Genomic Comparison of *Salmonella enterica* Serovars Enteritidis and Dublin

D. Matthews¹, R. Schmieder¹, J. Busch¹, N. Cassman¹, M. Doherty¹, D. Green², B. Matolock², B. Heffernan², G. Olsen², L. Farris³, D. Schifferli², S. Maloy¹, E. Dinsdale¹, and R. Edwards¹,²

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

### Introduction

- *Salmonella enterica* consists of >2,600 serovars that differ in host-range and virulence
- Variation in gene content (prophages, pathogenicity 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 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 Assembler² and scaffolded around the respective reference genomes using BLASTn³ and the Nucmer module of Mummer⁴,⁵
- Draft genome sequences were assembled using a custom script with gaps between contigs filled with Ns
- Draft genome sequences were then annotated by RAST⁶, and visually inspected using Artemis⁷
- Genomic alignments were performed using progressiveMauve⁸
- SNPs and indels between genome sequences were identified using snpalign from Nucmer⁴,⁵

### Results

#### Scaffolding Summary

<table>
<thead>
<tr>
<th>Reference</th>
<th>Enteritidis LK5</th>
<th>Dublin SARB12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference genome size</td>
<td>4,681,848 bp</td>
<td>4,842,808 bp</td>
</tr>
<tr>
<td>Total # of contigs</td>
<td>49</td>
<td>64</td>
</tr>
<tr>
<td># of used contigs</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>--Mean contig size</td>
<td>159,249 bp</td>
<td>132,657 bp</td>
</tr>
<tr>
<td>--Contig size range</td>
<td>3,964,910-584 bp</td>
<td>2,649,475-984 bp</td>
</tr>
<tr>
<td># of unused contigs</td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>--Plasmid</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>--n-mer ooper</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>% of gapped bases (% of ref. total)</td>
<td>(1.38%)</td>
<td>(3.13%)</td>
</tr>
<tr>
<td># of gaps</td>
<td>64,824</td>
<td>63,320</td>
</tr>
<tr>
<td>Mean gap length</td>
<td>26 bp</td>
<td>31 bp</td>
</tr>
<tr>
<td>Gap length range</td>
<td>47-6129 bp</td>
<td>1-6313 bp</td>
</tr>
</tbody>
</table>

#### Enteritidis P125109 vs. LK5

- **Single Nucleotide Polymorphisms**
  - **Indels**
    - Total: 482
    - **HP tracts**: 117
    - **Transitions**: 339
    - **Coding Total**: 392
      - **Synonymous**: 153
      - **Non-synonymous**: 239
    - **DNA**: 11
    - **Intergenic**: 89
  - **Total**: 97

#### Dublin CT 02021853 vs. SARB12

- **Single Nucleotide Polymorphisms**
  - **Indels**
    - Total: 730
    - **HP tracts**: 177
    - **Transitions**: 527
    - **Coding Total**: 583
      - **Synonymous**: 236
      - **Non-synonymous**: 347
      - **Unknown**: 14
      - **Intergenic**: 124
    - **Total**: 99

### 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 andPullorum genomic sequences in comparative analyses

### Acknowledgements

Alejandro Prieto-Davo
Keri Elkins
Greg Boecker
Victor Seguritan
Tara Atigh
Archantha Srinivas

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

### References