A *Salmonella* Geno-Serotyping Array (SGSA) for the Rapid Classification of Serovars

Catherine Yoshida\(^1\), Erika J. Lingohr\(^1\), Kristyn Franklin\(^1\), Levente Bodrossy\(^2,5\), Muna Anjum\(^3\), Clifford G. Clark\(^4\), Andrew M. Kropinski\(^1\)

\(^1\)Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Guelph, ON
\(^2\)Austrian Institute of Technology, Seibersdorf, Austria
\(^3\)CSIRO Marine and Atmospheric Research, Hobart, Tasmania, Australia (current address)
\(^4\)Animal Health Veterinary Laboratories Agency, New Haw, Addlestone, Surrey, United Kingdom
\(^5\)National Laboratory for Enteric Pathogens, Winnipeg, MB, Canada

CONTACT: catherine.yoshida@phac-aspc.gc.ca
PHAC – *Division of Science and Technology*

- Develop genomic tools and other interventions
- Comparative genome-based research to advance knowledge and innovations on pathogens within different environments
- Leverage genome sequence data

**GOAL:**

Enhance mitigation efforts of infectious diseases caused by zoonotic enteric pathogens
Salmonella Identification

• More than 2500 serotypes identified by unique antigenic formulae as characterized by the White-Kauffmann-Le Minor scheme
• Serologically distinguished by somatic (O) and flagellar (H) antigens
• Phage Typing and PFGE used to further classify certain serovars (including S. Typhimurium and S. Enteritidis)
Importance of Classification

1. Key for surveillance and outbreak detection
   - Serotypes can be closely associated with specific disease
   - Provides a database that is well understood by physicians, researchers and epidemiologists
   - Identification of infection reservoirs as many serotypes have unique host ranges (ex. Cattle: S. Dublin, Chickens: S. Gallinarum)

2. Key for population biology of the organism
   - Identification of host-specific genes
   - Trace horizontal gene transfer (O:58 antigen gene cluster in both subsp. IIIb (diarizonae), and subsp. II (salamae))
Traditional Serotyping

- White-Kauffmann-Le Minor scheme
- Determination of an antigenic formula
- Based on somatic (O) and flagellar (H) surface antigens

- Costly and laborious requiring highly trained technicians
- Subjective results possible
- Limited commercially available anti-sera
- Variation in in-house preparations
Global need for molecular typing method

**Phenotypic** → **Genotypic**

- Improved source attribution and molecular epidemiology
- Increased need for subtyping
- Increased whole genome sequence data

- **Molecular typing and subtyping tools:**
  - PFGE
  - MLST
  - MLVA
  - SNP
  - Microarray
  - PCR
  - Pyrosequencing
  - CRISPR
Our array – *Salmonella* Geno-Serotyping Array (SGSA)

Developed to provide a rapid molecular method to “antigenically” type *Salmonella*

- Mimic the White-Kauffmann-Le Minor scheme for continuity
- A robust system requiring only 1 day
- Adjustable for high throughput analysis
- Economical
- Non-subjective results
Somatic Antigens (O)

Chromosomally encoded between \textit{galF} and \textit{hisF}

- Glycosyltransferases - addition of sugars to the O antigen oligosaccharide
- \textit{wzx} flippase - translocates O antigen oligosaccharide across the membrane
- \textit{wzy} polymerase - adds entire O antigen unit to the LPS

\textit{wzx} and \textit{wzy} are highly variable and considered specific to serogroup
Flagellar Antigens (H)

Chromosomally encoded in two different locations

- *fliC* encodes the phase 1 flagellin
- *fljB* encodes the phase 2 flagellin
- Genes are highly conserved at their 5’ and 3’ end

Central region is variable and therefore specific to serotype
Salmonella

LPS

Salmonella

O antigen (rfb cluster)

H1 flagellar antigen (fliC)

H2 flagellar antigen (fljB)
Probe Design

Align sequences to identify SNPs, or sequence variations of gene targets

Design probes with as much unique sequence as possible, and with a 5’ SNP
Initial Group

Prototype microarray and extensive sequence databases. Continued development.

Improved labelling method for the detection of SNPs.

Expertise with the Alere ArrayTube™ system for typing and commercial applications.
## 43 Most Prevalent Serovars

20 most frequently typed human, animal, and environmental *Salmonella* serovars in the UK, Austria, and North America

| 1,4,5,12:i:- | Indiana | Paratyphi B var. Java |
| 61:k:1,5,(7) | Infantis | Pullorum |
| Abony       | Javiana  | Rissen |
| Agona       | Kedougou | Saintpaul |
| Anatum      | Kentucky | Schwarzengrund |
| Braenderup  | Kiambu   | Senftenberg |
| Cerro       | Kottbus  | Stanley |
| Corvallis   | Mbandaka | Stanleyville |
| Derby       | Mississippi | Tennessee |
| Dublin      | Montevideo | Thompson |
| Enteriditis | Muenchen | Typhi |
| Gallinarum  | Newport  | Typhimurium |
| Give        | Oranienburg | Virchow |
| Hadar       | Orion var. Binza | |
| Heidelberg  | Paratyphi A | |
SGSA Specifications

98 oligonucleotide probes

- Somatic targets
- Flagellar targets
- Flagellar complexes
- Serovar specific targets
- Caspular (Vi)
- *Salmonella* specific

All probes printed in triplicate on the array
<table>
<thead>
<tr>
<th>Somatic Group</th>
<th>Flagellar Phase 1</th>
<th>Flagellar Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A and D&lt;sub&gt;1&lt;/sub&gt;/D&lt;sub&gt;2&lt;/sub&gt;</td>
<td>a, i</td>
<td>1,2</td>
</tr>
<tr>
<td>B</td>
<td>b, k</td>
<td>1,2,7</td>
</tr>
<tr>
<td>C1</td>
<td>c, l,v and l,z&lt;sub&gt;13&lt;/sub&gt;</td>
<td>1,5</td>
</tr>
<tr>
<td>C2</td>
<td>d, l,z&lt;sub&gt;13&lt;/sub&gt;,z&lt;sub&gt;28&lt;/sub&gt;</td>
<td>1,5,7</td>
</tr>
<tr>
<td>E&lt;sub&gt;1&lt;/sub&gt;/E&lt;sub&gt;4&lt;/sub&gt;</td>
<td>e,h, l,z&lt;sub&gt;28&lt;/sub&gt;</td>
<td>1,6</td>
</tr>
<tr>
<td>G</td>
<td>f,g, m,t</td>
<td>1,7</td>
</tr>
<tr>
<td>H</td>
<td>f,g,s, m,t,p,u</td>
<td>e,n,x, e,n,x,z&lt;sub&gt;15&lt;/sub&gt;</td>
</tr>
<tr>
<td>J</td>
<td>f,g,t, r</td>
<td>e,n,x,z&lt;sub&gt;15&lt;/sub&gt;</td>
</tr>
<tr>
<td>K</td>
<td>f,g,m,t, r,[i]</td>
<td>enz&lt;sub&gt;15&lt;/sub&gt;</td>
</tr>
<tr>
<td>L</td>
<td>g,m,p,s, Y</td>
<td>z&lt;sub&gt;6&lt;/sub&gt; and z&lt;sub&gt;67&lt;/sub&gt;</td>
</tr>
<tr>
<td>M</td>
<td>g,m,s</td>
<td>Z</td>
</tr>
<tr>
<td>O</td>
<td>g,m,t</td>
<td>Z&lt;sub&gt;10&lt;/sub&gt;</td>
</tr>
<tr>
<td>P</td>
<td>g,p</td>
<td>Z&lt;sub&gt;29&lt;/sub&gt;</td>
</tr>
<tr>
<td>V</td>
<td>g,p,s</td>
<td>Z&lt;sub&gt;38&lt;/sub&gt;</td>
</tr>
<tr>
<td>Y</td>
<td>g,q and g,m,q</td>
<td>z&lt;sub&gt;4&lt;/sub&gt;,z&lt;sub&gt;23&lt;/sub&gt;</td>
</tr>
<tr>
<td>O:58</td>
<td>g,t and g,s,t</td>
<td>z&lt;sub&gt;4&lt;/sub&gt;,z&lt;sub&gt;24&lt;/sub&gt;</td>
</tr>
<tr>
<td>O:61</td>
<td>g,z&lt;sub&gt;51&lt;/sub&gt;</td>
<td></td>
</tr>
</tbody>
</table>

**Antigens Detectable on the SGSA**
SGSA Typing Capabilities

- **20/46** somatic serogroups
- **47/114** flagellar antigens
- In combination, probes have the potential to identify 1026 / 1532 ssp I serotypes 1358 of all *Salmonella* serotypes!

**>98% of annual Canadian isolates**
SGSA – Molecular Typing Method

**Overnight**

- Single Colony Isolation
- gDNA

**Morning**

- Multiplex PCR
- SSELO Labelling

**Afternoon**

- Hybridization
- Detection
- Data Analysis

---

Public Health Agency of Canada | Agence de la santé publique du Canada
S. Typhimurium  B: i :1,2
<table>
<thead>
<tr>
<th>Somatic Group</th>
<th>Flagellar Phase 1</th>
<th>Flagellar Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A and D&lt;sub&gt;1&lt;/sub&gt;/D&lt;sub&gt;2&lt;/sub&gt;</td>
<td>a i</td>
<td>1,2</td>
</tr>
<tr>
<td>B</td>
<td>b k</td>
<td>1,2,7</td>
</tr>
<tr>
<td>C&lt;sub&gt;1&lt;/sub&gt;</td>
<td>c l,v and l,z&lt;sub&gt;13&lt;/sub&gt;</td>
<td>1,5</td>
</tr>
<tr>
<td>C&lt;sub&gt;2&lt;/sub&gt;</td>
<td>d l,z&lt;sub&gt;13&lt;/sub&gt;,z&lt;sub&gt;28&lt;/sub&gt;</td>
<td>1,5,7</td>
</tr>
<tr>
<td>E&lt;sub&gt;1&lt;/sub&gt;/E&lt;sub&gt;4&lt;/sub&gt;</td>
<td>e,h l,z&lt;sub&gt;28&lt;/sub&gt;</td>
<td>1,6</td>
</tr>
<tr>
<td>G</td>
<td>f,g m,t</td>
<td>1,7</td>
</tr>
<tr>
<td>H</td>
<td>f,g,s m,t,p,u e,n,x</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>f,g,t r e,n,x,z&lt;sub&gt;15&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>f,g,m,t r, [i] enz&lt;sub&gt;15&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>g,m,p,s y z&lt;sub&gt;6&lt;/sub&gt; and z&lt;sub&gt;67&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>g,m,s z</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>g,m,t z&lt;sub&gt;10&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>g,p z&lt;sub&gt;29&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>g,p,s z&lt;sub&gt;38&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>g,q and g,m,q z&lt;sub&gt;4&lt;/sub&gt;,z&lt;sub&gt;23&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>O:58</td>
<td>g,t and g,s,t z&lt;sub&gt;4&lt;/sub&gt;,z&lt;sub&gt;24&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>O:61</td>
<td>g,z&lt;sub&gt;51&lt;/sub&gt;</td>
<td></td>
</tr>
</tbody>
</table>

**EXAMPLES:**
- C<sub>1</sub>:k:<sub>1</sub>,<sub>5</sub> Thompson
- B:r:<sub>1</sub>,<sub>2</sub> Heidelberg
- D:g,p: - Dublin
Uses for the SGSA

• Alternative typing system for public and private laboratories and research institutions

• Rapid, high throughput analysis, ie. outbreaks

• Preliminary step to serotyping in reference laboratories

• Validation of difficult samples
Limitations of Molecular Typing

- Difficulty differentiating similar strains, and biovars from each other
- Not indicative of gene expression (genotype vs phenotype)
- Presence of prophages and plasmids can change the serotype
- Mutations or SNPs in probe specific regions can cause discrepant typing results
Plan:

1. Preliminary Validation
2. Validation
   - 3 locations (PHAC, AIT, AHVLA)
   - 3712 isolates
5. Accreditation
7. Commercialization
8. Implementation
Preliminary Validation - *Part 1*

- 66 probes validated against 82 serotypes (in triplicate)
- In total: 133 Serovars, 287 samples
- Identification of probe patterns for the design of an EXCEL Macro which will automate result analysis

**92% Specificity, 99% Sensitivity and 100% Repeatability**
Preliminary Validation - Part 2: Blind Panel

• 100% identified as *Salmonella*

• 92% antigenic alleles detected

• 76% of serotypes yielded complete antigenic formulas

Monophasic / Rough Samples

**Monophasic**
- 16/25 classified as diphasic
- 36% correlation

**Rough**
- 5/8 identified
- 38% correlation

- Sequenced antigens detected using SGSA (phenotypically absent)
- Antigens not detected by SGSA are due to mutations, deletions or insertions
Plan:

1. Preliminary Validation

2. Validation
   - 3 locations (PHAC, AIT, AHVLA)
   - 3712 isolates

5. Accreditation

7. Commercialization

8. Implementation
Validation

• **Sensitivity**
  – Testing all antigens on SGSA x3
  – Greater amount of testing for prevalent strains

• **Specificity**
  – Non-target *Salmonella* strains (>300 strains)
  – Parallel testing at 3 locations (PHAC, AIT, AHVLA)

• 3712 strains total
Plan:

1. Preliminary Validation

2. Validation – in progress...
   - 3 locations (PHAC, AIT, AHVLA)
   - 3712 isolates

5. Accreditation

7. Commercialization

8. Implementation
Whole Genome Sequence Analysis

• Additional coverage for SGSA
  – Rare strains from international sources
  – New targets: antigen/serovar specific

• Led to international collaborations:
  – Development of bioinformatics pipelines (LFZ +AHVLAL)
  – Comparative genomics (LFZ+Cornell+AHVLAL)
    • Enteritidis vs Nitra, colonial variants
Example: Serovar-Specific Probes

- S. Enteritidis vs S. Nitra
- Deletion of the pepT gene (found in S. Enteritidis, absent in S. Nitra)
- Confirmed with 60 S. Enteritidis strains and 2 S. Nitra strains
Whole Genome Sequencing Projects

**Additional O Serogroups**
- 27 strains - 454 sequencing
- 100% coverage!

**A/D Serogroups**
- 11 strains – 454 / Illumina sequencing
- Define differences b/w these very similar serogroups

**Additional H Antigens**
- 4 strains - 454 sequencing
- 45% coverage

- **38 Salmonella WGS in pipeline!**
- **Improve H Antigen coverage**
  - Targeted gene sequencing
Improvements and Advancements

- Expansion to incorporate remaining serogroups and flagellar antigens
- Serovar-specific probes
- Individual somatic factors
- Detection of vaccine strains, antimicrobial resistance markers, sub-typing schemes
Summary - SGSA

- Developed molecular typing tool for *Salmonella*
- SGSA has ability to detect over 1300 serovars
- SGSA reduces overall costs by 70% compared to traditional serotyping
- SGSA reduces identification time from 4 days to 1 day

We are hoping to commercialize in the near future!
Molecular Diagnostics Workshop: Linking Research to Practice

Date: February 28, 2012

Objectives

• Determine how best to optimize molecular platforms to meet the needs of all the stakeholders while simultaneously fostering new research partnerships.
• Identify the key drivers of molecular platforms.
• Examine major barriers with the potential of inhibiting the implementing of molecular platforms.
• Develop effective communication strategies for disseminating research results to the public health sector.
Linking Research to Practice

Survey

• To be administered prior to the workshop
• Used to gain information on the current state of affairs with regards to the use of molecular technologies
• Results will be used to tailor the workshop discussions

Contact: kristyn.franklin@phac-aspc.gc.ca
Acknowledgements

- John Nash, Andre Villegas – PHAC-LFZ
- Stephanie Murphy – PHAC-LFZ
- Salmonella Reference Laboratory – PHAC-LFZ
- Animal Health Veterinary Laboratory Agency – UK
- Austrian Institute of Technology – Austria
- Alere Technologies (formerly Clondiag Chip Technologies) – Germany

Thank you!