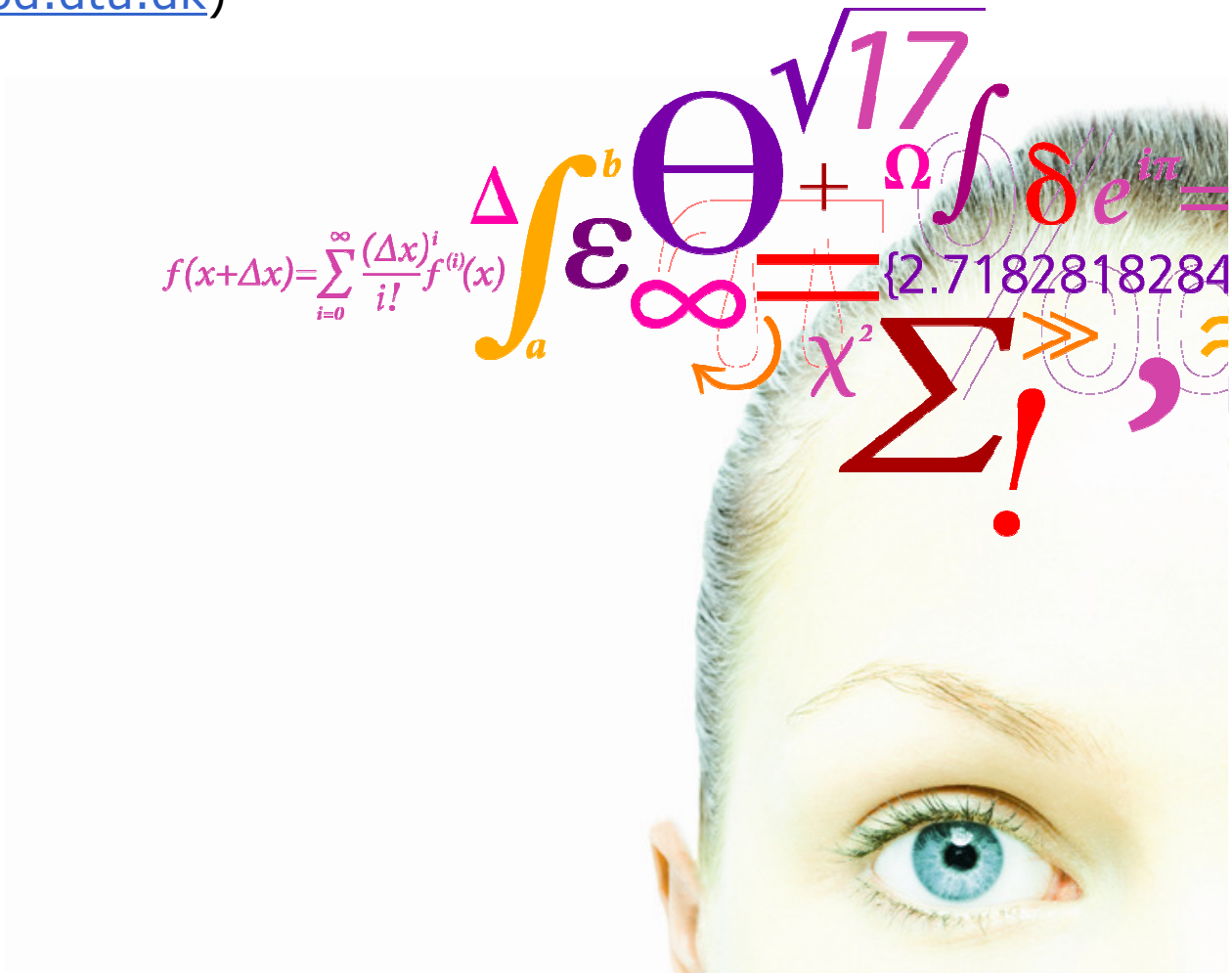
Global Data for Real-time Detection and Prevention of Outbreaks and Emerging Diseases

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

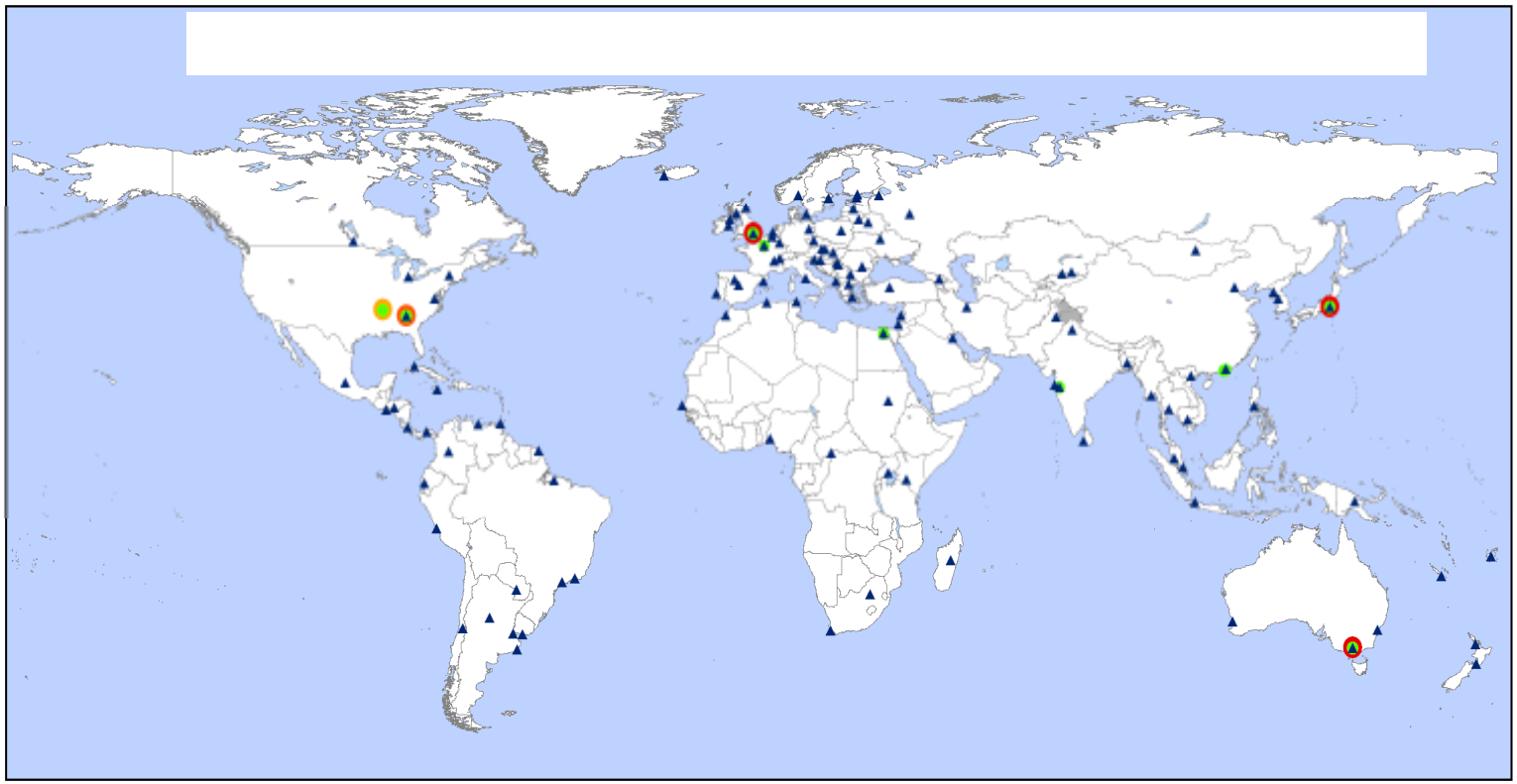


Center for Genomic Epidemiology
www.genomicepidemiology.org

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

World Health Organization

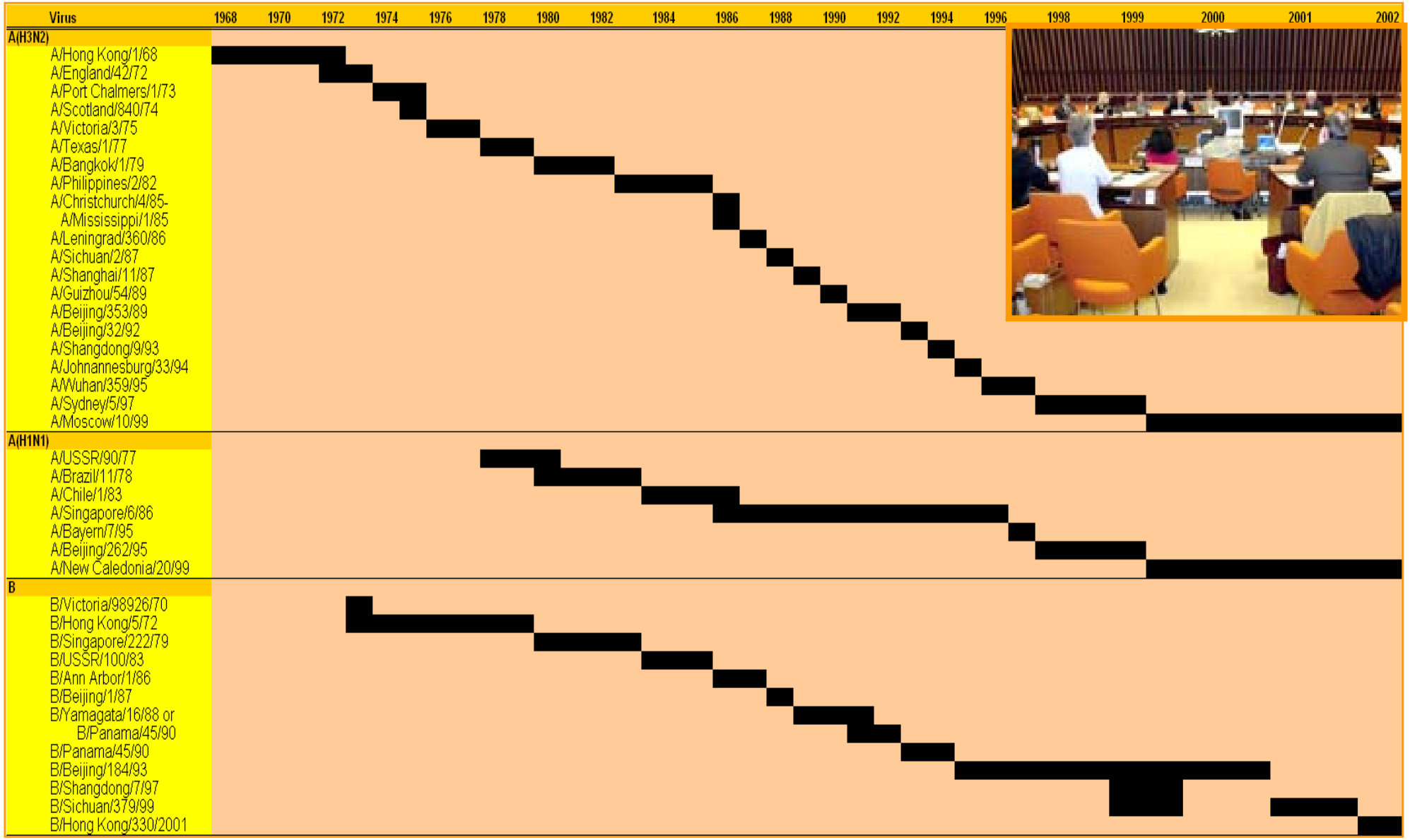
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers 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

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



NEWS & ANALYSIS

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*

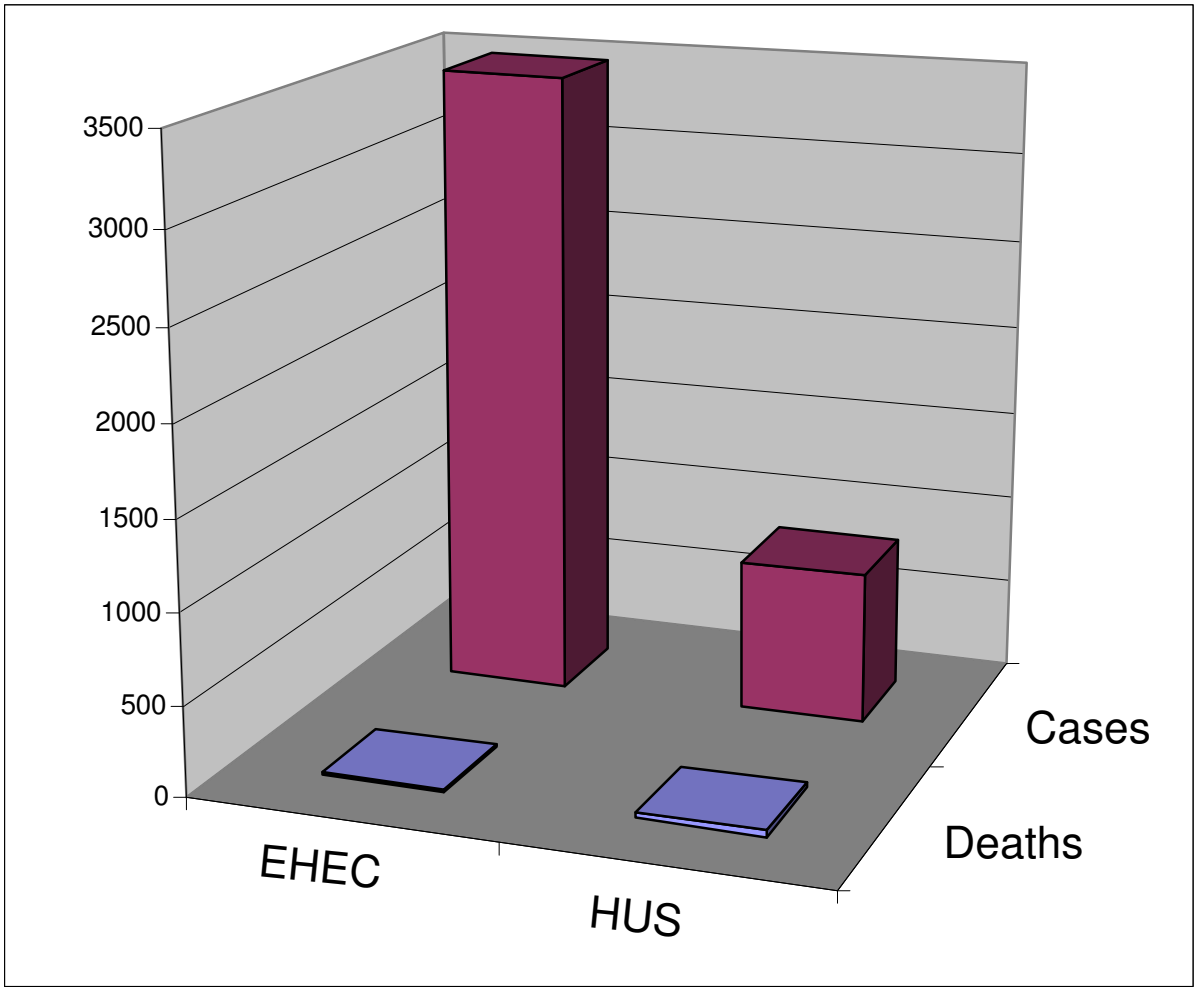
When cholera raged in the German port city of Hamburg in 1892 and killed thousands of people, famous epidemiologist Robert Koch pinpointed contaminated drinking

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

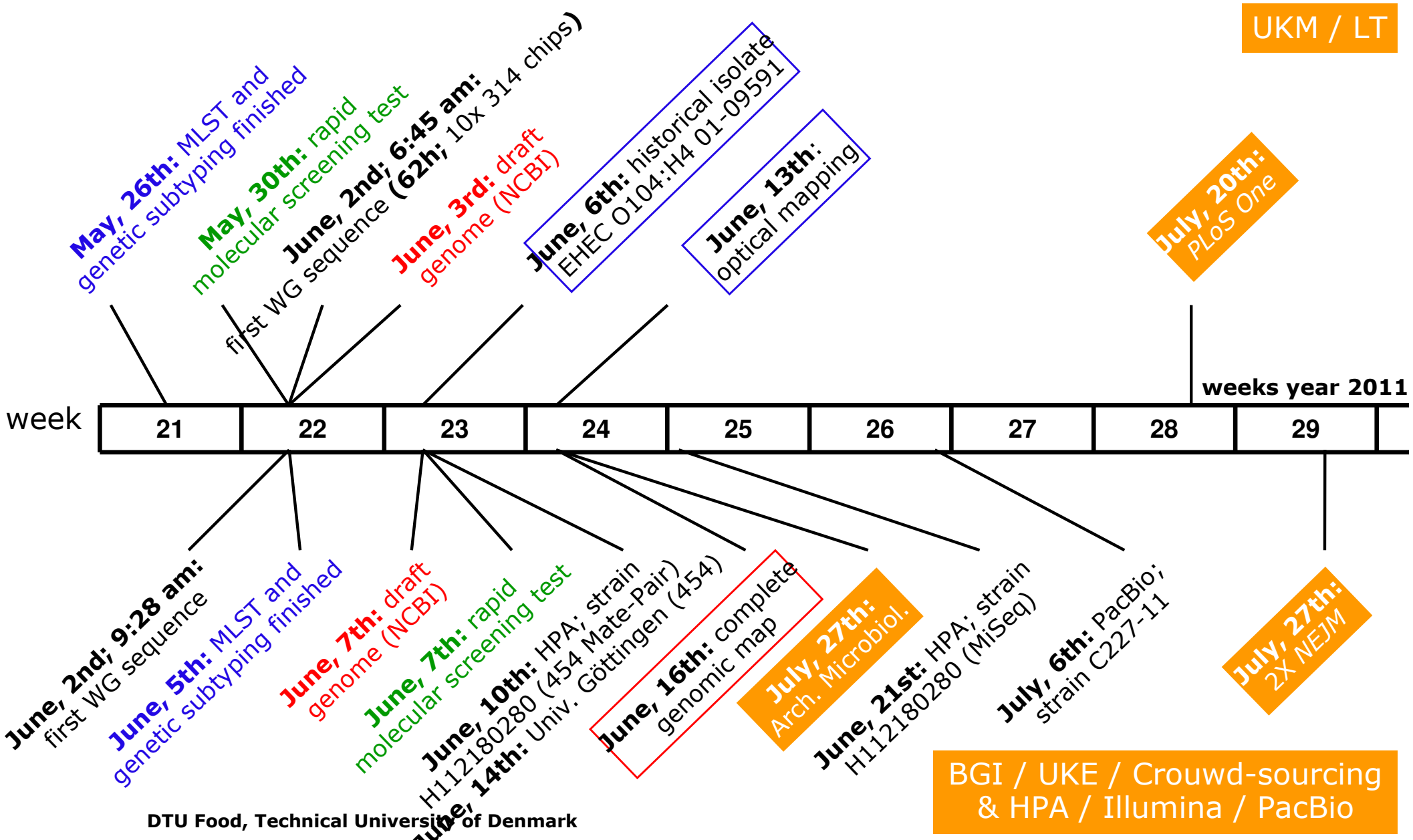
World Largest HUS Epidemic Due to EHEC

- 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



RKI, July 26th, 2011, adapted [http://www.rki.de/]

Achievements Prospective Genomic Epidemiology



The Challenge

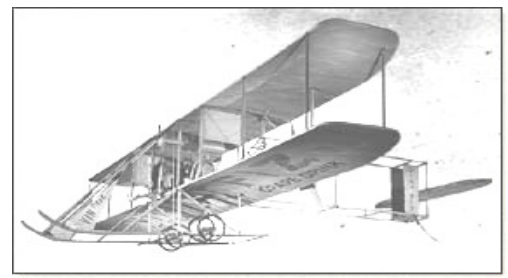
- Continue to increase the power of surveillance using molecular diagnostics
- Develop diagnostics that can be used as far down the health system as possible
- Continue to provide the benefits of that surveillance to all countries

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

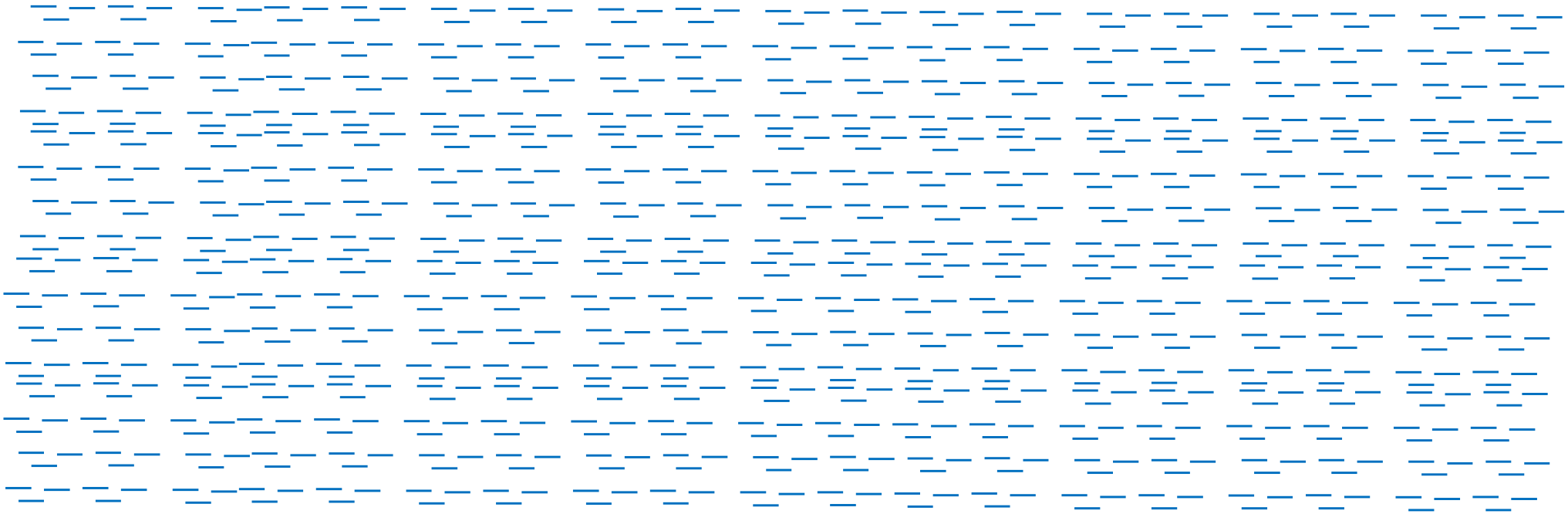


niversity of Denmark

Second generation sequencing

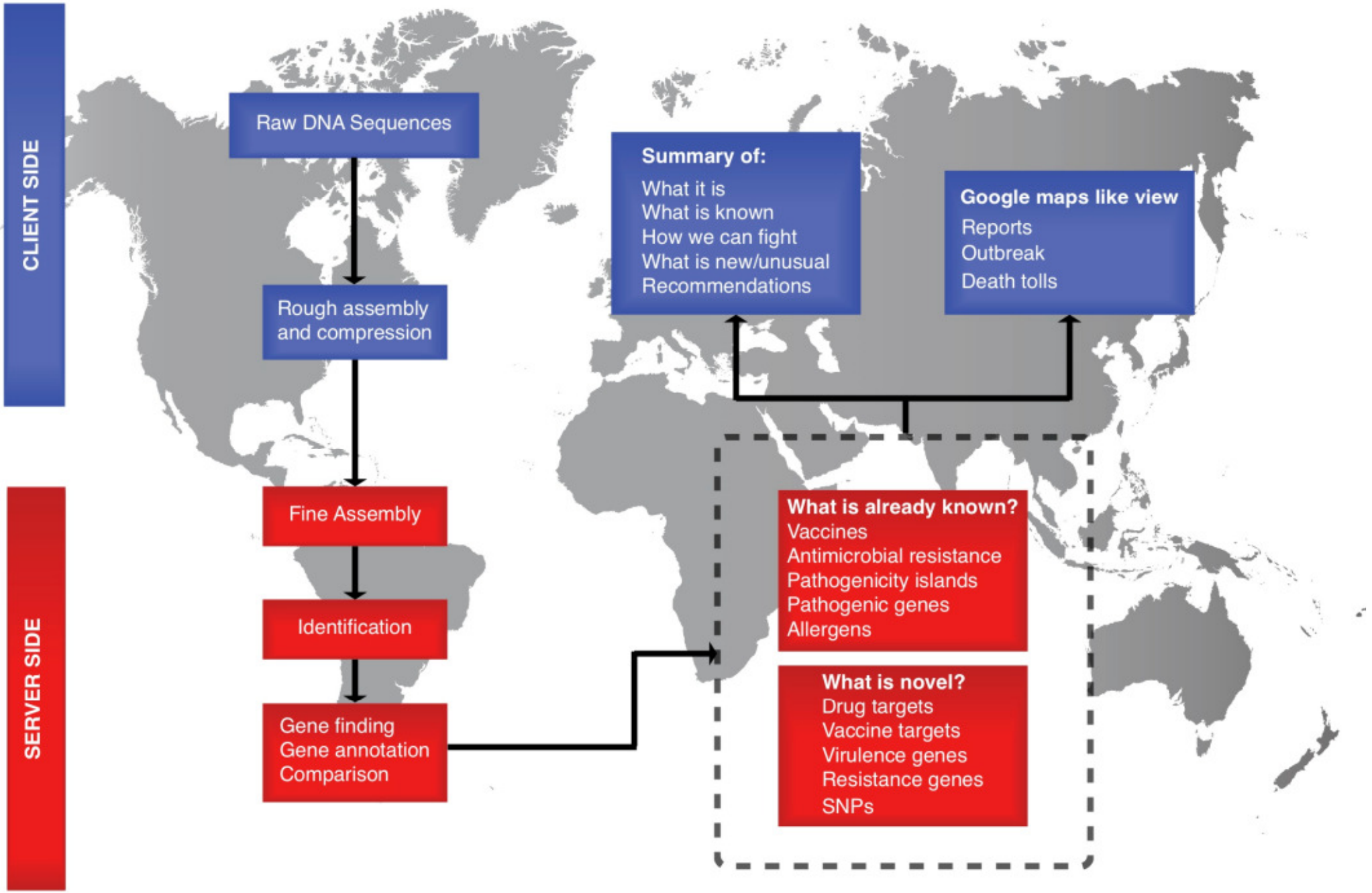
NGS output

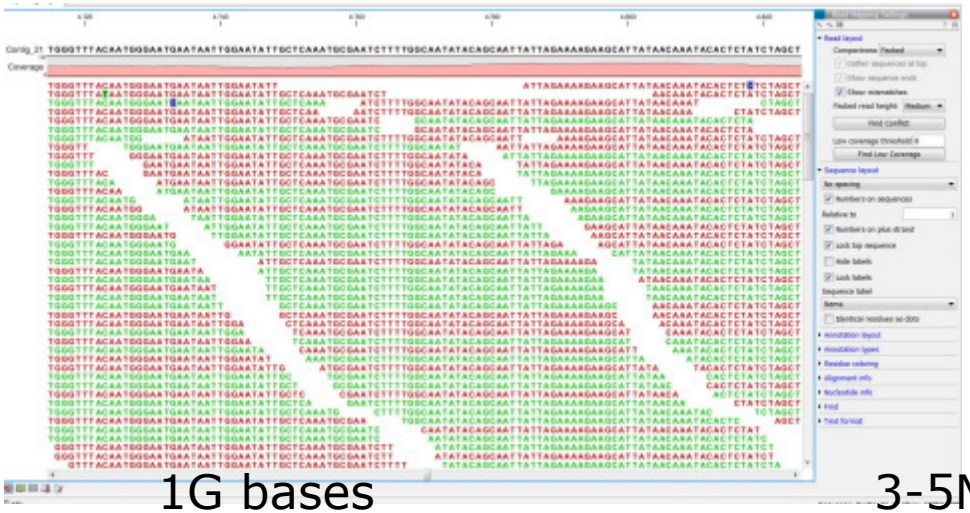
Huge numbers of small fragments (35-500 bp)



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
SERVER SIDE
Raw DNA Sequences
Rough assembly and compression
Fine Assembly
Identification
Gene finding
Gene annotation
Comparison
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
What is already known?
Vaccines
Antimicrobial resistance
Pathogenicity islands
Pathogenic genes
Allergens
What is novel?
Drug targets
Vaccine targets
Virulence genes
Resistance genes
SNPs



1G bases

3-5M bases

NGS
Illumin
PacBi
454..

Assembly pipeline

SNP* based typing

Allele 1

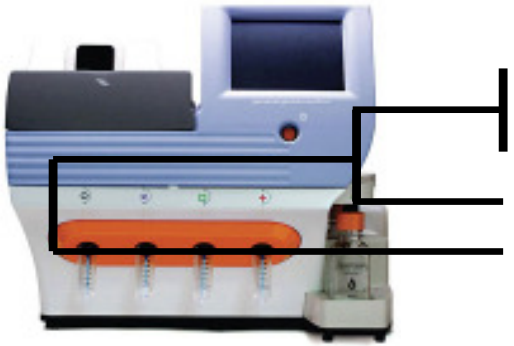
Allele 2

Allele 3

Allele 4

Allele 5

ST



List of genes (100% or >95%)
Theoretical resistance phenotype

AAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAA
AAAATAAAAAAAAAAA
AATAAATAATAATAAA

Virulence genes

*SNP - Single Nucleotide Polymorphism (extreme MLST)

Examples – MLST and Resfinder

MLST (Multilocus Sequence Typing)

Instructions | Output format | Article abstract

Browse | Remove | Clear

Uploads

Total files: 0 (N/A).

Select MLST configuration

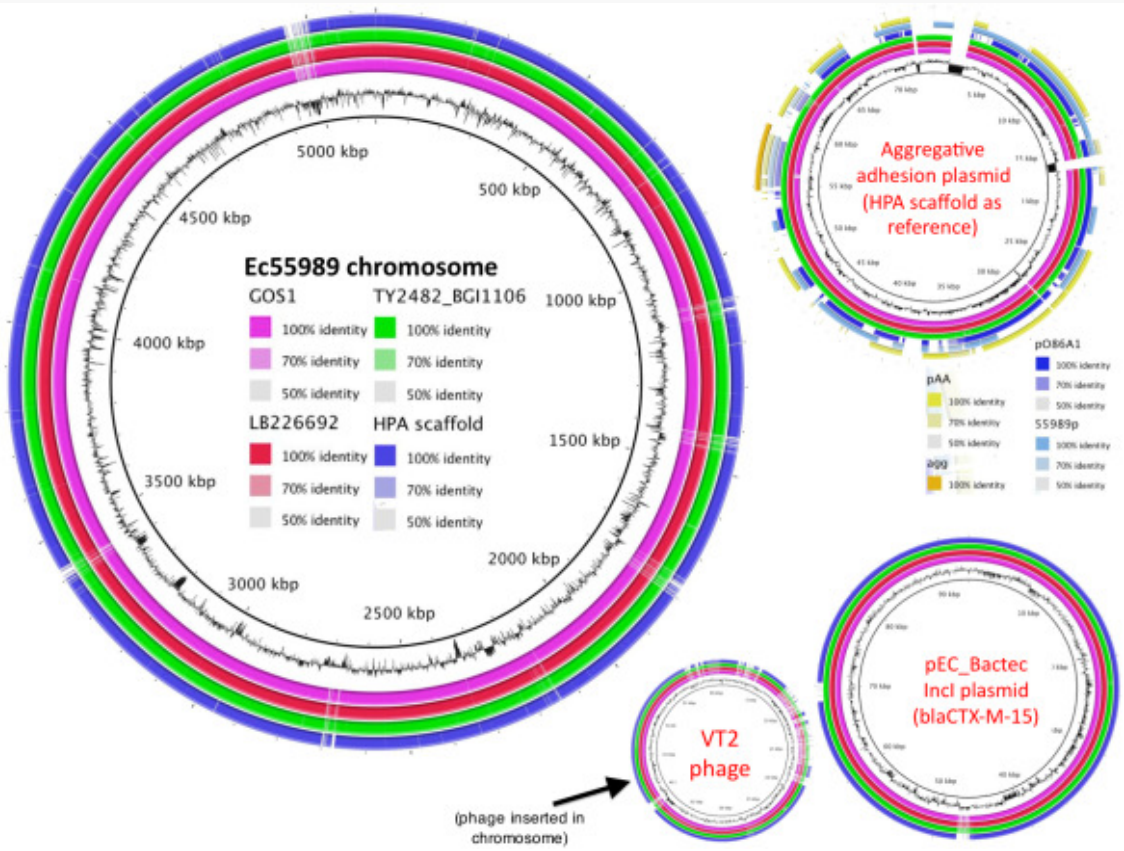
Escherichia coli#1

Select type of your reads

Assembled Genome/Contigs*

Submit | Clear fields

VTEC O104:H4 outbreak strain



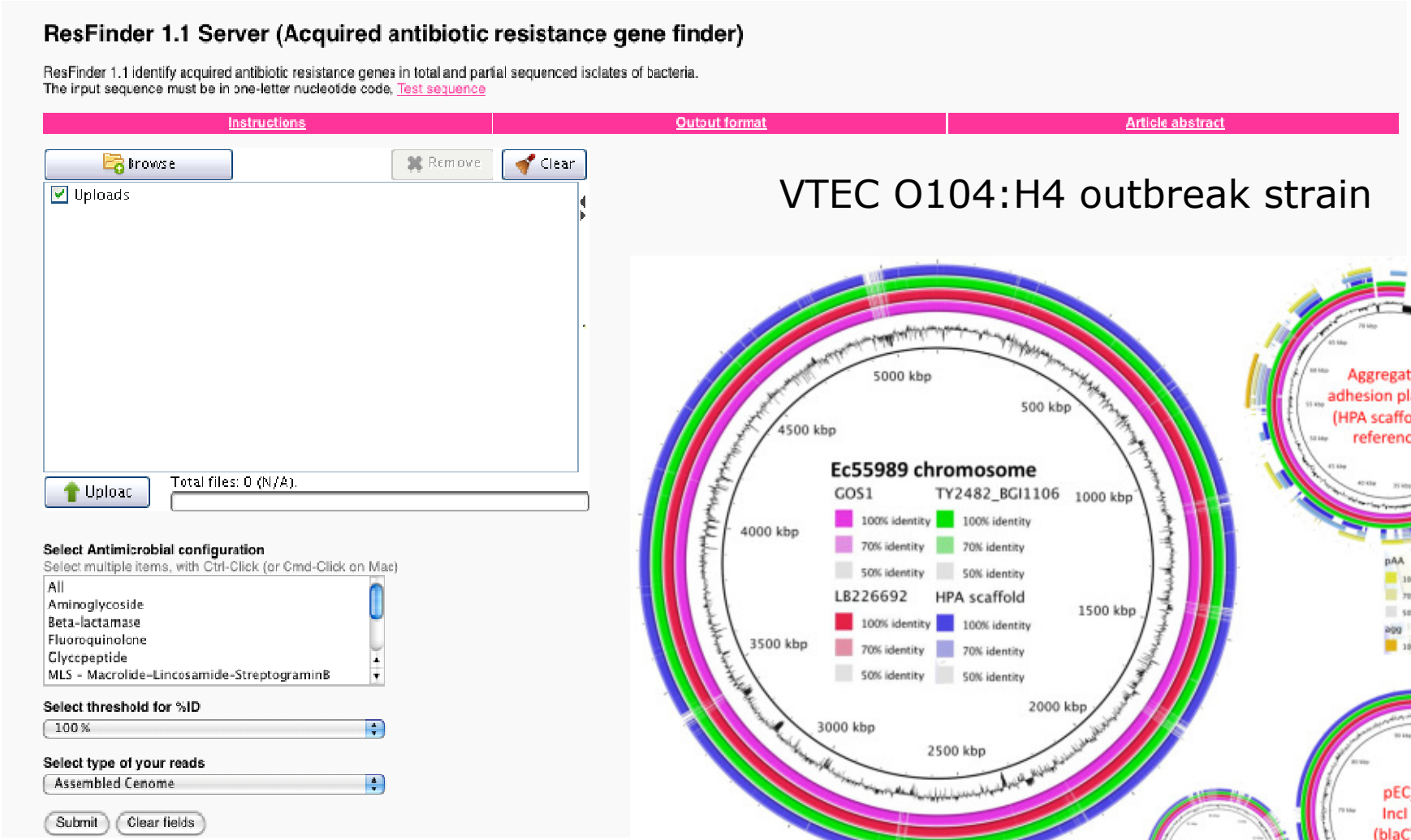
MLST Results

Sequence Type: *ST-678*

SETTINGS:
Organism: *Escherichia coli*
MLST Profile: *ecoli*
Genes in MLST Profile: *7*

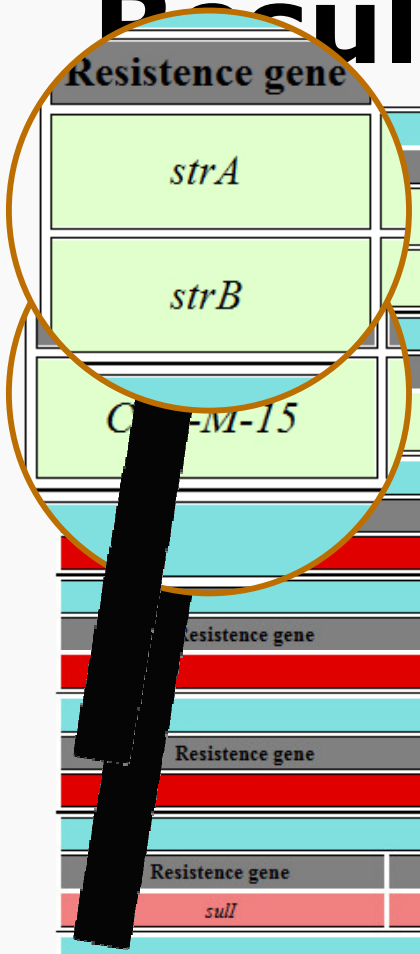
Locus	%Identity	Allele Length/HSP Length	Gaps	Allele
adk	100%	536/536	0	*adk-6*
fumc	100%	469/469	0	*fumc-6*
gyrb	100%	460/460	0	*gyrb-5*
icd	100%	518/518	0	*icd-136*
mdh	100%	452/452	0	*mdh-9*
pura	100%	478/478	0	*pura-7*
reca	100%	510/510	0	*reca-7*

Examples – MLST and Resfinder



Results

ResFinder Results



Aminoglycoside

Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
strA	100.00%	804/804	Aminoglycoside resistance Alternate name; aph(3")-Ib	AF321551
strB	100.00%	837/837	Aminoglycoside resistance Alternate name; aph(6)-Id	FJ474091

Beta-lactam

Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
C -M-15	100.00%	876/876	Beta-lactamase resistance Alternate name; UOE-1	DQ302097

Fluoroquinolone

Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
No resistence genes found.				

MLS - Macrolide-Lincosamide-StreptograminB

Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
No resistence genes found.				

Phenicol

Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
No resistence genes found.				

Sulphonamide

Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
sulI	100.00%	840/761	Sulphonemide resistance	AY224185

Tetracycline

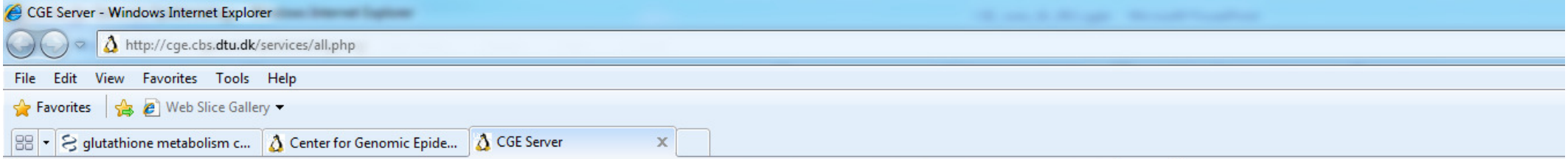
Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
tet(A)	100.00%	1200/1200	Tetracycline resistance	AJ517790

Trimethoprim

Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
dfrA7	100.00%	474/474	Trimethoprim resistance	JF806498

Glycopeptide

Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
No resistence genes found.				



Center for Genomic Epidemiology

Username
Password
New login
Login

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

Open Access Research

BMJ open accessible medical research

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.

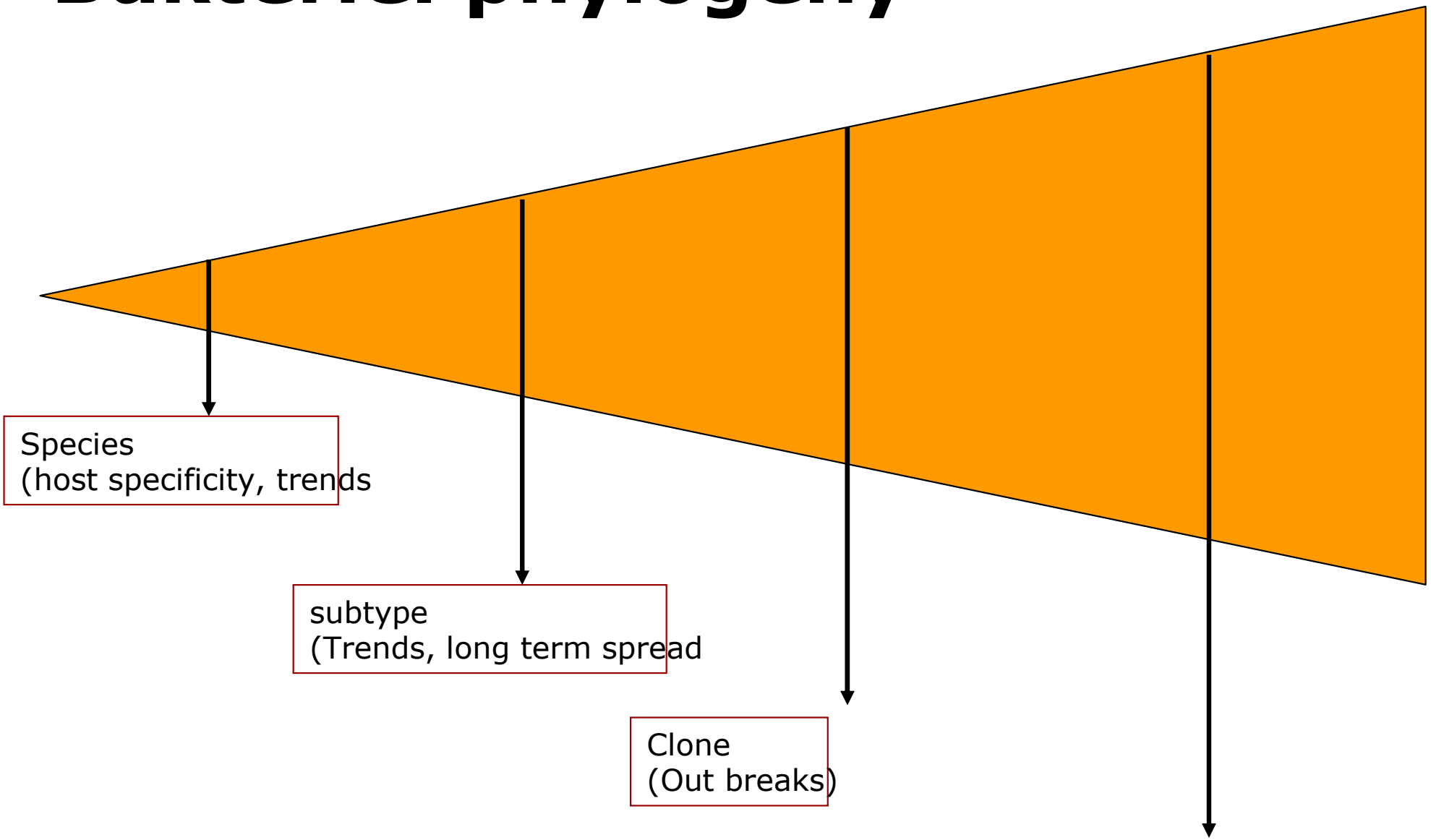
Is a global solution feasible?

- 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)

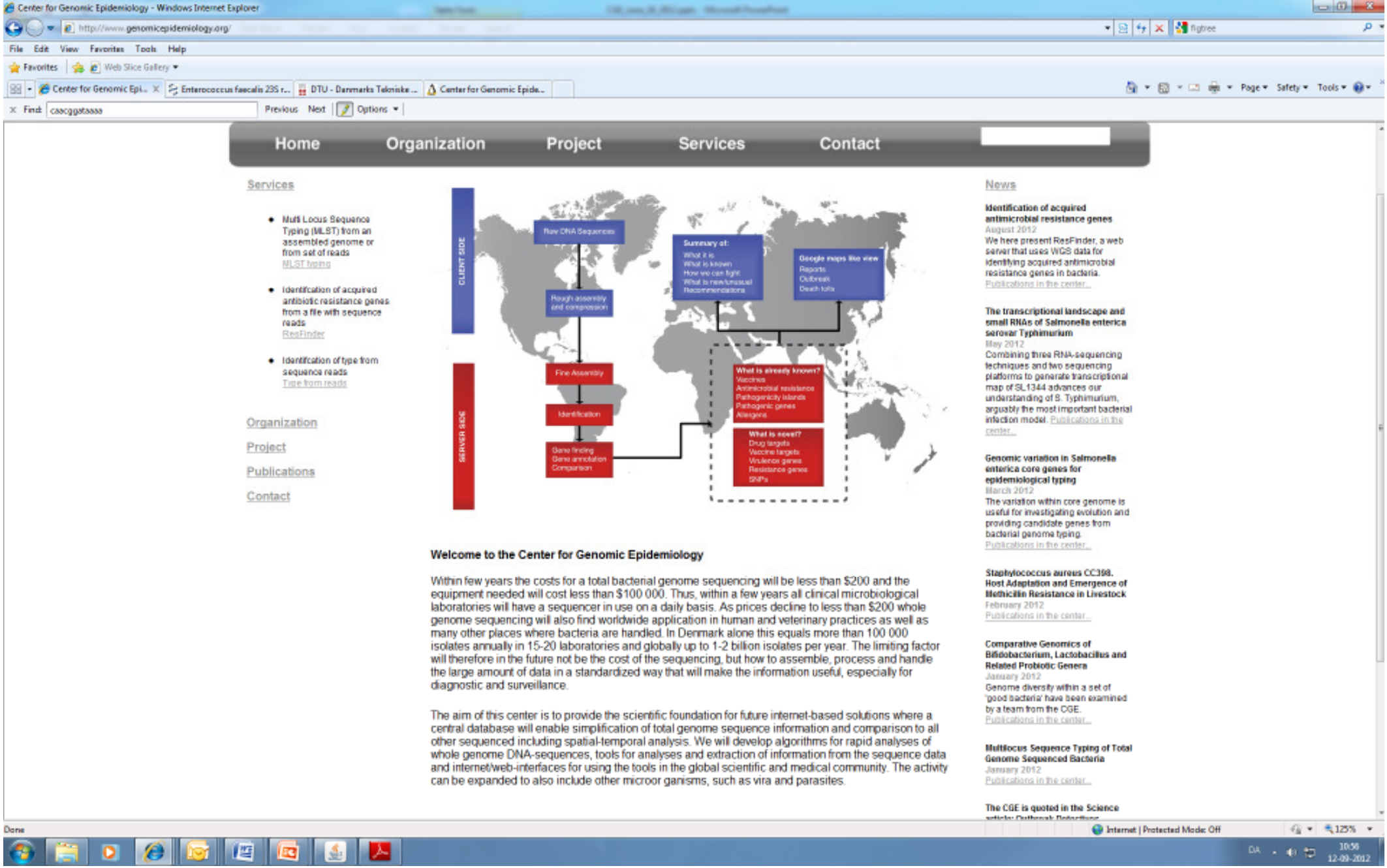
>100 participants covering, governmental institutions, universities, hospitals and companies (FDA, CDC, FBI, NCBI, DoD, USDA, PHAC, Harvard, Maryland, Virginia, Los Alamos, TGen, Aligent, Broad, EBI, eCDC, Sanger, Oxford, DTU, RIVM, FZB, BGI, DDBJ, etc)

- Challenges:
  - Meta-data
  - Publications (selfish scientists)
  - National authorities (need to know before the press)
  - Legally (laywers)


Kolera From UN

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)



www.genomicepidemiology.org