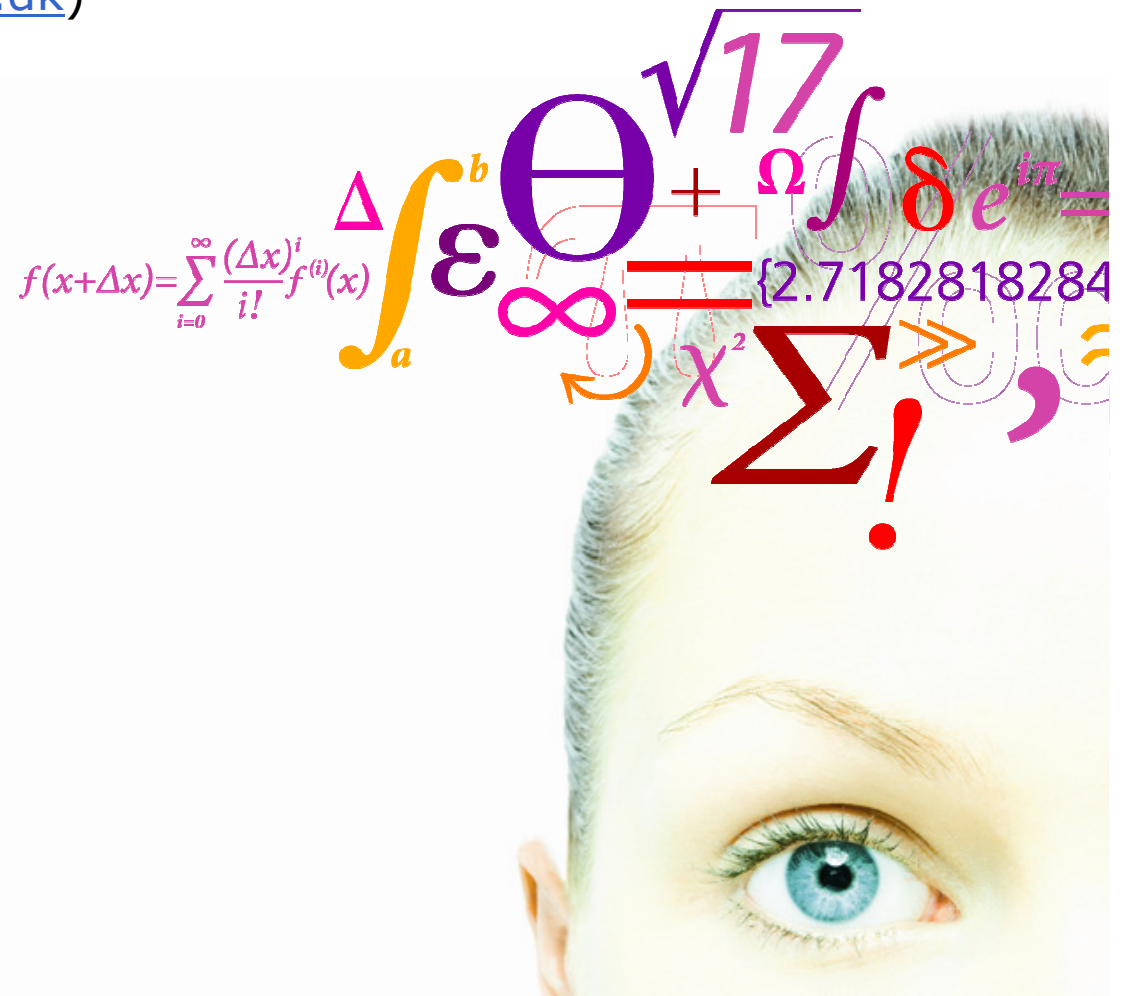


Global Data for Real-time Detection and Prevention of Outbreaks and Emerging Diseases

Frank M. Aarestrup (fmaa@food.dtu.dk)

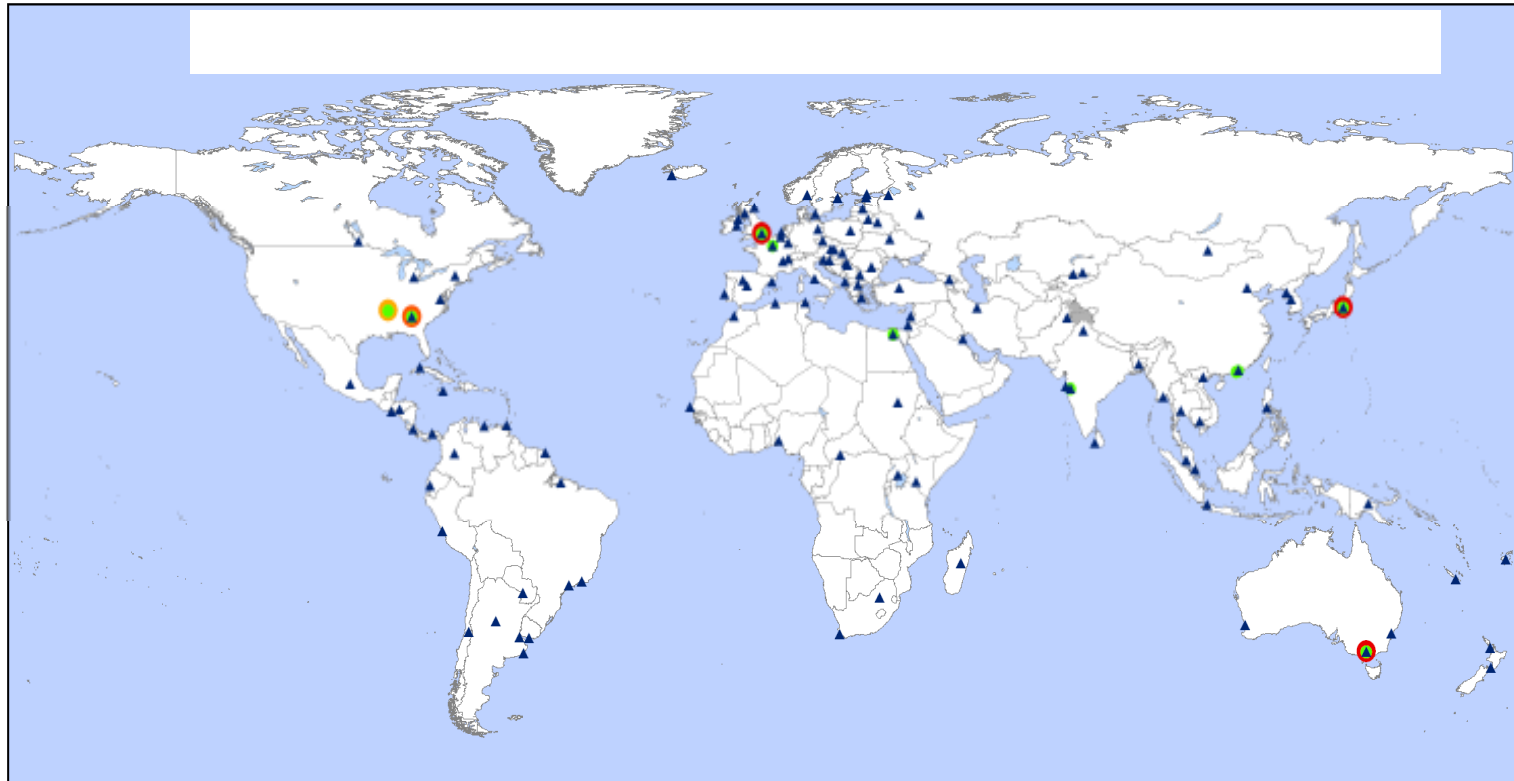
Center for Genomic Epidemiology
www.genomicepidemiology.org

DTU Food
National Food Institute



Global influenza surveillance network

The WHO Global Influenza Surveillance Network (GISN), July 2008



25 July 2008

- ▲ National Influenza Centres
- H5 Reference Laboratories
- WHO Collaborating Centre for Studies on the Ecology of Influenza in Animals
- WHO Collaborating Centre for the Surveillance, Epidemiology and Control of Influenza
- WHO Collaborating Centres for Reference and Research on Influenza



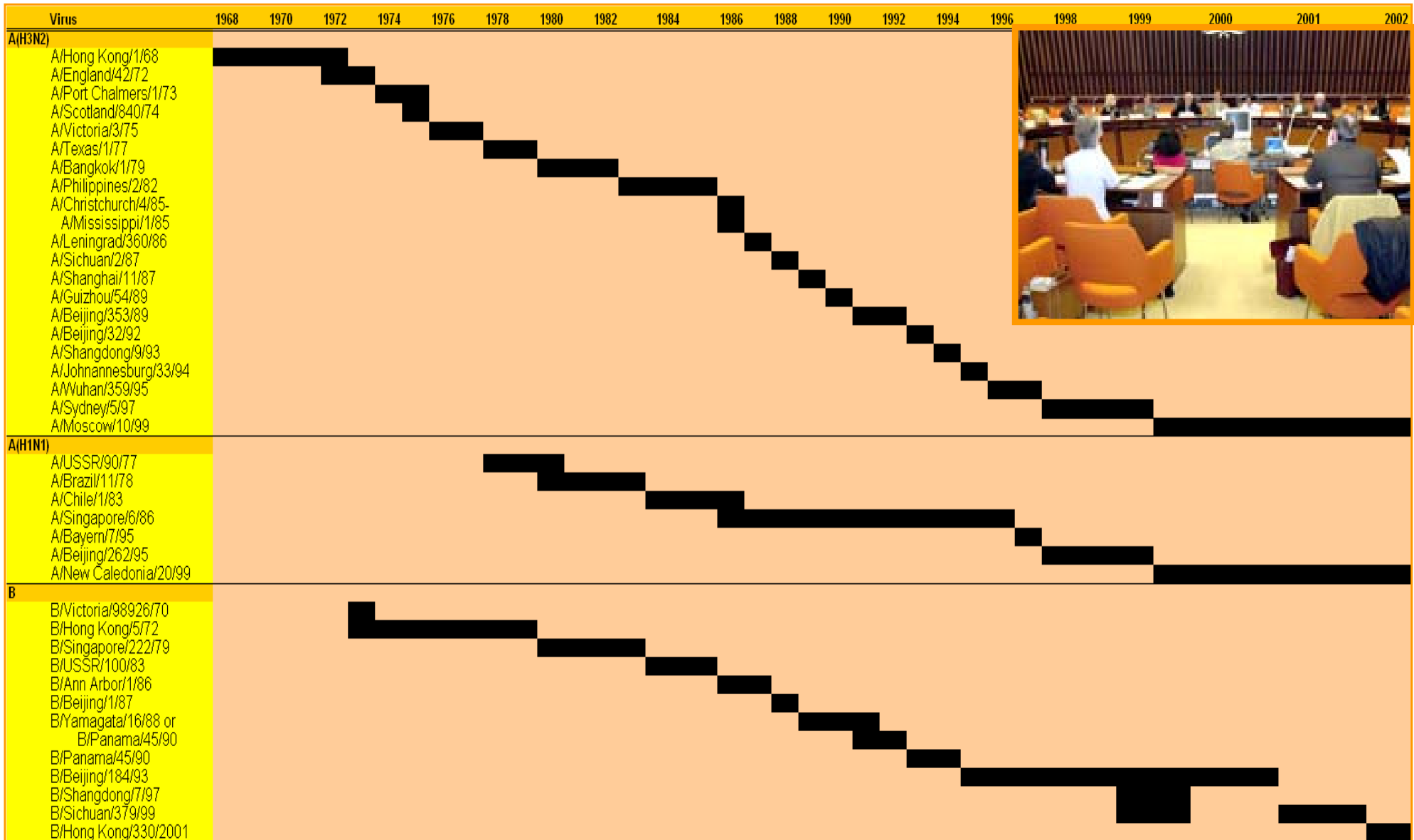
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.
Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

Data Source: WHO FluNet, GISN
Map Production:
HSE/EPR/GIP, HSE/EPR/GIS
World Health Organization
© WHO 2008. All rights reserved

Seasonal influenza: risk assessment and management

- Clinical specimens: >200,000 each year processed by National Influenza Centres to diagnose seasonal influenza
- Genetic & antigenic characterization: viruses classified and most predominant strains identified for vaccine development
- Selection & development of vaccine candidate viruses: necessary for vaccine development and production
- Provision of candidate vaccine viruses for vaccine development: to any qualified vaccine producer
- Development diagnostic tests provided at not cost to all National Influenza Centres

Antigenic shift and drift of seasonal influenza virus: vaccine composition





Contaminated food for thought

If it is to deal effectively with outbreaks of infectious diseases, Germany must streamline its convoluted systems for reporting and communication.

GERMANY

Scientists Rush to Study Genome of Lethal *E. coli*

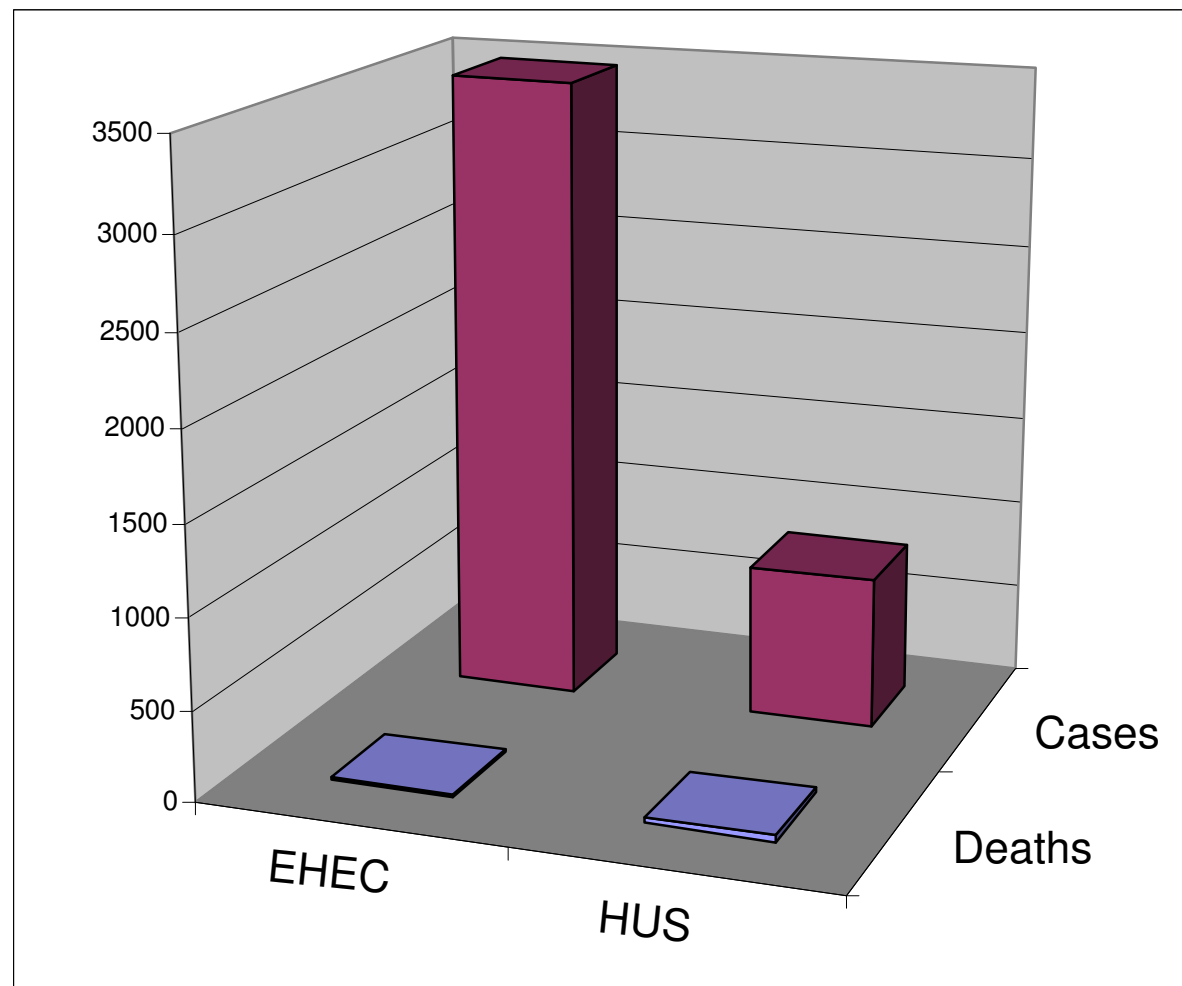
When cholera raged in the German port city of Hamburg in 1892 and killed thousands of people, famous epidemiologist Robert Koch pinpointed contaminated drinking

the dangerous Shiga toxin that enters the cells lining the gut and inhibits protein synthesis. The resulting cellular destruction leads to abdominal cramping and eventually bloody

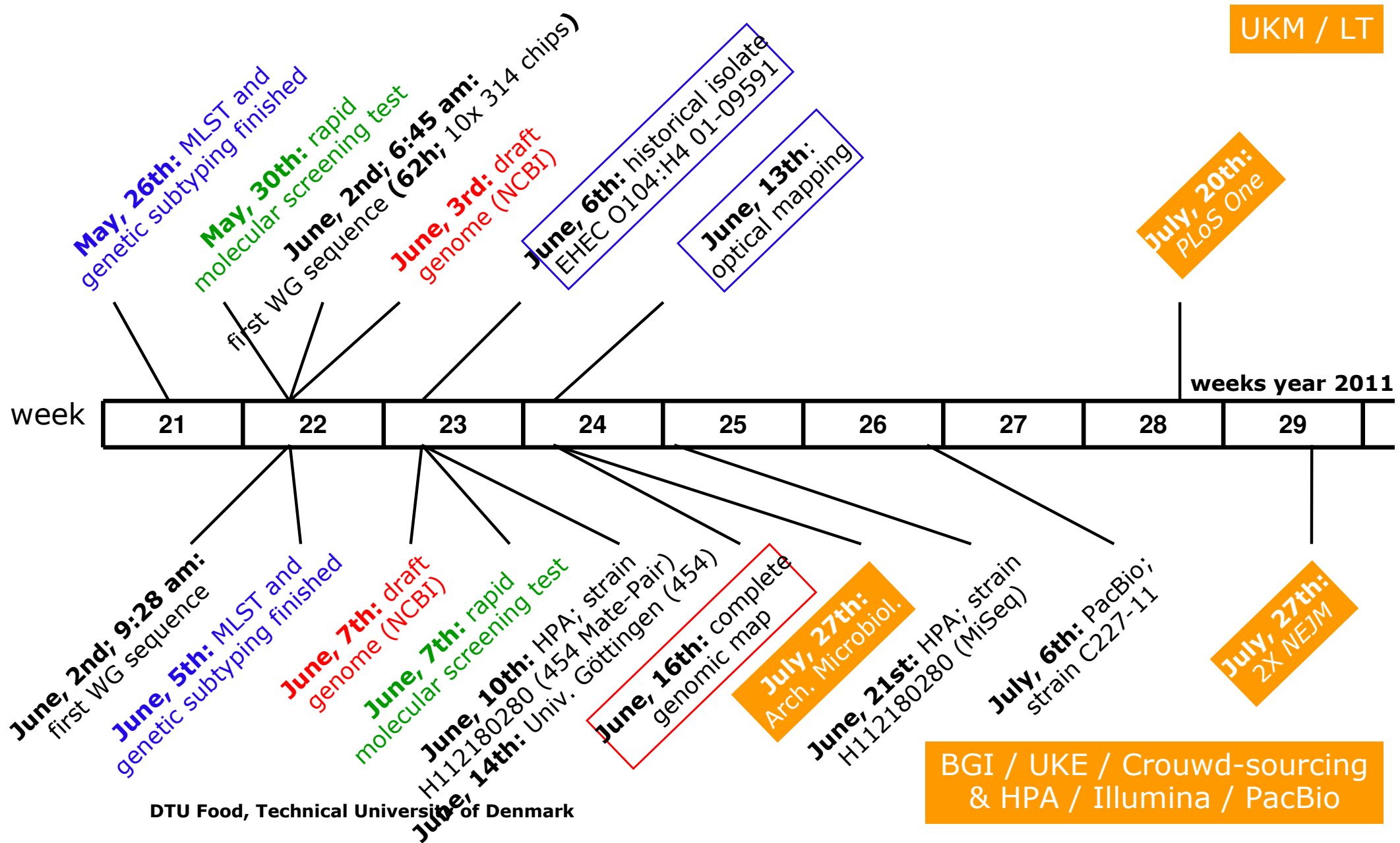
that they had deciphered the microbe's entire 5.2-million-base-pair genome and immediately made the DNA sequence available for researchers to download. Scores of scientists all over the world started poring over the data, assembling sequence fragments generated by BGI into a coherent genome, and comparing it to reference genomes for *E. coli* and other bacteria. The same day, a collaboration between the University of Münster and Life Technologies Corp., which

World Largest HUS Epidemic Due to EHEC

- Germany (RKI, July, 26th, 2011)
 - EHEC
 - 3,481 cases
 - 18 deaths
 - HUS
 - 852 cases
 - 32 deaths
- Europe / North America (WHO, July 21st, 2011)
 - EHEC
 - 89 cases
 - no deaths
 - HUS
 - 52 cases
 - 2 deaths



Achievements Prospective Genomic Epidemiology



The Challenge

- Continue to increase the power of surveillance using molecular diagnostics
- Develop diagnostics that can be used as far down the health system as possible
- Continue to provide the benefits of that surveillance to all countries

The Evolving Scale of Science:

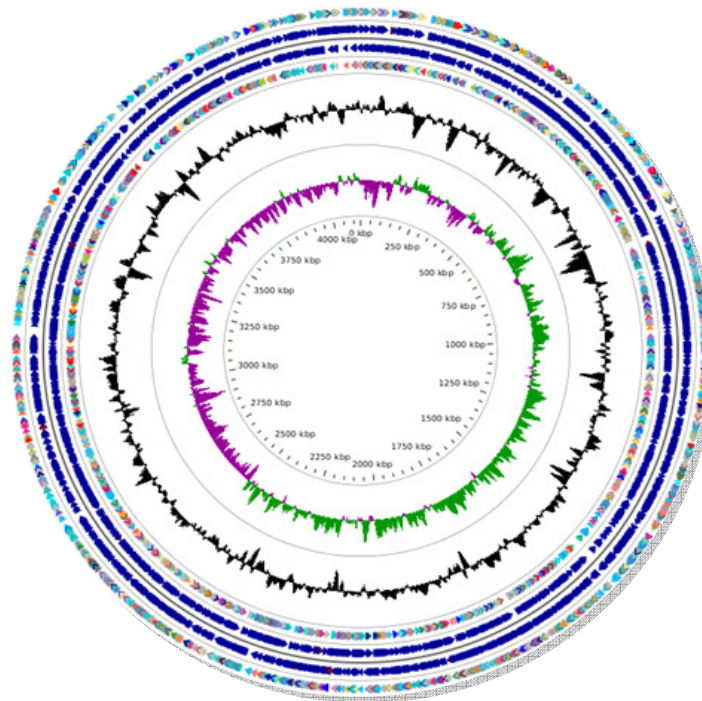
1980s:

One gene
One technician
One project



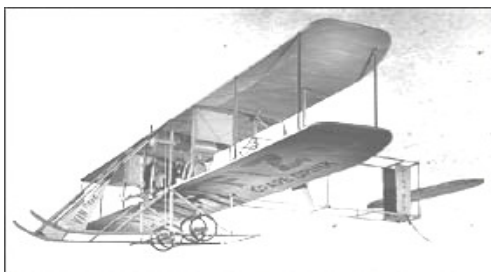
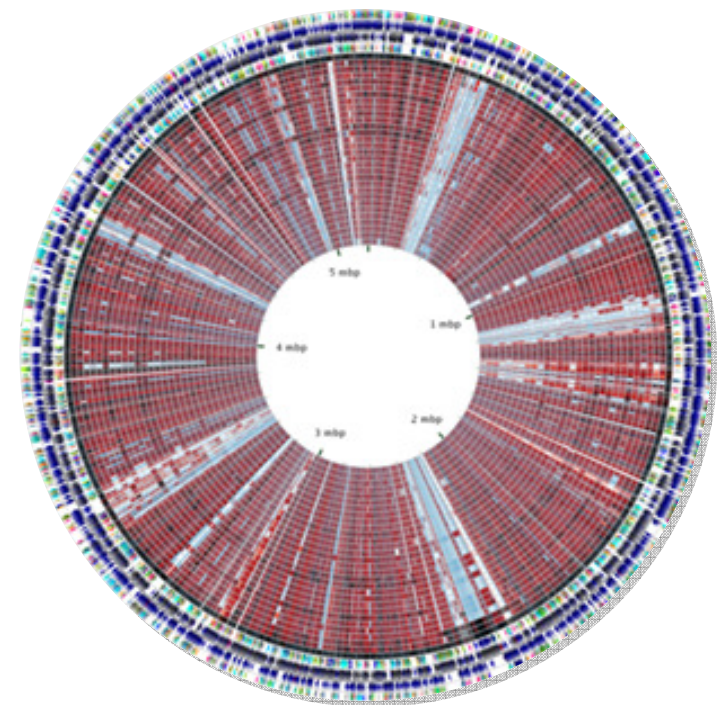
1990s:

Whole-genome sequencing of single *Reference* strains



2000s:

Whole-genome sequencing of multiple strains



University of Denmark

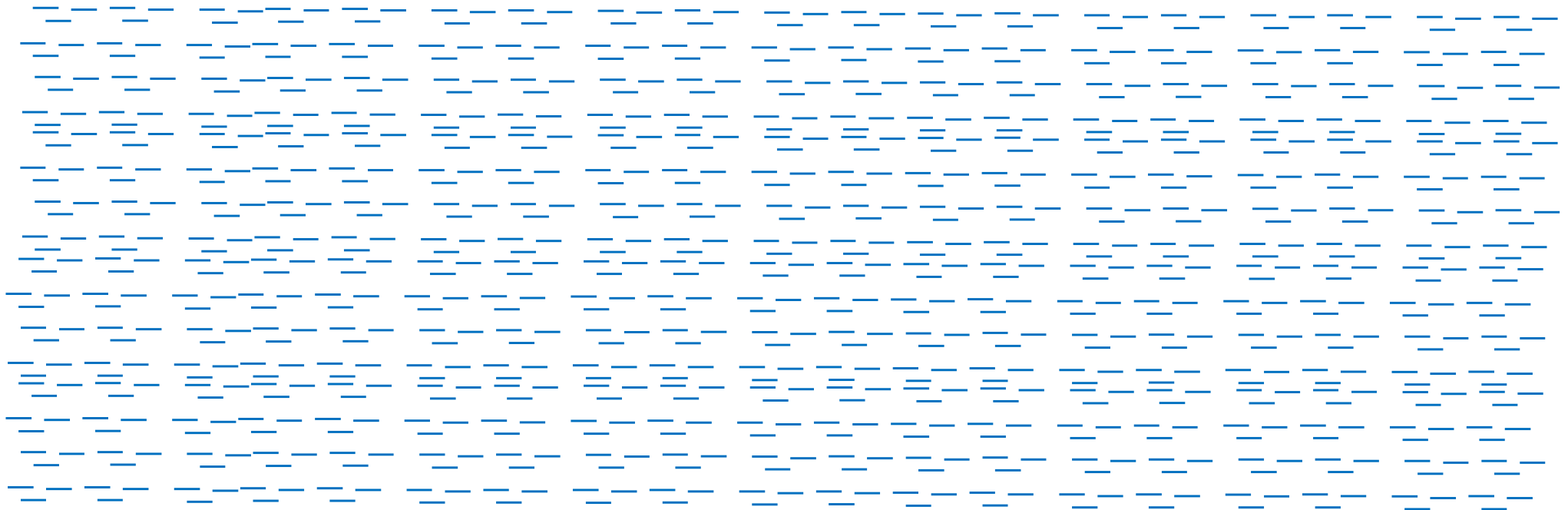


Second generation sequencing



NGS output

Huge numbers of small fragments (35-500 bp)



Purpose of Center for Genomic Epidemiology

- Provide a proof of concept of combining bioinformatics with global epidemiology in real-time
- And provide a useful facility for frontline users

CLIENT SIDE

Raw DNA Sequences

Rough assembly and compression

Summary of:
What it is
What is known
How we can fight
What is new/unusual
Recommendations

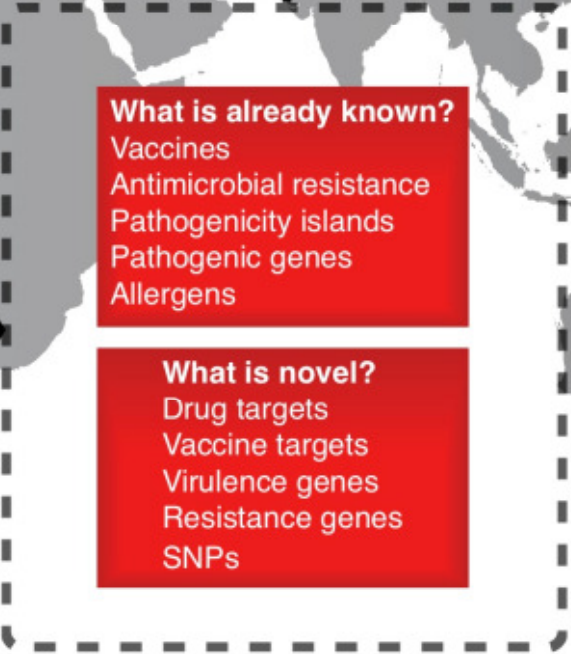
Google maps like view
Reports
Outbreak
Death tolls

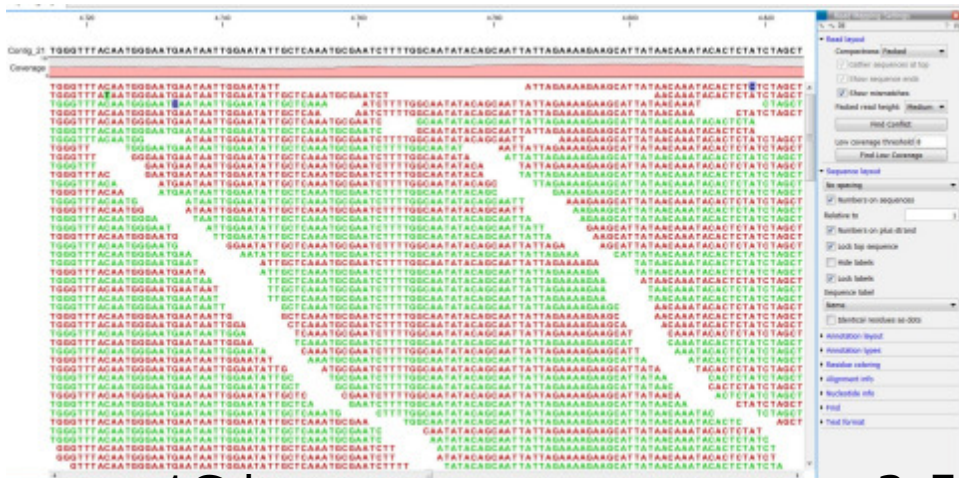
SERVER SIDE

Fine Assembly

Identification

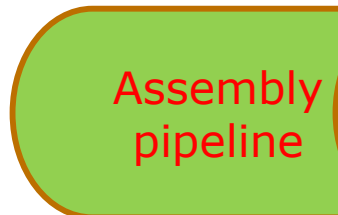
Gene finding
Gene annotation
Comparison



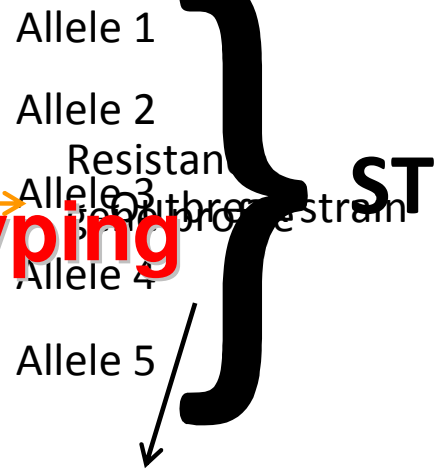


1G bases

3-5M bases



SNP* based typing



List of genes (100% or >95%)
Theoretical resistance phenotype
 Virulence genes
 AAAAAA AAAAAA AAAAAA
 AAAAAA AAAAAA AAAAAA
 AAAAT AAAAAA AAAAAA
 AATAAA TAATAATAA

* SNP Single Nucleotide Polymorphism (extreme MLST)

Examples – MLST and Resfinder

MLST (Multilocus Sequence Typing)

[Instructions](#)
[Output format](#)
[Article abstract](#)

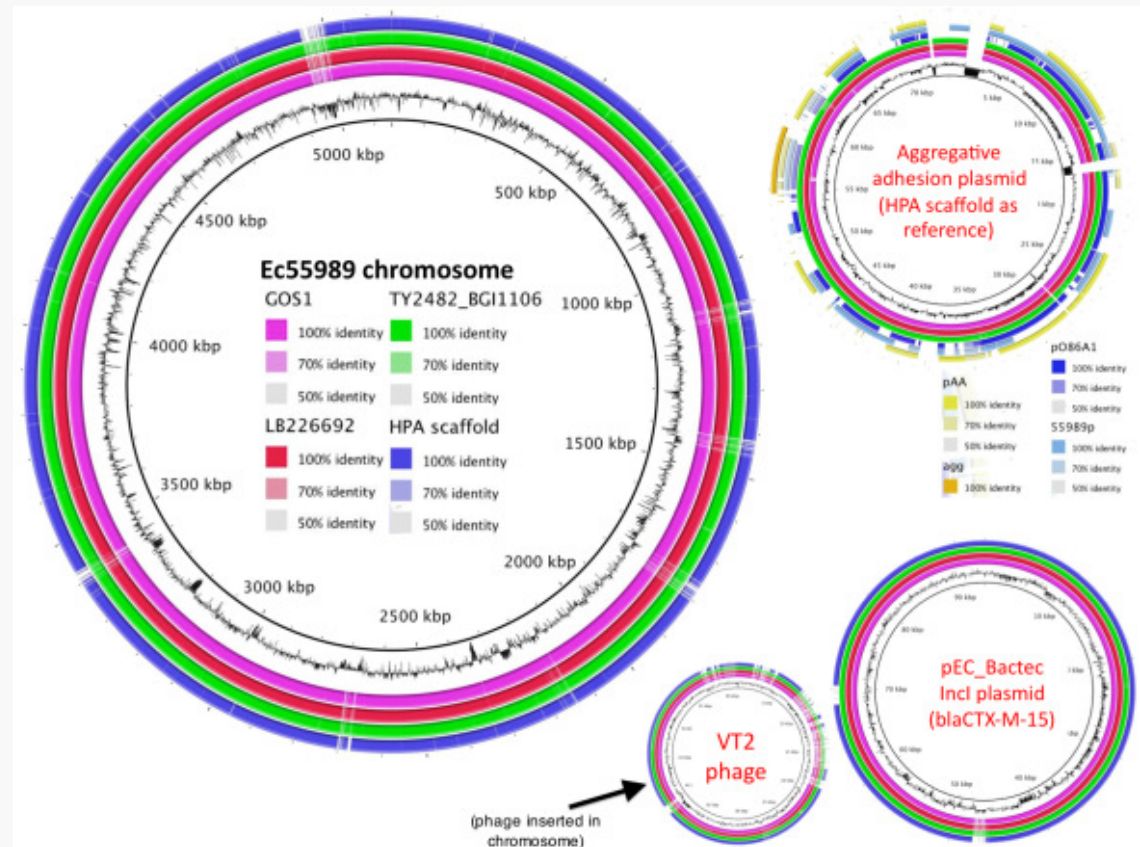
Uploads

Total files: 0 (N/A).

Select MLST configuration
 Escherichia coli#1

Select type of your reads
 Assembled Genome/Contigs*

VTEC O104:H4 outbreak strain



MLST Results

Sequence Type: **ST-678**

SETTINGS:

Organism: *Escherichia coli*

MLST Profile: *ecoli*

Genes in MLST Profile: 7

Locus	%Identity	Allele Length/HSP Length	Gaps	Allele
<i>adk</i>	100%	536/536	0	<i>adk-6</i>
<i>fumc</i>	100%	469/469	0	<i>fumc-6</i>
<i>gyrb</i>	100%	460/460	0	<i>gyrb-5</i>
<i>icd</i>	100%	518/518	0	<i>icd-136</i>
<i>mdh</i>	100%	452/452	0	<i>mdh-9</i>
<i>pura</i>	100%	478/478	0	<i>pura-7</i>
<i>reca</i>	100%	510/510	0	<i>reca-7</i>

Examples – MLST and Resfinder

ResFinder 1.1 Server (Acquired antibiotic resistance gene finder)

ResFinder 1.1 identify acquired antibiotic resistance genes in total and partial sequenced isolates of bacteria. The input sequence must be in one-letter nucleotide code, [Test sequence](#)

Instructions
Output format
Article abstract

Browse
 Remove
 Clear

Uploads

Upload
Total files: 0 (N/A)

Select Antimicrobial configuration
Select multiple items, with Ctrl-Click (or Cmd-Click on Mac)

- All
- Aminoglycoside
- Beta-lactamase
- Fluoroquinolone
- Glycopeptide
- MLS - Macrolide-Lincosamide-StreptograminB

Select threshold for %ID

Select type of your reads

Submit
Clear fields

VTEC O104:H4 outbreak strain

The figure displays several genomic maps for the VTEC O104:H4 outbreak strain. The largest map is the **Ec55989 chromosome**, a circular map of approximately 5000 kbp. It features a central track of sequence data and outer tracks representing antibiotic resistance genes. A legend for the Ec55989 chromosome identifies genes and their identity percentages:

- GOS1**: 100% (magenta), 70% (light magenta), 50% (grey)
- TY2482_BGI1106**: 100% (green), 70% (light green), 50% (grey)
- LB226692**: 100% (red), 70% (light red), 50% (grey)
- HPA scaffold**: 100% (blue), 70% (light blue), 50% (grey)

Other smaller maps include:

- Aggregative adhesion plasmid (HPA scaffold as reference)**: A circular plasmid map.
- pO86A1**: A circular plasmid map with a legend for 100%, 70%, and 50% identity levels.
- pAA**: A circular plasmid map with a legend for 100%, 70%, and 50% identity levels.
- p55989p**: A circular plasmid map with a legend for 100%, 70%, and 50% identity levels.
- pGG**: A circular plasmid map with a legend for 100% identity.
- pEC_Bactec Incl plasmid (blaCTX-M-15)**: A circular plasmid map.
- VT2 phage**: A small circular map of the phage, with an arrow pointing to its location on the chromosome and the note "(phage inserted in chromosome)".

Results

ResFinder Results

Resistance gene				
Aminoglycoside				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
<i>strA</i>	100.00%	804/804	Aminoglycoside resistance Alternate name; aph(3'')-Ib	AF321551
<i>strB</i>	100.00%	837/837	Aminoglycoside resistance Alternate name; aph(6)-Id	FJ474091
Beta-lactam				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
<i>CM-15</i>	100.00%	876/876	Beta-lactamase resistance Alternate name; UOE-1	DQ302097
Fluoroquinolone				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
No resistance genes found.				
MLS - Macrolide-Lincosamide-StreptograminB				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
No resistance genes found.				
Phenicol				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
No resistance genes found.				
Sulphonamide				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
<i>sulI</i>	100.00%	840/761	Sulphonemide resistance	AY224185
Tetracycline				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
<i>tet(A)</i>	100.00%	1200/1200	Tetracycline resistance	AJ517790
Trimethoprim				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
<i>dfrA7</i>	100.00%	474/474	Trimethoprim resistance	JF806498
Glycopeptide				
Resistance gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
No resistance genes found.				

Center for Genomic Epidemiology

Username Password [Home](#)[Services](#)[User Home](#)

NEWS: Dear all CGE service users, [Show](#)

Overview of Services (Experimental)

Main Service - Pipeline

[CGE](#) (In development)

Sequence Typing

[MLST](#) (Works)

[pMLST](#) (Works)

Resistance - Virulence - Plasmids

[ResFinder](#) (Works)

[PlasmidFinder](#) (Works)

[VirulenceFinder](#) (Works)

Phylogenetic Tree

[snpTree](#) (Works)

Species Finding

[KmerFinder](#) (Works)

[SpeciesFinder](#) (Works)

[TaxonomyFinder](#) (This program is in development)

[Read2Type](#) (This service is not implemented on the new server)

[Tapir](#) (This service is not implemented on the new server)

Genome Assembly

[Assembler](#) (Works)

[Support](#)[Technical problems](#)

When can WGS replace all other techniques?

- The - £ \$ € - question
- WGS – today 100 € going to 50 or 10?
- Traditional:
 - Identification – 10 €
 - Serotyping – 25 €
 - Susceptibility testing – 15-25 €
 - PFGE – 50 €
 - MLST – 250 €
 - Molecular characterisation – 10 – 1000 €

Already competitive



A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance

David W Eyre,^{1,2} Tanya Golubchik,^{2,3} N Claire Gordon,^{1,2} Rory Bowden,^{2,3,4} Paolo Piazza,⁴ Elizabeth M Batty,^{2,3} Camilla L C Ip,^{2,3} Daniel J Wilson,^{1,4} Xavier Didelot,^{2,3} Lily O'Connor,^{1,5} Rochelle Lay,⁵ David Buck,⁴ Angela M Kearns,⁶ Angela Shaw,⁷ John Paul,⁸ Mark H Wilcox,⁹ Peter J Donnelly,⁴ Tim E A Peto,^{1,2,5} A Sarah Walker,^{1,2,10} Derrick W Crook^{1,2,5}

Also:

Köser et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med. 2012 Jun 14;366(24):2267-75.

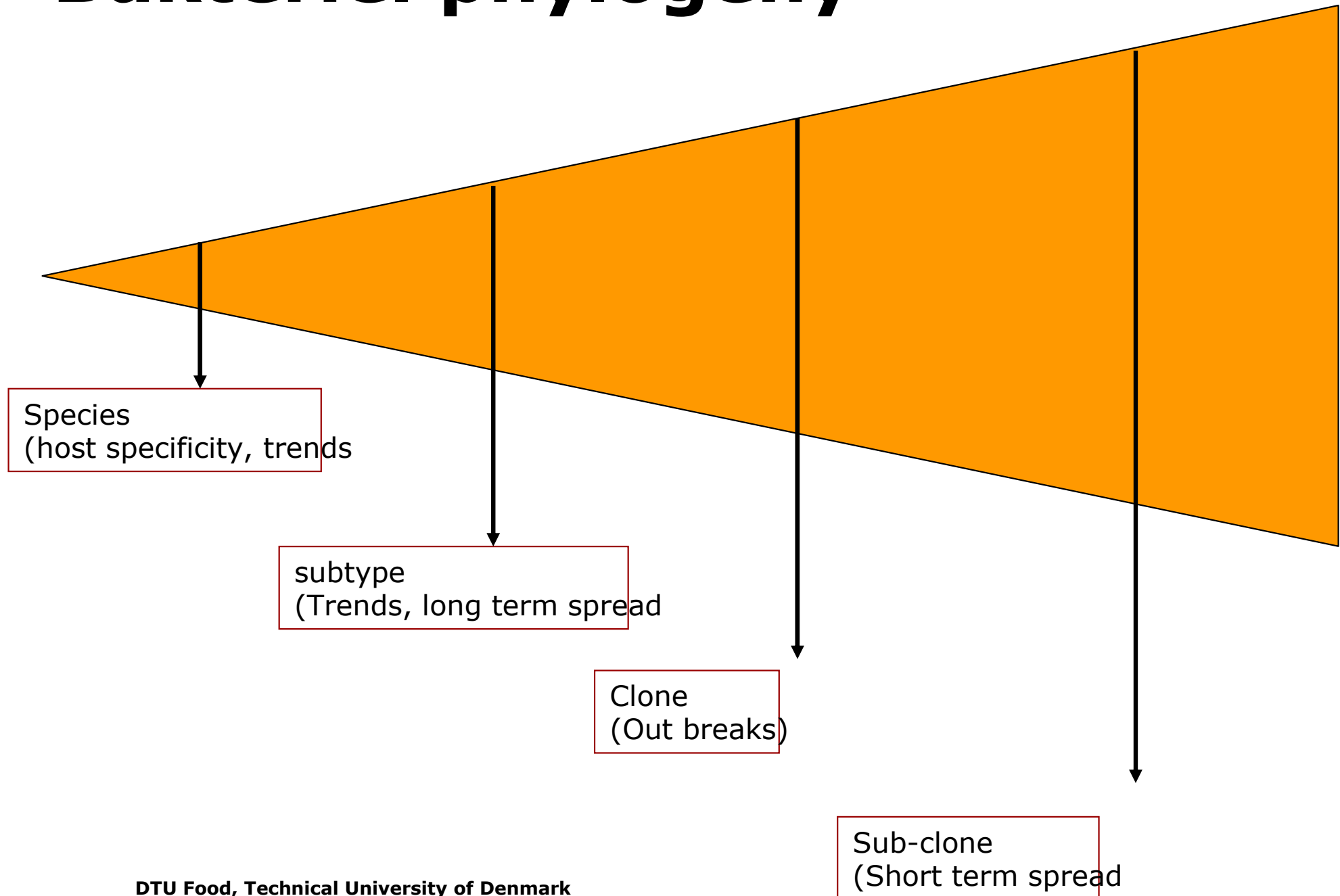
Is a global solution feasible?

- Technically
- Scientific
- Politically

Computing needs

- Denmark (1 million sequences/year):
 - 2h CPU time + 10mb storage for 1,000,000 genomes per year in 6 years
= 230 Cores, 60TB storage needed
- EU (100 million sequences/year)
 - 2,300 cores, 6PB storage
- Global needs (1 billion sequences/year)
 - 23,000 cores, 60PB storage ~ 30th biggest computer (Smaller than Airbus')

Bakteriel phylogeny



Scientific challenges

- Standardized analysis and output
- Interpretation of data
- Combining WGs with epidemiology
 - The need for classical epidemiology will remain the same
 - There will be a new need for modelers and epidemiologists working with real.-time data and combining phylogeny with spatial and temporal data

Building global partnership

- Global consensus meetings
 - Bruxelles (Sep. 2011)
 - Washington DC (March 2012)
- >100 participants covering, governmental institutions, universities, hospitals and companies (FDA, CDC, FBI, NCBI, DoD, USDA, PHAC, Harvard, Maryland, Virginia, Los Alamos, TGen, Aligent, Broad, EBI, eCDC, Sanger, Oxford, DTU, RIVM, FZB, BGI, DDBJ, etc)
- Challenges:
 - Meta-data
 - Publications (selfish scientists)
 - National authorities (need to know before the press)
 - Legally (lawyers)



Conclusions

- **There is a need and we believe it is possible**
- **Need to broaden discussion to a larger forum**
- **Need for special groups working on**
 - Global participation (208 countries!!!)
 - Repository (NCBI, EBI, DDBJ)
 - Access (Political, legal)
 - Output (simple & advanced, diagnostic & epidemiology, genes & species)
 - Taxonomy
 - Epidemiology, statistics and modelling (we need more scientists)

Center for Genomic Epidemiology - Windows Internet Explorer
 http://www.genomicepidemiology.org/

File Edit View Favorites Tools Help

Center for Genomic Epi... x Enterococcus faecalis 235 r... DTU - Danmarks Tekniske... Center for Genomic Epi...

x Find: caocggatassa Previous Next Options

Home Organization Project Services Contact

Services

- Multi Locus Sequence Typing (MLST) from an assembled genome or from a set of reads
[MLST typing](#)
- Identification of acquired antibiotic resistance genes from a file with sequence reads
[ResFinder](#)
- Identification of type from sequence reads
[Type from reads](#)

Organization
[Project](#)
[Publications](#)
[Contact](#)

News

Identification of acquired antimicrobial resistance genes
 August 2012
 We here present ResFinder, a web server that uses WGS data for identifying acquired antimicrobial resistance genes in bacteria.
[Publications in the center...](#)

The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium
 May 2012
 Combining three RNA-sequencing techniques and two sequencing platforms to generate transcriptional map of SL1344 advances our understanding of S. Typhimurium, arguably the most important bacterial infection model.
[Publications in the center...](#)

Genomic variation in Salmonella enterica core genes for epidemiological typing
 March 2012
 The variation within core genome is useful for investigating evolution and providing candidate genes from bacterial genome typing.
[Publications in the center...](#)

Staphylococcus aureus CC398. Host Adaptation and Emergence of Methicillin Resistance in Livestock
 February 2012
[Publications in the center...](#)

Comparative Genomics of Bifidobacterium, Lactobacillus and Related Probiotic Genera
 January 2012
 Genome diversity within a set of 'good bacteria' have been examined by a team from the CGE.
[Publications in the center...](#)

Multilocus Sequence Typing of Total Genome Sequenced Bacteria
 January 2012
[Publications in the center...](#)

The CGE is quoted in the Science
 article: Outbreaks, Detection

Welcome to the Center for Genomic Epidemiology

Within few years the costs for a total bacterial genome sequencing will be less than \$200 and the equipment needed will cost less than \$100 000. Thus, within a few years all clinical microbiological laboratories will have a sequencer in use on a daily basis. As prices decline to less than \$200 whole genome sequencing will also find worldwide application in human and veterinary practices as well as many other places where bacteria are handled. In Denmark alone this equals more than 100 000 isolates annually in 15-20 laboratories and globally up to 1-2 billion isolates per year. The limiting factor will therefore in the future not be the cost of the sequencing, but how to assemble, process and handle the large amount of data in a standardized way that will make the information useful, especially for diagnostic and surveillance.

The aim of this center is to provide the scientific foundation for future internet-based solutions where a central database will enable simplification of total genome sequence information and comparison to all other sequenced including spatial-temporal analysis. We will develop algorithms for rapid analyses of whole genome DNA-sequences, tools for analyses and extraction of information from the sequence data and internet/web-interfaces for using the tools in the global scientific and medical community. The activity can be expanded to also include other microorganisms, such as virus and parasites.

Done

Internet | Protected Mode: Off

100% 10:56 12-09-2012

www.genomicepidemiology.org