WGS and source tracking investigations: Adressing critical knowledge gaps for improved data interpretation

Leen Baert
Food Safety Microbiology, Nestlé Research
WGS: example how it can serve the food industry

- *Salmonella* detected in product.
- *Salmonella* detected during laboratory monitoring.
- Laboratory performed proficiency testing prior to this event.
- Serotyping indicated *S.* Hadar for E and product sample.
- PT strain was *S.* Hadar.
- Hypothesis of laboratory cross-contamination raised.

Product  =  E  =  PT

WGS workflow consists of several steps

Sampling & DNA extraction
- Microbiology laboratory with dedicated areas
- Trained technicians

Sequencing
- IT infrastructure
- Software tools
- Assessment of quality
- Bioinformatics expertise

Bioinformatics
- Molecular laboratory
- Temperature stable area
- Sequencing platform
- Robots if large sample throughput

Results

Interpretation:
- Genomic expertise
- Microbiology expertise
- Operational expertise

WGS workflow: which bioinformatics approach

Quality assessment is key to obtain reliable hqSNP analysis

Example of quality checks carried out within the bioinformatics analysis:

1. Data reception
   - Ensure data integrity

2. Raw data QC
   - Ensure sequence data reach the quality criteria

3. Isolate taxonomy identification
   - Ensure genus/species is the expected one

4. Read trimming, genome assembly
   - Verify genome assembly quality

5. Specific genes identification
   - Add information about isolate characteristics in the metadata

6. First grouping
   - Identify 1st groups of related genomes for SNP analysis

7. SNP detection
   - Resolve isolate genomes relatedness

---

Validation of all components of the WGS workflow is required.

Interpretation of WGS results for source tracking: persistence

Evolution of *Listeria monocytogenes* in a Food Processing Plant Involves Limited Single-Nucleotide Substitutions but Considerable Diversification by Gain and Loss of Prophages

Anna Sophia Harrand,¹ Balamurugan Jagadeesan,² Leen Baert,² Martin Wiedmann,² Renato H. Orsi²

¹Department of Food Science, Cornell University, Ithaca, New York, USA
²Nestlé Institute of Food Safety and Analytical Sciences, Nestlé Research, Vers-ches-les-Blanc, Lausanne, Switzerland

*L. monocytogenes* evolution in a cold-smoked salmon processing facility indicated:
- ~0.35 changes per genome per year.
- Rapid diversification of prophages.

→ Useful to interpret WGS results for a root cause analysis where isolates might be persistent.

Interpretation of WGS results for source tracking: stress

*S.* Agona and *S.* Mbandaka repeatedly exposed to heat (90°C for 5 min) in a low water activity and high fat matrix.

<table>
<thead>
<tr>
<th>Cycle</th>
<th><em>S.</em> Agona Allele/SNP</th>
<th><em>S.</em> Mbandaka Allele/SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9/6</td>
<td>2/6</td>
</tr>
<tr>
<td>2</td>
<td>11/12</td>
<td>2/4</td>
</tr>
<tr>
<td>3</td>
<td>12/16</td>
<td>4/5</td>
</tr>
<tr>
<td>4</td>
<td>14/17</td>
<td>5/8</td>
</tr>
<tr>
<td>5</td>
<td>22/23</td>
<td>4/9</td>
</tr>
<tr>
<td>6</td>
<td>20/23</td>
<td>4/10</td>
</tr>
<tr>
<td>7</td>
<td>19/25</td>
<td>8/19</td>
</tr>
<tr>
<td>8</td>
<td>21/27</td>
<td>8/11</td>
</tr>
<tr>
<td>9</td>
<td>26/32</td>
<td>8/11</td>
</tr>
<tr>
<td>10</td>
<td>28/38</td>
<td>8/10</td>
</tr>
</tbody>
</table>

- No increased fitness.
- The genetic changes increased with every heat treatment.
- Genetic changes appeared randomly.

→ Useful to interpret WGS results for a root cause analysis where isolates are exposed to stress conditions.
Interpretation of WGS results for source tracking: mobile elements


→ Median MGE count: 14

~2.5% of average S. enterica genome.
Interpretation of WGS results for source tracking: mobile elements

≥1 plasmid in 69% of *S. enterica* genomes (n=990).

Median length 94 kb.

Median plasmid count/genome: 2.

Interpretation of WGS results for source tracking: mobile elements

SNP analysis of a S. Typhimurium outbreak cluster of 17 isolates:

→ Plasmids as part of the reference for SNP analysis distort the inference of phylogeny and SNP distances among closely related isolates.

Interpretation of WGS results for source tracking: importance of metadata

→ Closely related *L. monocytogenes* strains may have no apparent common source.
→ Metadata is important for interpretation of WGS data.

Summary

WGS analysis can serve the food industry: decision tree when to use it.

Understanding of WGS workflow:
• Quality assessment
• Validation

Interpretation of WGS results:
• Evolution in processing environment
• Impact of mobile genetic elements
• Metadata

Thank you!

Evolution of Listeria monocytogenes in a Food Processing Plant Involves Limited Single-Nucleotide Substitutions but Considerable Diversification by Gain and Loss of Prophages

Anna Sophia Harrand, Balamurugan Jagadeesan, Leen Baert, Martin Wedmann, Renato H. Orsi

Implications of Mobile Genetic Elements for Salmonella enterica Single-Nucleotide Polymorphism Subtyping and Source Tracking Investigations

Shaojun Li, Shaokang Zhang, Leen Baert, Balamurugan Jagadeesan, Catherine Ngom-Bru, Taylor Griswold, Lee S. Katz, Heather A. Carleton, Xiaoyu Deng

Opinion

Whole Genome Sequencing Applied to Pathogen Source Tracking in Food Industry: Key Considerations for Robust Bioinformatics Data Analysis and Reliable Results Interpretation

Caroline Barretto, Cristian Rincón, Anne-Catherine Portmann

Whole genome sequencing used in an industrial context reveals a Salmonella laboratory cross-contamination

Katia Rousou-Seynash, Caroline Barretto, Corine Fournier, Deborah Moine, Johan Gimenez, Leen Baert

Identification of Closely Related Listeria monocytogenes Isolates with No Apparent Evidence for a Common Source or Location: A Retrospective Whole Genome Sequencing Analysis

Rennato H. Orsi, Balamurugan Jagadeesan, Leen Baert, and Martin Wedmann

Guidance document on the use of whole genome sequencing (WGS) for source tracking from a food industry perspective

Leen Baert, Peter McClarren, Aneri Winkler, James Karr, Maritijn Be, Adrienne Klijn

Thank you!