Subtyping the top 30 *Salmonella* serotypes using a combination of CRISPR elements and virulence genes: Salmonella CRISPR-MLVST

Heather Carleton, MPH, Ph.D.

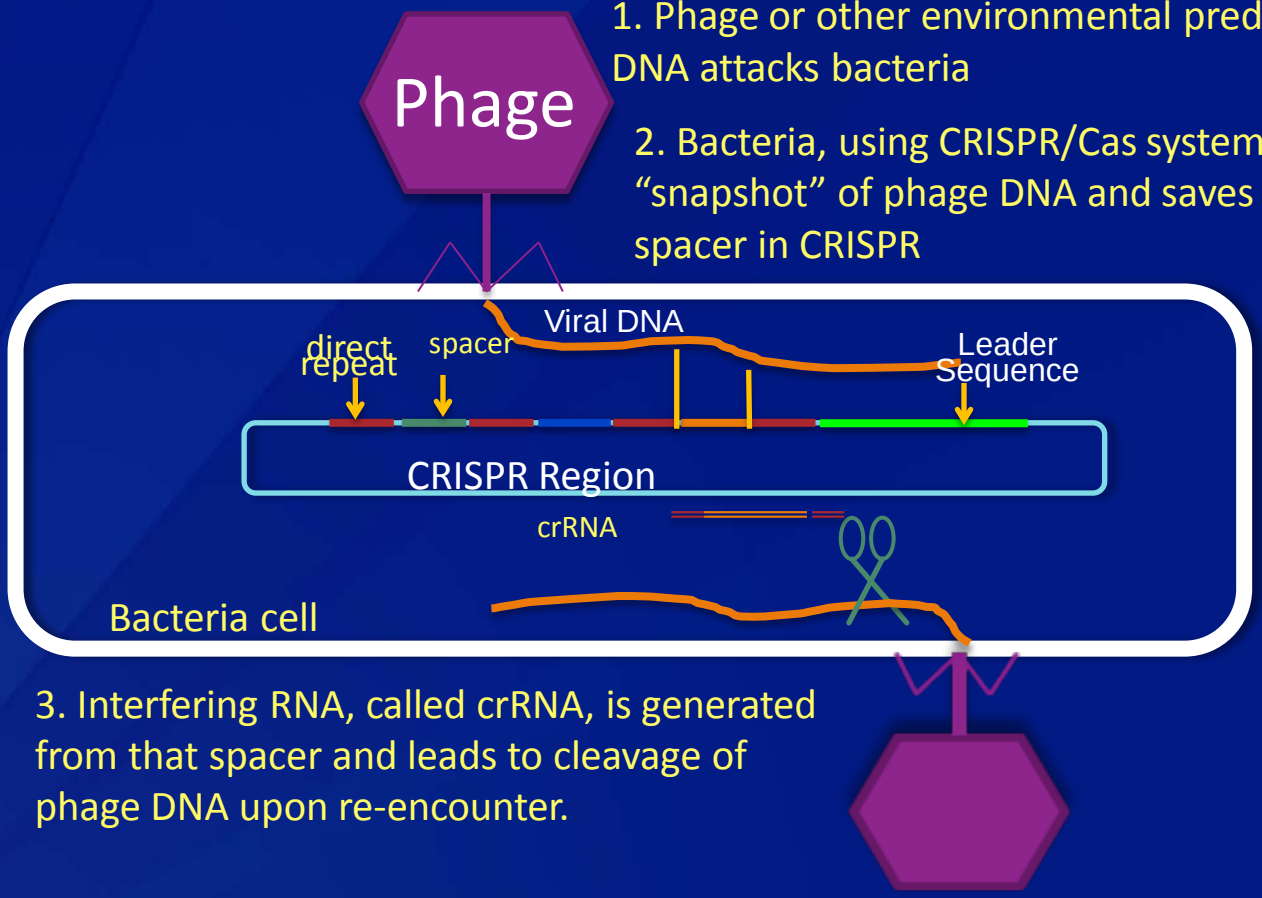
InFORM 2013

National Center for Emerging and Zoonotic Infectious Diseases
Division of Foodborne, Waterborne, and Environmental Diseases

Objectives

- Introduce the CRISPR/cas system
- Present subtyping scheme combining CRISPR and virulence genes (CRISPR-MLVST)
- Preliminary Data
- Challenges and Future Activities

Overview of CRISPR/Cas System



1. Phage or other environmental predatory DNA attacks bacteria

2. Bacteria, using CRISPR/Cas system, takes a "snapshot" of phage DNA and saves it as a spacer in CRISPR

3. Interfering RNA, called crRNA, is generated from that spacer and leads to cleavage of phage DNA upon re-encounter.

CRISPR/Cas System History

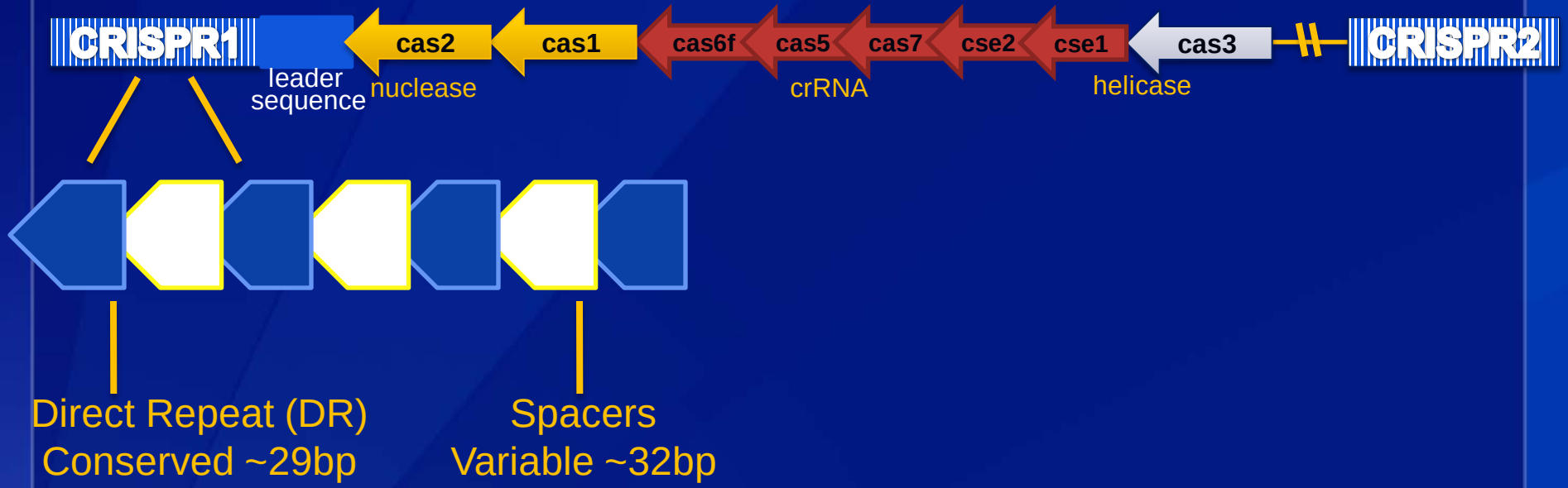
- First described in 1987 by Ishino et al in *E. coli*.
- "CRISPR" coined in 2000 by Mojica et al.

	Genomes Analyzed	CRISPR Loci Found
Archaea	150	557 (125)
Bacteria	2480	3355 (1126)
Total	2630	3912 (1251)

- CRISPR studied for subtyping different bacteria:
  - *Salmonella enterica*
  - *Campylobacter jejuni*
  - *Escherichia coli and Shigella*
  - *Vibio cholerae*
- Level of discrimination depends on organism

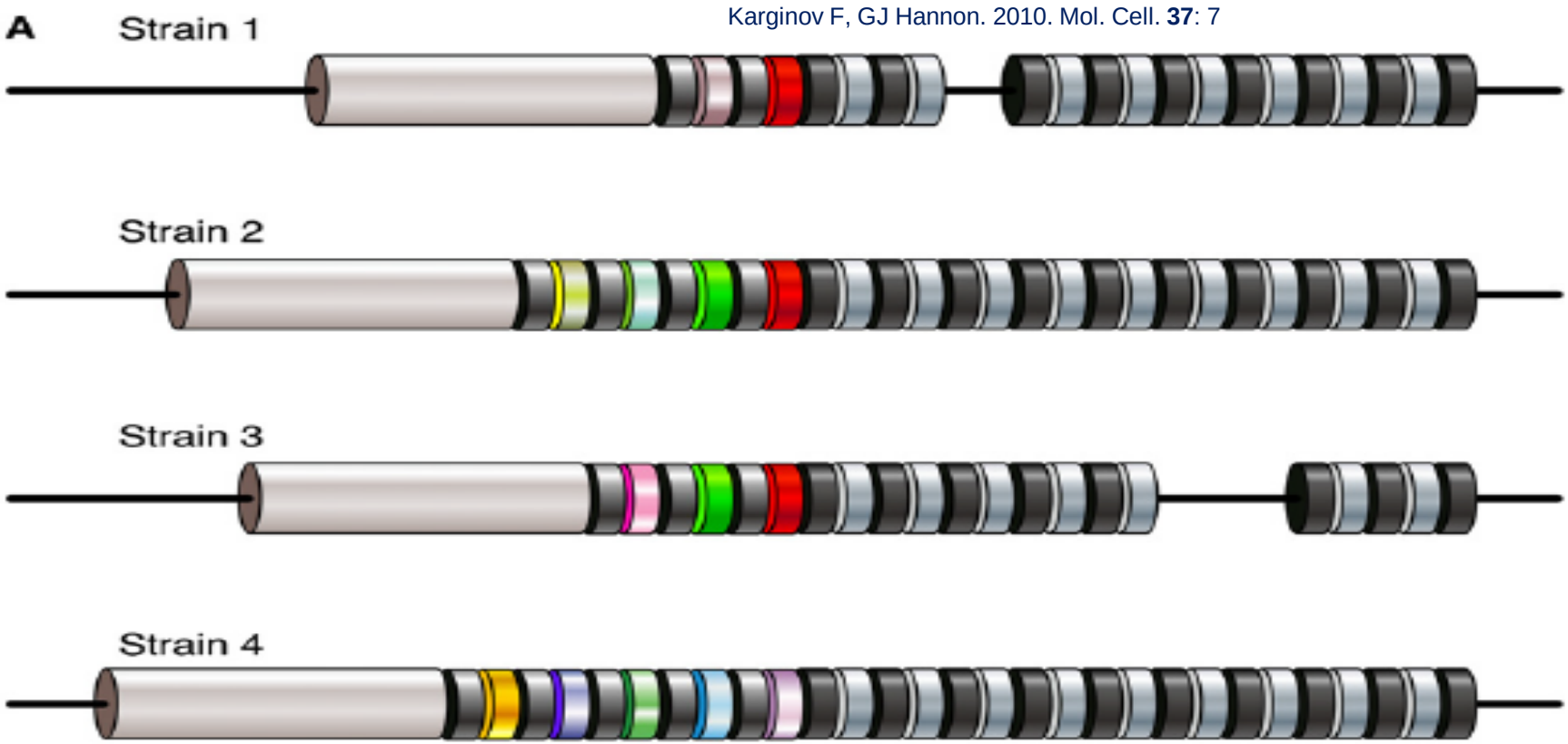
CRISPRFinder (crispr.u-psud.fr), current as of 11/5/2013

CRISPR/Cas System



- New spacers added next to leader sequence
- Spacers a reflection of predatory DNA in environment and can be used for subtyping – bacteria from same environment will have same spacers

Subtyping using CRISPR region

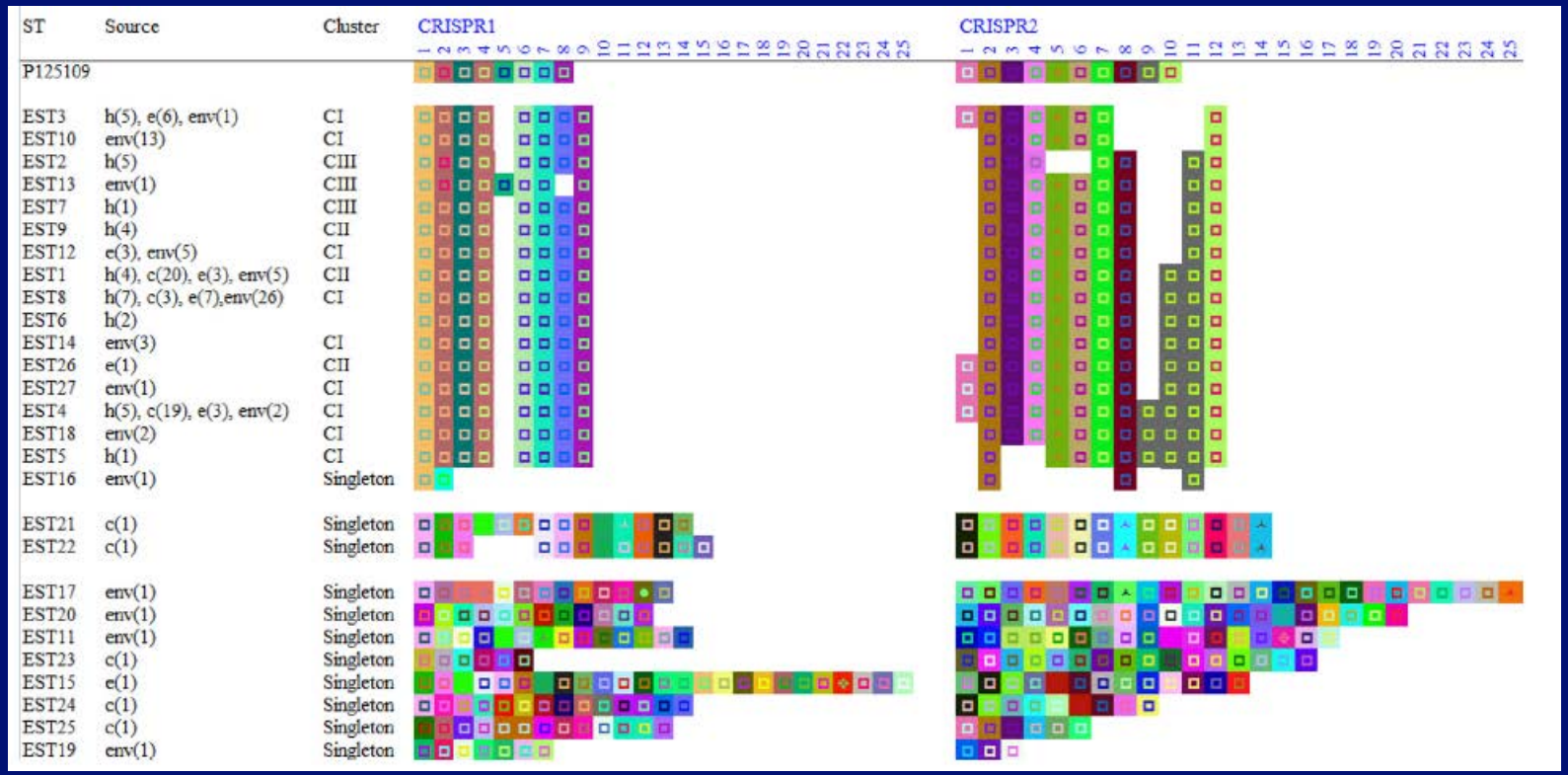


Karginov F, GJ Hannon. 2010. Mol. Cell. **37**: 7

- Subtyping based on the addition and deletion of spacer regions and variation between spacers

Subtyping using CRISPR Regions in *Salmonella*

- *Salmonella* has 2 CRISPR regions
- Ability to discriminate between serotypes is limited

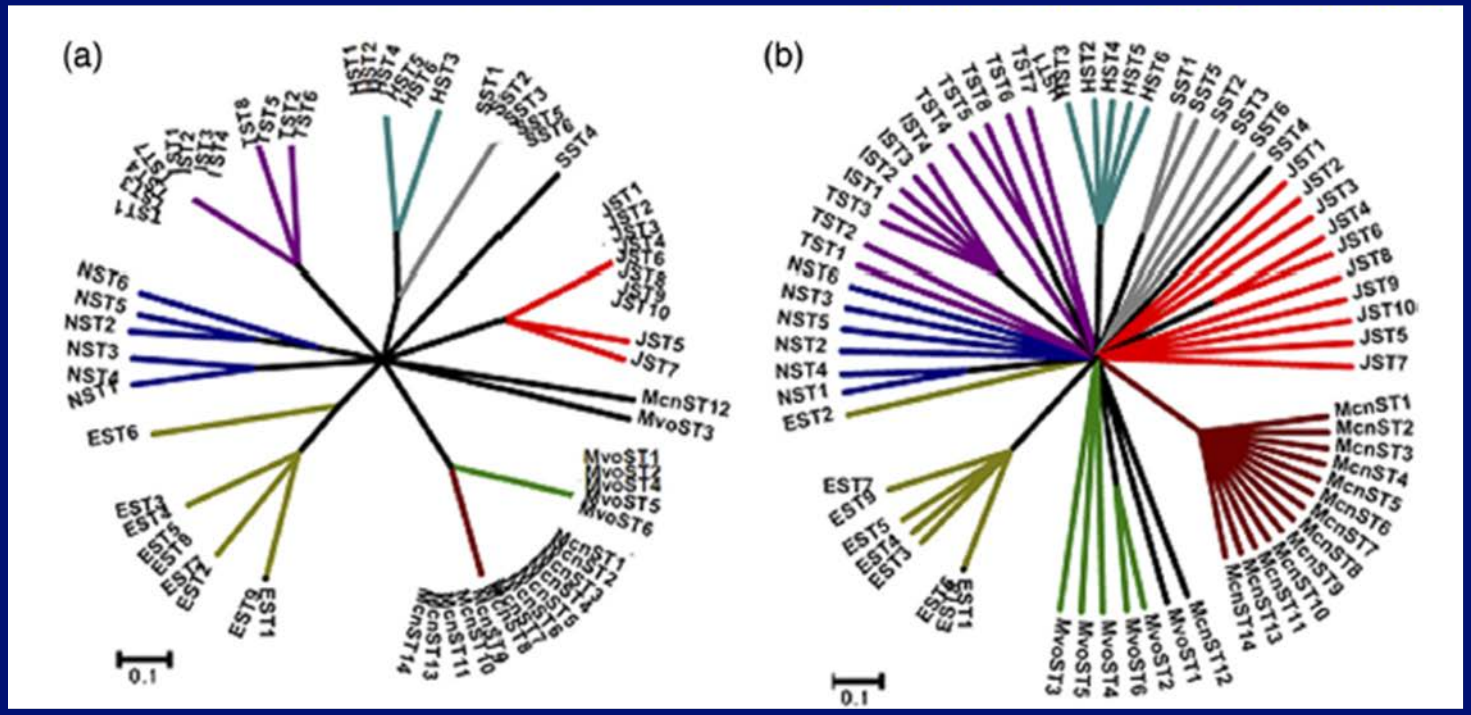


Liu F, et al. 2011. Appl. Environ. Microbiol. **77**: 4520-4526.

Subtyping with CRISPR and Virulence Genes in *Salmonella*

- **Virulence genes group isolates by serotype while CRISPR 1 and 2 identify clusters of epidemiologically related isolates within a serotype**

fimH and *sseL* CRISPR1 and CRISPR2

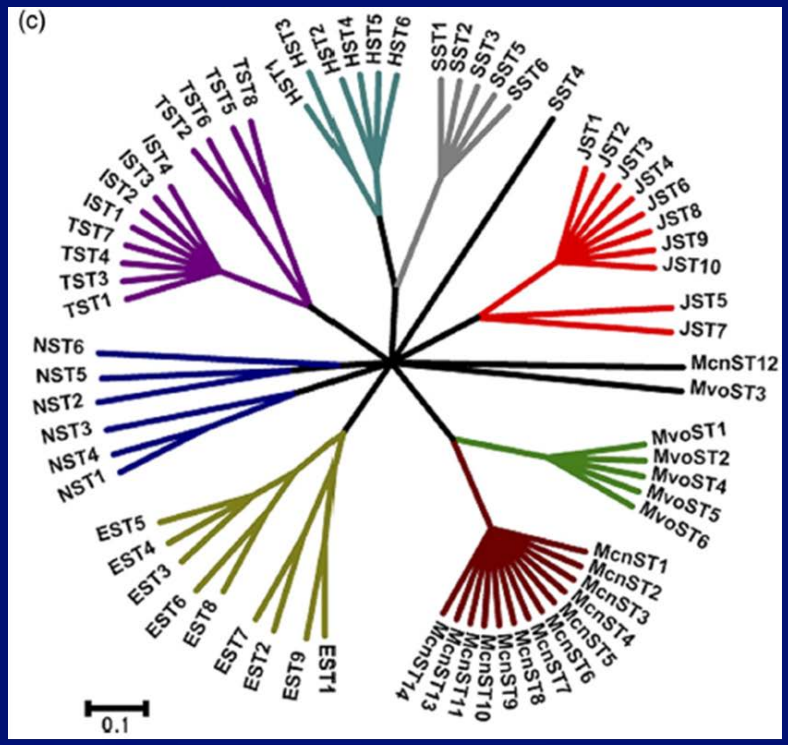


Liu F, et al. 2011. Appl. Environ. Microbiol. **77**: 1946 – 1956.

Salmonella CRISPR-MLVST

- **Virulence genes group isolates by serotype while CRISPR 1 and 2 identify clusters of epidemiologically related isolates within a serotype**

sseL, *fimH*, and CRISPR1-CRISPR2



Liu F, et al. 2011. Appl. Environ. Microbiol. **77**: 1946 – 1956.

Salmonella CRISPR-MLVST study design

- Perform CRISPR-MLVST on both sporadic and epidemiologically associated isolates from top 30 serotypes of *Salmonella enterica*
- Determine whether this method is comparable to other subtyping methods including PFGE and MLVA

Methods for CRISPR-MLVST

- **Perform Sanger sequencing on all 4 targets using previously designed primers from Penn State collaborators and Pasteur Institute Primers**
- **Develop a database using the Bionumerics software package to:**
  - Assign CRISPR types based on variable spacer regions only
  - Assign allele types to the virulence genes *sseL* and *fimH*

Assigning Allele Numbers

- **For MLVST: A new allele is defined as at least single base pair difference in the coding region for *sseL* or *fimH* that has not been previously identified**
- **For CRISPR1 and 2:**
  - New spacer is defined as at single base pair difference from previous identified spacers
  - New allele is a new spacer order not previously identified
- **Overall sequence type based on all four targets and is a combination of all 4 allele numbers (*fimH:sseL*:CRISPR1:CRISPR2)**

Preliminary Results

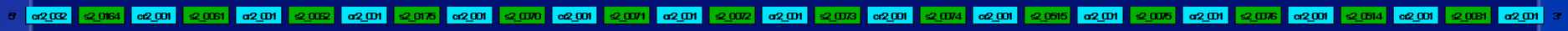
Category	Number
Total number of isolates	958
Total number of serotypes	32
Number of *fimH* alleles	68
Number of *sseL* alleles	76
Number of CRISPR 1 alleles	189
Number of CRISPR 2 alleles	134
Number of isolates sequenced at all 4 alleles	106

CRISPR 1 Region

- 43 different repeat sequences identified ranging from 28-29 bp
- 478 different spacers identified ranging from 11-64 bp (average 32bp)
- Number of spacers per allele: range 2 to 32

CRISPR 2 Region

- 48 direct repeats identified ranging from 26-30bp (average 29)
- 547 spacers identified that were 28-54bp long (average 32bp)
- Number of spacers per allele: range 2 to 26



fimH and *sseL* Alleles

- Most common differences between isolates are single nucleotide polymorphisms
- Most serotypes only have 1-2 allele types

CRISPR-MLVST Results for Serotype Saint Paul: Low Epidemiological Concordance

fimH	*sseL*	CRISPR1	CRISPR2	Oubreak associated
17	20	28	39	0806MAJN6-1c
17	54	105	45	Sporadic
17	20	28	130	Sporadic
17	20	20	134	Sporadic
17	20	20	31	Sporadic
17	20	20	31	0805NMJN6-1c
17	20	20	31	0805NMJN6-1c
17	20	27	40	Sporadic
17	20	27	40	0902NEJN6-1
17	20	27	40	0902NEJN6-1
17	20	27	40	0902NEJN6-1
17	20	20	41	Sporadic
17	20	20	41	Sporadic
17	20	20	41	Sporadic
17	20	20	41	Sporadic
17	20	20	41	0805NMJN6-1c
17	20	20	41	0805NMJN6-1c
17	20	20	41	0805NMJN6-1c
17	20	28	42	0805NMJN6-1c

CRISPR-MLVST Results for Serotype Newport: High Epidemiological Concordance

fimH	*sseL*	CRISPR1	CRISPR2	Outbreak-associated
5	6	126	11	1005SDCJJP-1
5	6	126	11	1005SDCJJP-1
5	6	126	11	1005SDCJJP-1
5	6	126	11	1005SDCJJP-1
5	6	126	11	1005SDCJJP-1
5	6	126	11	1005SDCJJP-1
5	6	126	11	1005SDCJJP-1
5	6	126	11	1005SDCJJP-1
5	6	126	11	1005SDCJJP-1
5	8	7	12	1008MDJJP-1
5	8	7	12	1008VAJJP-1
5	8	7	12	1008VAJJP-1
5	6	103	11	1005UTJJP-1
5	6	103	11	1005UTJJP-1
3	5	90	10	Historical isolate

Challenges

- **Difficult to design universal primers for all *Salmonella* serotypes**
- **The same primer design does not work for all isolates within the same serotype**
- **CRISPR region variation and size**
  - As more spacers are added the CRISPR region becomes larger, pushing the limits of what we can feasibly Sanger sequence
  - CRISPR regions vary requiring design of serotype specific primers

Future Plans

- Characterize additional isolates at all four targets to further understand diversity at CRISPR and virulence genes

- Include CRISPR regions and virulence genes as part of the genetic information extracted from whole genome sequences as use of NGS increases

Acknowledgments

- **EDLB**
  - Efrain Ribot
  - Eija Trees
  - Peter Gerner-Smidt
  - Amber Schmidtke
  - Dana Castillo
  - Ashley Sabol
  - *Salmonella* Reference Lab
  - PulseNet database team
- **CDC Biotechnology Core Facility**
- **DFBMD Genomics Group**
- **External Collaborations**
  - Ed Dudley and Nikki Shariat (Penn State Univ)
  - Laetitia Fabre, Francois-Xavier Weill (Pasteur Institute)
  - Dolf Michielsen and Applied Maths (CRISPR plug-in)
- **ASM-CDC Postdoctoral Fellowship Program**