Genomic approaches to reconstruct the landscape of microbial contamination in aquatic systems

GLOBAL WATER FOOD SAFETY SUMMIT

Prof. Jaime Martinez-Urtaza
jaime.martinez-urtaza@cefas.co.uk
j.martinez-urtaza@exeter.ac.uk
Cefas is an executive agency within the UK Government.

We currently employ over 500 staff at two main specialist laboratories and operate our own ocean going research vessel.

**Lowestoft**: international fisheries science and management, coastal and marine ecosystem processes, environmental radioactivity assessment, contaminant analysis, regulatory advice for coastal activities and information technology services

**Weymouth**: research and advice for fish health, disease diagnosis, fish and shellfish hygiene/food safety, ecotoxicology, evaluation of products used in aquaculture
Populations dynamics and diseases

Two Models:

**Salmonella**
- Foodborne pathogen

**Vibrio**
- Waterborne pathogen
  - *V. cholerae*
  - *V. parahaemolyticus*
Classic Typing Techniques → Whole Genome Sequencing

MLVA
PFGE
MLST → Genomic Epidemiology
Core genome
- Mutations
- Recombination

Accessory Genome
- Horizontal Gene

Phylogeny

Local Adaptation
AB Resistance
Virulence
**STRAIN 1**

**STRAIN 2**

---

**When?**
Date of emergence of clones and variants

**Where?**
Place of emergence and routes of dispersal

**Introduction?**
Route of entry for the pathogen

**Adaptation?**
Likelihood of becoming endemic
Biological ‘corridors’ of disease?

1) Isolation of (*) strains

2) Genome sequencing and SNP identification

3) Comparative sequence analysis and phylogenetic determination
Analysing disease emergence using a combination of approaches

- Epidemiology and risk
- Remote sensing
- Molecular biology
- Oceanography & climate science
Phylodynamics of contamination and transmission of Salmonella in aquatic systems

Application of genome-wide analysis for tracking the dispersion of *Salmonella enterica* strains in rivers of Sinaloa, Mexico

Salmonella

• Pathogen of human and animals

• One of the major causes of foodborne infection at global scale
Sinaloa, Mexico

- Mexico is characterized by a **high morbidity** associated with salmonellosis.

- 119,374 cases in 2010, representing **106 cases per 100,000 inhabitants**

- Sinaloa, nearly twice the number of cases, **192 cases per 100,000 inhabitants**

- United States, **16.7 cases per every 100,000 population**
Sinaloa

Phylodynamics of Salmonella contamination in the river Culiacán

Surface waters

Coastal water

Waste water

Livestock
Despite the growing importance of non-host ecology in resolving the epidemic dynamics of Salmonella, investigations of this pathogen in natural settings has been typically constrained by the low occurrence of this organism in environmental samples, typically below 5%.

The rare presence of Salmonella in the environment has made extremely difficult the identification of patterns of contamination in a specific region or ecosystem, limiting any possibility for a reliable traceback from original sources which ultimately allows for the delineation of the routes of dispersal.

The low number of strains identified over the course of environmental surveys typically belong to a broad range of serovars with highly diverse genetic backgrounds.

This situation limits the comparison of strains at a serovar level and any effort to compare strains of different serovars would result inconclusive.
Prevalence and genetic diversity of *Salmonella* spp. in a river in a tropical environment in Mexico

Maribel Jiménez, Jaime Martinez-Urtaza, Maria Xose Rodriguez-Alvarez, Josefina Leon-Felix, Cristobal Chaidez

Journal of Water and Health Dec 2014, 12 (4) 874-884; DOI: 10.2166/wh.2014.051
Tropical Climate
Pulsed-field gel electrophoresis

PFGE

Geographical and Temporal Dissemination of Salmonellae Isolated from Domestic Animal Hosts in the Culiacan Valley, Mexico

Maribel Jiménez · Jaime Martínez-Urtaza · Cristobal Chaidez
Prevalence and genetic diversity of *Salmonella* spp. in a river in a tropical environment in Mexico
Maribel Jiménez, Jaime Martínez-Urtaza, María Xose Rodríguez-Alvarez, Josefina Leon-Felix and Cristobal Chaidez
AIM: to test the efficiency of WGS to identify the point source of contamination and draw the routes of dispersal of Salmonella in a specific geographic area (Sinaloa, Mexico).

Are these populations resident in the river?

Could we use genomics to track the dispersion?
The Study

- A total of **150** *Salmonella enterica* strains (Jiménez et al., 2011, 2014) were sequenced. Only **61** local *Salmonella enterica* genomes belonging to the two prevailing serovars -Oranienburg and Saintpaul- were selected for the study.

- **Six sampling sites** were selected to cover the study area (named A, B, C, D, E and F). Sites A and B were located on mountain sides; Site C was located in Culiacan City; Sites D and E were located in the valley next to the city limit; and Site F was located on the Pacific coast.

- Strains were isolated from **river water** and **animal feces** (cow, goat and chicken).
## Strains

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oranienburg</td>
<td>19</td>
</tr>
<tr>
<td>Give</td>
<td>11</td>
</tr>
<tr>
<td>Saintpaul</td>
<td>9</td>
</tr>
<tr>
<td>Minnesota</td>
<td>8</td>
</tr>
<tr>
<td>Anatum</td>
<td>4</td>
</tr>
<tr>
<td>Luciana</td>
<td>3</td>
</tr>
<tr>
<td>Thompson</td>
<td>3</td>
</tr>
<tr>
<td>Weltevreden</td>
<td>3</td>
</tr>
<tr>
<td>Agona</td>
<td>3</td>
</tr>
<tr>
<td>Newport</td>
<td>2</td>
</tr>
<tr>
<td>Soahanina</td>
<td>2</td>
</tr>
<tr>
<td>Albany</td>
<td>1</td>
</tr>
<tr>
<td>Montevideo</td>
<td>1</td>
</tr>
<tr>
<td>Muenchen</td>
<td>1</td>
</tr>
<tr>
<td>Gaminara</td>
<td>1</td>
</tr>
<tr>
<td>Cayar</td>
<td>1</td>
</tr>
<tr>
<td>Javiana</td>
<td>1</td>
</tr>
<tr>
<td>Meleagridis</td>
<td>1</td>
</tr>
<tr>
<td>Enteritidis</td>
<td>1</td>
</tr>
<tr>
<td>Typhimurium</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>76</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oranienburg</td>
<td>23</td>
</tr>
<tr>
<td>Saintpaul</td>
<td>10</td>
</tr>
<tr>
<td>Minnesota</td>
<td>7</td>
</tr>
<tr>
<td>Give</td>
<td>5</td>
</tr>
<tr>
<td>Anatum</td>
<td>5</td>
</tr>
<tr>
<td>Muenster</td>
<td>5</td>
</tr>
<tr>
<td>Pomona</td>
<td>5</td>
</tr>
<tr>
<td>Infantis</td>
<td>5</td>
</tr>
<tr>
<td>Poona</td>
<td>5</td>
</tr>
<tr>
<td>Weltevreden</td>
<td>4</td>
</tr>
<tr>
<td>Montevideo</td>
<td>4</td>
</tr>
<tr>
<td>Newport</td>
<td>3</td>
</tr>
<tr>
<td>Bovismorbificans</td>
<td>2</td>
</tr>
<tr>
<td>Sandiego</td>
<td>2</td>
</tr>
<tr>
<td>Javiana</td>
<td>2</td>
</tr>
<tr>
<td>Sundsvall</td>
<td>2</td>
</tr>
<tr>
<td>Luciana</td>
<td>2</td>
</tr>
<tr>
<td>Agona</td>
<td>2</td>
</tr>
<tr>
<td>Seftenberg</td>
<td>1</td>
</tr>
<tr>
<td>Braenderup</td>
<td>1</td>
</tr>
<tr>
<td>Kiambu</td>
<td>1</td>
</tr>
<tr>
<td>Soahanina</td>
<td>1</td>
</tr>
<tr>
<td>Texas</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>98</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>Strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>2</td>
</tr>
<tr>
<td>Brazil</td>
<td>2</td>
</tr>
<tr>
<td>China</td>
<td>3</td>
</tr>
<tr>
<td>Colombia</td>
<td>10</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>2</td>
</tr>
<tr>
<td>Denmark</td>
<td>1</td>
</tr>
<tr>
<td>Egypt</td>
<td>1</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1</td>
</tr>
<tr>
<td>Filipinas</td>
<td>2</td>
</tr>
<tr>
<td>France</td>
<td>4</td>
</tr>
<tr>
<td>Ghana</td>
<td>1</td>
</tr>
<tr>
<td>Honduras</td>
<td>1</td>
</tr>
<tr>
<td>India</td>
<td>26</td>
</tr>
<tr>
<td>Indonesia</td>
<td>2</td>
</tr>
<tr>
<td>Jamaica</td>
<td>1</td>
</tr>
<tr>
<td>Lebanon</td>
<td>1</td>
</tr>
<tr>
<td>Malaysia</td>
<td>2</td>
</tr>
<tr>
<td>Malta</td>
<td>1</td>
</tr>
<tr>
<td>Mexico</td>
<td>23</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>1</td>
</tr>
<tr>
<td>Tailand</td>
<td>1</td>
</tr>
<tr>
<td>Turkey</td>
<td>8</td>
</tr>
<tr>
<td>UK</td>
<td>80</td>
</tr>
<tr>
<td>USA</td>
<td>117</td>
</tr>
<tr>
<td>Vietnam</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>295</strong></td>
</tr>
</tbody>
</table>
Sequencing strategy

• Strains were sequenced by MiSeq (Illumina)

• Minimum coverage of 40–120X.

• Libraries were prepared with the Nextera XT DNA sample preparation kit (Illumina).

PacBio (FDA)

• S_Muenster_CFSAN001301
• S_Give_CFSAN024229
• S_Poona_ATCC_BAA_1673
• S_Infantis_CFSAN003307
• S_Oranienburg_CFSAN001285
• S_Saintpaul_CFSAN004173
• S_Minnesota_CFSAN017963
• S_Pomona_ATCC_10729_CFSAN000720
Raw sequences

Process sequences (QA, adapter removal)  
*Trimomatic*

Genome assembly  
*A5 pipeline*

Genome alignment  
*Harvest tools*

Core genome

Recombination analysis  
*ClonalframeML*

Remove gaps  
*Trimal*

Remove recombinant regions  
*cutseq*

Core SNPs

Remove gaps

Extract SNPs in coding regions  
*SnpEff*

Extract SNPs in coding regions

Phylogenetic analysis  
*RAxML*

Molecular clock analysis  
*Beast*
Global phylogeny with ALL Serotypes
Global phylogeny of *Salmonella* Oranienburg. Maximum likelihood tree of the full dataset showing the phylogenetic relations between the core genome of 92 global and 40 local *S*. Oranienburg isolates.
Phylogeny of *Salmonella* Oranienburg ST23 comprising all the isolates from Sinaloa.
Global phylogeny of *Salmonella* Saintpaul. Maximum likelihood tree showing the phylogenetic relations between the core genome of 93 global and 19 local isolates.
Phylogeny for the specific clade of ST50 where all the isolates from Sinaloa were included.
Local phylogeny and resistome cluster analysis of *Salmonella* Oranienburg.

Maximum Likelihood tree of local *S.* Oranienburg dataset showing the phylogenetic relations between the core genome 40 local isolates.
Local phylogeny and resistome cluster analysis of *Salmonella* Saintpaul.

Maximum Likelihood tree of local *S. Saintpaul* dataset showing the phylogenetic relations between the core genome 19 local isolates.
Identification of Potential Niche-Specific Genes within the Accessory Genome of *Salmonella* Oranienburg

Alec Ko Bailey
Identification of Potential Niche-Specific Genes within the Accessory Genome of *Salmonella* Saintpaul

Prophage occurrence within isolates of the three major clades

<table>
<thead>
<tr>
<th>Isolate Name</th>
<th>Sample date</th>
<th>Sample location</th>
<th>Source</th>
<th>Clade #</th>
<th>Gifsy-2 (47)</th>
<th>Gifsy-1 (27)</th>
<th>Fels-1 (23)</th>
<th>Burkho (17)</th>
<th>SEN4 (20)</th>
<th>SEN5 (20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>47351</td>
<td>26/08/08</td>
<td>El Castillo</td>
<td>River</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47363</td>
<td>21/10/08</td>
<td>Jotagua</td>
<td>River</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51344</td>
<td>12/08/08</td>
<td>Agua</td>
<td>Chicken</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51346</td>
<td>09/09/08</td>
<td>Agua</td>
<td>Chicken</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51347</td>
<td>23/09/08</td>
<td>Agua</td>
<td>Chicken</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47396</td>
<td>25/02/09</td>
<td>Puente</td>
<td>River</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47397</td>
<td>03/10/09</td>
<td>Puente</td>
<td>River</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47361</td>
<td>10/07/08</td>
<td>Jotagua</td>
<td>River</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47402</td>
<td>26/08/08</td>
<td>Jotagua</td>
<td>River</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51336</td>
<td>26/08/08</td>
<td>Jotagua</td>
<td>Cow</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONCLUSIONS

According to results from MLST, isolates of S. Oranienburg and S. Saintpaul from Sinaloa belongs to a single STs which also included isolates from other countries.

A deeper analysis onto the phylogenetic relationships within each ST revealed a more complex structure with clear differences between both serovars:

• **S. Saintpaul** isolates from Sinaloa were identified as member of a single cluster only comprised by Mexican isolates and **without connections** with isolates from anywhere in the world.

• **S. Oranienburg** populations from Sinaloa **shared position** in the phylogenetic tree with isolates from the USA.
ANTIMICROBIAL RESISTANCE

- The diversity and prevalence of AMR genes detected within the genomes of serovars Saintpaul and Oranienburg analyzed in this study, indicated that S. enterica isolates from environmental sources **retain the potential for multidrug resistance.**

- Due to the similarity of the antibiotic resistance profile found among S. Oranienburg and S. Saintpaul isolates, the AMR gene profiles have been shown to be of **limited applicability** for source tracking purposes.

- The detection of similar patterns of resistance in isolates from animals and water suggests that the **extensive use or misuse of antimicrobial agents.**
SOURCE TRACKING

- Local populations of S. Oranienburg and S. Saintpaul in Sinaloa are composed by multi-resistant clonal groups moving frequently between asymptomatic domestic animals and non-host environments with no clear geographical barriers within sampling area investigated in this study.

- Domestic and farm animal activities near rivers are primary contributors to the persistence of specific clones found in environmental settings. The presence of strains with a high level of genetic similarity at different sampling sites revealed an effective circulation of Salmonella populations in the area potentially mediated by the movement of domestic animals and the dispersal through river flow.
The application of a new generation of tools for microbial source tracking based on the use of genomic data identification of resistance and virulence genes, genomic characterization, pangenome construction,

... but would it be possible to reconstruct the biogeography of the dispersal.

RECONSTRUCTING THE LANSCAPE OF SALMONELLA CONTAMINATION IN RIVERS USING A GENOMIC BAYESIAN FRAMEWORK