Clinical metagenomics in urinary tract infections

Frank M. Aarestrup
DTU – Food
www.genomicepidemiology.org
www.compare-europe.eu
Metagenomic sequencing directly on the samples can reduce time and improve result.

Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples

Henrik Hasman, Dhany Saputra, Thomas Slicheritz-Ponten, Ole Lund, Christina Aaby Svendsen, Niels Frimodt-Møller, Frank M. Aarestrup

National Food Institute, Technical University of Denmark, Lyngby, Denmark; Systems Biology, Technical University of Denmark, Lyngby, Denmark; Hvidovre Hospital, Hvidovre, Denmark

Whole-genome sequencing (WGS) is becoming available as a routine tool for clinical microbiology. If applied directly on clinical samples, this could further reduce diagnostic times and thereby improve control and treatment. A major bottleneck is the availability of fast and reliable bioinformatic tools. This study was conducted to evaluate the applicability of WGS directly on clinical samples and to develop easy-to-use bioinformatic tools for the analysis of sequencing data. Thirty-five random urine samples from patients with suspected urinary tract infections were examined using conventional microbiology, WGS of isolated bacteria,
The study

- 35 random urinary samples

- Routine culturing
  - All isolated bacteria WGS

- Direct sequencing
  - Reads mapped to databases
Results - ID

• Cultures from 19 (pure from 17)
  – WGS improved identification

• Metagenomics from 23
  – Four culture negative (G. vaginalis, L. iners, Prevotella, E. coli/E. faecalis)
  – Two mixed cultures (E. coli, E. coli/E. faecalis)
Online bioinformatic tool

Assemble pipeline

Resfinder

Resistance gene profile

Center for Genomic Epidemiology

ResFinder usage, August 2012 - December 2015

Total use: 78500 runs

Distribution among countries 2015

49 different countries in total

United States, 18%

Netherlands, 12%

United Kingdom, 10%

Other, 9%

Austria, 1%
### Phenotypic Resistance

<table>
<thead>
<tr>
<th></th>
<th>Resistant</th>
<th>Susceptible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted resistant</td>
<td>475</td>
<td>7</td>
</tr>
<tr>
<td>Predicted susceptible</td>
<td>16</td>
<td>2553</td>
</tr>
</tbody>
</table>

99.2% concordance

---

### Retest

<table>
<thead>
<tr>
<th></th>
<th>Resistant</th>
<th>Susceptible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistant</td>
<td>475</td>
<td>7</td>
</tr>
<tr>
<td>Susceptible</td>
<td>0</td>
<td>2569</td>
</tr>
</tbody>
</table>

99.8% concordance

Spectinomycin in *E. coli*
Number of reads mapping to tetracycline resistant and tetracycline susceptible Enterococci and E. coli
Agreement to resistance

<table>
<thead>
<tr>
<th>Number</th>
<th>Species</th>
<th>Phenotypic resistance</th>
<th>WGS predicted</th>
<th>Metagenomic predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>E. coli</td>
<td>S</td>
<td>S</td>
<td>ESBL</td>
</tr>
<tr>
<td>21</td>
<td>E. coli</td>
<td>AMP, CIP, GEN, NAL</td>
<td>AMP, GEN, TET</td>
<td>AMP, GEN, TET</td>
</tr>
<tr>
<td>27</td>
<td>E. coli</td>
<td>S</td>
<td>S</td>
<td>TET</td>
</tr>
</tbody>
</table>
Improving the methods
Bias the data

Estimate the depth of the sample

Error k-mers

Set threshold according to estimated depth
143 isolates from Oxford University Hospital, comprising 858 phenotypic susceptibility tests, most on beta-lactams.

193 isolates from Danish pig farms, comprising 2,547 phenotypic susceptibility tests, covering a broad spectrum of antibiotic classes.
Sequence quality and resistance determination

<table>
<thead>
<tr>
<th></th>
<th>SRST2</th>
<th>ResFinder</th>
<th>Kmer</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE</td>
<td>95</td>
<td>94</td>
<td>97</td>
</tr>
<tr>
<td>SP</td>
<td>96</td>
<td>96</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Down sampled</strong></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>29</td>
<td>17</td>
<td>95</td>
</tr>
<tr>
<td>SP</td>
<td>98</td>
<td>98</td>
<td>97</td>
</tr>
</tbody>
</table>
Metagenomic assignment
metamerfinder

\[ P(T < t_i) = 1 - e^{-\lambda_i t_i} \quad , \quad \lambda = \frac{1}{\mu_i} \quad \land \quad i \in (Depth; Coverage) \]

Equation 2; The exponential survival function, \( \mu \): background expectation, \( t \): measured quality
## Agreement to resistance

<table>
<thead>
<tr>
<th>No</th>
<th>Species</th>
<th>Phenotypic resistance</th>
<th>WGS predicted</th>
<th>Metagenomic predicted</th>
<th>Metamerfinder</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>E. faecalis</td>
<td>TET</td>
<td>TET</td>
<td>TET</td>
<td>E. faecalis (TET) E. coli (S) L. crispatus (ERY)</td>
</tr>
<tr>
<td>10</td>
<td>E. coli</td>
<td>S</td>
<td>S</td>
<td>ESBL</td>
<td>ESBL</td>
</tr>
<tr>
<td>21</td>
<td>E. coli</td>
<td>AMP, CIP, GEN, NAL</td>
<td>AMP, GEN, TET</td>
<td>AMP, GEN, TET</td>
<td>AMP, GEN, TET</td>
</tr>
<tr>
<td>27</td>
<td>E. coli</td>
<td>S</td>
<td>S</td>
<td>TET</td>
<td>TET, SUL, AMP, STR, TMP</td>
</tr>
</tbody>
</table>
E. coli in Urine samples

A: ATCC 8739 reference

SNP tree

_d = Direct sequencing on urine
_i = sequencing of isolate from urine
Why a central public repository?

*Besides the language and altruistic issues*

- The data comparison problem
- Allowing easy transfer between levels of access including public
- Allow access to bioinformatics for the frontline and frontline data for bioinformaticians
- Allowing for constantly improving the analytic pipelines
Establishing and improving surveillance

• Rapid sharing
  – Online bioinformatic tools
    • >1,000 jobs per day
  – Facilities for rapid sharing

• Natural reservoirs
  – Major issues with individual samples

The Rio Convention
The Nagoya Protocol
Copenhagen according to sewage

Very preliminary data

Drug use the previous year
Our vision: one system serves all

Guiding principles:
- Cross sector, cross domain, open source (not commercial)
- Interaction with the rest of the world (all inclusive)
- Data for action (actionable outputs)
- Central repository (ENA, DDJ, NCBI) (bring the tools to the data)

There can be no real-time detection & surveillance without real-time data sharing