Introduction to Whole Genome Sequencing & Food Safety Applications

Jørgen Schlundt

Michael Fam Chair Professor
Director NTU Food Technology Centre
Nanyang Technological University

New Opportunities?

- History

1953
1970
1977
1990
2003
2004

454 starts the next generation sequencing era

- History

2004

Next Generation Sequencing

- 454 Life Sciences: Parallelized pyrosequencing

Reduced costs 6 fold at the time – now much cheaper
Paradigm Shift  
from Pasteur and Koch to Watson and Crick

“It is likely that in 5 to 10 years all clinical microbiological laboratories will have a DNA sequencer.. complete bacterial genome might be less than 50 EURO”

The capacity to exchange – and manage - large data quantities over web-based systems has likewise increased dramatically over recent years

Enabling the potential creation of global databases of DNA-codes of all relevant microbiological strains”

Statement, int. expert meet on microbiol. genomic id. systems, Bruxelles: GMI1 (2011)

Faster - Better - Cheaper

First bacterial genome:

Haemophilus influenzae  
- published 1995  
- > one year process, > 1 million US$ price-tag

Today (with just one sequencer)

- 1 or a few genomes can be generated in 1 day  
- 800-3200 genomes can be generated in 2-3 weeks for < $50 each

The $10 microorganism genome will soon be a reality

(Randall J. Olsen, Director Molecular Diagnostics Laboratory, Texas, USA)  
American Society for Microbiology, 1st meeting on NGS (Sep. 2015)

Passing Past Pasteur

- Old school
  – One to several weeks to perform full typing  
  – Very different typing systems for microorganisms  
  – Very specialized epi. for different microorganisms

- New school
  – DNA sequence hold all information of a microorganism  
  – Hours (12-24) to perform sequencing + typing  
  – One test fits all (virus, bacteria, fungi, parasites)  
  – Same-Same microbiology (human, animal, environ)  
  – Same-Same epidemiology
Moving faster
– reaching further
• Longer reads – faster – simpler
“In a few years’ time, people who may be several steps removed from basic genomic research, like teachers in a classroom, could be using this device to teach science in new, exciting ways….”
Camilla Ip, Oxford University (EBI Press Release, Oct. 2015)

• Faster bioinformatics – big data solutions
“….mobile sequencing with real-time data streaming… with a high-speed Internet connection, the first dataset could arrive 20 minutes after the DNA is loaded,”
Guy Cochran, EBI (European Bioinformatics Institute)

Whole Genome Sequencing
advantages:
Diagnosis of microorganisms will improve
Outbreak investigation will improve
Surveillance and Prevention will improve
Microbiological Research will improve

One Example:
Source Attribution
Before WGS: Salmonella only
After WGS: all Microorganisms (potentially)

How can typing link lab id. and epidemiology:
Linking disease and food:
Definition: Source attribution - partitioning human disease burden to specific sources (Pires et al., 2009).
Integrated food safety approach

Burden of illness/disease
Total cases / DALYs

Source attribution
Most important sources

Analytical epidemiological
Most important risk factors

Risk Assessment
Identify effective interventions

Risk factors
- Animal feed
- Breeding herds
- Production herds
- Slaughterhouses
- Processing
- Retail
- Consumer

Food safety continuum

Example: Salmonella source attribution

Salmonella attribution

Example: Salmonella attribution

Source attribution to guide policy intervention
Salmonellosis in Denmark, 1988-2013

Source: Danish Zoonoses Centre, National Food Institute

Humans cases

Source: Danish Zoonoses Centre, National Food Institute

Arrows indicate implementation of sector-specific intervention
Attrition of *Salmonella* to sources, Latin America, 1990 and 2000

![Bar graph showing attribution of *Salmonella* to sources in Latin America, 1990 and 2000.](image)

What causes the problems?

**REVIEW OF RISK FACTOR STUDIES**

**SOURCE ATTRIBUTION STUDIES**

- What are the risk factors
- What are the sources
- What is the absolute and/or relative importance of sources/risk factors?


Sources of human salmonellosis

**REVIEW OF RISK FACTOR STUDIES**

**SOURCE ATTRIBUTION STUDIES**

- Pork (N=411)
- Beef (N=37)
- Eggs (N=1,474)
- Broilers (N=137)
- Turkeys (N=9)
- Ducks (N=17)
- Im. pork (N=38)
- Im. beef (N=32)
- Im. duck (N=31)
- Im. turkey (N=50)
- Im. chicken (N=221)
- Travel (N=752)
- Unknown (N=711)

What are the options for intervention?

**Farm-to-fork continuum**

- Animal feed
- Breeding herds
- Production herds
- Slaughterhouses
- Processing
- Retail
- Consumer

- A specific food-hazard combination is identified as a risk
- What can we do about it?
- Which interventions strategies should be selected?
  - Where in the food-production chain should they be implemented?
  - Where do we get most effect for least costs?

What are the options for intervention?

**RISK ASSESSMENTS**

**COST-EFFECTIVENESS /BENEFIT ANALYSES**

- Animal feed
- Breeding herds
- Slaughterhouses
- Processing
- Retail
- Consumer

- A specific food-hazard combination is identified as a risk
- Where in the food-production chain should they be implemented?
- Where do we get most effect for least costs?
New source attribution potential from WGS

Salmonella can be separated because of:
- Serotyping (> 2,500 serotypes)
- Phage typing
- Antimicrobial susceptibility testing

- So far genotyping has not been used much
- But the future potential of sequence-based typing through WGS is enormous!

Meaning that source attribution could be possible for all microorganisms in future

Global Microbial Identifier GMI

A global system enables three lines of action:
- Simple identification of all microorganisms through faster, cheaper, more correct characterization
- A DNA database of all microbiological strains globally, enabling real-time global (and national) surveillance of disease and resistance
- A DNA database enabling a giant resource for genomic knowledge of all microorganisms

FUTURE Global Microbial Identifier

1st global tool to identify Microorganisms and Resistance

Is something happening already?
All publicly available DNA sequence data deposited in the International Collaboration

America (NCBI)
Hosting Genbank

Europe (EBI)

Asia (DDBJ)

4 TeraBytes Inbound Daily
24 Hour Exchange
But why soon?

- Different researchers and different sectors are already starting to build separate databases with separate interphases and algorithms.
- If too many separate / different systems are built it will be increasingly difficult to agree to common format and common understanding.
- And we will – again – leave developing countries behind.

Moving on

- GMI 5, Copenhagen, Feb 2013
  Preparing a Road Map
- GMI 6, California, Sep 2013
  Agreeing a Charter
- GMI 7, York UK, Sep 2014
  Constructing an Organization
- GMI 8, Beijing, 11-13 May 2015
  Moving on - > 400 participants
- GMI 9, Rome, May 2016 – in FAO, Rome
  Including developing countries
- GMI 10, Mexico, May 2017
  Strengthening health involvement

NGS leap-frog potential in developing countries:

- New diagnostics avoids need for specialized training (testing for Salmonella and Norovirus the same).
- New diagnostic systems reaching further
  - with real-time characterization of microorganisms
  - in decentralized labs w sequencers + internet
- NGS enable uniform lab-based surveillance systems
- Same-same for humans & animals - One Health
Refusing global sharing – why?

• Concerns about national security and safety (including use of sequences for deliberate enhancement of pathogenicity)

• Outbreak situation, data not public - potential for prosecution

• Potential misuse by others, including IP-related risks (drugs, diagnostic methods, vaccines)

• How to share when you only have part of the data (food / human / animal / ?)

• Institutional and management barriers

• “My scientific publication more important than principle of sharing”

In my opinion:

In the future sharing of such data will occur, no matter what

The most damage (in relation to health, but also politically) is done when data is concealed as part of a cover-up

Experience shows that (even dangerous) data always gets out

There is so much to gain (in relation to health, food safety, but also to economy) if we share data

One for all
All for one

www.globalmicrobialidentifier.org