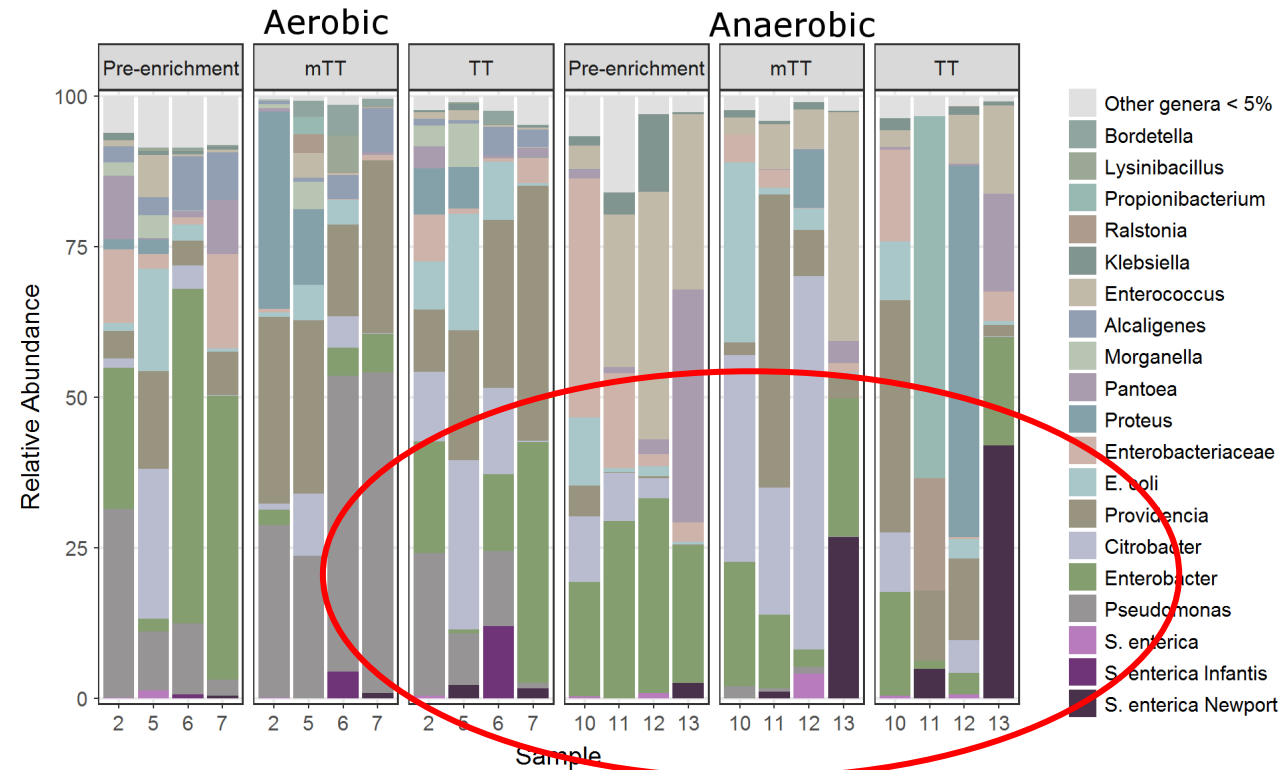
<https://wwwn.cdc.gov/norsdashboard/>

Quasi metagenomics- Microbiological steps to recover individual isolates of *Salmonella*

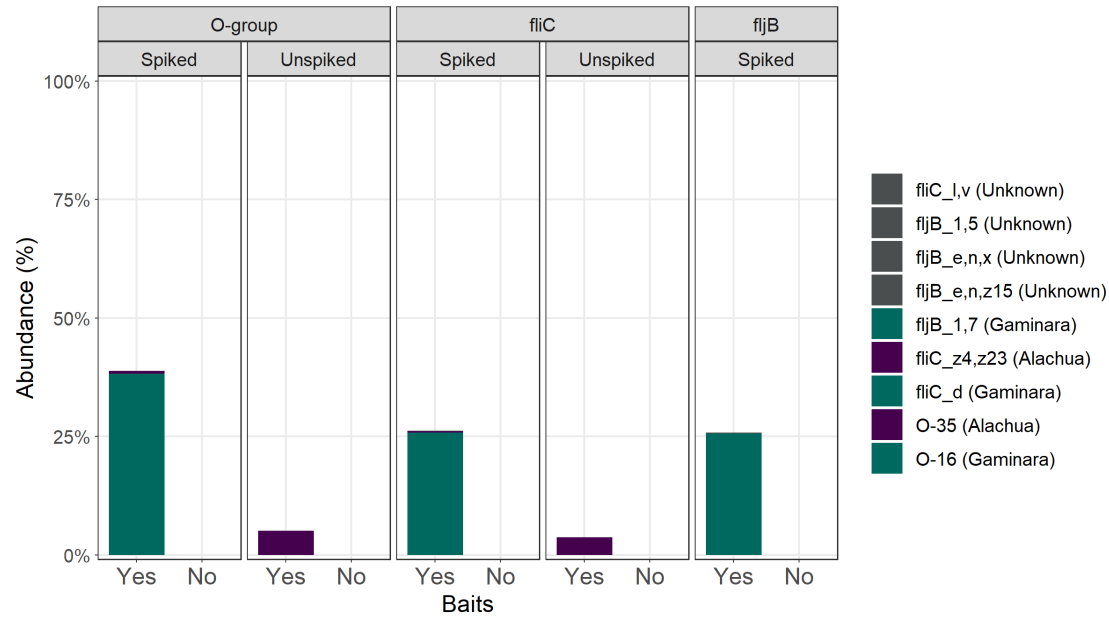


Challenges on metagenomics dataset on *Salmonella* serotyping

Papaya Outbreak- 2017

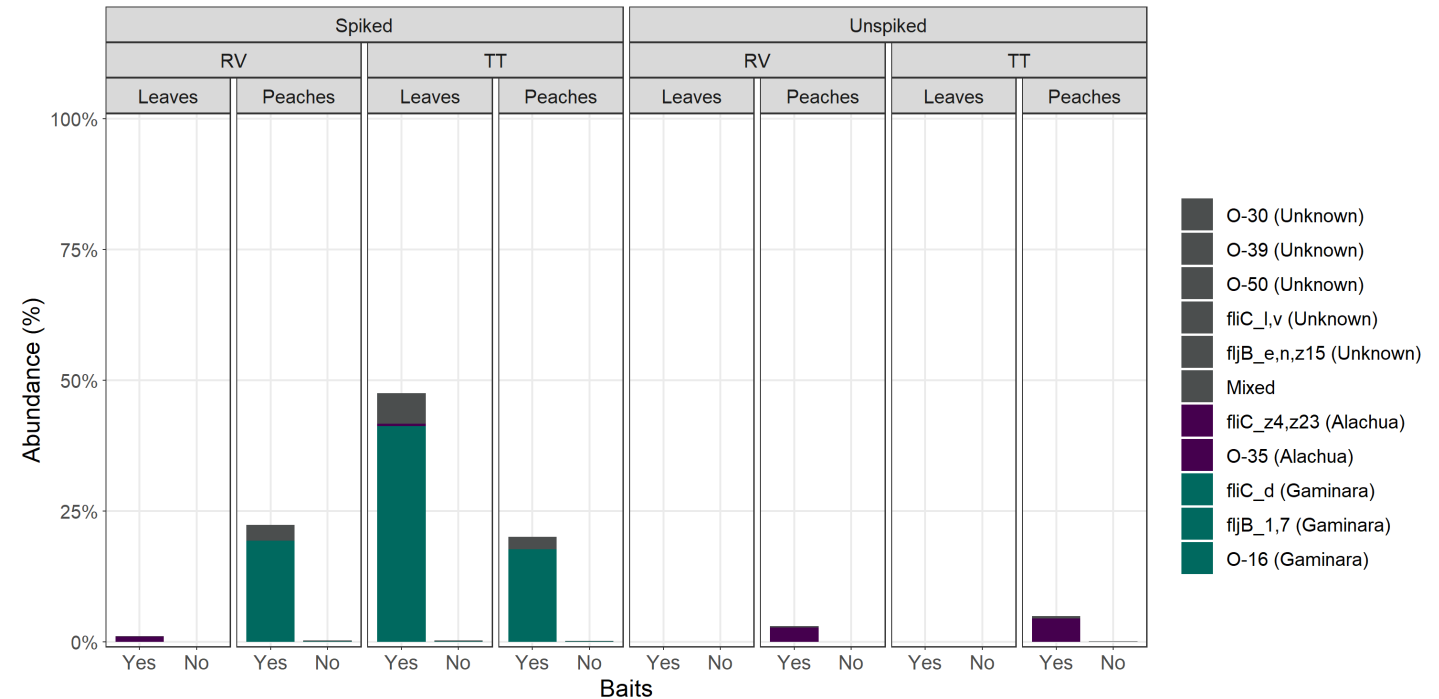


Kallisto abundance

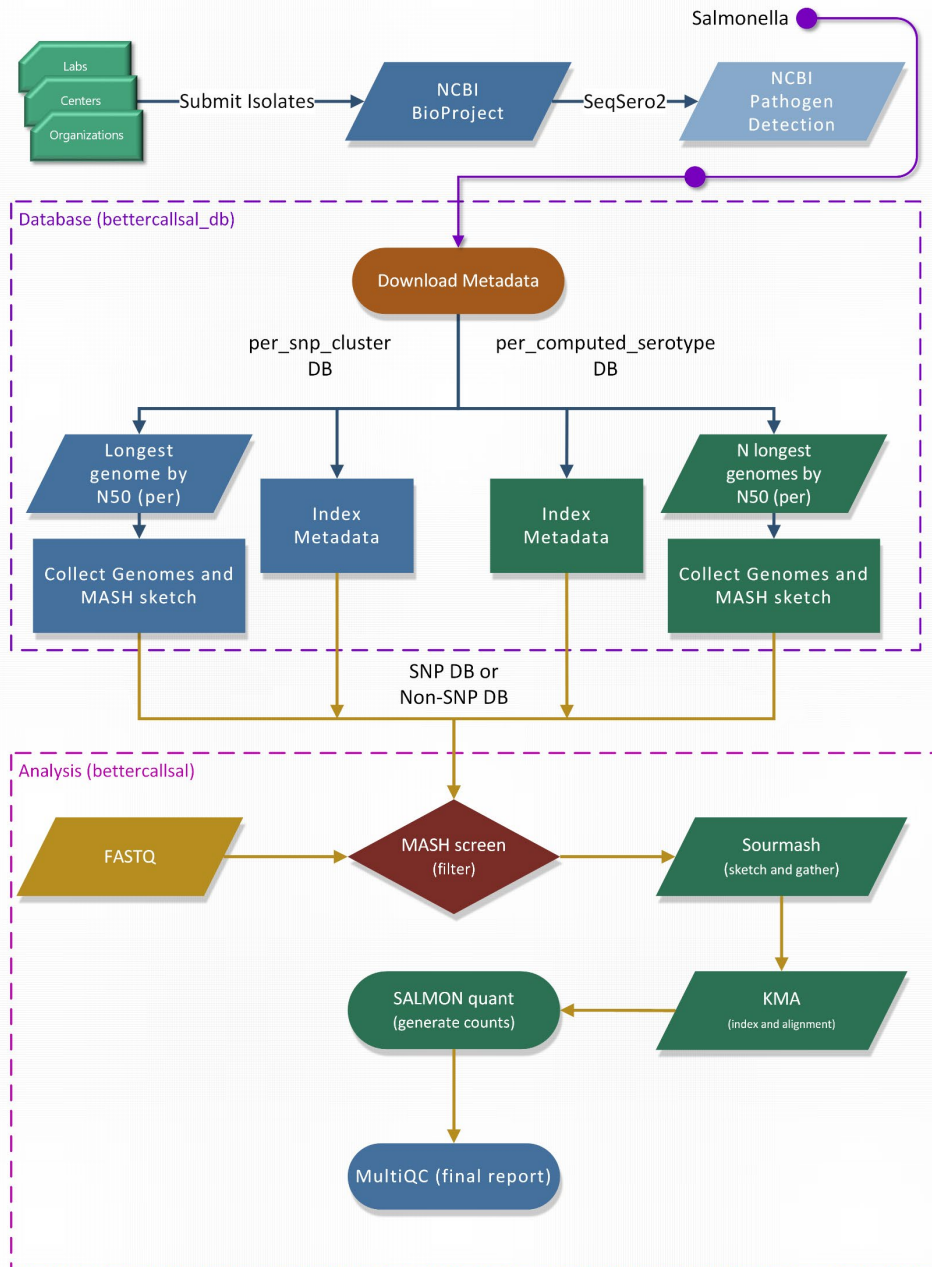


- Pulling *Salmonella* reads -
 - Assembly based approaches
 - SeqSero2 on the *Salmonella* only reads
 - SeqSero2 on the assembly
 - SISTR/SRST2
 - Many other tools

SERO-BLAST



bettercallsal



Nextflow – Domain Specific Language (BCS – HTML) Platform, portability layer between logic and execution, checkpoint tracking, execution resumed at any step.

MASH – Genome/metagenome distance estimation using MinHash, generates sketch to represent large datasets.

KMA – kmer alignment, ConClave score to address ties, reference assembly and statistics. Good for highly redundant databases.

SALMON – quantification of reads, performance superior to kallisto, G+C content and positional bias adjustment.

bettercallsal



Obtain precise *Salmonella* serotype/serovar information for each sample *insilico*.



Characterize presence of potential multiple *Salmonella* serotypes in a single sample.



Work well with both metagenomics and quasi-metagenomics (enriched) datasets. Very fast.



Leverage *Salmonella* WGS isolate data made available (FDA GenomeTrakr) at NCBI.



Automated workflows. v0.4.1:

<https://github.com/CFSAN-Biostatistics/bettercallsal>



Latest version is on Galaxytrkr- bcs 0.6.0.

<https://galaxytrkr.org/root/login?redirect=%2F>

BETTERCALLSAL

Copy Print Excel CSV

Sample	Agona 4:f,g,s:-	Gaminara 16:d:1,7	Newport 8:e,h:1,2	Senftenberg 1,3,19:g,s,t:-	Thompson 7:k:1,5
FarmA-Papaya2-RV	-	167404 (0.79%)	-	-	131418 (0.62%)
FarmA-Papaya2-TT	157566 (0.76%)	117713 (0.57%)	-	249136 (1.21%)	170751 (0.83%)
FarmC-Papaya13-mTT	-	-	3109383 (37.51%)	-	-

NCBI Pathogens Isolate Browser	
PDT001117224.1	PDT000027683.3
PDT001297912.1	PDT000234048.2
PDT000344927.1	PDT000520012.1
PDT000559817.1	

1

Next

The results table shown here is a collection from all samples.

Copy

Print

Excel

CSV

Search:

FILE	SCHEME	ST	aroC	dnaN	hemD	hisD	purE	sucA	thrA
Farma-Papaya2-RV	senterica_achtman_2	-	190?	1275?	3	1823?	~1082	1004?	~979
Farma-Papaya2-TT	senterica_achtman_2	-	7	13	8	1289?	1267	8	~149
FarmaC-Papaya13-mTT	senterica_achtman_2	2370	16	2	45	43	36	39	547

Showing 1 to 3 of 3 entries

[Previous](#)

1

Next

The results table shown here is a collection from all samples.

Copy

Print

Excel

CSV

Search:

[illegible]

Papaya Outbreak 2017

https://cfsan-pub-xfer.s3.amazonaws.com/Kranti.Konganti/bettercallsal/manuscript/papaya_outbreak_results.html

Filters

Page 1 of 6 Records per Page 5 Choose columns Download Hide plus AMR genotypes Expand all Displaying 1 - 5 of 26

#	Organism group	Strain	Isolate identifiers	Serovar	Isolate	Create date	Location	Isolation source	Isolation ...	Min-same	Min-diff	BioSample	Assembly	Computed types
1	Salmonella ente...	MD-DOH-1...	CFSAN067219 MD-DOH-18-00004 SRS2426088		PDT000234048.2	2017-08-18	USA:MD	papaya	environm...	0	0	SAMN07456254	GCA_005758125.1	Serotype: Senftenberg Antigen formula: 1,3,19:g,s,t:-
2	Salmonella ente...	FDA101965...	CFSAN066286 FDA00012036 FDA1019658-S006- SRS2424094		PDT000233861.2	2017-08-16	Mexico	papaya	environm...	1	1	SAMN07429348	GCA_006136055.1	Serotype: Senftenberg Antigen formula: 1,3,19:g,s,t:-
3	Salmonella ente...	FDA101965...	CFSAN066295 FDA00012045 FDA1019657-S003- SRS2424096		PDT000233863.2	2017-08-16	Mexico	papaya	environm...	1	1	SAMN07429357	GCA_007045205.1	Serotype: Senftenberg Antigen formula: 1,3,19:g,s,t:-
4	Salmonella ente...	FDA101965...	CFSAN066288 FDA00012038 FDA1019658-S008- SRS2424098		PDT000233866.2	2017-08-16	Mexico	papaya	environm...	0	0	SAMN07429346	GCA_006135955.1	Serotype: Senftenberg Antigen formula: 1,3,19:g,s,t:-
5	Salmonella ente...	FDA101965...	CFSAN066291 FDA00012041 FDA1019658-S012- SRS2424102		PDT000233869.2	2017-08-16	Mexico	papaya	environm...	0	0	SAMN07429361	GCA_006136075.1	Serotype: Senftenberg Antigen formula: 1,3,19:g,s,t:-

Share Hide Table Labels Load Labels Neighbors Common subtree Watch Watched isolates

environmental/other, 2017-08-25, Mexico, papaya, FDA1019661-S020-001, PDT000236323.2
 environmental/other, 2017-08-16, Mexico, papaya, FDA1019657-S008-001, PDT000233871.2
 environmental/other, 2018-07-12, Mexico, papaya, Sa U75, PDT000347469.1
 environmental/other, 2017-08-25, Mexico, papaya, FDA1019661-S004-001, PDT000236309.2
 environmental/other, 2017-08-25, Mexico, papaya, FDA1019661-S023-001, PDT000236321.2
 environmental/other, 2017-08-25, Mexico, papaya, FDA1019661-S007-001, PDT000236329.2
 environmental/other, 2019-05-31, USA DE, environmental, CFSAN074579, PDT000510570.1
 environmental/other, 2017-08-16, Mexico, papaya, FDA1019657-S003-001, PDT000233863.2
 environmental/other, 2017-08-25, Mexico, papaya, FDA1019661-S015-001, PDT000236333.2
 environmental/other, 2017-08-25, Mexico, papaya, FDA1019661-S022-001, PDT000236319.2
 clinical, 2017-09-22, USA, PNUSAS023300, PDT000242694.2
 environmental/other, 2017-08-25, Mexico, papaya, FDA1019661-S026-001, PDT000236320.2
 environmental/other, 2017-08-16, Mexico, papaya, FDA1019658-S006-001, PDT000233861.2
environmental/other, 2017-08-18, USA MD, papaya, MD-DOH-18-000049-8, PDT000234048.2
 environmental/other, 2017-08-16, Mexico, papaya, FDA1019658-S008-001, PDT000233866.2
 environmental/other, 2017-08-16, Mexico, papaya, FDA1019658-S012-001, PDT000233869.2
 environmental/other, 2017-08-25, Mexico, papaya, FDA1019661-S020-001, PDT000236323.2



U.S. FOOD & DRUG
ADMINISTRATION



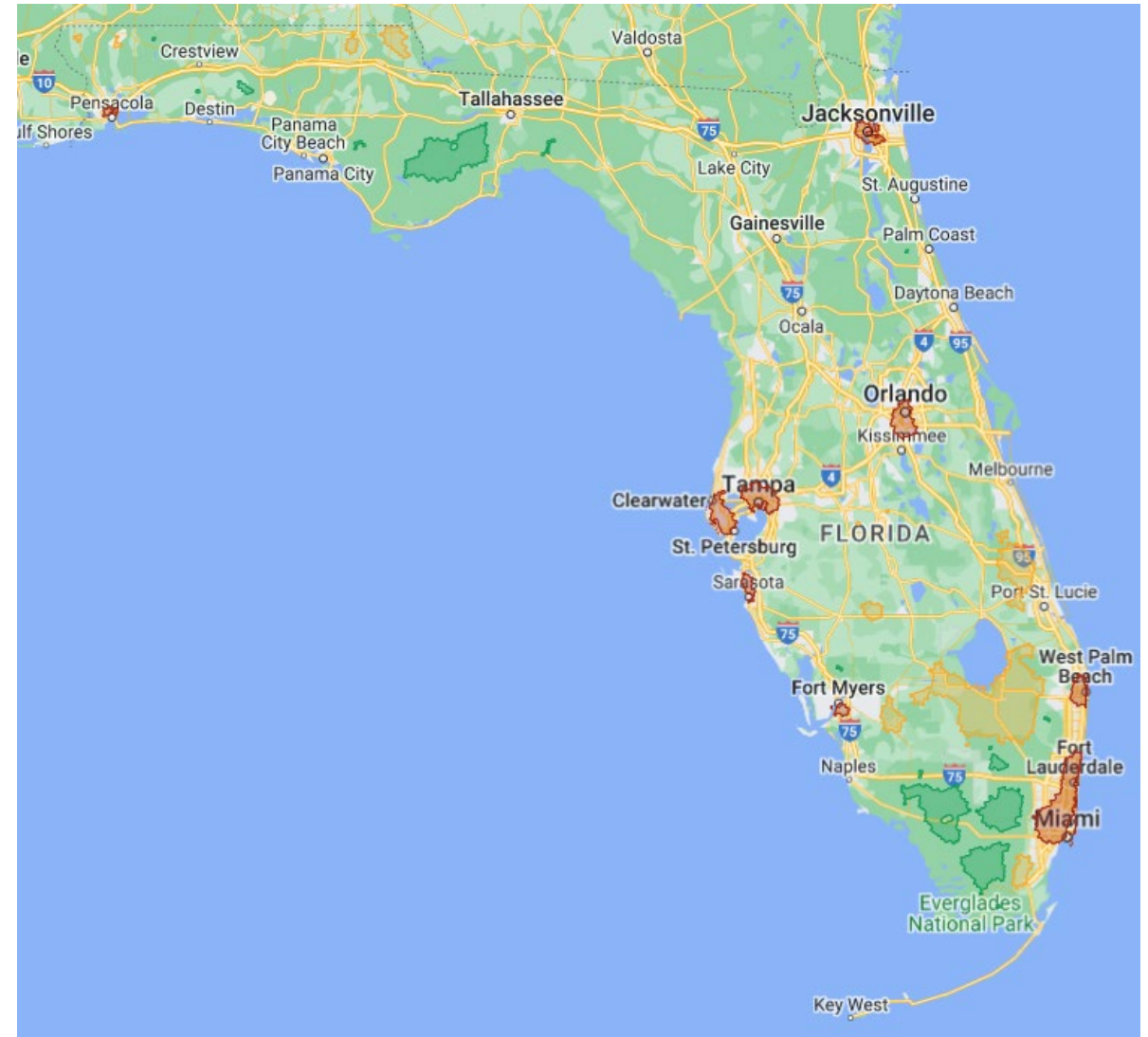
Environmental sampling- application of bettercallsal

Chief Scientist Grant
CVM

Brandon Kocurek, Andrea Ottesen, and Errol Strain

Florida

- A total of 9 sites were sampled across the State of Florida
- Sites were selected from Developed, Agriculture & Natural hotspots
 - 3 Developed
 - 3 Agriculture
 - 3 Natural
- 3 treatment groups were collected, prepared, sequenced & analyzed:
 - Culture Independent (CI)
 - General Enrichment (2x BPW)
 - *Salmonella* selective enrichment (RV/TT)

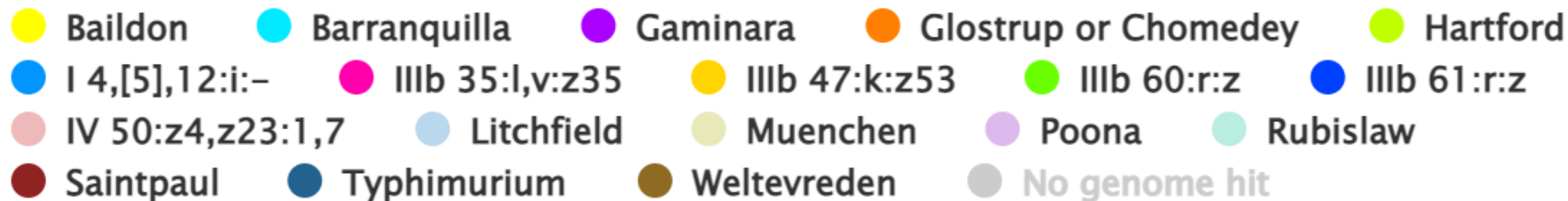
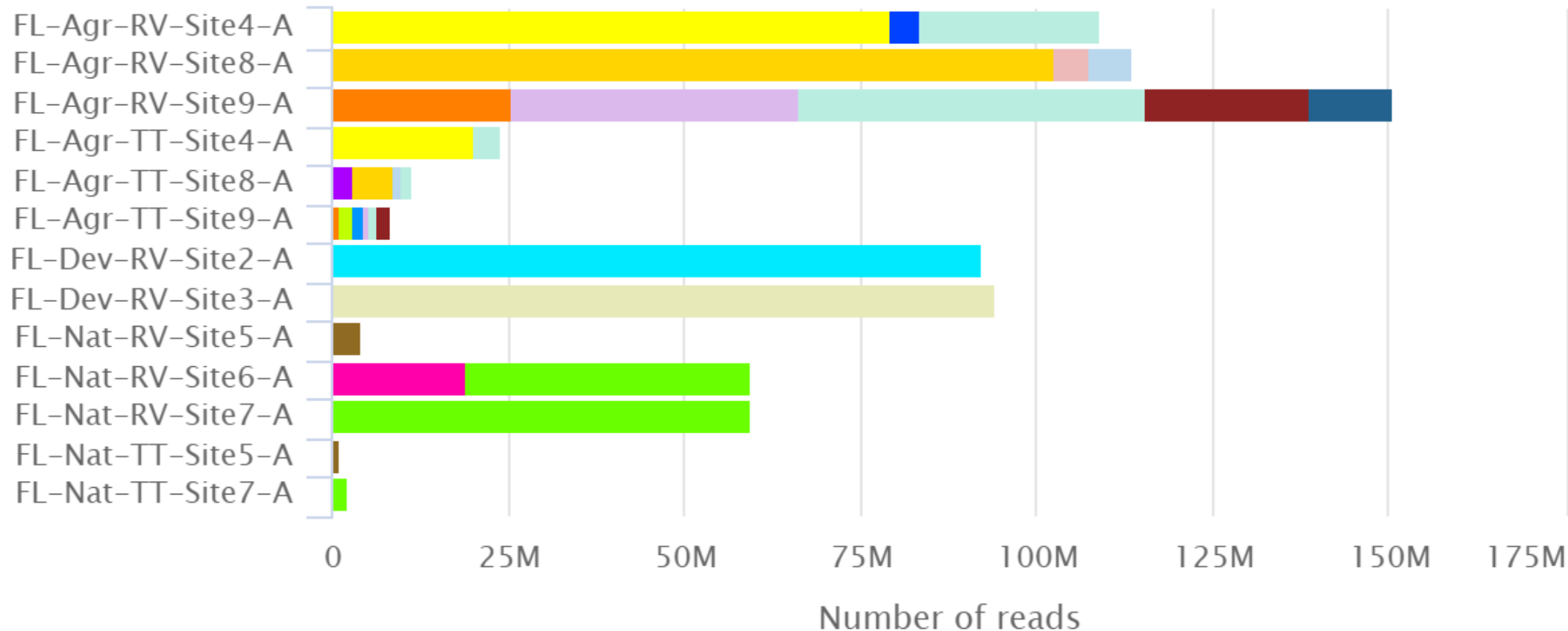


Developed

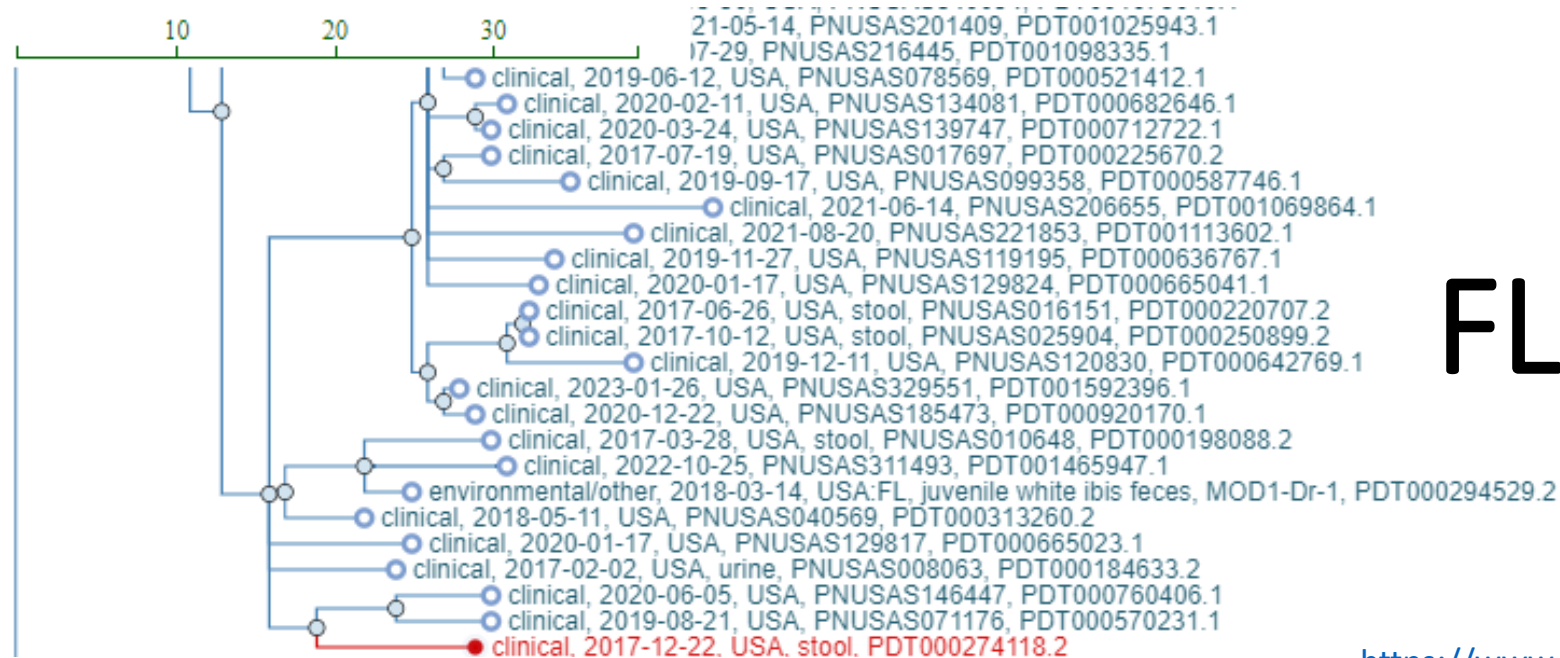
Agriculture

Natural

Salmon: Read counts



Organism group	Strain	Isolate identifiers	Serovar	Isolate	Create date	Locat...	Isolation source	Isolation ...	Min-same	Min-diff	BioSample	Assembly	Computed types
Salmonella ente...		PNUSAS027067 SRS2783364		PDT000274118.2	2017-12-22	USA	stool	clinical	19	22	SAMN07830261	GCA_010133595.1	Serotype: Baildon Antigen formula: 9,46:a:e,n,x
Salmonella ente...	FL_FLDACS...	FL_FLDACS-0926 SRS470635		PDT000000428.2	2013-08-28	USA:FL	alligator meat	environm...	10	5	SAMN02253011		Serotype: Baildon Antigen formula: 9,46:a:e,n,x
Salmonella ente...	NY_IDR110...	MDP - 10-00066 NY_IDR1100001 SRS528562	Baildon	PDT000002519.5	2018-06-26	USA:...	food	environm...	13	12	SAMN01902381	GCA_007773065.1	Serotype: Baildon Antigen formula: 9,46:a:e,n,x
Salmonella ente...	FL-MDD-75	FL-MDD-75 SRS580059	Baildon	PDT000024367.2	2014-03-26	USA:FL	surface water	environm...	33	22	SAMN02664723	GCA_010656465.1	Serotype: Baildon Antigen formula: 9,46:a:e,n,x
Salmonella ente...	FLDOH-38	CFSAN018028 FLDOH-38 SRS649114		PDT000034001.2	2014-06-27	USA:FL		clinical	22	23	SAMN02802284	GCA_011444175.1	Serotype: Baildon Antigen formula: 9,46:a:e,n,x



FL-Agr-RV-Site4-A

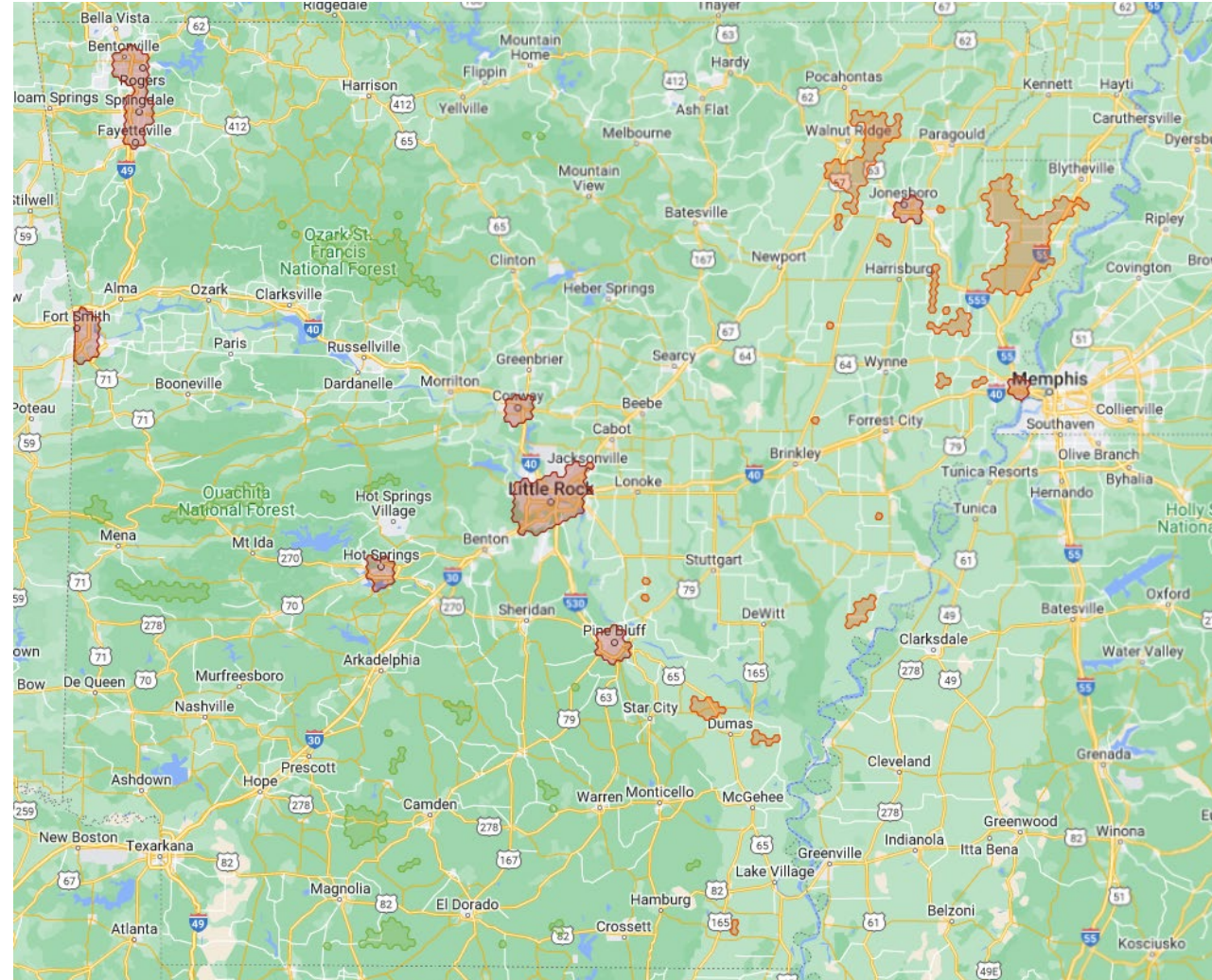
<https://www.ncbi.nlm.nih.gov/pathogens/isolates/#PDT000274118.2>

Comparing bettercallsal to WGS

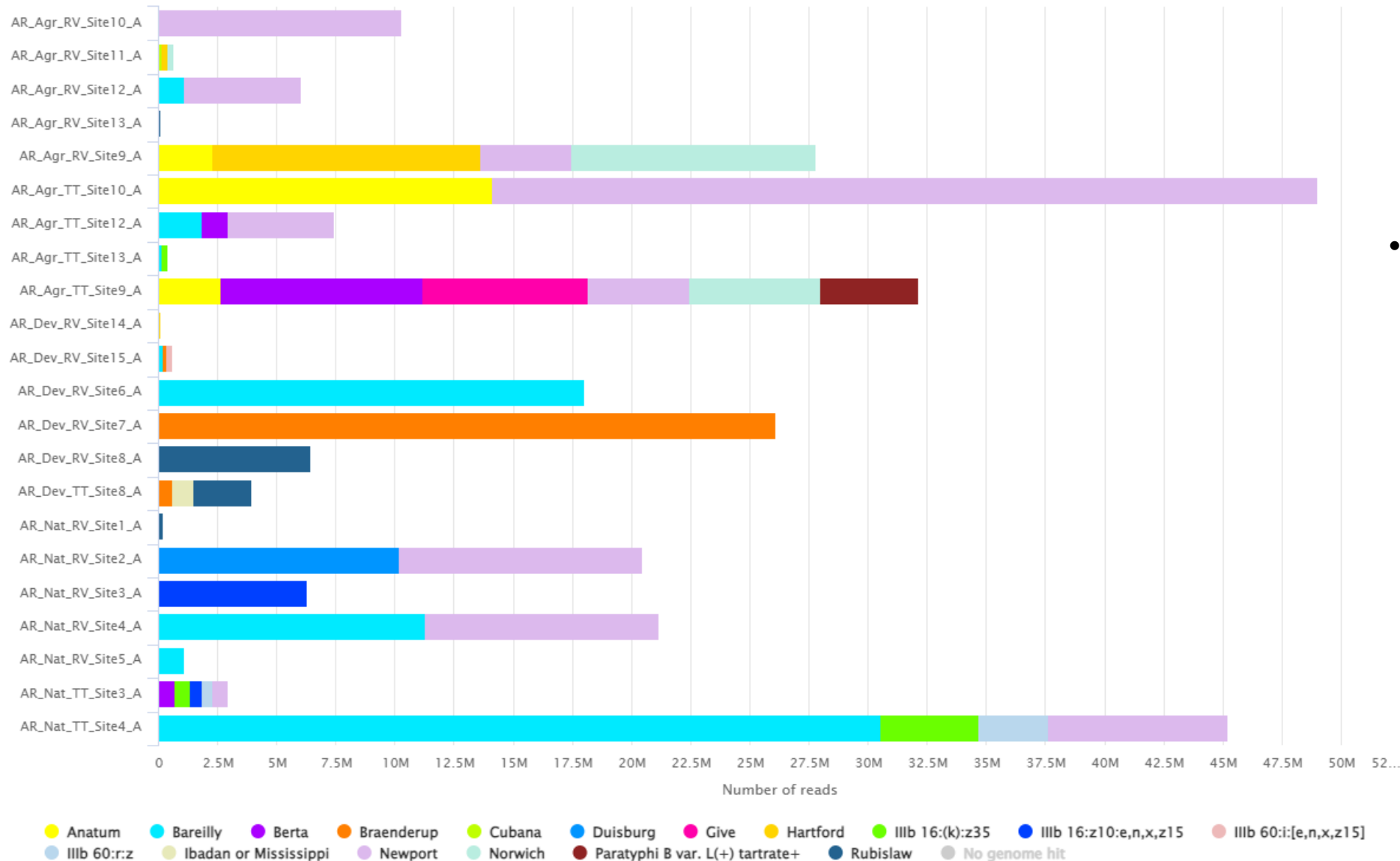
Group	Site	RV/TT	Bettercallsal (RV/TT QMGS)	Isolates
Developed	1	RV	NA	NA
		TT	NA	NA
	2	RV	Barranquilla*	Gaminara
		TT	Gaminara	Gaminara
	3	RV	Muenchen	Muenchen
		TT	Muenchen	Muenchen
Natural	5	RV	Weltevreden	NA
		TT	Weltevreden	Na
	6	RV	IIIb 35:l,v;z35 IIIb 60:r:z	IIIb 60:r:z
		TT	NA	NA
	7	RV	IIIb 60:r:z	IIIb 60:r:z
		TT	IIIb 60:r:z	IIIb 60:r:z
Agriculture	4	RV	Baildon IIIb 61:r:z Rubislaw	Baildon Rubislaw
		TT	Baildon Rubislaw	Baildon
	8	RV	IIIb 47:k:z53 IV 50:z4,z23:1,7 Litchfield	IIIb 47:k:z53
		TT	Gaminara IIIb 47:k:z53 Litchfield Litchfield	Gaminara IIIb 47:k:z53 Rubislaw
	9	RV	Glostrup or Chomedey Poona Rubislaw Saintpaul Typhimurium	Rubislaw Abaetetuba IIIb 47:k:z53
		TT	Glostrup or Chomedey Hartford I 4,[5],12:i:- Poona Rubislaw Saintpaul	Rubislaw Saintpaul

Arkansas

- A total of 15 sites were sampled across the State of Arkansas
- Sites were selected from Developed, Agriculture & Natural hotspots
 - 5 **Developed**
 - 5 **Agriculture**
 - 5 **Natural**
- 3 treatment groups were collected, prepared, sequenced & analyzed:
 - Culture Independent (CI)
 - General Enrichment (2x BPW)
 - Salmonella selective enrichment (RV/TT)



Salmonella serovars (Bettercallsal)



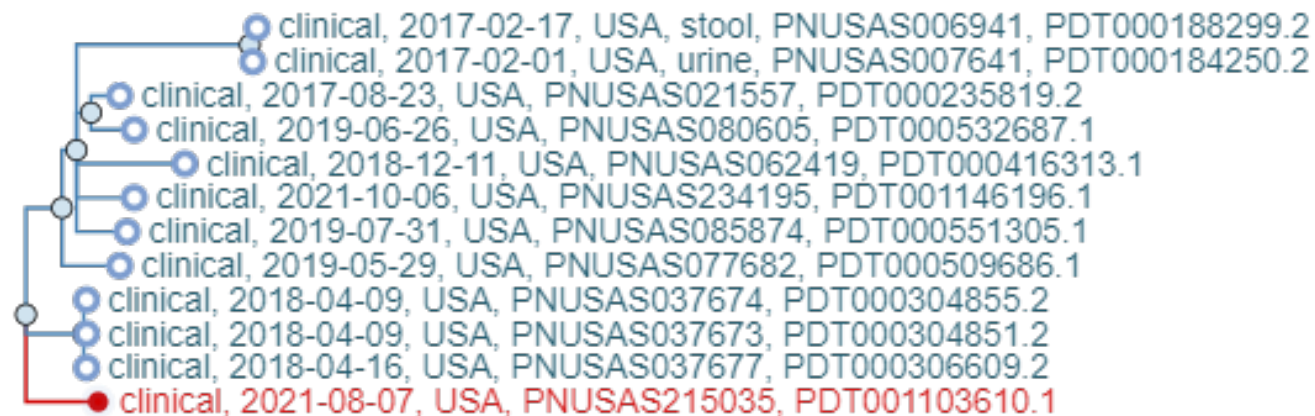
Culture Positive

- Culture positive Salmonella found in the following sites:

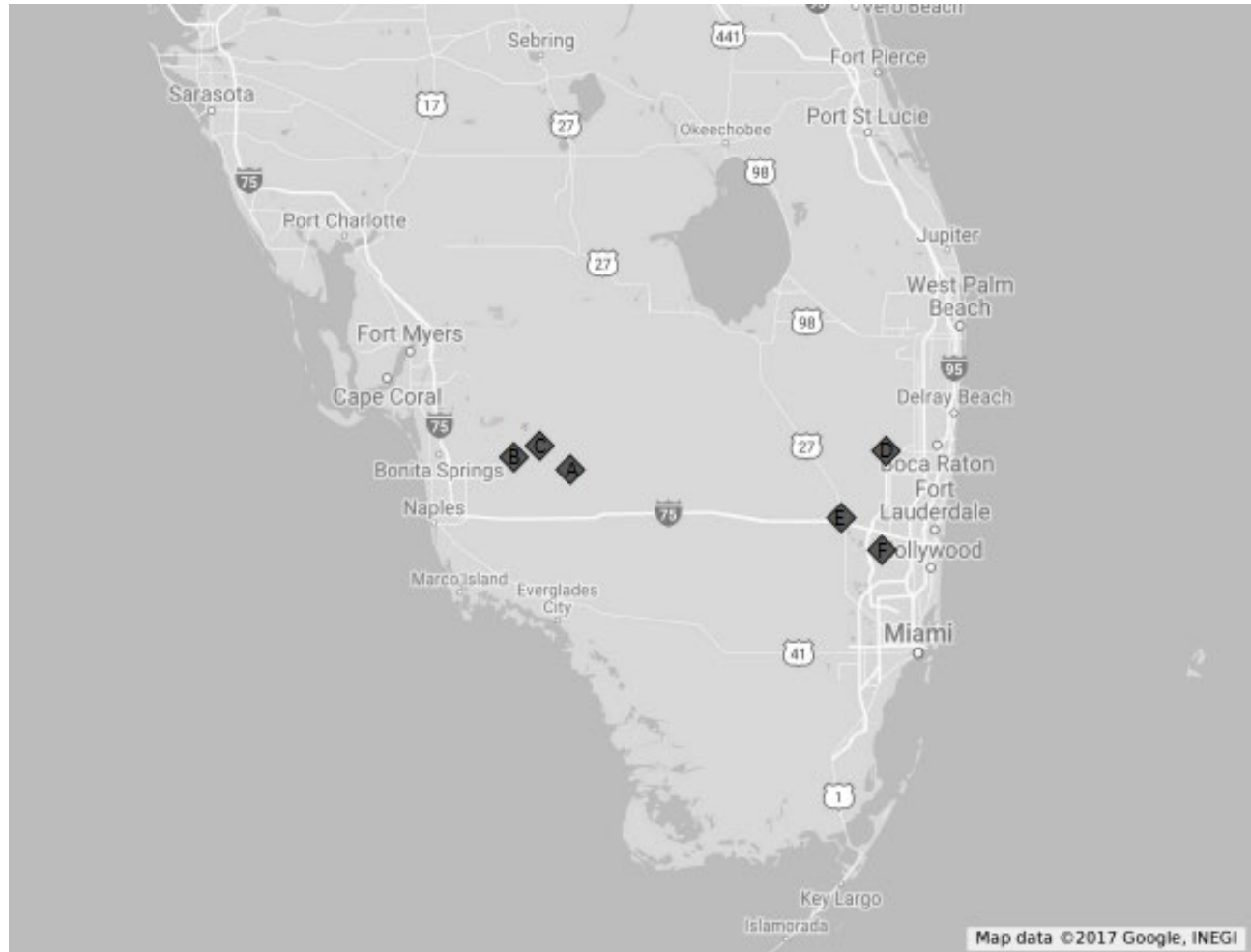
Site	Dev/Agr/Nat
2	Natural
3	Natural
4	Natural
6	Developed
7	Developed
8	Developed
9	Agriculture
10	Agriculture
12	Agriculture

Agr RV Site10- Arkansas

#	Organism group	Strain	Isolate identifiers	Serovar	Isolate	Create date	Locat...	Isolation source	Isolation ...	Min-same	Min-diff	BioSample	Assembly	Computed types
1	Salmonella ente...	PNUSAS215...	PNUSAS215035 SRS9693875		PDT001103610.1	2021-08-07	USA		clinical	18	n/a	SAMN20600798	GCA_019485005.1	Serotype: Newport Antigen formula: 8:e,h:1,2
2	Salmonella ente...	PNUSAS007...	PNUSAS007641 SRS1952141	Newport	PDT000184250.2	2017-02-01	USA	urine	clinical	1	n/a	SAMN06281875	GCA_008483905.1	Serotype: Newport Antigen formula: 8:e,h:1,2
3	Salmonella ente...	PNUSAS006...	PNUSAS006941 SRS1989728	Newport	PDT000188299.2	2017-02-17	USA	stool	clinical	1	n/a	SAMN06320658	GCA_009123385.1	Serotype: Newport Antigen formula: 8:e,h:1,2
4	Salmonella ente...	PNUSAS021...	PNUSAS021557 SRS2455325		PDT000235819.2	2017-08-23	USA		clinical	10	n/a	SAMN07524561	GCA_006140275.1	Serotype: Newport Antigen formula: 8:e,h:1,2
5	Salmonella ente...	PNUSAS037...	PNUSAS037673 SRS3145496		PDT000304851.2	2018-04-09	USA		clinical	0	n/a	SAMN08899048	GCA_006319465.1	Serotype: Newport Antigen formula: 8:e,h:1,2
6	Salmonella ente...	PNUSAS037...	PNUSAS037674 SRS3145500		PDT000304855.2	2018-04-09	USA		clinical	0	n/a	SAMN08899243	GCA_006319005.1	Serotype: Newport Antigen formula: 8:e,h:1,2



CFSAN- Florida sampling- Culture Independent



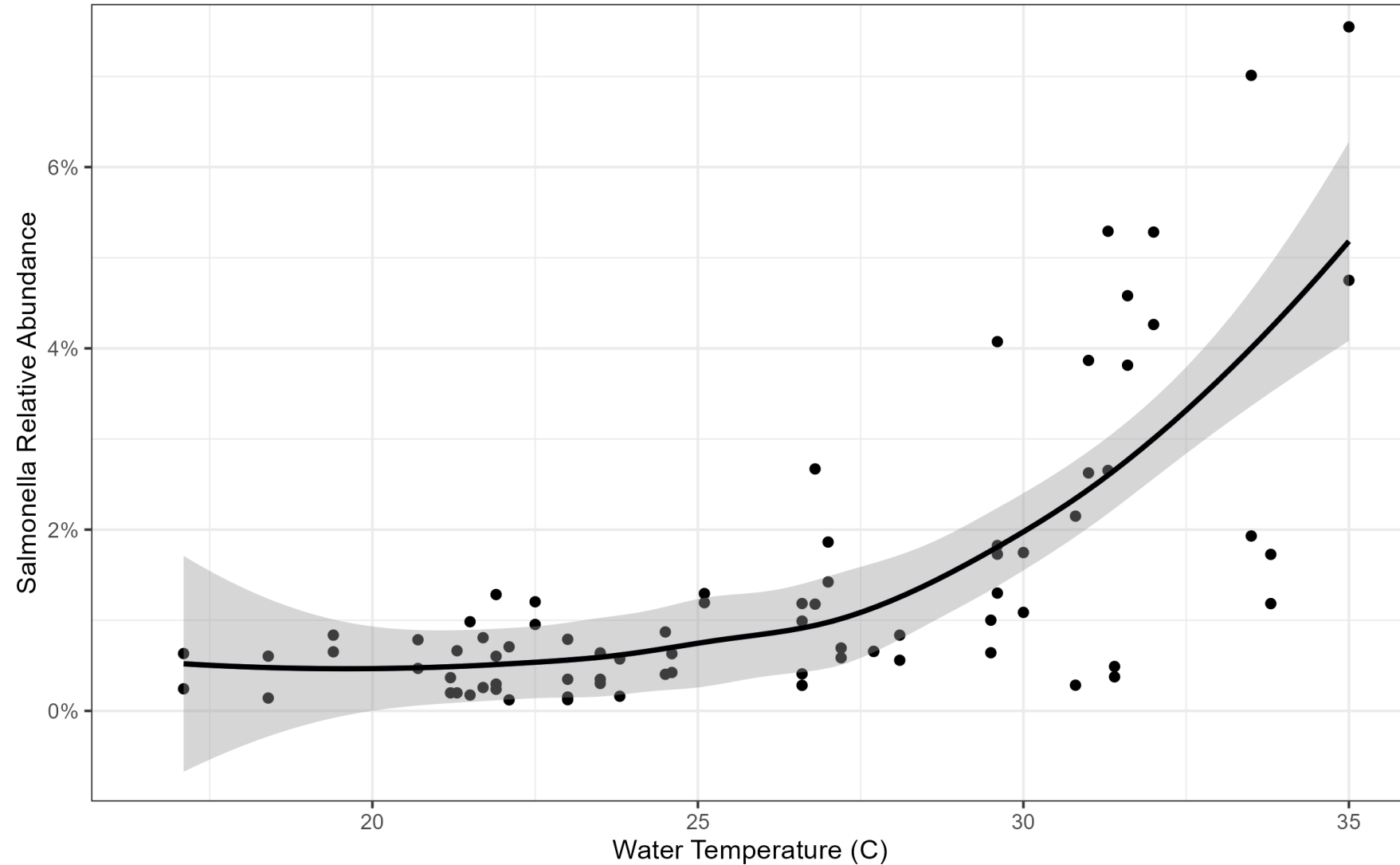
6 Sites- South Florida

Dr. Rebecca Bell and Dr Eric Brown

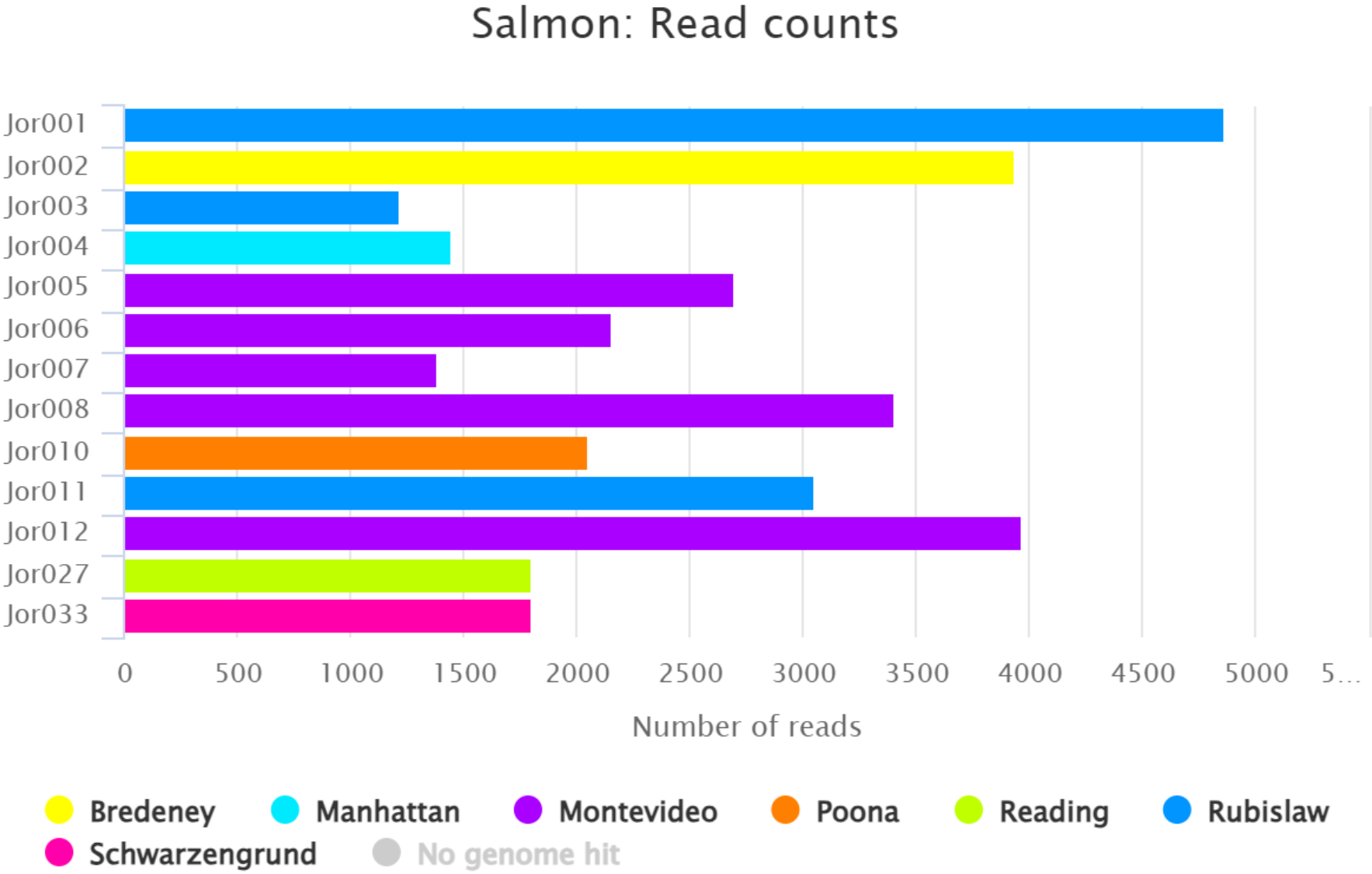
Samples collected- Aug 2015 to July 2016

Site	Type
A	Canal Near Agriculture
B	Canal Near Agriculture
C	Natural Swamp
D	Natural Swamp
E	Man Made Canal
F	Man Made Canal

Correlation of relative abundance of Salmonella to water temperature



Culture independent samples- 1% genome coverage



Conclusions and next steps

- Bettercallsal helps in detection of multi serovars of Salmonella in metagenomics dataset. Could be used in WGS when serotype information is not available.
- Many check points to avoid false positives. CFSAN longitudinal study on environmental samples, we observed close to ~20% samples are dual serovars, ~10% are 3 serovars
- Easy to access/navigate report with Sequence typing and computed serotype information

Next steps

- **mashtree + iTOL for the main workflow and then gunc + checkm2 check for the db creation**
- **Expand the workflow for other organisms. We are already working on Cronology for Cronobacter.**
- **Use bettercallsal for WGS data and make some custom changes.**

Acknowledgements

CFSAN

Kranti Konganti

Chris Grim

Amanda Windsor

Mark Mammel

Elizabeth Reed

Rebecca Bell

Sandra Tallent

Eric Brown

CVM

Brandon Kocurek

Andrea Ottesen

GenomeTrakr labs and NCBI PD

