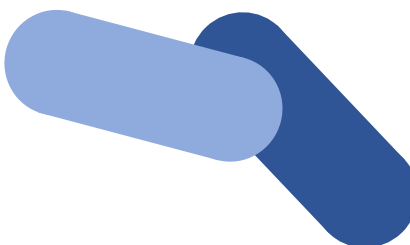




False-colour transmission electron micrograph of a plasmid of bacterial DNA. A plasmid is a molecule of DNA that can exist apart from the chromosome in a bacterium, & which can replicate on its own. This plasmid, designated pSC101, was the carrier molecule in the first genetic engineering experiments, carried out at Stanford University by S. Cohen & H. Boyer. Foreign DNA was spliced to it & the composite or "chimeric" plasmid was then introduced into an *Escherichia coli* bacterium, where it propagated & expressed the genetic information carried by the transplanted DNA. These experiments made possible the biotechnology revolution. Magnification: x47,000 at 6x4.5cm size.



Blockkurs BIO 296

Microbial bioinformatics: sequencing technologies to pathogen analysis

-

Introduction

Prof. Adrian Egli, MD, PhD, FAMH
Institute for Medical Microbiology, University of Zurich

16.03.2023

Email:



aegli@imm.uzh.ch

@AppliedMicrobi2



**Universität
Zürich**^{UZH}



Acknowledgment



Dr. Helena Seth-Smith



Dr. Tim Roloff



Dr. Fanny Wegner

... and the PhD students: Elisa, Srinithi, Zoey, Yukino

“We can always learn something new, provided we believe we can”, Virginia Satir

Goals for this course

- **Hands on** experience with whole genome sequencing using Oxford Nanopore Technologies methods.
- **Comparison** with Illumina data: theory and practise.
- **Bioinformatic analysis** of resulting data using online tools, custom pipelines and licensed softwares.
- **Critical comparison** of results from different sources.
- **Interpretation of results** in relevant / clinical context.
- **Development** of potential diagnostic assay

Please get ready... You are in for a ride!

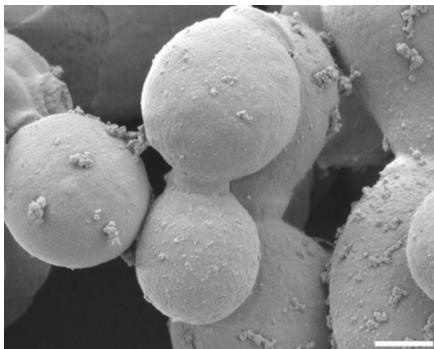


www.youtube.com

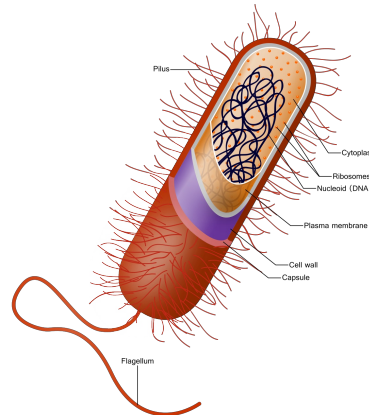
What are the tasks of a microbiological diagnostic laboratory...

Three main tasks:

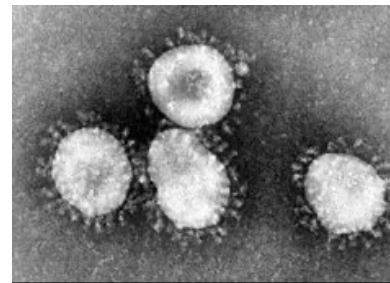
- **Identification** of pathogens (what makes you sick?)
- Testing of **antimicrobial susceptibilities** (what works?)
- Investigation of outbreaks (who infected who?)



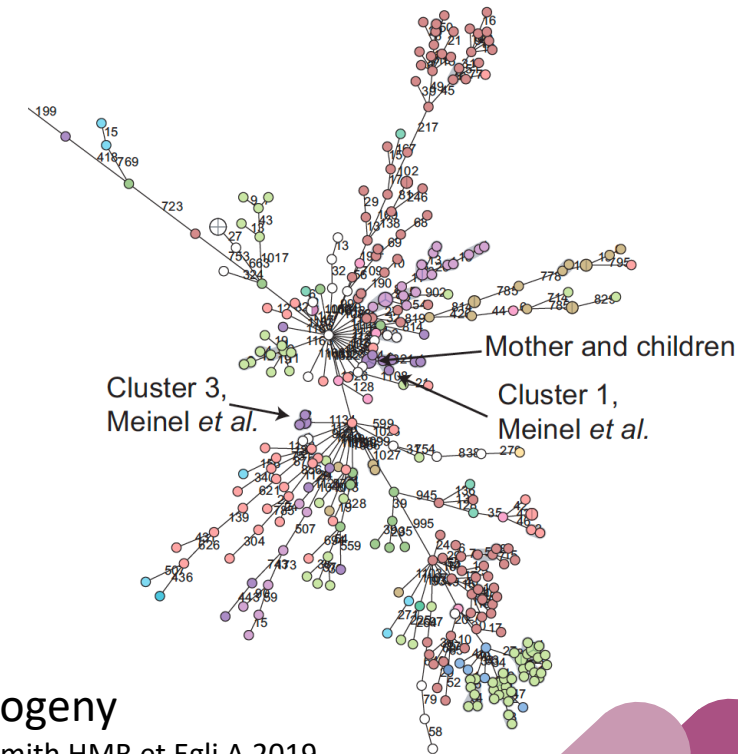
Fungi (5-10 μ m)



Bacteria (1 μ m)



Virus (10 – 400nm)



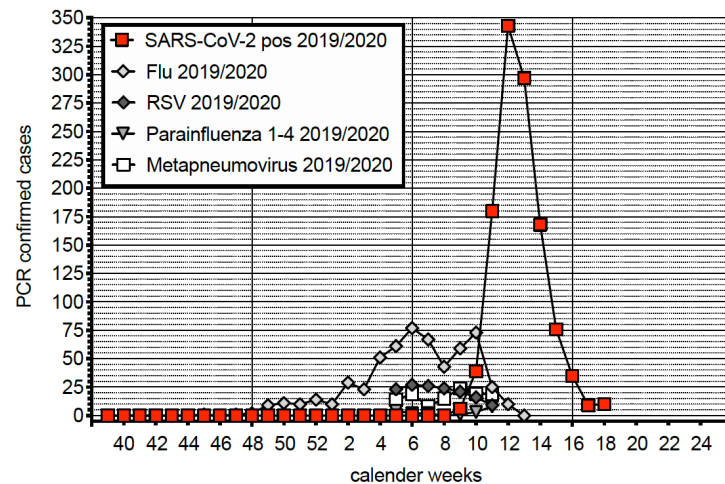
Phylogeny

Seth-Smith HMB et Egli A 2019

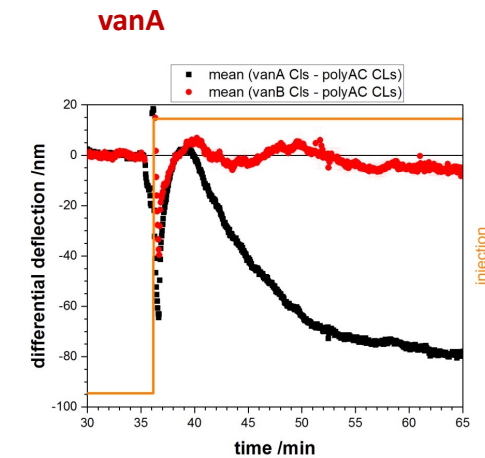
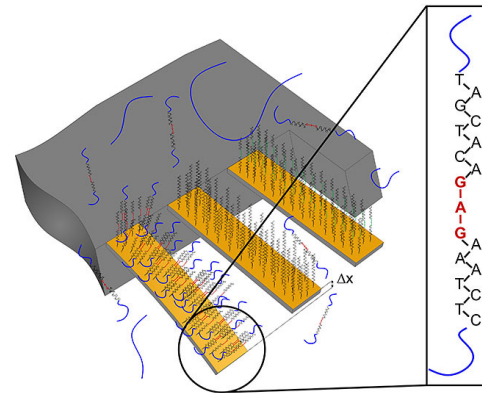
What else does a microbiological diagnostic laboratory

Two more things:

- Epidemiology (Surveillance)
- Development of new tests



Influenza over the year



New diagnostics with an ultra-sensitiven nano-needle

(Cantilever technology, with Dep. Physics University of Basel, E. Meyer & Ch. Gerber)

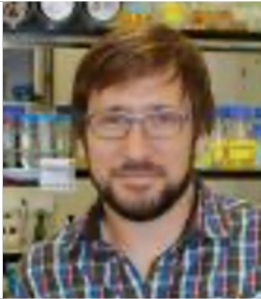
Huber F, Lang HP et al. Biosensors 2022; Huber F, Lang HP et al. Glob Chall 2020

The team: the best of two worlds

Routine diagnostics



Dr. V. Hinic
Bacteriology



Dr. F. Imkamp
Mol. Diagnostics



Dr. B. Schulthess
Mycobacteriology

NGS



Dr. T. Roloff
Tech. leader NGS

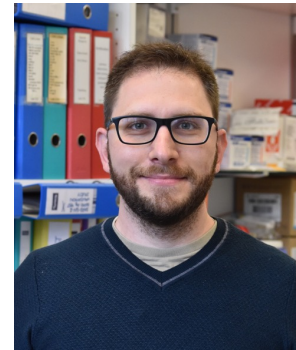


Dr. H. Seth-Smith
Bioinformatics

more staff:

2 microbiologists
3 microbiologists in training
5 technical team leaders

Research group



Dr. Marco Meola
Metagenomics



Dr. Fanny Wegner
Bioinformatics

Current members:

4 senior scientists
5 PhD students
4 laboratory technicians
1 Master student

Focus:

Translational
New technologies
Bioinformatics and AI
Host-pathogen interaction

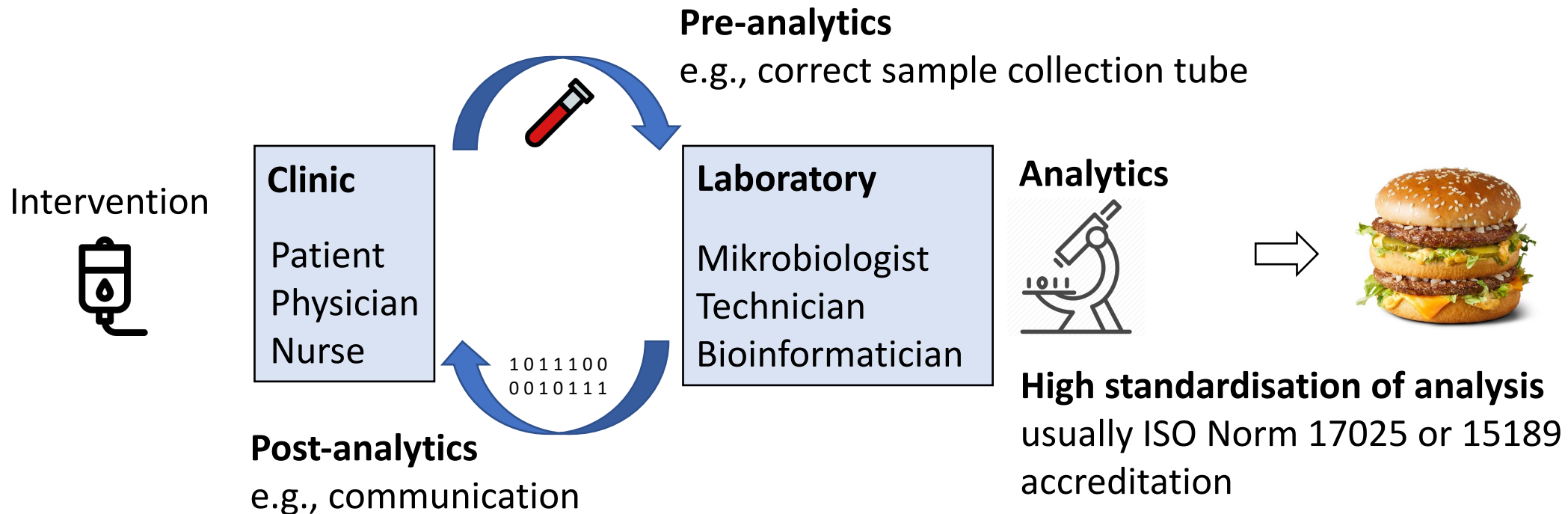
The FAMH Education: A good and interesting job!



Federatio **A**nalyticorum **M**edicorum **H**elveticorum

- „Laboratory Specialist“ with focus on diagnostics (4 years specialization)
- **5 areas:**
 - Medical microbiology (bacteriology, mycology, virology, parasitology)
 - Clinical chemistry
 - Diagnostic hematology
 - Diagnostic immunology
 - Human genetics
- Also „biologists“ can get a FAMH title e.g., molecular biology
- More info: www.famh.ch

Clinics – Laboratory - Clinics

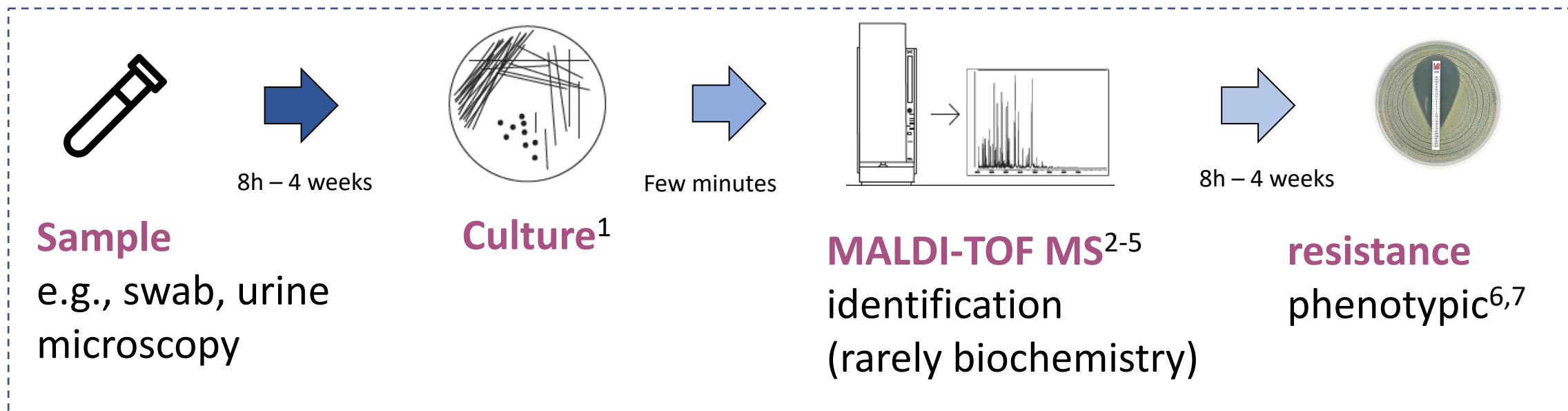


Turn around time: Time from sample arrival in the lab until result leaves the lab

Brain-to-Brain time: Time from Patient examination to intervention



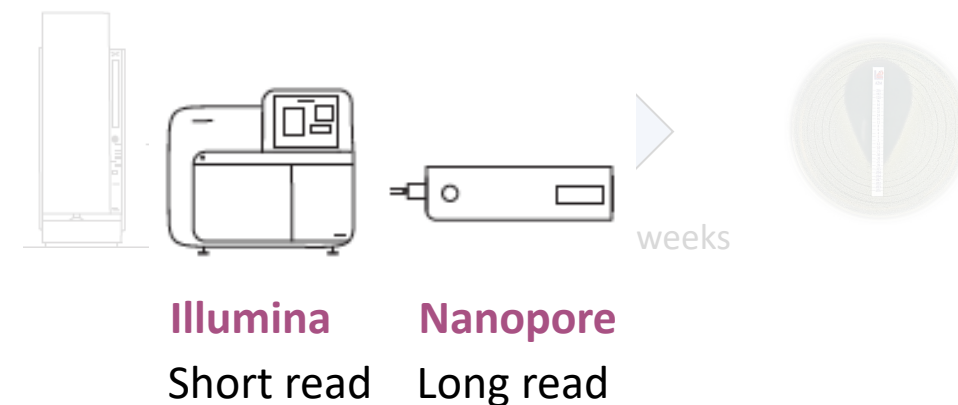
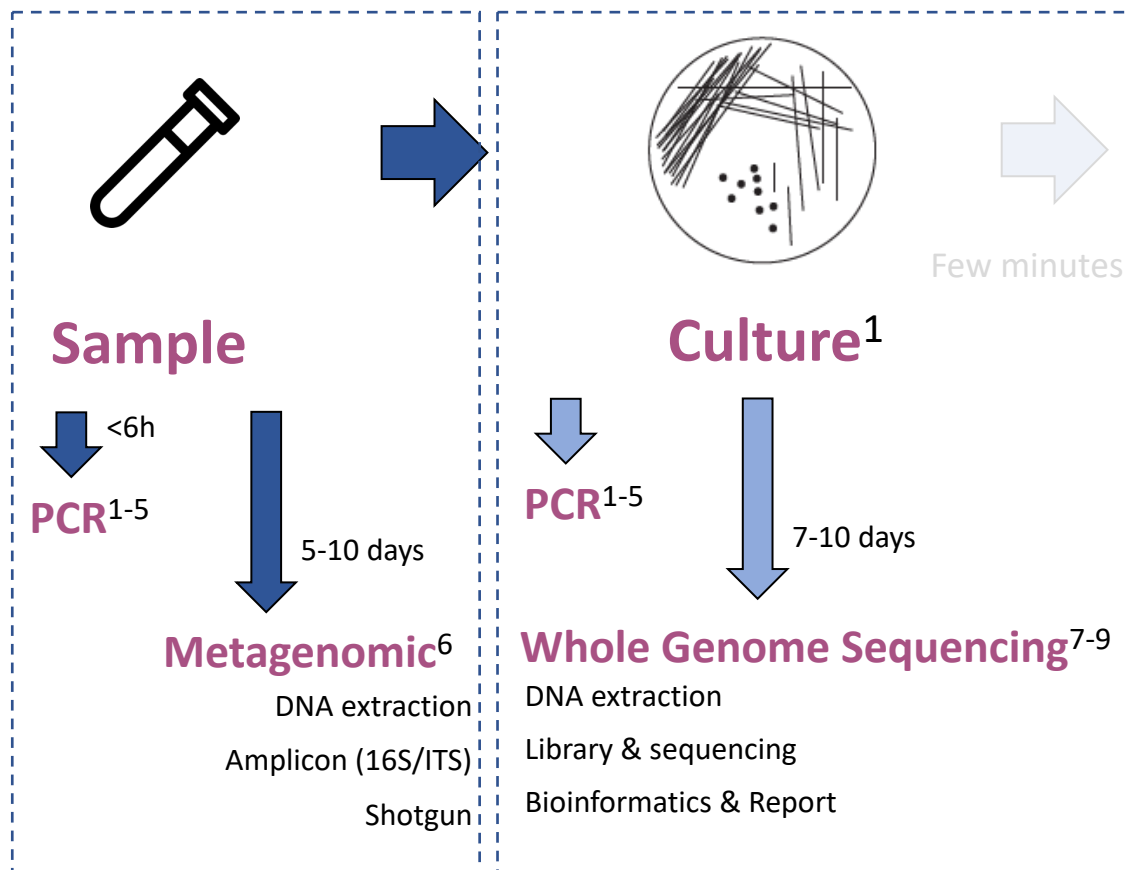
Phenotypic axis of diagnostic



¹ Hinic V, Amrein I, ... et Egli A J Micro Met 2017; ² Dierig A, Frei R, Egli A, Ped Infect Dis J 2015 ³ Egli A et al. Transpl Infect Dis 2015; ⁴ Osthoff M, ... et Egli, A. Clin Microbiol Infect 2017; ⁵ Weis C, ... Egli A, Borgwardt K, Bioinformatics 2020; ⁶ Egli A, Schmid H, et al. Clin Microbiol Infect 2017; ⁷ Hinic V, Reist J, Egli A J Microbiol Met 2018



Genotypic axis of diagnostic



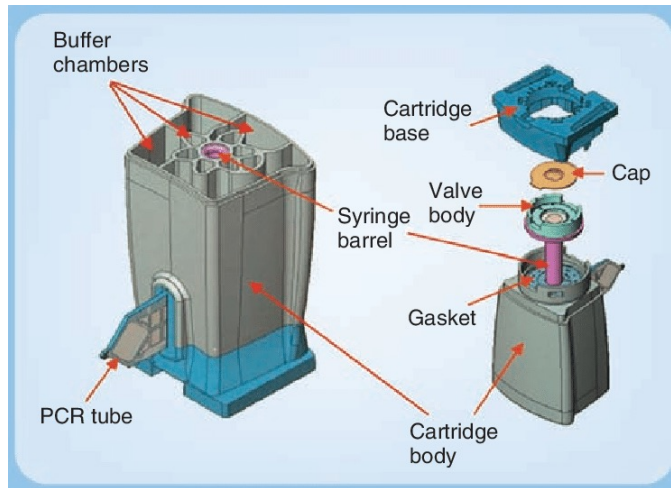
→ **Two sequencing techniques**

¹ Hinic V, Amrein I, ... et Egli A J Micro Met 2017; ² Dierig A, Frei R, Egli A, Ped Infect Dis J 2015 ³ Egli A et al. Transpl Infect Dis 2015; ⁴ Osthoff M, ... et Egli, A. Clin Microbiol Infect 2017; ⁵ Weis C, ... Egli A, Borgwardt K, Bioinformatics 2020; ⁶ Egli A, Schmid H, et al. Clin Microbiol Infect 2017; ⁷ Hinic V, Reist J, Egli A J Microbiol Met 2018

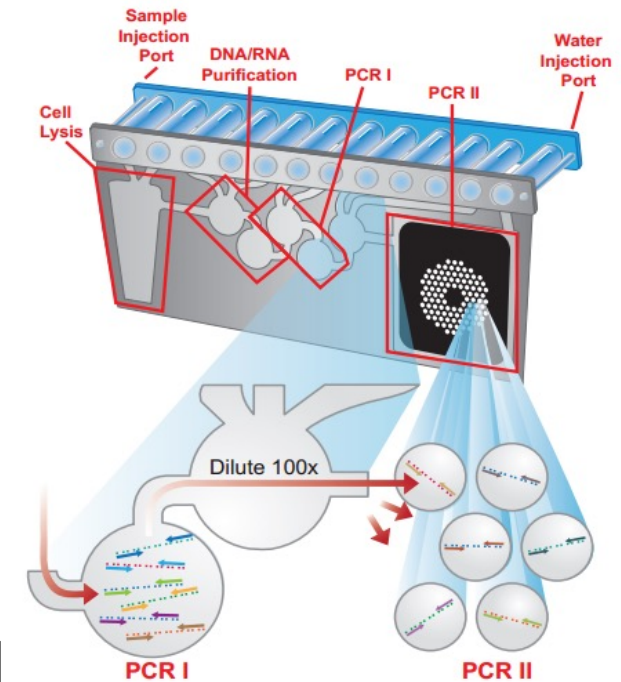
Examples of commercial PCR systems



GeneXpert

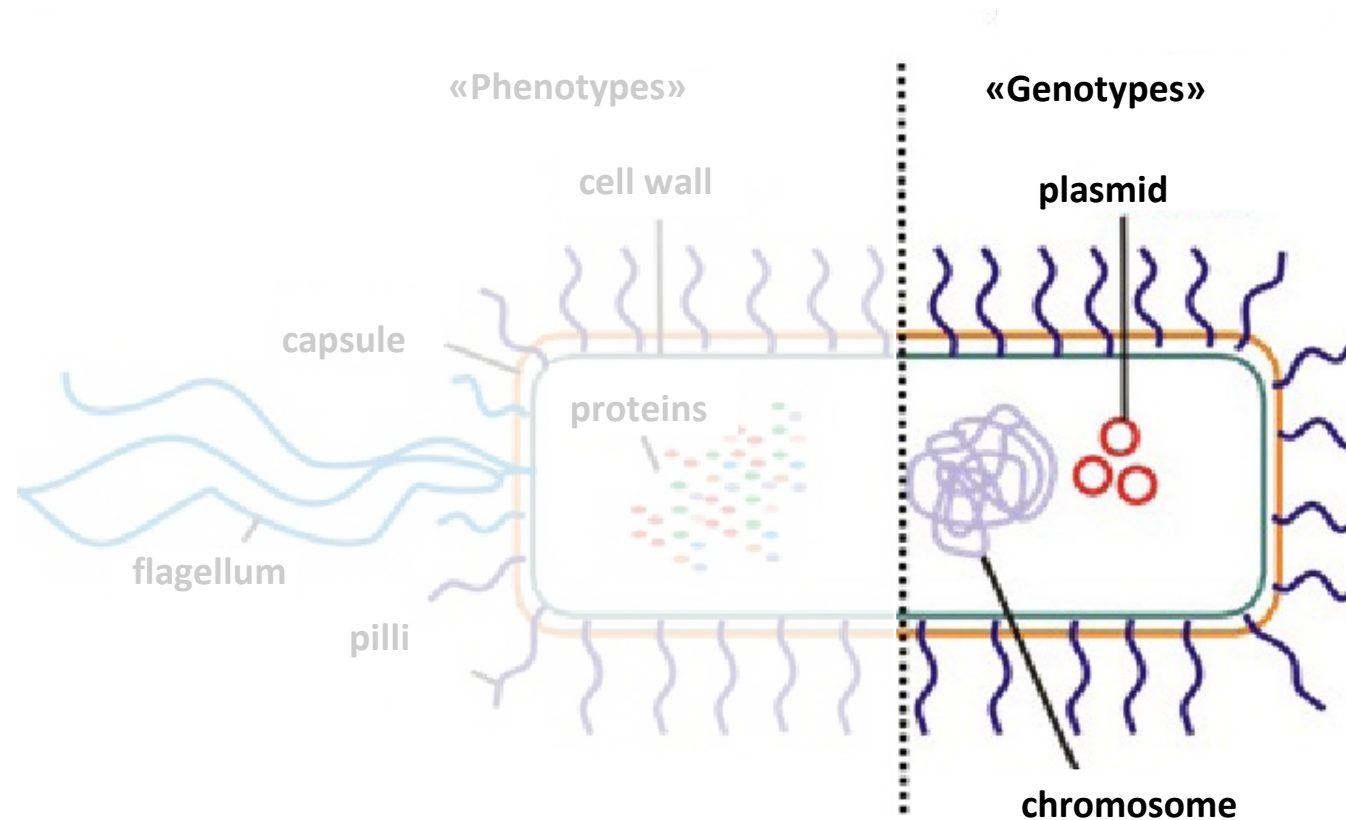


Biofire Filmarray



→ Fast, simple, relatively broad
... but expensive

Focus on the genotype



Why try to infer phenotypes from genotypes

= plan for the Ikea cupboard
→ Imagine how it looks like

How large is the genome of a bacterium?

A: It fits on a A4 page with Arial size 8

B: The genome reaches the moon

C: Has as many characters as "The Lord of the Rings"

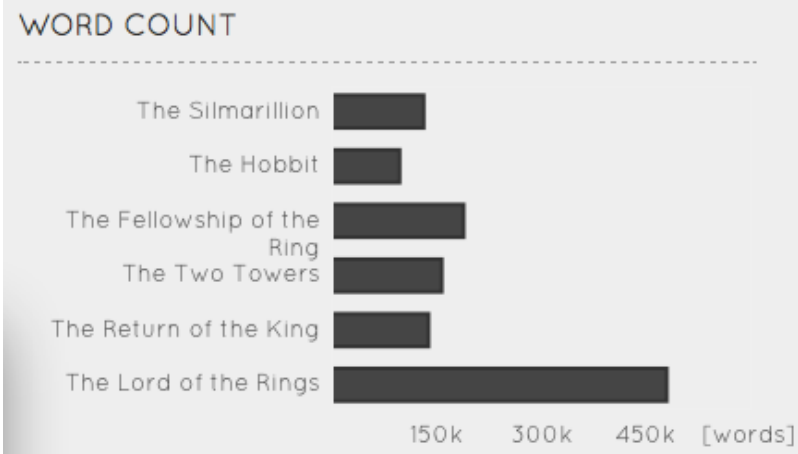
D: Extracted DNA weights 1mg

E: What again is DNA?



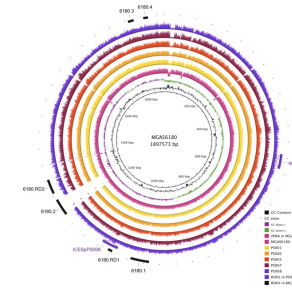
The genome size matches “The Lord of the Rings”

Approx.
3,534,000 characters
on 1,178 pages



An analysis of
Tolkien's books

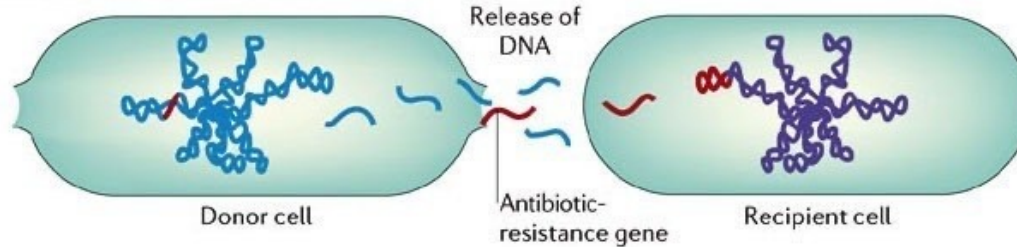
Comparison



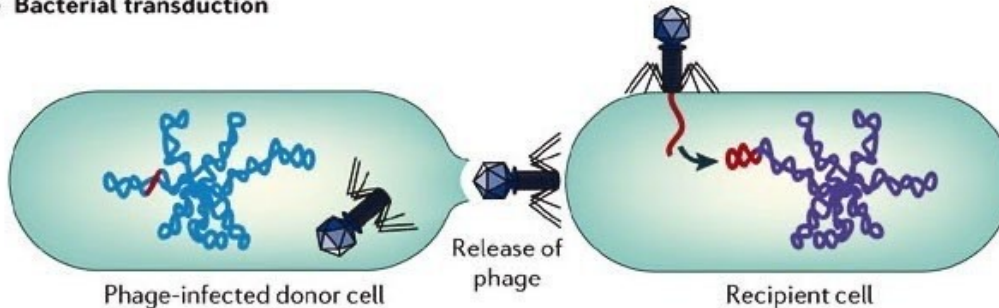
Organism	Num. nucleotides	Num. genes	Reference
HIV1	9,700	9	GenBank: AF033819.3
Influenza A	14,000	11	Jalovaara P et al. Genome Announc 2015
EBV	170,000	80	Tso KKY et al. Infect Agent Cancer, 2013
<i>Helicobacter pylori</i>	1,700,000	1,600	Tomb JF et al. Nature 1997
<i>Staphylococcus aureus</i>	2,900,000	2,700	Baba T et al. J Bacteriol 2008
Lord of the Rings	3,534,000	-	http://lotrproject.com/statistics/books/wordcount
<i>E. coli</i>	4,600,000	4,300	Blattner FR et al. Science 1997
<i>Homo sapiens</i>	3,200,000,000	21,000+	https://www.ncbi.nlm.nih.gov/grc/human
<i>Paris japonica</i> (Pflanze)	150,000,000,000	unknown	Hidalgo O et al. Trends Plant Sci 2017

Variability is essential – gene exchange

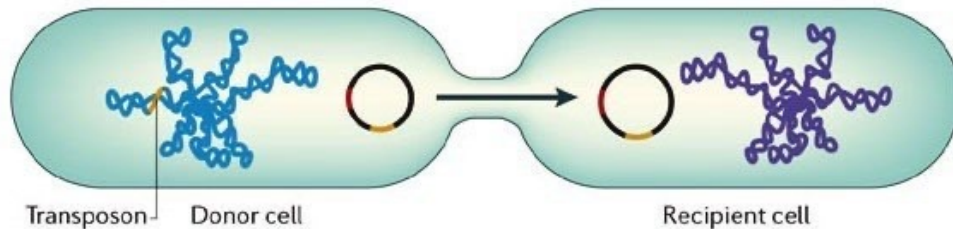
a Bacterial transformation



b Bacterial transduction



c Bacterial conjugation



- Transformation: Transfer of cell-free or naked DNA from one cell to another.
- Transduction: Transfer of genes from one cell to another via a bacteriophage.
- Conjugation: Transfer of genes between cells that are in physical contact with each other.

What is exchanged?

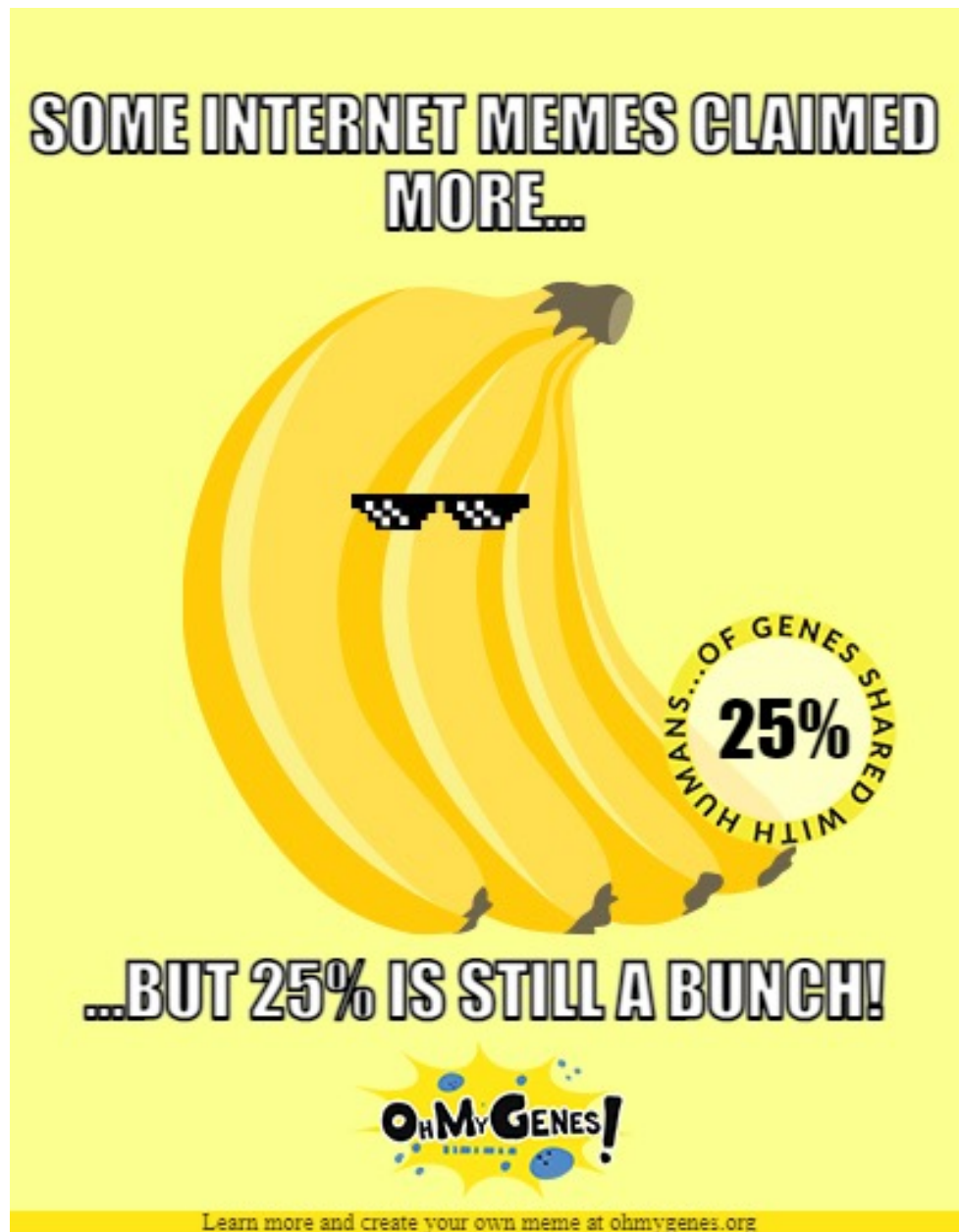
- **Mobile genetic elements:** Plasmids, transposons, and genome fragments
- Plasmid: A plasmid is a small, ring-shaped, extrachromosomal and usually double-stranded DNA molecule found in many bacteria. It contains additional genetic information and is replicated autonomously.
- Transposon: Transposons are coding DNA sequences that can change their location within the DNA (so-called 'jumping genes'). Transposons occur in all organisms. They can contain one or more genes.
- These mobile genetic elements **confer important properties**:
 - Antibiotic resistance and pathogenicity and virulence factors.
 - E.g., penicillinases (antibiotic cleavage)
 - E.g., adhesins and invasins for adherence to host cells

Diagnostic usage: Genotypic characteristics

- **Detection of a single gene** (identification or resistance)
 - PCR e.g., *blaZ* or diphtheria toxin
 - Isothermal amplification e.g., *Oxa48*
 - **Ribosomal rRNA** (identification)
 - 16S ribosomal RNA gene (bacteria)
 - ITS1/2 gene (fungi)
 - Multi-locus Sequencing typing
 - Core genome multi-locus sequencing typing
 - Whole Genome Sequencing typing
- } typing (e.g., transmission studies)

Why do we want to sequence the whole genome?

- Multi-locus Sequencing typing
 - Core genome multi-locus sequencing typing
 - Whole Genome Sequencing typing
- } typing (e.g., transmission studies)
- Answer important questions in biology and medicine!
e.g., how much of a Banana are you?



OMG!

~~Oh my god!~~

Oh **My Genes!**

- Banana^{1,2}: 25% similarity, 36,439 genes,
- Common ancestor 1.5 billion years
- 24th November “Happy Evolution Day”
(Charles Darwin’s On the Origin of Species)²

If you want to learn more about evolution of humans and bananas:

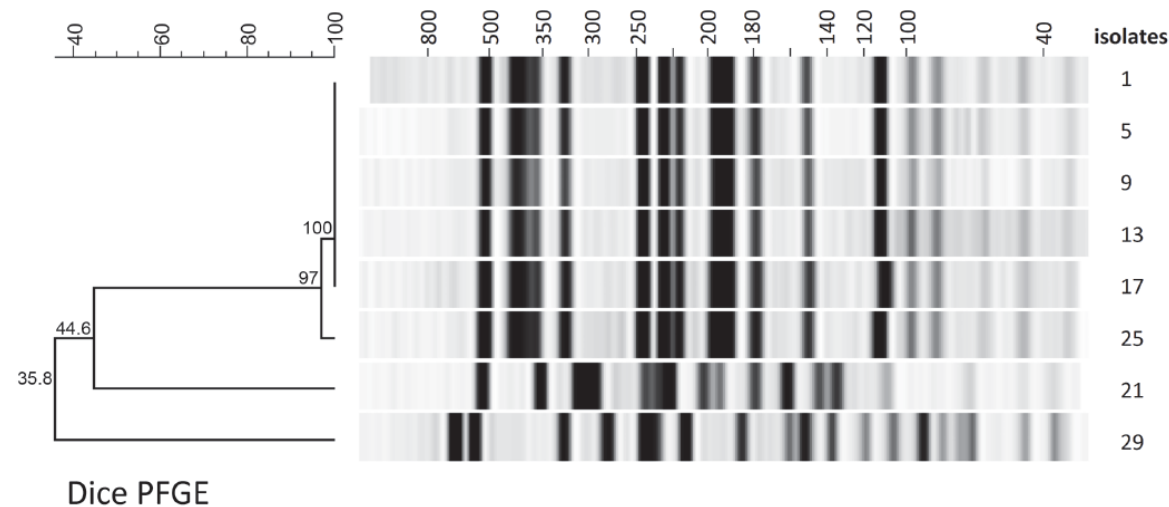
<https://lightofevolution.org/en/banana-split/>

1 D’Hont A, Denoeud F, et al. Nature 2012:

2 <https://lightofevolution.org/en/banana-split/>

How we used to type...

- Comparison of band patterns using “**Pulsed Field Gel Electrophoresis**”¹
- Algorithm impacts the phylogeny²
- Few databases available³ and was the reference standard for many years^{4,5,6}
 - Disadvantage: low reproducibility between centers⁷



Egli A et al. Plos one, 2015

¹ Tenover FC, et al. J Clin Microbiol, 1995; ² Duck WM et al. J Clin Microbiology, 2003; ³ McDougal LK, et al. J Clin Microbiology, 2003; ⁴ Arbeit RD et al. J Infect Dis 1990; ⁵ Prevost G et al. J Hosp Infect 1991; ⁶ Tosh PK et al. Infect Control Hosp Epidemiol 2011; ⁷ van Belkum et al. J Clin Microbiology, 1998



Further typing methods

- “Amplified fragment length polymorphisms” (**AFLP**)¹
- “Random amplification of polymorphic DNA” und “arbitrarily primed polymerase chain reaction” (**RAPD**)²
- “Repetitive-element polymerase chain reaction” (**rep-PCR**) e.g. DiversiLab^{3,4}
- “Variable-number tandem repeat” (**VNTR**) Typisierung⁵
- “Single locus sequence typing” (**SLST**)⁶
- “Staphylococcus aureus protein A gene”-typing (**spa**)^{7,8}
- “Multi-locus sequence typing” (**MLST**)^{9,10}

1 Vos P et al. Nucleic Acids Res 1995; 2 Lanini S et al. Plos one 2011; 3 Versalovic J et al. Methods Mol Cell Biol. 1994; 4 Babouee B et al. J Clin Microbiol 2011; 5 Sabat A et al. J Clin Microbiol 2003; 6 Beall B et al. J Clin Microbiol 1996; 7 Frenay HM et al. Eur J Clin Microbiol Infect Dis 1996; 8 Malachowa N et al. J Clin Microbiol 2005; 9 Selander RK et al. Appl Environ Microbiol 1986; 10 Maiden MC et al. PNAS 1998;

Evolution of sequencing technologies

MLST

Sanger Sequencing
(first generation)



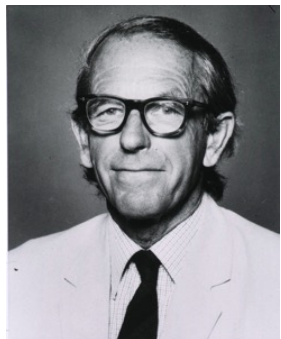
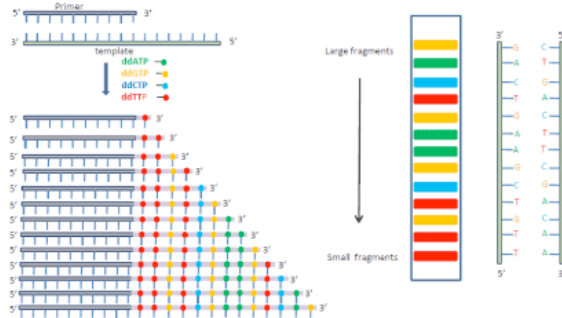
cgMLST

Short read sequencing
(next generation)

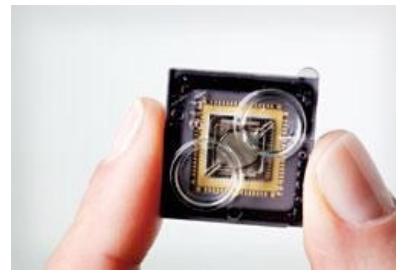


cgMLST & more

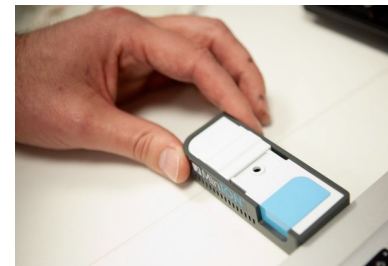
Long read sequencing
(next next generation)



Frederick Sanger
(1918-2013)



IonTorrent



MinIon



MiSeq Illumina



PacBio


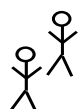


Why does resolution matter?

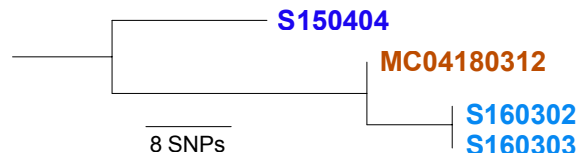
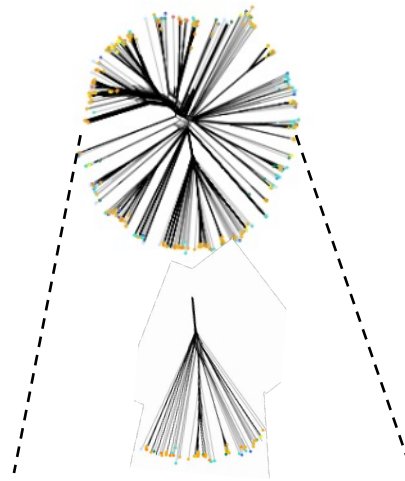
- Because if we do not care, then it will always be the chicken's fault!



This picture is not from Switzerland, but in Switzerland per day about 200,000 chickens are killed for meat production.

Resolution matters!

MLST 67% (232/344)	
cgMLST ≤20 loci difference 19% (65/344)	
SNP Probable ≤ 4 SNPs 10% (33/344) Probable + Possible ≤ 12 SNPs 14% (46/344)	 

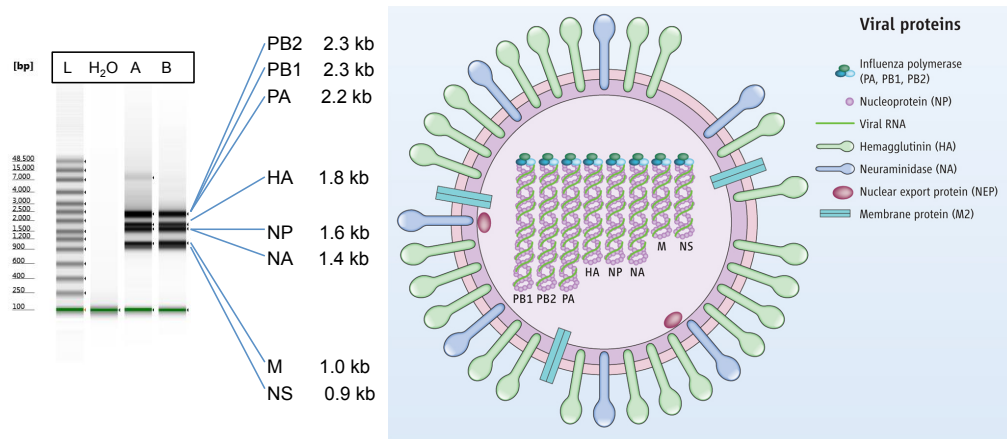


Resolution:
MLST < cgMLST < SNP

→ Chicken is not the only source
of *Campylobacter* spp.
infection

Reist J, Seth-Smith HMB, Egli A, unpublished

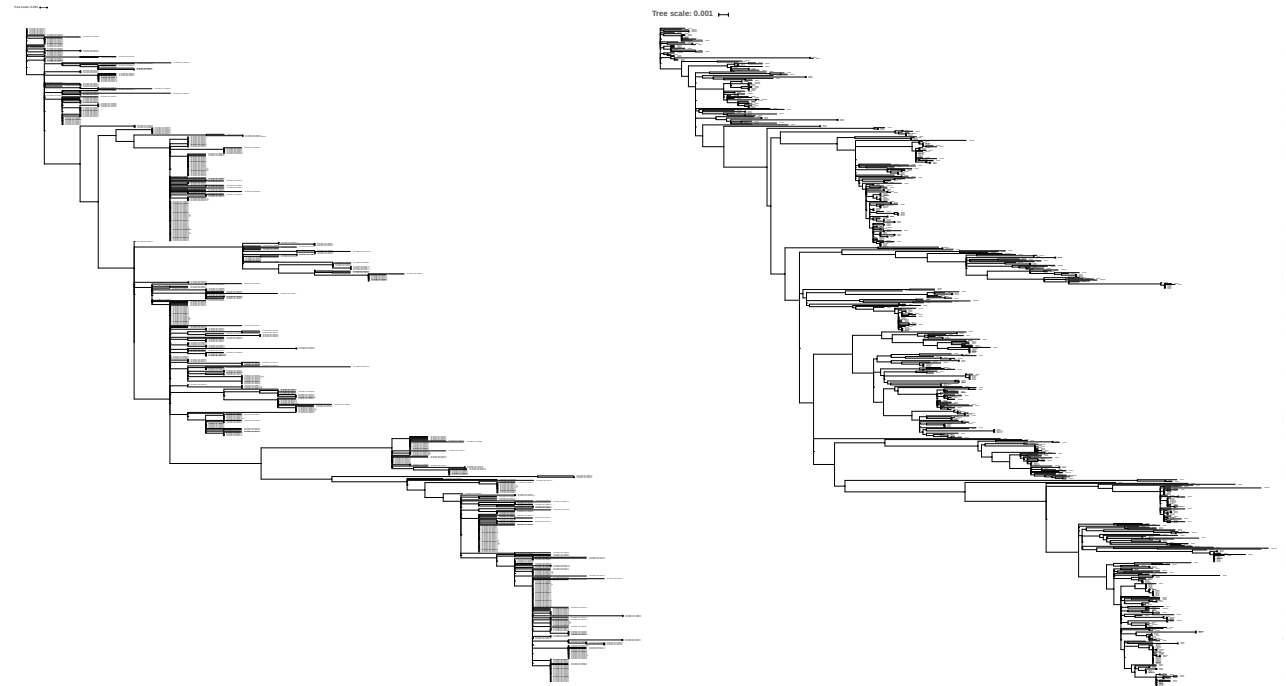
With more genetic information, the resolution increases



whole genome size of 13.5 kb

Influenza transmission in the City of Basel.

Mueller NF, Wüthrich D, et al. Plos Pathogen 2020



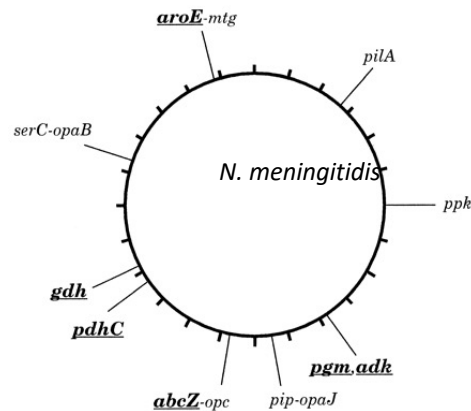
Haemagglutinin

n=766

alle 8 segments

Multi-locus sequencing typing

- Comparison of **6-7 core genes**¹; generation of data with Sanger sequencing
- Dependent on the bacterial species
 - E.g. *Neisseria meningitidis*: *abcZ*, *adk*, *aroE*, *gdh*, *pdhC*, and *pgm*²
 - E.g. *Staphylococcus aureus*: *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqi*³
 - Typing scheme: bacteria n=96 and eucaryotes n=9⁴

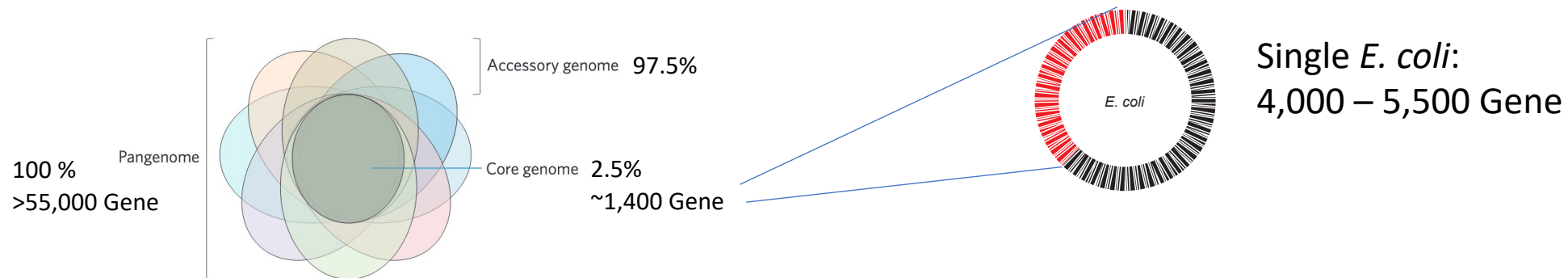


¹ Urwin R and Maiden MCJ, Trends in Microbiology, 2003; ² Maiden MCJ et al. PNAS, 1998;

³ Jolley KA et Maiden MCJ, BMC Bioinformatics, 2010; ⁴ pubmlst.org/databases

The core genome

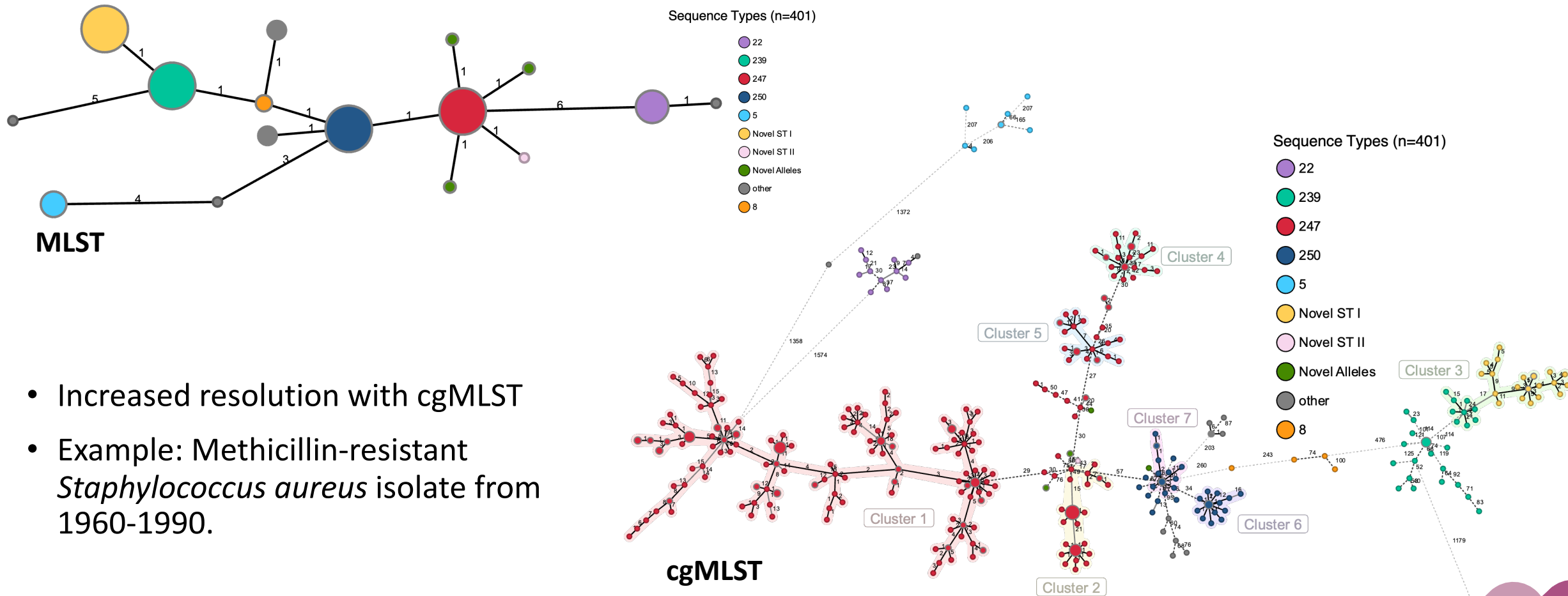
- The core genome of a bacterial species is the commonly shared list of genes.
- The core genome is according to the species a few 100 to 1500 genes.
- Example:
 - *E. coli* has a huge pan-genome^{1, 2} (Exchange with other *E. coli*) with 1400 core genes.



- *Chlamydia trachomatis* has a small core genome with 888 genes³

¹ McInerney JO et al. *Nature Microbiology* 2, no. 4 (March 28, 2017); ² Horesh G. et al. *Microbial Genomics*, 15 (2021); ³ Versteeg B, et al. *BMC Genomics* 2018

Difference in resolution: MLST vs cgMLST



- Increased resolution with cgMLST
- Example: Methicillin-resistant *Staphylococcus aureus* isolate from 1960-1990.

Benvenga V et al. unpublished

Clinical applications of NGS

- **Typing of outbreaks**
 - Methicillin resistant *Staphylococcus aureus*¹⁻⁵
 - Vancomycin resistant *Enterococcus faecium*⁶⁻⁷
 - Carbapenemase producing Enterobacteriaceae⁸⁻¹⁰
 - *Mycobacterium tuberculosis*^{11,12,13}
- **Virulence factors**^{14,15,16}
- **Resistance profiles**^{17,18,19}
 - known resistance mechanisms
 - new or unknown resistance mechanisms

¹ Mellmann et al. JCM 2016; ² Bosch et al. EuroSurv 2016; ³ Köser et al. NEJM 2012; ⁴ Tong et al. Genome Res 2015; ⁵ Stegger et al. Mbio 2014; ⁶ McGann et al. Diagn Microbiol ID 2016; ⁷ Broderick et al. Genome Med 2016; ⁸ Marsh et al. PlosOne 2015; ⁹ Jiang et al. CMI 2015; ¹⁰ Mathers et al. AAC 2015; ¹¹ Stucki et al. JCM 2016; ¹² Merker et al. Nat Genet 2015; ¹³ Comas et al. Nat Genet 2011; ¹⁴ Nunvar et al. Plos one 2016; ¹⁵ Meinel et al. CMI 2016; ¹⁶ Lindsay et al. Front Microbiol 2016; ¹⁷ Walker et al. Lancet ID 2015; ¹⁸ Metcalf et al. CMI 2016; ¹⁹ Hornsey et al. JAC 2011

Who you gonna call?

- Bioinformatician¹
- Data scientists
- New professions in the modern diagnostic laboratory.
- Analysis of complex and large data with statistical approaches and visualization tools.
- Swiss Institute of Bioinformatics (SIB; <https://www.sib.swiss/>)



Google Search Term: “Bioinformatiker”

“To understand God's thoughts, we must study statistics, for these are the units of measure, that illustrate his intentions”, Florence Nightingale

1 Vincent AT et Charette SJ, Front in Genetics 2015

A vision for the lab

→ Microbiology of the future

- “Hypothesis free diagnostics” and “digital microbiology”¹
- Large, well-curated databases, which allow to interpret the data
- Step 1: Combination of bacteria and viruses (DNA ist DNA).
- Step 2: Prediction of functions
- Step 3: Combination of pathogen and host
 - Characterizing the hosts response z.B. inflammatory patterns, prognosis
- Step 4: Expression of virulence, resistance and immunity
 - Transcriptomics

¹ Egli A, Schrenzel J, Greub G; CMI 2021



Take home message

- The routine diagnostic laboratory focuses on three tasks:
identification, susceptibility tests of antimicrobial drugs, and typing
- Genotypes can be determined with molecular techniques e.g. PCR, but also sequencing
- The workflow is high standardized and regulated to reach accreditation quality
- The “turn-around-time” and “brain-to-brain” time is critical -> Actionability
- Resolution matters for the interpretation of transmission events
MLST << cgMLST < WGS

Thank you for your attention

- Contact:

Prof. Adrian Egli, MD, PhD, FAMH

Institut für Medizinische Mikrobiologie, Universität Zürich

Email: aegli@imm.uzh.ch



@AppliedMicrobi2



**Universität
Zürich**^{UZH}

