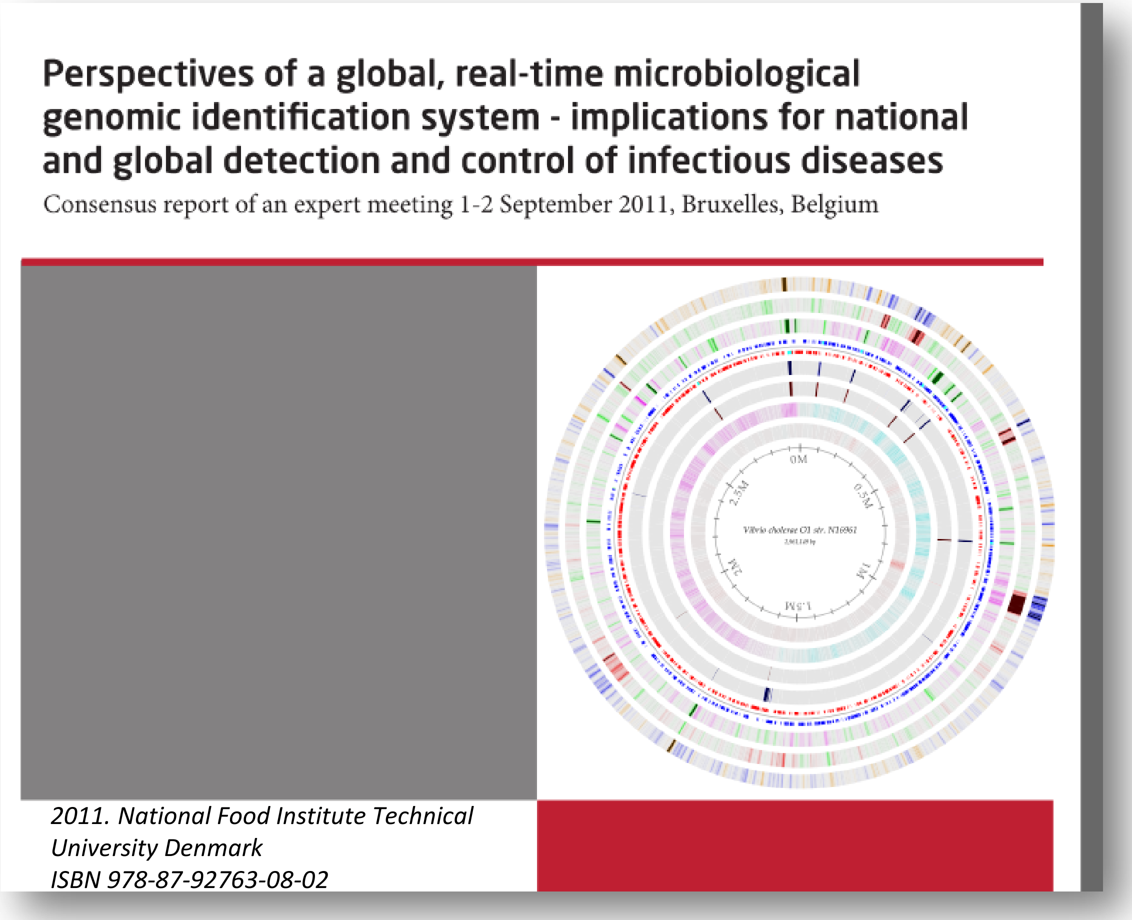
Working group 3
Analytical approaches

Marion Koopmans, Marc Allard



Perspectives of a global, real-time microbiological genomic identification system - implications for national and global detection and control of infectious diseases

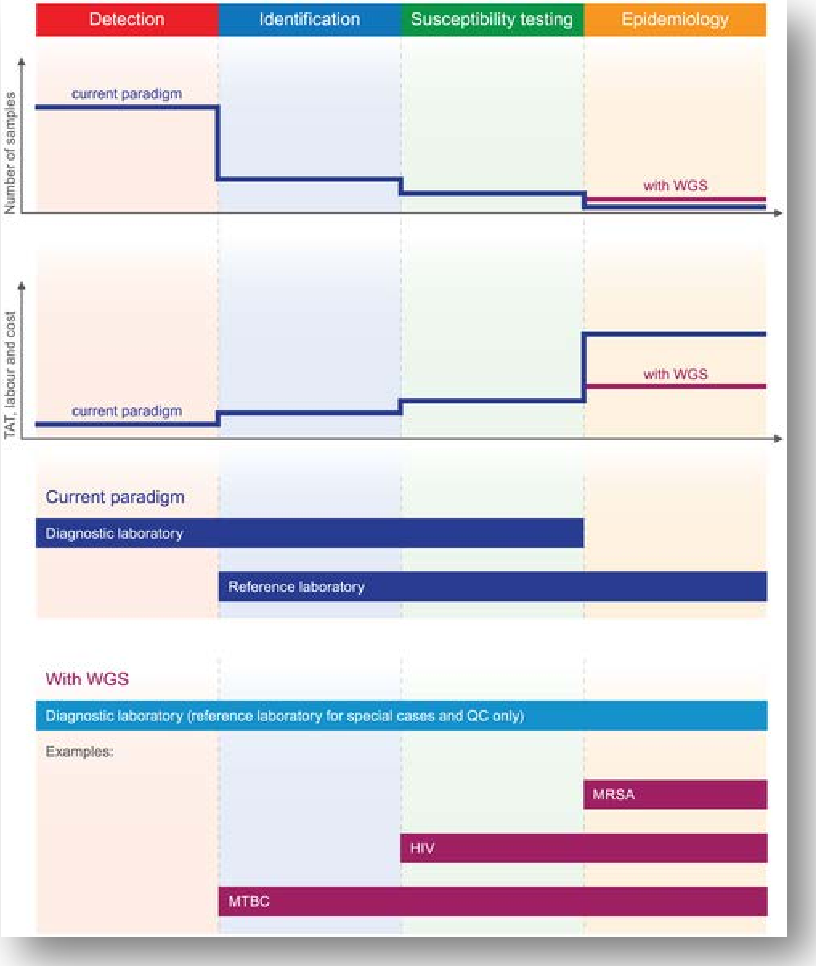
Consensus report of an expert meeting 1-2 September 2011, Bruxelles, Belgium

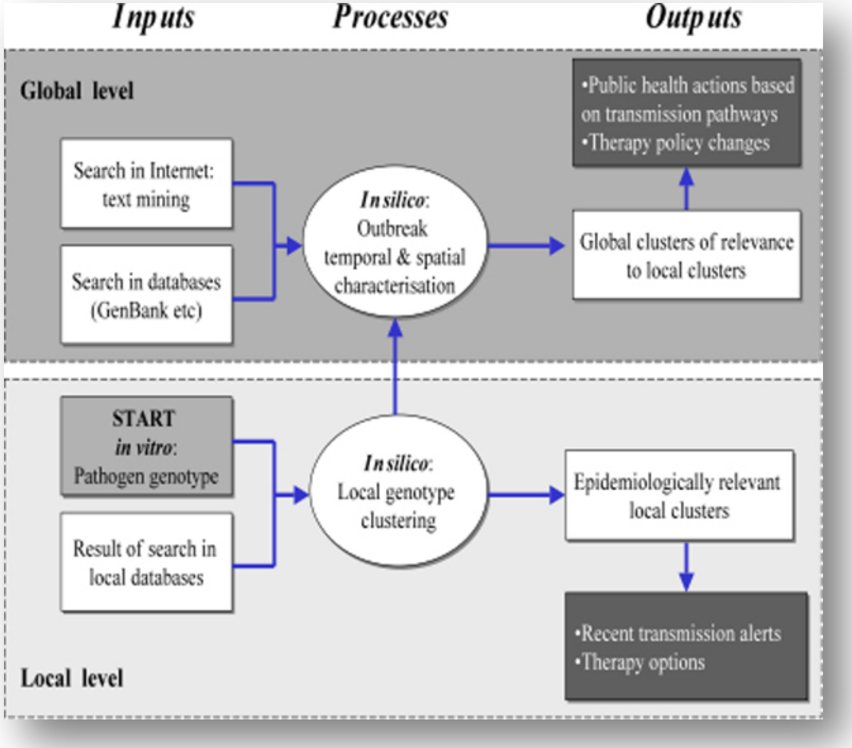
2011. National Food Institute Technical University Denmark
ISBN 978-87-92763-08-02

Consensus statement Brussels meeting

- A global system within a five to ten years horizon
- Merging access to laboratory data across traditional disciplines (‘virology, bacteriology, parasitology’ or ‘animal, food, human’)
- translational activities to provide information to stakeholders and end-users, including clear interpretation of all outcomes
- automated reports such as treatment guidelines, molecular typing information as well as spatial and temporal data
- A detailed road map for further next steps should be developed and agreed upon by all parties involved, from the scientific to the public health and political levels. The meeting organizers will work on developing a broad, inclusive list of participants in the road-map creation, along with “consultative” status entities
- Emphasis on free release of data
- It is also important that sufficient emphasis be placed on the ability to use information immediately in case of public health emergencies

Potential impact of whole genome sequencing in diagnostic and public health microbiology.





Köser ET AL., Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. PLoS Pathogens 2012
Sintchenko et al., Towards bioinformatics assisted infectious disease control. BMC bioinformatics 2009

Translational bioinformatics

- Dr. Elias Zerhouni (previous NIH director)….

…. to more effectively move the discoveries and innovations in the laboratory to the bedside, leading to improved diagnosis, prognosis, and treatment

Consensus statement Brussels meeting

- A global system within a five to ten years horizon
- Merging access to laboratory data across traditional disciplines (‘virology, bacteriology, parasitology’ or ‘animal, food, human’) **> sharing experience, joint workplan**
- translational activities to provide information to stakeholders and end-users, including clear interpretation of all outcomes **> involve end-users in clinic and public health**
- automated reports such as treatment guidelines, molecular typing information as well as spatial and temporal data **> involve end-users in clinic and public health**
- A detailed road map for further next steps should be developed and agreed upon by all parties involved, from the scientific to the public health and political levels. The meeting organizers will work on developing a broad, inclusive list of participants in the road-map creation, along with “consultative” status entities **> this meeting**
- Emphasis on free release of data **> how to get there realistically?**
- It is also important that sufficient emphasis be placed on the ability to use information immediately in case of public health emergencies **> robustness, validation**

Disease Outbreak Detection in the Genomics Era: a roadmap forward, and the road less travelled

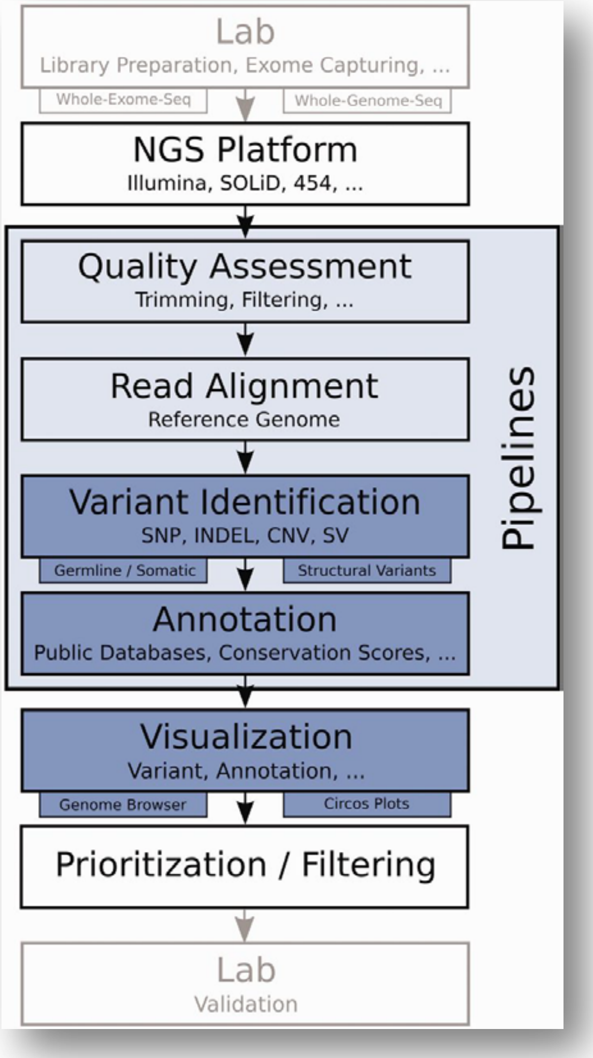
Washington, March and September 2012, hosted by FDA and NIH

- Appropriate metadata
- Computer resources required for global implementation
- Standards for optimal data analysis
- Political, legal, global health challenges
- Interpreting results and formatting of reports for point of care utility
- Contributions of industry

- ➢ WGS is already practised widely (US)
- ➢ Fast developing field, including informatics, ICT
- ➢ Good examples available
- ➢ Industry commitment

- ➢ Some reluctance in clinical and public health arena
- ➢ Mostly focussed on bacteriology
- ➢ Needs more engagement from other parts of the world

A survey of tools for variant analysis of next-generation genome sequencing data



11 packages

17 packages

29 packages

74 packages

40 packages

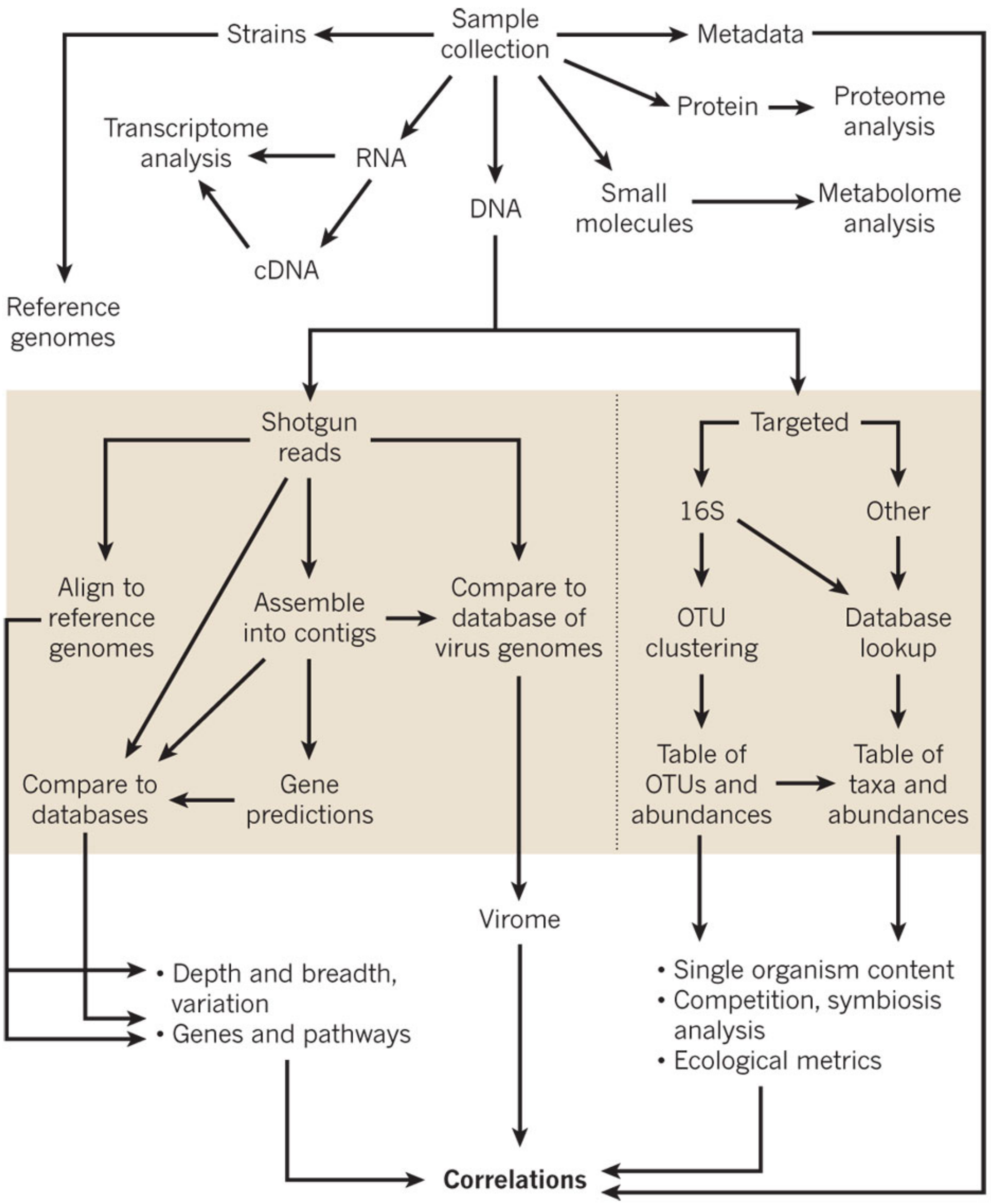
13 pipeline packages

How to choose?
Is it important to choose?
Who compares?
How to keep up to date?

Pabinger S et al. Brief Bioinform 2013

Characteristic	Bacteria	Viruses	Eukaryotic microbes
Genome size	0.5–10 megabases	1–1,000 kilobases	10–50 megabases
Number of taxa in the human microbiome	At least thousands	Unknown, but could be as many as bacteria	Unknown, but may be fewer than bacteria
Relative abundances	Highly variable	Highly variable	Unknown
Targeted **detection** methods	Sequencing of genes such as **5S and 16S rRNA**	**No universal method** for genes, but virus-specific polymerase chain reaction assays for some	Sequencing of **18S rRNA** gene Spacer region in rRNA
Shotgun approach to analyses	**Alignment to reference** genomes or database comparison	**Database comparison**	**Alignment to reference** genomes or database comparison
Subspecies or strain diversity	**Modest sequence variation** Horizontal gene transfer also contributes	**High sequence variation**	Unknown

Some universal needs, at analytical level also specific needs



Sample collection

Strains

Reference genomes

Metadata

Protein

Proteome analysis

RNA

Transcriptome analysis

cDNA

DNA

Small molecules

Metabolome analysis

Shotgun reads

Align to reference genomes

Assemble into contigs

Compare to database of virus genomes

Compare to databases

Gene predictions

Virome

- Depth and breadth, variation
- Genes and pathways

Targeted

16S

Other

OTU clustering

Database lookup

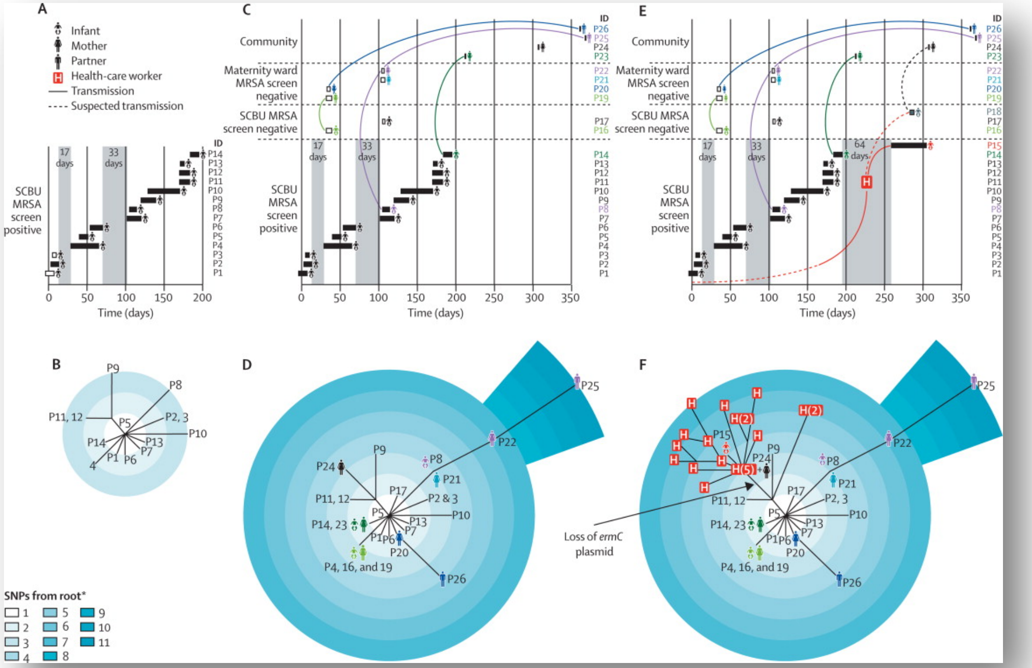
Table of OTUs and abundances

Table of taxa and abundances

- Single organism content
- Competition, symbiosis analysis
- Ecological metrics

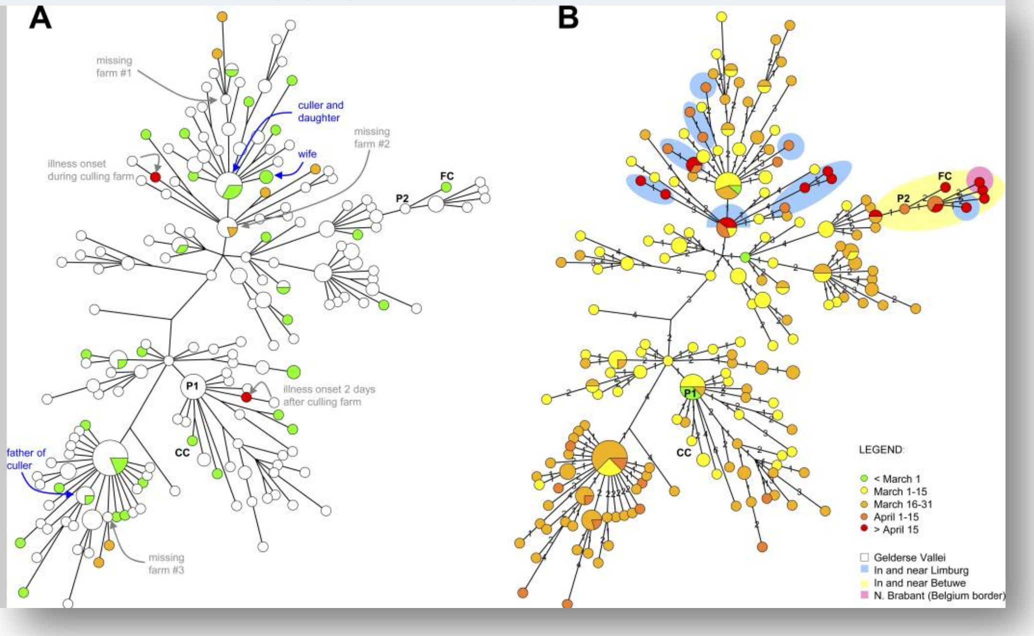
Correlations

Weinstock et al., 2012



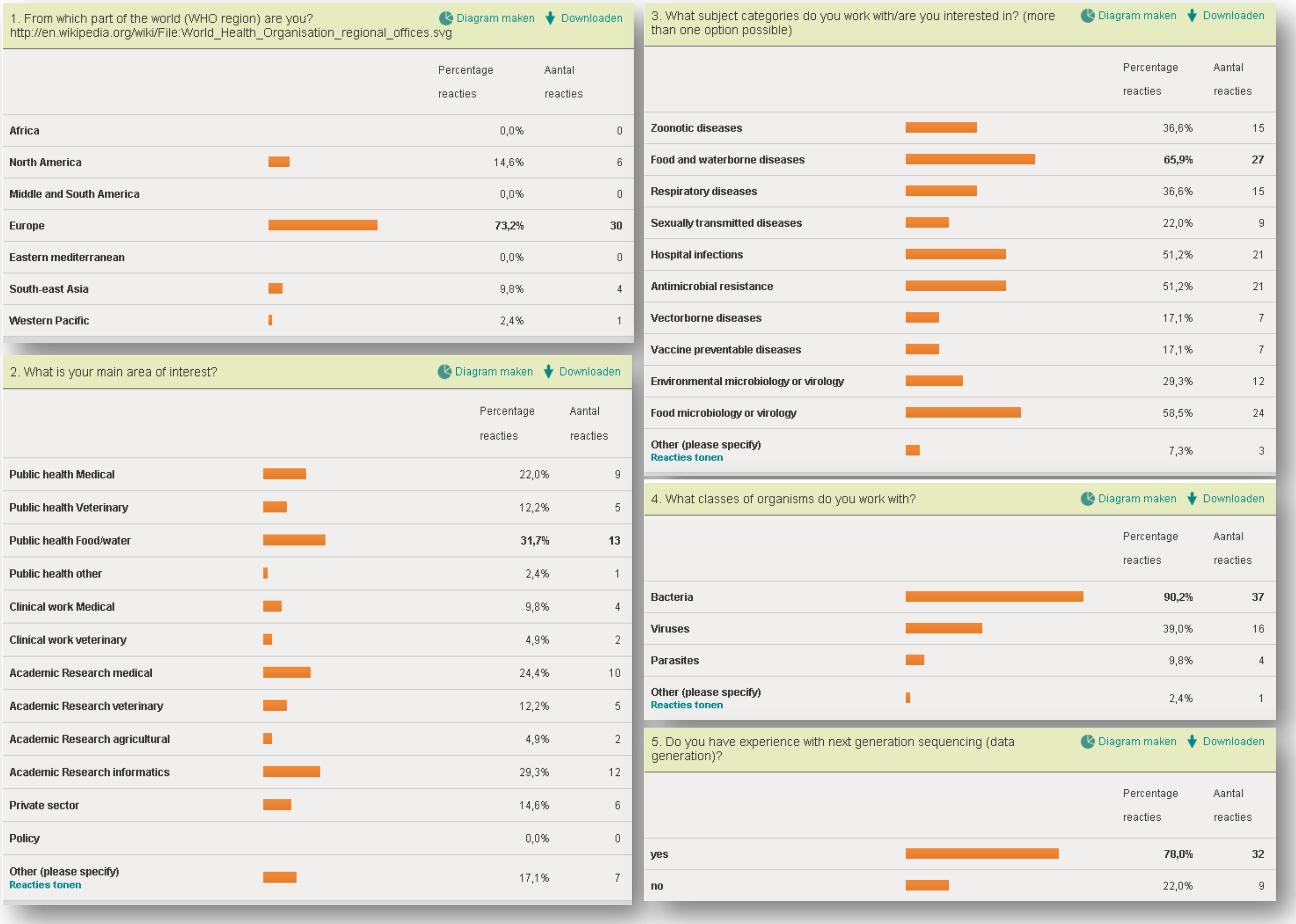
Analysis of an MRSA outbreak

Harris et al., 2013



Analysis of an avian influenza outbreak

Jonges et al., 2011



1. From which part of the world (WHO region) are you? http://en.wikipedia.org/wiki/File:World_Health_Organisation_regional_offices.svg

	Percentage reacties	Aantal reacties
Africa	0,0%	0
North America	14,6%	6
Middle and South America	0,0%	0
Europe	**73,2%**	**30**
Eastern mediterranean	0,0%	0
South-east Asia	9,8%	4
Western Pacific	2,4%	1

2. What is your main area of interest?

	Percentage reacties	Aantal reacties
Public health Medical	22,0%	9
Public health Veterinary	12,2%	5
Public health Food/water	**31,7%**	**13**
Public health other	2,4%	1
Clinical work Medical	9,8%	4
Clinical work veterinary	4,9%	2
Academic Research medical	24,4%	10
Academic Research veterinary	12,2%	5
Academic Research agricultural	4,9%	2
Academic Research informatics	29,3%	12
Private sector	14,6%	6
Policy	0,0%	0
Other (please specify) **Reacties tonen**	17,1%	7

3. What subject categories do you work with/are you interested in? (more than one option possible)

	Percentage reacties	Aantal reacties
Zoonotic diseases	36,6%	15
Food and waterborne diseases	**65,9%**	**27**
Respiratory diseases	36,6%	15
Sexually transmitted diseases	22,0%	9
Hospital infections	51,2%	21
Antimicrobial resistance	51,2%	21
Vectorborne diseases	17,1%	7
Vaccine preventable diseases	17,1%	7
Environmental microbiology or virology	29,3%	12
Food microbiology or virology	58,5%	24
Other (please specify) **Reacties tonen**	7,3%	3

4. What classes of organisms do you work with?

	Percentage reacties	Aantal reacties
Bacteria	**90,2%**	**37**
Viruses	39,0%	16
Parasites	9,8%	4
Other (please specify) **Reacties tonen**	2,4%	1

5. Do you have experience with next generation sequencing (data generation)?

	Percentage reacties	Aantal reacties
yes	**78,0%**	**32**
no	22,0%	9

Work Group Goal:

We will provide guidance for the development of analytical tools for optimal positioning and functioning of the GMI platform.

Suggested Work Group Themes:

End-user requirements: To define requirements for GMI functioning from the perspective of end-users (clinical, public health, research)

Applications: To define requirements for GMI functioning in terms of applications (identification, outbreak detection etc.)

Priority targets/diseases: To define requirements for GMI functioning in terms of priority targets / diseases

Analytical options: To map current analytical options and solutions against the needs of GMI end-users

R&D / implementation gaps: Possible R&D or implementation gaps, and projects that may address these gaps