Optical DNA mapping for identification and characterization of bacteria
My Nyblom  
Chalmers University of Technology

Anna Johnning  
Fraunhofer-Chalmers Centre  
Chalmers University of Technology  
University of Gothenburg
Introduction
Bacterial identification - Motivation

- A variety of bacteria can be found in patients with bloodborne bacteria, skin infections, urinary tract infections – increasingly drug resistant
- Rapid on demand detection critical to:
  - Limit the spread of disease
  - Control outbreaks
  - Improve patient outcomes
- Identification of the infection cause is critical for control of antibiotic resistance rates
Aims

• Develop a optical DNA mapping-based method for bacterial identification
• Evaluate the performance of the method for clinically relevant species
  • Cultured samples
  • Clinical samples
  • Mixed samples
• Extend the method to characterize plasmids that encode antimicrobial resistance genes
Species identification
Methodology overview

One-step labeling

Confinement in nanofluidic channels

Automated analysis for Size + ID + Gene Detection
1. DNA extraction

• Extraction by plug lysis
• Protocol working for both G+ and G-
• Very gentle compared to other DNA extraction methods
  • Little fragmentation

• Very long DNA molecules extracted
2. Sequence-specific staining

DNA stained with two molecules

- **Netropsin**
  - AT-specific
  - Non-fluorescent
  - Minor groove-binding

- **YOYO**
  - Fluorescent
  - Bis-intercalating, inbetween the bases
2. Sequence-specific staining

DNA stained with two molecules

• Netropsin
  • AT-specific
  • Non-fluorescent
  • Minor groove-binding

• YOYO
  • Fluorescent
  • Bis-intercalating, inbetween the bases

Intensity variation reflecting the underlying sequence
3. Nanofluidics

- Microscopy
  - Fluorescence - **YOYO**
- Nanofluidic chip
  - 150 nm x 100 nm
- Typically 100 frames of each DNA molecule
3. Nanofluidics

- Microscopy
  - Fluorescence - YOYO
- Nanofluidic chip
  - 150 nm x 100 nm
- Typically 100 frames of each DNA molecule
3. Nanofluidics

- Microscopy
  - Fluorescence - YOYO
- Nanofluidic chip
  - 150 nm x 100 nm
- Typically 100 frames of each DNA molecule
4. Obtaining the intensity profile
5. Profile alignment

>Ref_1
AGTCATCGATGCTA
GATCGATCGATCGA

>Ref_2
CTAGCTAGCCATGC
TACGGTTTGTGTAC

<table>
<thead>
<tr>
<th>seqID</th>
<th>score</th>
<th>stretch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref_1</td>
<td>0.89</td>
<td>1.01</td>
</tr>
<tr>
<td>Ref_2</td>
<td>0.52</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Score
Pearson correlation coefficient (CC)
5. Profile alignment: example

![Graph showing profile alignment examples](image-url)
5. Profile alignment: Reference database

- All complete bacterial genomes in RefSeq
- Excluded sequences
  - <500 kb
  - Annotated as “plasmid”
- Total of 10,527 sequences
  2,455 different species

Why all of RefSeq?
- High-quality genome for almost all pathogens and commensals
- Broad applications
- Lower risk for false positives
  - Computationally heavy profile alignment
  - “Even RefSeq has its spots”
5. Profile alignment: Blacklist

- Alignments against blacklisted reference sequences are ignored in the analysis
- Easy to update
- Increased flexibility

Examples:
- Non-complete assemblies
- *Candidatus* genomes
6. Match filtering

- Input: Table of results from profile alignment against the reference database

  for each profile:
  - if profile is too short: discard
  - else:
    - identify high quality matches
    - if discriminative match to a species:
      - if best match has ok quality: output
      - else: discard
  - Output: Species distribution of discriminative profiles

<table>
<thead>
<tr>
<th>Profile</th>
<th>Reference</th>
<th>Score</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td><em>E. coli</em> strain A</td>
<td>0.90</td>
<td>615</td>
</tr>
<tr>
<td>P1</td>
<td><em>E. coli</em> strain B</td>
<td>0.88</td>
<td>611</td>
</tr>
<tr>
<td>P1</td>
<td><em>E. coli</em> strain C</td>
<td>0.86</td>
<td>611</td>
</tr>
<tr>
<td>P1</td>
<td><em>S. enterica</em> strain A</td>
<td>0.70</td>
<td>625</td>
</tr>
<tr>
<td>P2</td>
<td><em>E. coli</em> strain C</td>
<td>0.70</td>
<td>455</td>
</tr>
<tr>
<td>P2</td>
<td><em>S. enterica</em> strain A</td>
<td>0.69</td>
<td>450</td>
</tr>
<tr>
<td>P2</td>
<td><em>E. cloacae</em> strain A</td>
<td>0.60</td>
<td>461</td>
</tr>
<tr>
<td>P2</td>
<td><em>E. cloacae</em> strain B</td>
<td>0.58</td>
<td>461</td>
</tr>
<tr>
<td>P3</td>
<td><em>E. coli</em> strain C</td>
<td>0.30</td>
<td>501</td>
</tr>
<tr>
<td>P3</td>
<td><em>E. coli</em> strain A</td>
<td>0.29</td>
<td>489</td>
</tr>
<tr>
<td>P4</td>
<td><em>S. aureus</em> strain A</td>
<td>0.99</td>
<td>156</td>
</tr>
</tbody>
</table>
6. Match filtering: Parameters

1. Minimum profile length (Len_min: 250px)

2. Maximal score range between best and worst high quality matches (C_diff: 0.05)

3. Minimum acceptable score (C_thres: 0.5)

<table>
<thead>
<tr>
<th>Profile</th>
<th>Reference</th>
<th>Score</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td><em>E. coli</em> strain A</td>
<td>0.90</td>
<td>615</td>
</tr>
<tr>
<td>P1</td>
<td><em>E. coli</em> strain B</td>
<td>0.88</td>
<td>611</td>
</tr>
<tr>
<td>P1</td>
<td><em>E. coli</em> strain C</td>
<td>0.83</td>
<td>611</td>
</tr>
<tr>
<td>P1</td>
<td><em>S. enterica</em> strain A</td>
<td>0.70</td>
<td>625</td>
</tr>
<tr>
<td>P2</td>
<td><em>E. coli</em> strain C</td>
<td>0.70</td>
<td>455</td>
</tr>
<tr>
<td>P2</td>
<td><em>S. enterica</em> strain A</td>
<td>0.69</td>
<td>450</td>
</tr>
<tr>
<td>P2</td>
<td><em>E. cloacae</em> strain A</td>
<td>0.60</td>
<td>461</td>
</tr>
<tr>
<td>P2</td>
<td><em>E. cloacae</em> strain B</td>
<td>0.58</td>
<td>461</td>
</tr>
<tr>
<td>P3</td>
<td><em>E. coli</em> strain C</td>
<td>0.30</td>
<td>501</td>
</tr>
<tr>
<td>P3</td>
<td><em>E. coli</em> strain A</td>
<td>0.29</td>
<td>489</td>
</tr>
<tr>
<td>P4</td>
<td><em>S. aureus</em> strain A</td>
<td>0.99</td>
<td>156</td>
</tr>
</tbody>
</table>
Parameter evaluation
Varying $C_{\text{thres}}$ & $C_{\text{diff}}$
Fixed $C_{\text{thres}} = 0.5$
Fixed $C_{\text{thres}} = 0.5$ & $C_{\text{diff}} = 0.05$
Species identification results

$\text{Len}_{\text{min}} = 250 \text{ px}$

$C_{\text{thres}} = 0.5$

$C_{\text{diff}} = 0.05$
Pure cultures

- K. pneumoniae: 11
- P. aeruginosa: 7
- P. mirabilis: 10
- S. aureus: 8
- S. saprophyticus: 1
- Any other species: 6
Mixed cultures

- E. coli
- K. pneumoniae
- P. aeruginosa
- P. mirabilis
- S. aureus
- S. saprophyticus
- Any other species
Clinical urine samples

- K. pneumoniae: 13
- P. aeruginosa: 9
- P. mirabilis: 11

Infectious Diseases

Cultivation-Free Typing of Bacteria Using Optical DNA Mapping

Vilhelm Müller, My Nyblom, Anna Johnning, Marie Wrande, Albertas Drimnas, Sriram KK, Christian G. Giske, Tobias Ambjörnsson, Linus Sandgren, Erik Kristiansson, and Fredrik Westerlund

Cite This: ACS Infect. Dis. 2020, 6, 1076-1084
Sub-species identification
Existing schemas

• Phylogroups
• MLST
• More...
Sub-species groups

1. Identify core genome
2. Build phylogenetic tree
3. Divide reference genomes into sub-species groups (SSGs)
4. Test different sub-species taxonomic resolutions
Sub-species groups:

*Escherichia coli*
Parameter evaluation

Fixed $C_{\text{thres}} = 0.5$ & $L_{\text{min}} = 250$
Results: Cultured isolated
Results: Mixes & clinical urine samples

Mix 1

- ST131
- ST156
- ST355
- ST405
- 15
- 12
- 10?
- 13

Mix 2

- ST12
- ST131
- ST410
- ST448
- 19
- 18
- 7
- 21

TPR (%)

# disc fragments

Subspecies resolution: High, Medium, Low

ST

69 73 131 636
Plasmid identification
Simultaneous plasmid identification

CRISPR-Cas9 mediated AMR-gene cut followed by DNA-labelling with YOYO and netropsin

Optical DNA-mapping in nanofluidic device
Simultaneous plasmid identification
Plasmid results

**CTX-M-14 and CTX-M-15 directed Cas9 cuts**

- **P1**: Aligned 183 px plasmid with Cas9 cut
- **P2**: Aligned 208 px plasmid with Cas9 cut
- **P3**: Aligned 123 px plasmid with Cas9 cut

![Graph showing subspecies resolution](image)

- **TPR (%):**
- **Subspecies resolution:**
  - High
  - Medium
  - Low

<table>
<thead>
<tr>
<th>Samples</th>
<th>TPR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>1193</td>
</tr>
<tr>
<td>P1</td>
<td>131</td>
</tr>
<tr>
<td>P2</td>
<td>38</td>
</tr>
<tr>
<td>P3</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

• Optical DNA mapping can be used for species and subspecies identification of clinically relevant species

• The method can be used directly on clinical samples without prior culturing

• The method can detect multiple isolates in complex mixtures

• Simultaneous detection of antimicrobial resistance plasmids possible
Acknowledgments

• Fredrik Westerlund
  • Karolin Frykholm
  • Vilhelm Müller
  • Gaurav Goyal
  • Miriam Hitz
  • Sriram KK
• Erik Kristiansson

• Linus Sandegren
  • Marie Wrange
• Tobias Ambjörnsson
  • Albertas Dvirnas
• Christian Giske
  • Tseguye Sewunet