Overview of Molecular Subtyping of Methods for Bacterial Pathogens

Christine Keys

The U.S. Food & Drug Administration
Center for Food Safety and Applied Nutrition
Office of Regulatory Science
Acknowledgements

Dr. Rebecca Bell  (IBIS T5000)
Dr. Jie Zheng  (xMAP Technology)
Dr. John Callahan  (MALDI)
Alice E Hayford  (Pyrosequencing)

Dr. Eric Brown  Branch Chief, Molecular Methods and Subtyping Branch

Christine.Keys@FDA.HHS.GOV
Objectives

- Methods in Use at the FDA
- Future Methods – SNP-Based
- Mass Spectrometry-Based
How much Discrimination is the Right Amount of Discrimination?

It Depends on the Question You are Asking

All inclusive, but bacteria are not identified at all
How much Discrimination is the Right Amount of Discrimination?

It Depends on the Question You are Asking

All exclusive, all bacteria are different (whole genome sequencing)
How much Discrimination is the Right Amount of Discrimination?
Information vs. Throughput

Tests per Sample

- microarray
- multiplex RT-PCR
- northern blot

Sample throughput

- Mass Spec
- xMAP Technology
- Real-Time PCR
Objectives

• Methods in Use at the FDA

• Future Methods – SNP-Based

• Mass Spectrometry-Based
PFGE
Pulsed-Field Gel Electrophoresis

- Large DNA Fragments
- "Universal" Technique
- PulseNet Standardizations
- Usually Epidemiologically Relevant
The National Molecular Subtyping Network for Foodborne Disease Surveillance

[Map of the United States with states color-coded and laboratory locations marked]

West Mountain South Central North Central Midwest Mid-Atlantic Southeast Northeast

- USDA-Laboratories
- FDA-Laboratories
- Area Laboratories
- PulseNet Headquarters
The 2006 spinach outbreak had clear associations between human, food, and animal/environmental isolates, supported by PFGE.

Shortly after the spinach outbreak, there were two concurrent shredded lettuce outbreaks.
MLVA
Multi-Locus VNTR Analysis

VNTR
Variable-Number Tandem Repeat

(Variable copy Numbers of Tandem Repeats)
Tandem Repeat
A sequence that is made up of a tandemly repeated sequence motif (arranged head-to-tail without interruption)

4 bp MOTIF X 4 COPIES = 16 bp ARRAY

Tandem Repeats may also be called:
• Simple Sequences (SS)
• Short Tandem Repeats (STR)
• Microsatellites
Variable-Number Tandem Repeat

A tandem repeat that varies in the number of copies of the motif. Variation is caused when errors in copy number are made during replication.

TGATGC\textcolor{red}{\textbf{CATACATA}}CATA\textbf{CATAGGACTTTAGC}

(motif is lost)

TGATGC\textbf{CATA}CATACATA\textcolor{red}{\textbf{CATACATA}}\textbf{CATAGGACT}

(motif is gained)
When we combine the analysis of multiple VNTRs throughout the genome, we create a MLVA system.
Objectives

• Methods in Use at the FDA

• Future Methods – SNP-Based

• Mass Spectrometry-Based
SNP Discovery and Analysis

Xmap Technology

Pyrosequencing

IBIS T5000
SNP
Single Nucleotide Polymorphism

General Population Allele → TCACACTGGGATCA

Discriminating Allele → TCACACGTGGGATCA
Advantages to using xMAP technology

- Nearly solution phase - Fast, Reproducible

- Multiplex – up to 100 targets

- Once an assay is designed, simple to run

- Can use a variety of capture reagents: antibodies, peptides, or oligonucleotides
xMAP Technology
(Luminex/BioPlex)

100 xMAP microsphere sets

Infrared dye

red dye
Pyrosequencing

Pyrosequencing is a real-time DNA sequencing technique for rapid analysis of short sequences. Primarily used for SNP discovery and analysis, the technology quantitatively measures allele frequency in a heterogeneous population.

With selection of unique markers, at least 500 strains can be typed daily.

Advantages:
• High accuracy potential
• Ease of use
• Highly flexible
• Now emerging as a popular platform for microbial typing
Pyrosequencing™ is a rapid sequencing method that utilizes the pyrophosphate released upon nucleotide incorporation to generate ATP, which is used as a substrate for luciferase to emit a bioluminescent signal.
Highlights of the Ibis T5000

• Rapid identification (~ 5 hours)
• Broad identification of all microbes
  – Bacteria, Viruses, Fungi, Protozoa
  – No culturing
  – No DNA sequencing
• No need for a priori knowledge of the infectious organisms present in the sample or the suspected organisms’ DNA sequence
  – Emerging infectious disease
• Mixed populations of microbes
  – Quantitative
• High resolution genotyping, strain identification and antibiotic resistance determination
• Cost effective, rapid, high throughput
Ibis Process Part 1: Sample Prep and Broad Range PCR

**Microbe Mixture**

**Extract Nucleic Acids**

**Broad Range Primers**
- Primer Set 1
- Primer Set 2
- Primer Set 3

**PCR Amplification**

**Internal Calibrant**
- # of molecules known

**PCR Products**
- Calibrant
Ibis Process Part 2: MS Analysis and Signal Processing

**Mass Spectrometer**

**Spectral Signal**

**Signal Processing**
Masses to Base Compositions

<table>
<thead>
<tr>
<th>ORGANISM</th>
<th>Mass</th>
<th>Base Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus anthracis</td>
<td>35278.823</td>
<td>A26 G34 C27 T27</td>
</tr>
<tr>
<td>Borrelia burgdorferi</td>
<td>33770.606</td>
<td>A29 G31 C23 T26</td>
</tr>
<tr>
<td>Clostridium botulinum</td>
<td>35843.944</td>
<td>A29 G33 C30 T24</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>35641.855</td>
<td>A22 G39 C29 T25</td>
</tr>
<tr>
<td>Staph aureus</td>
<td>35240.807</td>
<td>A24 G35 C30 T25</td>
</tr>
<tr>
<td>Strep pneumoniae</td>
<td>35270.806</td>
<td>A24 G35 C28 T27</td>
</tr>
<tr>
<td>Strep pyogenes (Group A Strep)</td>
<td>35281.808</td>
<td>A23 G37 C30 T24</td>
</tr>
</tbody>
</table>
Ibis Process Part 3: Triangulation Using Multiple Primer Pairs

Primer #1

<table>
<thead>
<tr>
<th>Mass</th>
<th>Base Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>18234.970 A_{12}G_{17}C_{17}T_{13}</td>
</tr>
<tr>
<td>Blue</td>
<td>17948.926 A_{14}G_{12}C_{12}T_{18}</td>
</tr>
<tr>
<td>Blue</td>
<td>18610.017 A_{11}G_{15}C_{15}T_{15}</td>
</tr>
<tr>
<td>Blue</td>
<td>17936.912 A_{11}G_{17}C_{16}T_{14}</td>
</tr>
<tr>
<td>Blue</td>
<td>18877.118 A_{18}G_{15}C_{15}T_{13}</td>
</tr>
</tbody>
</table>

Bioinformatics!
Objectives

• Methods in Use at the FDA

• Future Methods – SNP-Based

• Mass Spectrometry-Based
MS based approaches

• High mass accuracy analysis of PCR amplicons

• Fatty acids (FAME), desorption electrospray of phospholipids and metabolites

• Peptide-based methods (bottom up): digest everything with enzymes, look for peptides

• Matrix-assisted Laser Desorption/Ionization (MALDI) of whole cell

• Intact protein LC/MS of protein extract
SAR A2 – *typhimurium* (LT2)

SAR A46 – *paratyphi B* (DMS 3254/7/81)
Advantages of MALDI

• Intact protein MS is an additional tool for differentiation of closely related *Salmonella* strains

• Provides identification of protein “targets” for detection and sub-typing of strains without genome sequencing

• Could be used in combination with “top-down methods” for rapid analysis of protein variants

• Method may be adapted to epidemiological and microbial forensics needs
Dr. Rebecca Bell  (IBIS T5000)
Dr. Jie Zheng  (xMAP Technology)
Dr. John Callahan  (MALDI)
Alice E Hayford  (Pyrosequencing)

Dr. Eric Brown  Branch Chief, Molecular Methods and Subtyping Branch

Christine.Keys@FDA.HHS.GOV
Working to Keep Food and Cosmetics Safe and Promote Good Nutrition