Genomic Comparison of *Salmonella enterica* Serovars Enteritidis and Dublin

D. Matthews[1], R. Schmieder[1], J. Busch[1], N. Cassman[1], M. Doherty[1], D. Green[2], B. Matolock[2], B. Heffernan[2], G. Olsen[2], L. Farris[3], D. Schiffeli[4], S. Maloy[1], E. Dinsdale[1], and R. Edwards[1, 5]

[1]San Diego State University, San Diego, CA, [2]University of Illinois at Urbana-Champaign, Urbana, IL, [3]University of Tennessee Health Science Center, Memphis, TN, [4]University of Pennsylvania School of Veterinary Medicine, Philadelphia, PA, [5]Argonne National Laboratory, Argonne, IL

Introduction

- *Salmonella enterica* consists of >2,600 serovars that differ in host-range and virulence
- Variation in gene content (prophages, pathogenecity islands and pseudogenes) are thought to be responsible for differences in host-range and virulence found between serovars and strains belonging to the same serovar
- Enteritidis and Dublin are closely related serovars
- Enteritidis is second leading cause of food-borne salmonellosis
- Dublin is cattle-adapted but also causes disease in humans
- Enteritidis strain P125109[1] and Dublin strain CT 02021853 genomes were previously sequenced
- We sequenced the genomes of Enteritidis strain LK5 and Dublin strain SARB12
- Using the previously sequenced genomes as references, we performed bioinformatic comparative analyses to determine differences in prophage and pathogenicity island content, as well as identify insertions and deletions (indels) and single nucleotide polymorphisms (SNPs) between strains to classify putative pseudogenes

Methods

- Genomic DNA was isolated using the Wizard Genomic DNA purification kit according to the manufacturer's instructions (Promega U. S., Madison, WI, USA)
- Sequencing was performed using Sanger and 454 sequencing platforms
- Contigs were assembled from sequencing reads using GS *De Novo* Assembler[2] and scaffolded around the respective reference genomes using BLASTn[3] and the Nucmer module of Mummer[4, 5]
- Draft genome sequences were assembled using a custom script with gaps between contigs filled with Ns
- Draft genome sequences were then annotated by RAST[6], and visually inspected using Artemis[7]
- Genomic alignments were performed using progressiveMauve[8]
- SNPs and indels between genome sequences were identified using snpalign from Nucmer[4, 5]

Results

Scaffolding Summary

	Enteritidis LK5	Dublin SARB12
Reference genome	Enteritidis P125109	Dublin CT 02021853
Reference genome size	4,685,848 bp	4,842,908 bp
Total # of contigs	49	64
# of used contigs	28	36
--Mean contig size	159,249 bp	132,657 bp
--Contig size range	3,964-910,584 bp	2,649-475,988 bp
# of unused contigs	21	28
--Plasmid	1	4
--*rrn* operon	3	8
# of gapped bases (% of ref. total)	64,824 (1.38%)	63,220 (1.31%)
# of gaps	26	31
Mean gap length	2493 bp	2109 bp
Gap length range	47-6129 bp	1-6313 bp

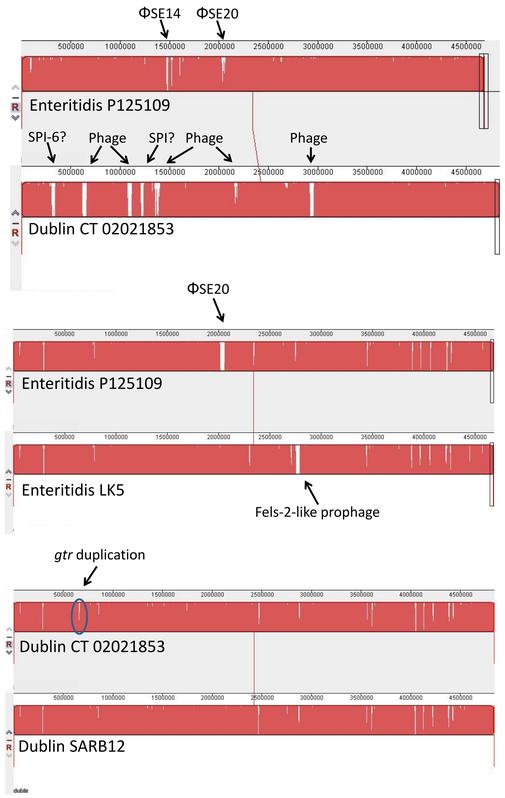
Enteritidis P125109 vs. LK5

Single Nucleotide Polymorphisms		Indels	
Total	**492**	**Total**	**97**
HP tracts	117	HP tracts	74
Transitions	339	≤ 2bp	82
Transversions	153	> 2 bp	15
Coding Total	**392**		
Synonymous	153		
Non-synonymous	239		
tRNA	11		
Intergenic	89		

Dublin CT 02021853 vs. SARB12

Single Nucleotide Polymorphisms		Indels	
Total	**720**	**Total**	**95**
HP tracts	177	HP tracts	72
Transitions	527	≤ 4bp	76
Transversions	193	> 4 bp	19
Coding Total	**583**		
Synonymous	236		
Non-synonymous	347		
Unknown	14		
Intergenic	124		

Genomic Alignments

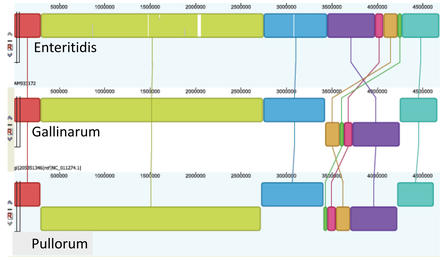


Conclusions

- Identified genomic differences between strains belonging to serovars Enteritidis and Dublin support the hypothesis that these differences contribute to host-range and virulence
- Higher SNP content between Dublin strains may be indicative of higher pseudogene content and host-adaptation of this serovar

Future Directions

- Validation of SNPs in homopolymeric tracts
- Determine differences in pseudogene content
- Include Gallinarum and Pullorum genomic sequences in comparative analyses



Acknowledgements

Alejandra Prieto-Davo
Keri Elkins
Greg Boecker
Victor Seguritan
Tara Atigh
Archana Srinivas

Funded by:
NSF grant DBI: 0850356 from the Division of Biological Infrastructure
NSF grant 1044453 from the Division of Undergraduate Research

References

[1] Thomson, N. R., Clayton, D. J. , Windhorst, D., *et al.* 2008. Comparative Genome Analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 Provides Insights into Evolutionary and Host Adaptation Pathways. *Genome Res.* 18:1624-1637.

[2] 454 Sequencing System Software Manual, version 2.5.3. 2010. 454 Life Sciences Corp., A Roche Company, Branford, CT

[3] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

[4] Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., and S. L. Salzberg. 2004. Versatile and Open Software for Comparing Large Genomes. *Genome Biology* 5:R12.

[5] Delcher, A. L., Phillippy, A., Carlton, J., and S. Salzberg. 2002. Fast Algorithms for Large-scale Genome Alignment and Comparison. *Nucleic Acids Res*. 30:2478-2483.

[6] Aziz R.K., Bartels, D., Best, A. A., DeJongh, M., *et. al.* 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75.

[7] Rutherford, K., Parkhill, J., Horsnell, T., Rice, P., Rajandream, M. A., and B. Barrell. 2000. Artemis: Sequence Visualization and Annotation. *Bioinformatics* 16:944-945.

[8] Darling, A. E. , Mau, B., and N. T. Perna. 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. *PLoS One* 5(6):e11147.