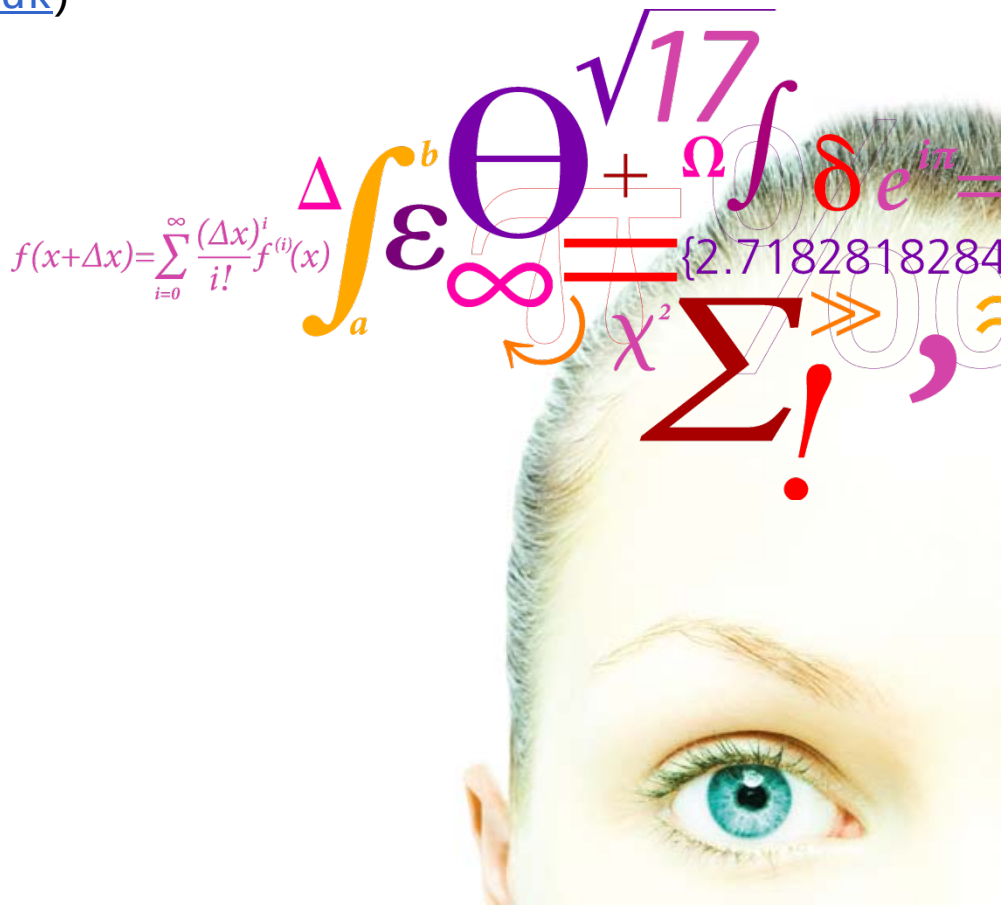
Global, real-time microbiological genomic identification and control project

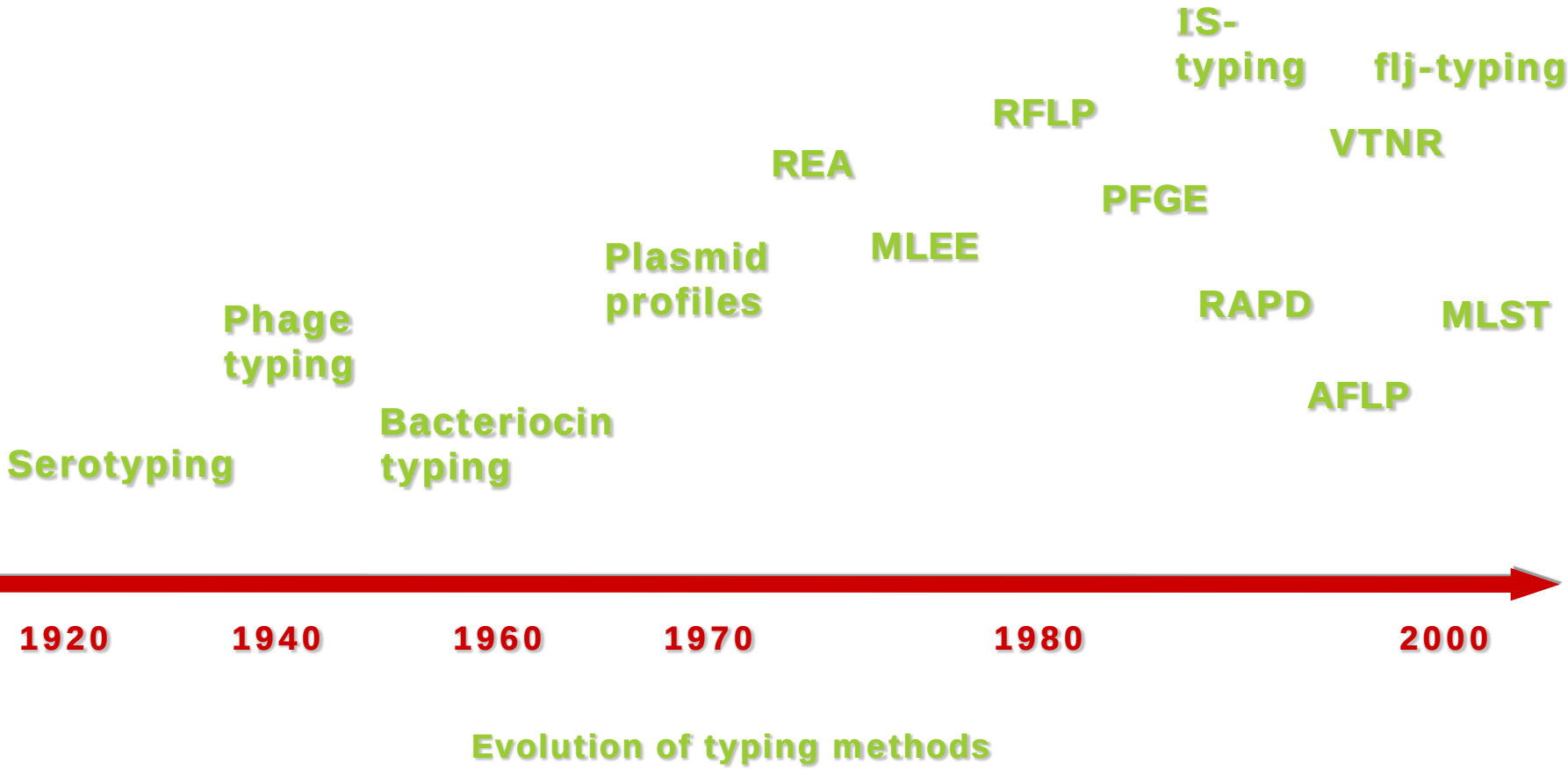
Frank M. Aarestrup (fmaa@food.dtu.dk)



Center for genomic Epidemiology
www.genomicepidemiology.org

DTU Food
National Food Institute

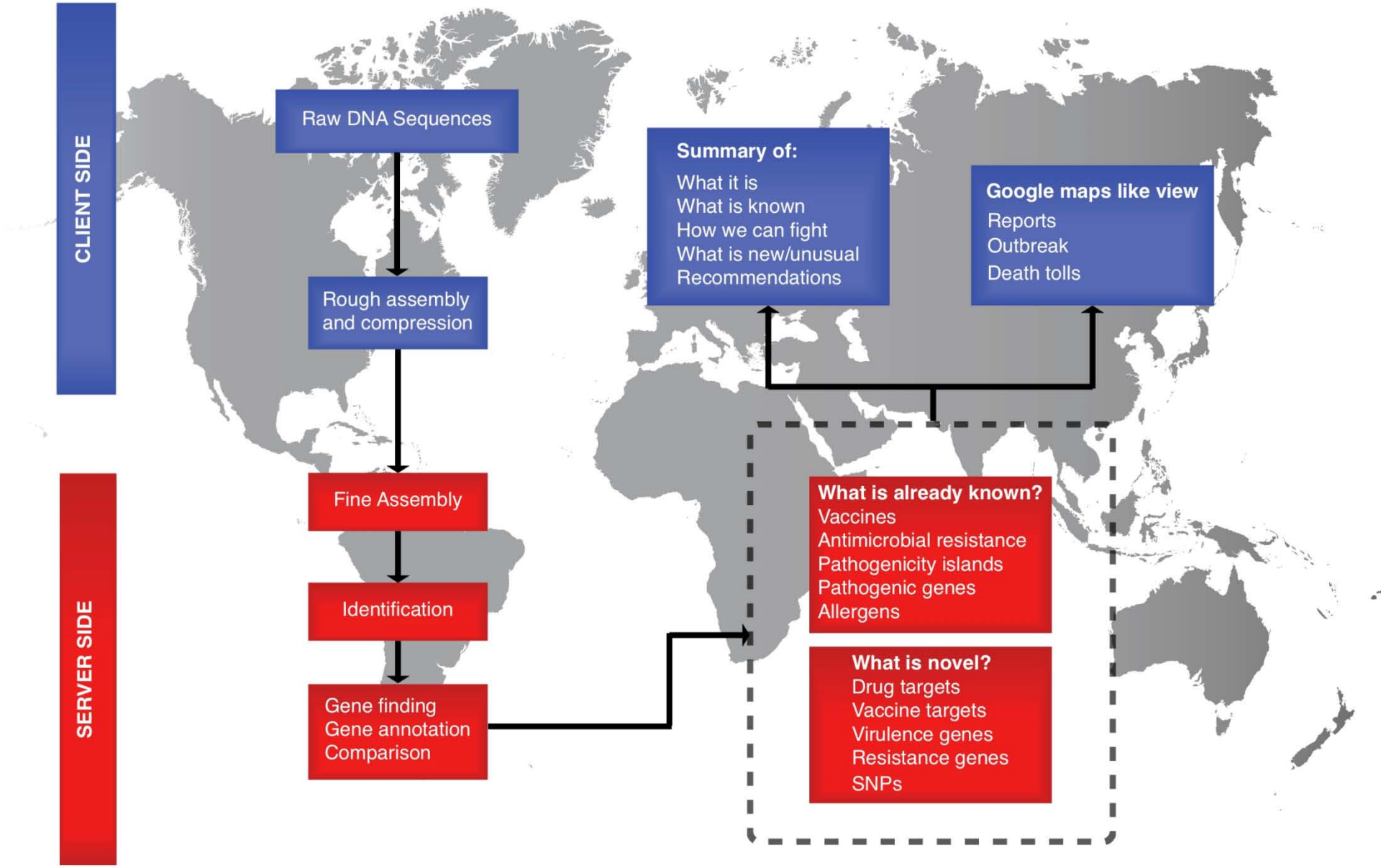
Bacterial identification and typing ...



Evolution of typing methods

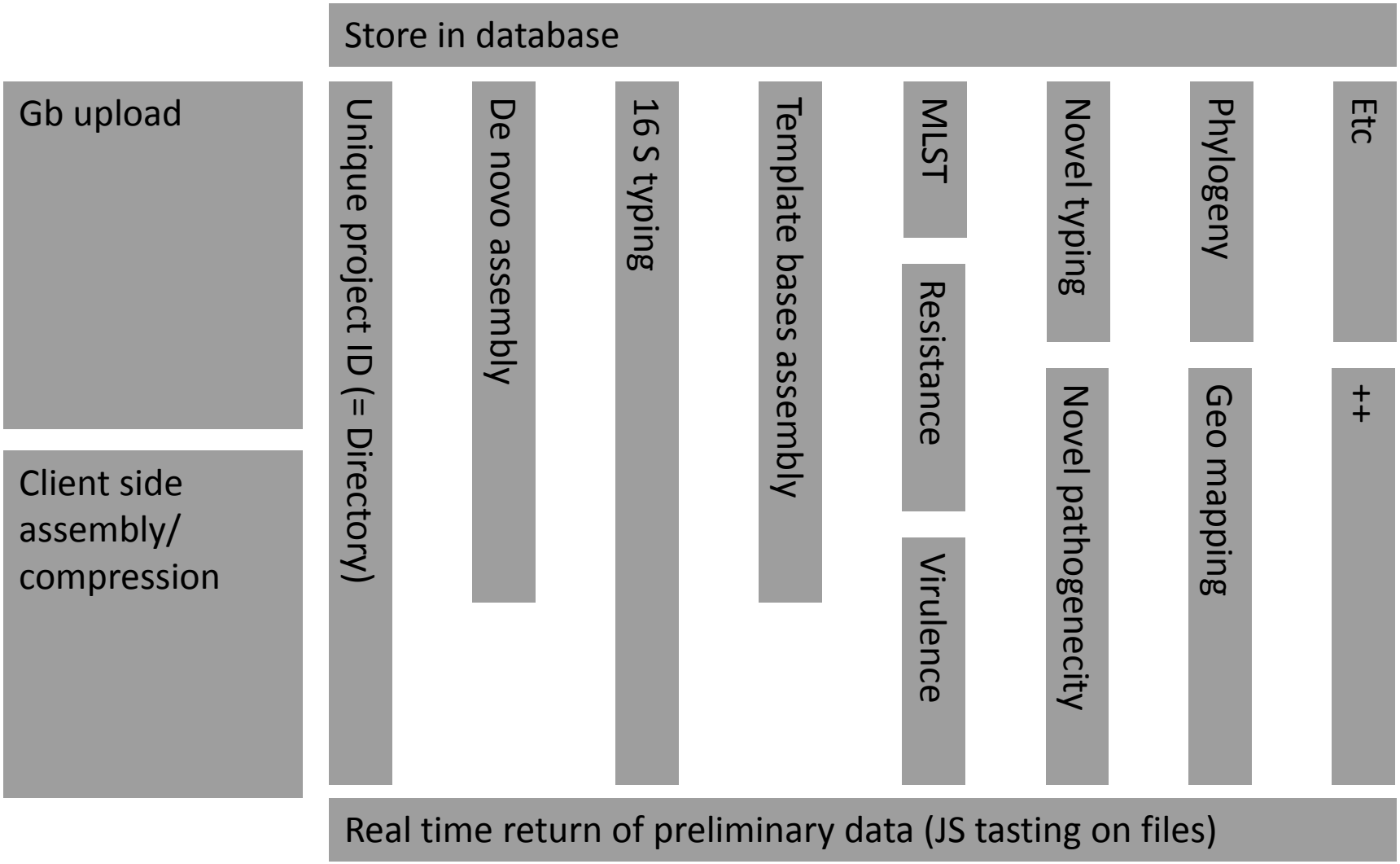
Purpose of Center for Genomic Epidemiology

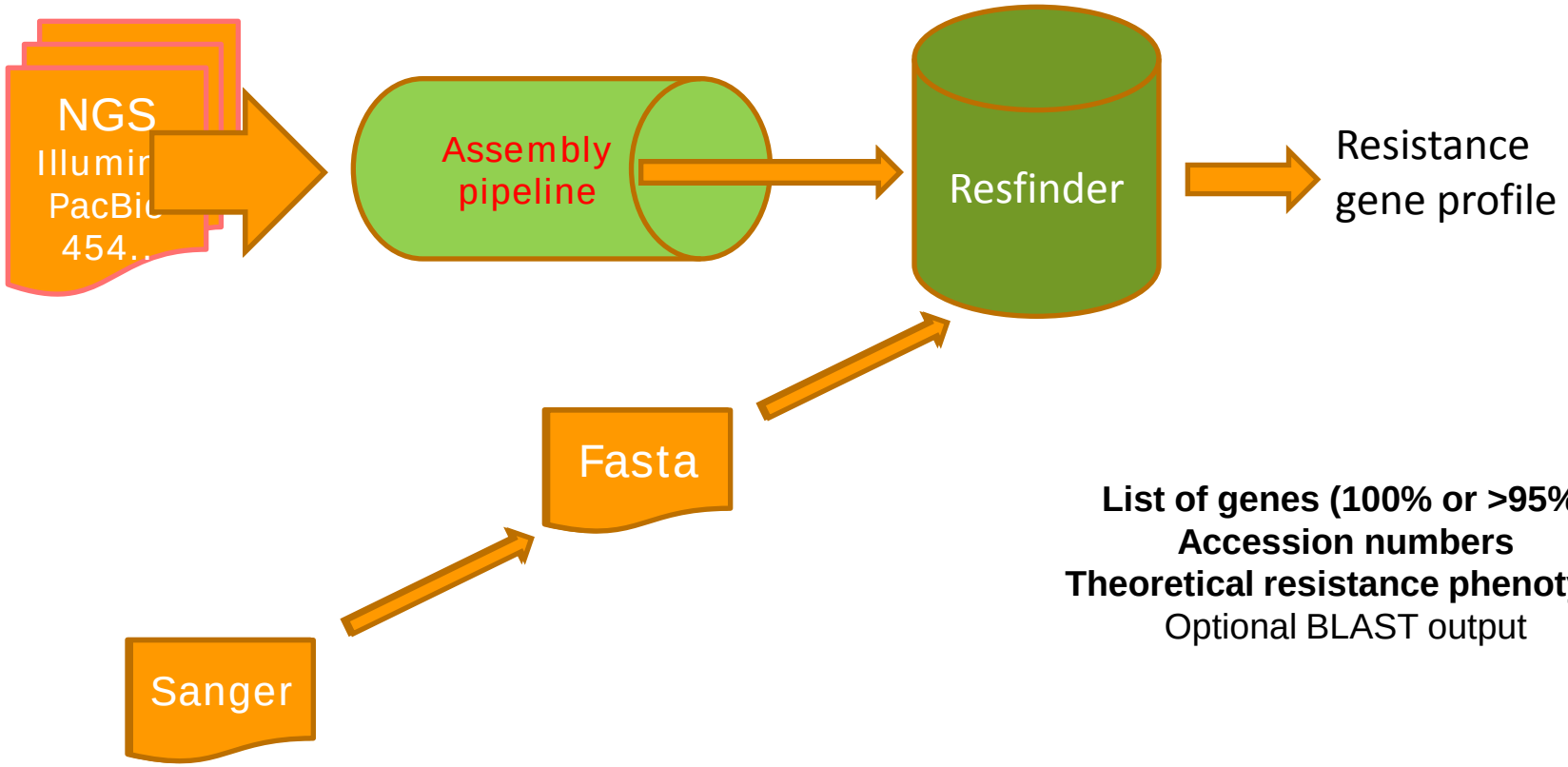
- Provide a proof of concept of combining bioinformatics with global epidemiology in real-time
- And provide a useful facility for frontline users


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

The project

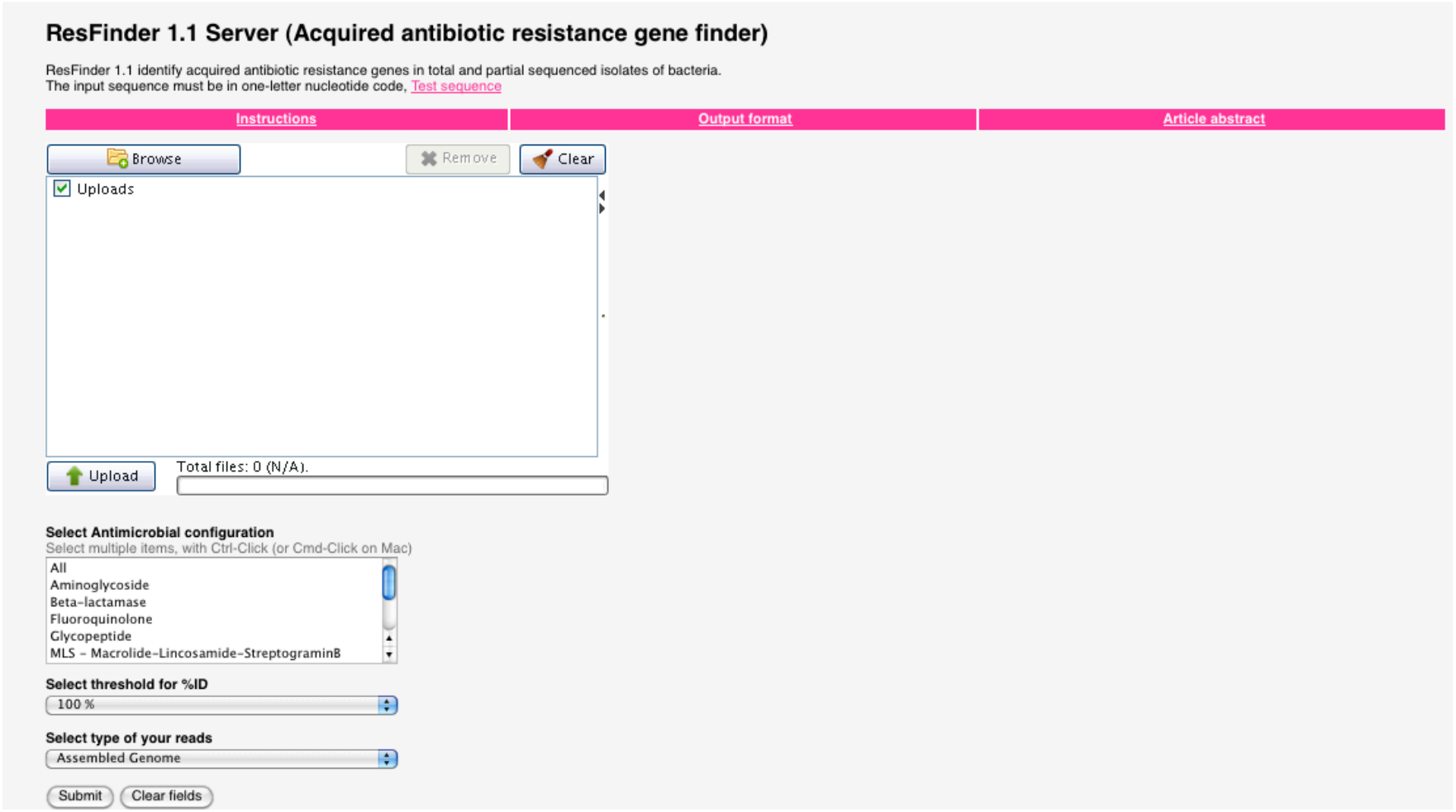
- WP 1 & 2 – Assembly, transfer and identification
- WP 3 & 4 – Rapid analyses for identification and characterization
- WP 5 – Special virulence features (incl. resistance)
- WP 6 – Geospatial
- WP 7 - interface


Store in database
Gb upload
Client side assembly/ compression
Unique project ID (= Directory)
De novo assembly
16 S typing
Template bases assembly
MLST
Resistance
Virulence
Novel typing
Novel pathogenecity
Phylogeny
Geo mapping
Etc
++
Real time return of preliminary data (JS tasting on files)



List of genes (100% or >95%)
Accession numbers
Theoretical resistance phenotype
Optional BLAST output

Front page


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
100 %
Select type of your reads
Assembled Genome
Submit
Clear fields

Results

CENTERFOR BIOLOGICAL SEQUENCE ANALYSIS CBS

ResFinder-1.0 Server - Results

Technical University of Denmark

Input Files: *Kleb-6-1-264y.fa*

ResFinder Results

SETTINGS:
Selected %ID threshold: *100.00*

Beta-lactamase				
Resistence gene	%Identity	Gene Length/HSP Length	Phenotype	Accession number
CTX-M-15	100.00%	876/876	Beta-lactamase resistance Alternate name; UOE-1	DQ302097
Oxa-30	100.00%	831/831	Beta-lactamase resistance	EF543148
Shv-100	100.00%	900/769	Beta-lactamase resistance	AM941846
Shv-28	100.00%	861/861	Beta-lactamase resistance	HM751101
Tem-1	100.00%	861/861	Beta-lactamase resistance Alternate name; RTEM-1	JF910132

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

Explain the output. Go back.

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')

Building global partnership

- Global consensus meetings
  - Bruxelles (Sep. 2011)
  - Washington DC (March 2012)

- Involving partners as curators of *in silico* arrays and/or specific species

- Proof of concept applications and projects
  - Case stories
    - Hospital diagnostics
    - V. cholerae from Nepal
    - MRSA CC398 phylo-evolution
    - Evalution on Campylobacter, E. coli, Salmonella

Washington – 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

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 (198 countries!!!)
  - Repository (NCBI, EBI, DDBJ)
  - Access (Political, legal)
  - Output (simple & advanced, diagnostic & epidemiology, genes & species)
  - Taxonomy

Can whole genome sequencing replace traditional susceptibility testing?

- Current methods
  - Will detect all phenotypic changes including those caused by increased transscription and several genes acting in combination
  - Problems with low and hetero-expression
  - Problems with performance
- Sequencing
  - Will detect what we look for (known genes)
  - Can also detect related genes

Need for development of *in silico* methods for potentialy modelling / predicting phenotypes basd on geno-typs

When will 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

Center for Genomic Epidemiology

http://www.genomicepidemiology.org/

Reader | Google

Homepage o...Lundegaard | Guest user | YouTube | Google Maps | Page via DTV proxy | DTU wireless | SBNInew | Predictors

Center for Genomic Epidemiology

Home | Organization | Project | Services | Contact

Services

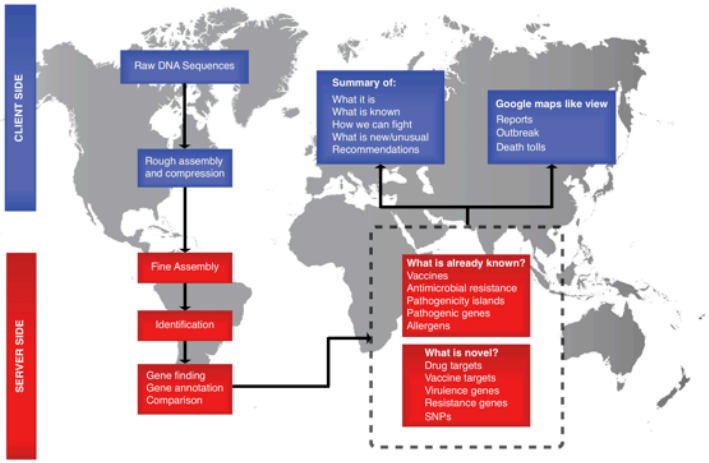
- Multi Locus Sequence Typing (MLST) from an assembled genome or from set of reads
  MLST typing
- Identifcation of acquired antibiotic resistance genes from a file with sequence reads
  Antibiotic resistance

Organization

Project

Publications

Contact



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 microor ganisms, such as vira and parasites.

News

Comparative Genomics of Bifidobacterium, Lactobacillus and Related Probiotic Genera
January 2012
Publications in the center...

Advisory Board Meeting
January 2012
February 2-3 2012, Copenhagen, Denmark.

Characterization of Salmonella enterica serovar Stanley isolates; a serovar endemic to Asia and associated with travel
December 2011
Link to article...

The CGE is quoted in the Science article: Outbreak Detectives Embrace The Genome Era
September 2011
Read more...

Denmark leads the way in the global battle against infectious diseases
September 2011
DTU organized an expert meeting on microbiological genomic identification systems in September in Bruxelles. Information and presentations...

Population Genetics of Vibrio cholerae from Nepal in 2010: Evidence on the Origin of the Haitian Outbreak
August 2011
In combination with evolutionary theory and advanced statistical methods, WGST represents the most powerful molecular approach imaginable and is setting a new standard for infectious disease epidemiology. More...

Defining the Pseudomonas Genus: Where do we draw the line with Azotobacter?
August 2011
Using several different genome-based methods, we find that the Azotobacter is most closely related to Pseudomonas, and perhaps is better understand as a different species within this genus. More...

22 publications

www.genomicepidemiology.org

Report Problems

Copyright DTU 2011 / All rights reserved
Center for Genomic Epidemiology, DTU, Kemitorvet, Building 204, 2800 Kgs. Lyngby, Denmark
Contact: Vibeke Dybdahl Hammer, Telephone: +45 3588 6420, E-mail: vdha@food.dtu.dk
Funded by: The Danish Council for Strategic Research