

wellcome  
connecting  
science

ACORN.

# Typing

May 30th 2024

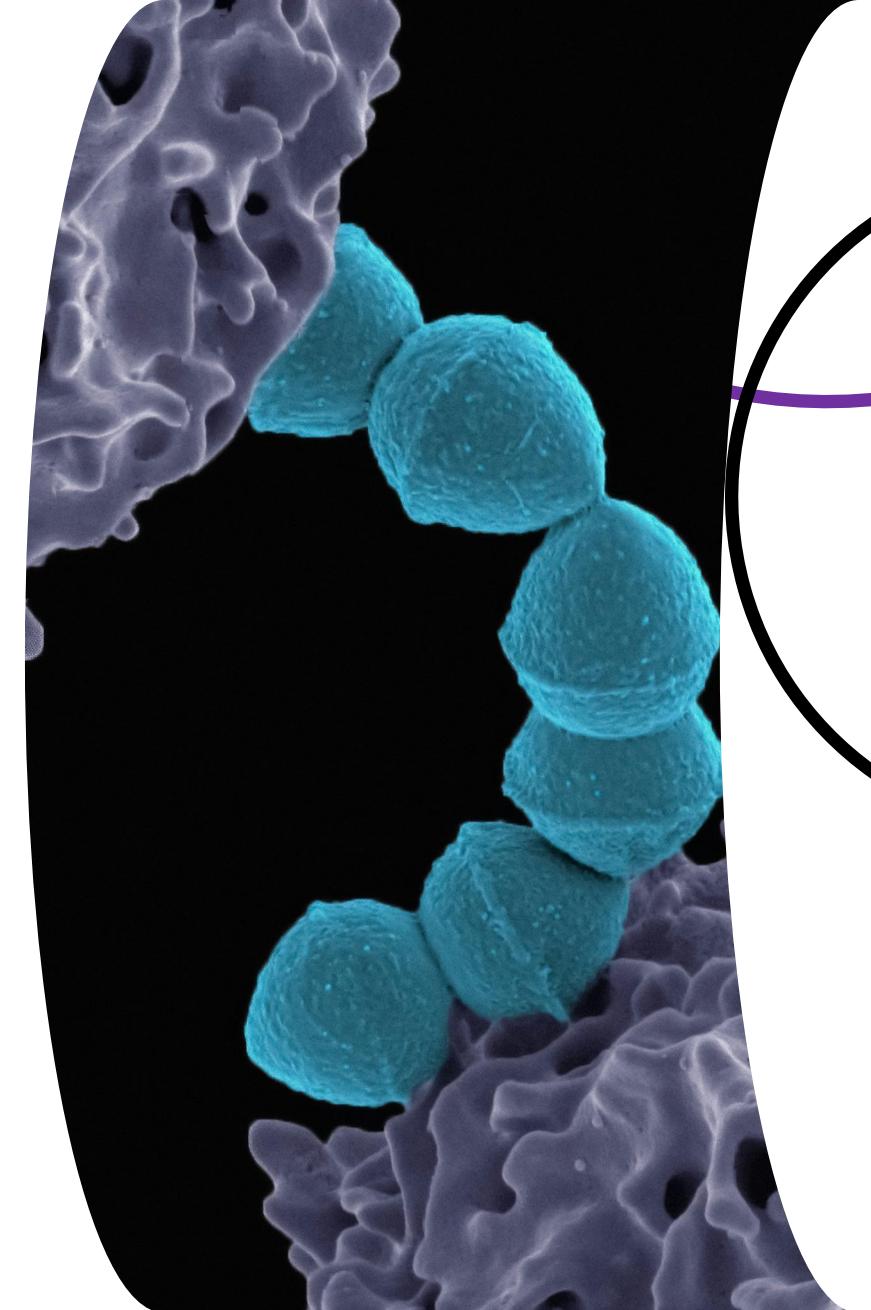
[Virtual, Across Africa and Asia]

Tung Trinh  
Bioinformatician – Oxford University Clinical Research Unit

WCS ACORN - Bioinformatics for  
Antimicrobial Resistance - Virtual Course



Streptococcus Pyogenes  
Photo by [National Institute of Allergy and Infectious Diseases](#) on [Unsplash](#)



# Etiology of infectious disease



Airplane!  
(1980)

Is this  
safe to  
eat ?

Look  
yummy to  
me

Which  
disease we  
may get ?



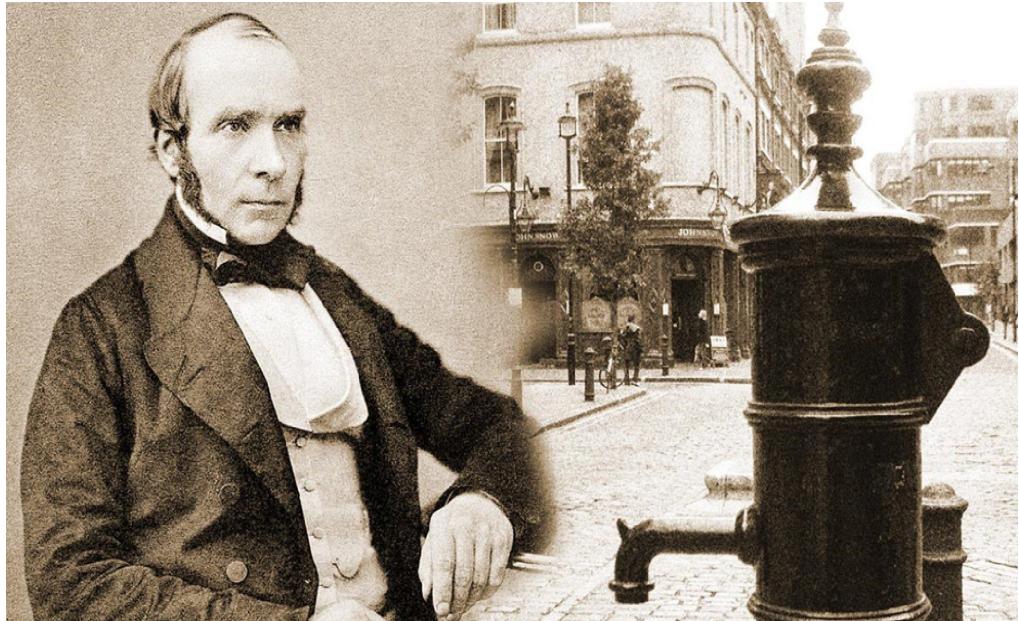
We are not identical. Neither is

- Distribution of disease
- The exposure to risk factors

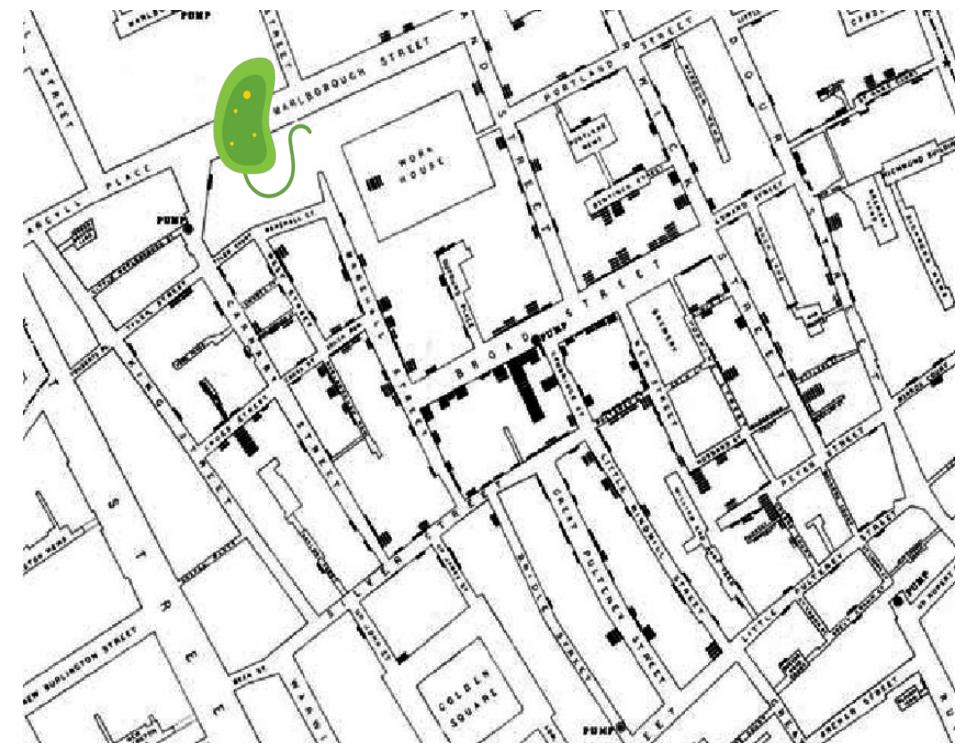


So how we identify an **disease outbreak** ?  
Are we infected from a **same pathogen** ?  
Is there **any link** between the cases ?

# The beginning of Epidemiology



*John Snow (1813-1858) Father of Modern Epidemiology*



*Map of London with Cholera cases in 1854*

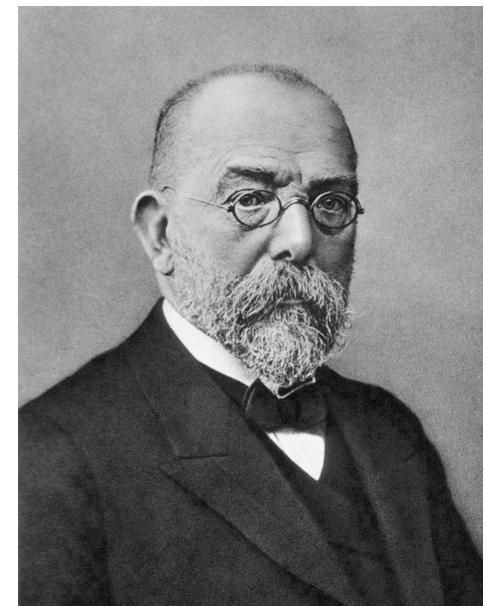
# However



So John Snow did statistical analysis “tying” cholera cases but do you think 100% of the case was *Vibrio cholerae* ?

John Snow linked the wastewater system with cholera cases ( not able to define the pathogen was responsible for)

Until Robert Koch, 1884 confirmation of the bacterial species *Vibrio cholerae* as the causative agent



Robert Koch (1843-1910)

# The early stage of **typing**



Serotyping  
since 1930s

Phagotyping  
since 1950s

DNA-based typing  
since 1980s

Initially, physiological, biochemical and other phenotypic properties served as markers for species identification.

Mainly for foodborne pathogen.

# Genotyping



**Sanger  
Sequencing  
(1977) for  
typing**



Cost a fortune

**pulsed-field gel  
electrophoresis  
(late 80s) for  
typing**



Scale-up able

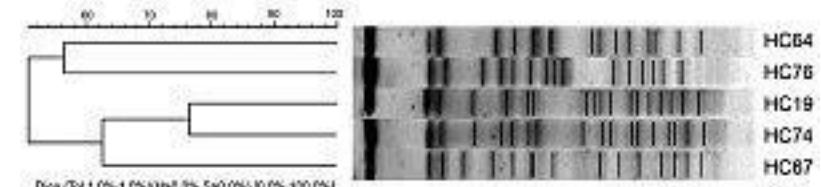
Gold standard  
method for next two  
decades



# Genotyping

90s – 00s, emergence of DNA-based typing methods

- Pulse Field Gel Electrophoresis (PFGE)
- Multiple-locus Variable number tandem repeat Analysis (MLVA)
- Amplified Fragment Length Polymorphisms (AFLP)
- **Multi Locus Sequence Typing (MLST)**
- YATMs (“Yet Another Typing Method” )



% Similarity	AFLP Type (No.)	Antimicrobial Resistance Profile <sup>a</sup>	IS1611 gene	IS1601 gene
0	A(1)	FEP	-	-
2	P(1)	FEP; IPM; SAM; T2P	+	-
4	R(1)	FEP; IPM; SAM; T2P	+	-
6	K(3)	FEP; IPM; SAM; T2P	-	+
8	L(3)	FEP; IPM; SAM; T2P	+	-
10	H(1)	FEP; IPM; SAM; T2P	+	-
12	N(3)	FEP; IPM; SAM; T2P	-	-
14	C(1)	FEP; IPM; SAM; T2P	+	-
16	M(2)	FEP; IPM;	-	-
18	E(2)	FEP; IPM; SAM; T2P	-	+
20	D(1)	FEP; IPM; SAM; T2P	+	-
22	P(1)	FEP; IPM; SAM; T2P	+	-
24	O(1)	FEP; IPM; SAM; T2P	+	+
26	B(1)	FEP; IPM; SAM; T2P	-	+
28	G(1)	FEP; IPM; SAM; T2P	-	+
30	J(1)	FEP; IPM; SAM; T2P	+	-

# MLST

- Analysis 7-9 loci
- Housekeeping genes
- 450-500 bp gene fragments
- Distributed across the genome
- Applicable with NGS

Combination of alleles at all loci ↗ Sequence Typing (ST)

# MLST

## Allelic profile

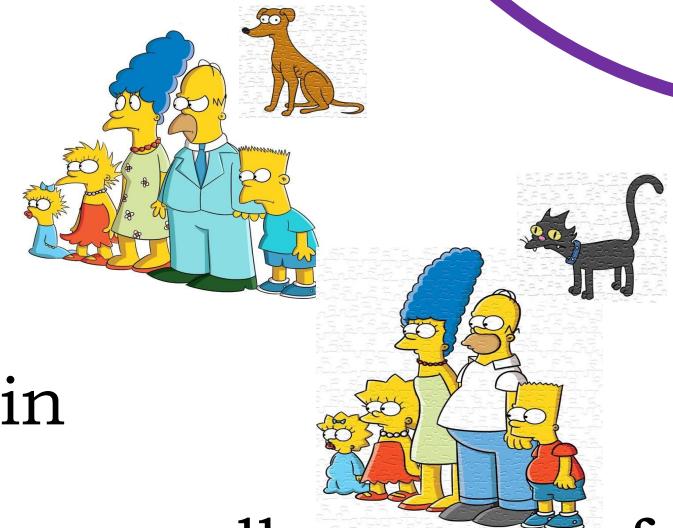
Leaf	Root	Flower	Seed	Twig	Fruit	Trunk	Sequence Type
1	2	2	3	4	5	6	1
...	...	...	...	...	...	...	...
1	4	3	2	4	2	3	65

Pub  
MLST

1,2,3, etc. are encoded from DNA fingerprint

# The problem with MLST

- Same ST but are genomically distinct
- 7 loci represents a tiny fraction of the data present in the whole genome of a strain
- Most MLST datasets are overrepresented with a small number of prevalent STs
- WGS can generate more data for in-silico typing and it cheaper ↗ cgMLST is applicable



# How we detect the gene of interest anyway ?



It's important also to detect molecular variation but it's not simple as it may seem

Similarity >> Identicalness

We need to **BLAST!!!** everything

*We learned in alignment lecture or not*

# Basic Local Alignment Search Tool (BLAST)

“Seed and extend”

1. Break the query into “words” – 3 AA, 10-12 nt
2. Look for exact matches between the words in the query and in each subject in the database.
3. For each query-subject match extend the alignment, calculating a score as you go.
4. Stop calculating for alignments where score goes below a certain threshold ( default is 10)



Google for genomic analysis

# BLAST – how we score



## Nucleotide

Just simple as

Match +1

Mismatch -1

Gap -1



## Amino Acid

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-1	-3	-3	-4	-3	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5				
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4



**BLOSUM** amino acid substitution matrix

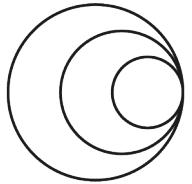
# BLAST - output

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">botulinum neurotoxin [Clostridium botulinum]</a>	2637	2637	100%	0.0	100.00%	<a href="#">AFV13854.1</a>
<input type="checkbox"/>	<a href="#">botulinum neurotoxin type A [Clostridium botulinum]</a>	2504	2504	100%	0.0	94.37%	<a href="#">WP_014520039.1</a>
<input type="checkbox"/>	<a href="#">botulinum neurotoxin type A [Clostridium botulinum]</a>	2503	2503	100%	0.0	94.37%	<a href="#">WP_078992015.1</a>
<input type="checkbox"/>	<a href="#">botulinum neurotoxin type A [Clostridium botulinum]</a>	2488	2488	100%	0.0	93.75%	<a href="#">WP_011948511.1</a>
<input type="checkbox"/>	<a href="#">botulinum neurotoxin type A [Clostridium botulinum]</a>	2488	2488	100%	0.0	93.75%	<a href="#">WP_061316836.1</a>
<input type="checkbox"/>	<a href="#">neurotoxin A [Clostridium botulinum]</a>	2488	2488	100%	0.0	93.67%	<a href="#">ABM73969.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Botulinum neurotoxin type A; Short=BoNT/A; AltName: Full=Bontoxilysin-A</a>	2487	2487	100%	0.0	93.67%	<a href="#">P0DPI0.1</a>

- BLAST provides an ‘E-score’ or ‘E-value’, E stands for Expectation
- It is the number of times you would expect to see an alignment with a similar score by chance
  - Lower is better;  $10^{-30}$  is a frequently used threshold
- E-value calculation depends on the size of the search space (the query and database size), and the score of the alignment

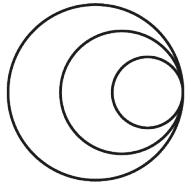


wellcome  
connecting  
science

# ACORN.



# Any Question ?



wellcome  
connecting  
science

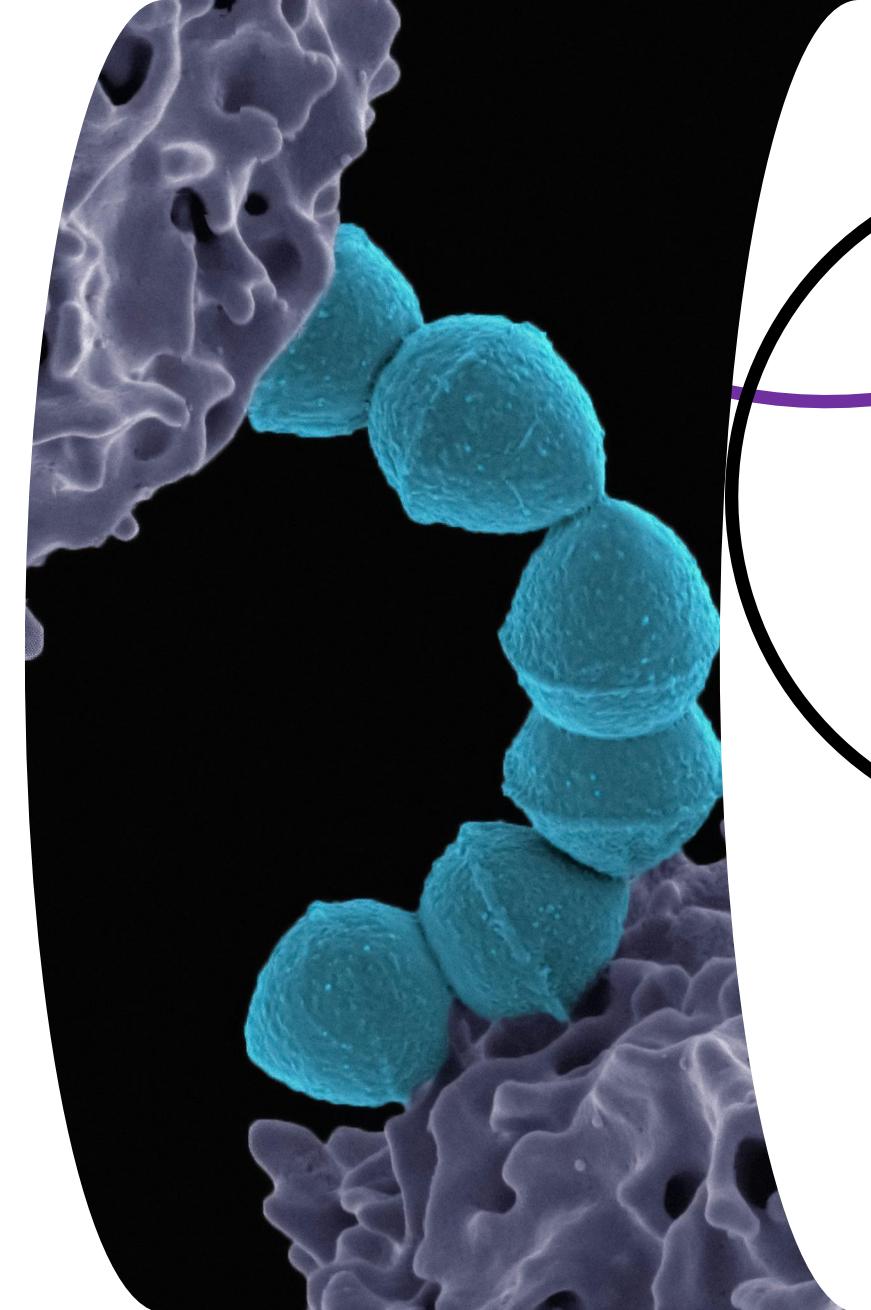
ACORN.

# Antibiotic Resistance (AMR)

WCS ACORN - Bioinformatics for  
Antimicrobial Resistance - Virtual Course



Streptococcus Pyogenes  
Photo by [National Institute of Allergy and Infectious Diseases](#) on [Unsplash](#)

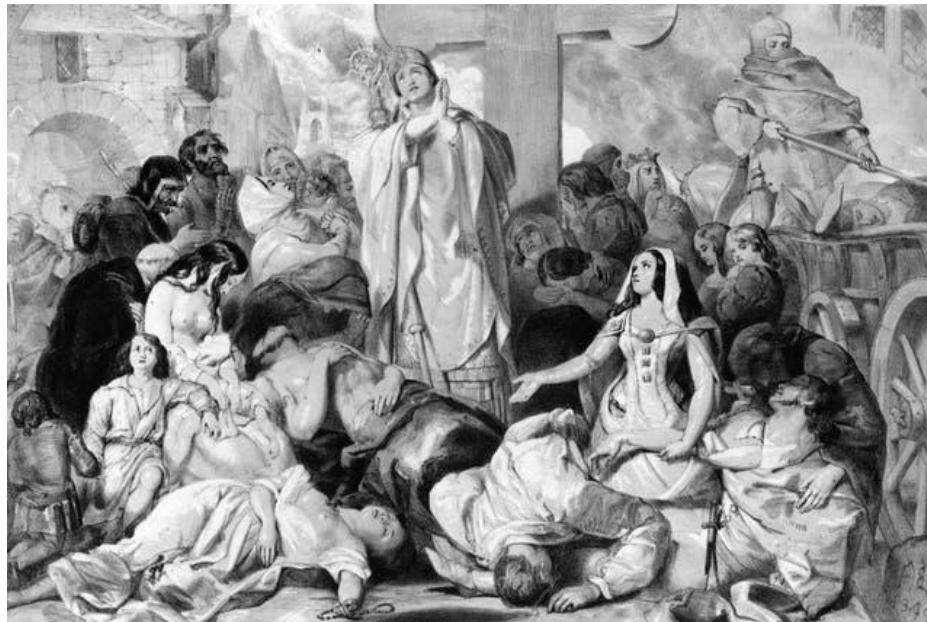


# Antibiotic – where we began

Antibiotic: a type of antimicrobial substance active against bacteria  
(Wikipedia)

**Penicillin** was discovered in 1928 by  
**Alexander Fleming.**

People before 1928



But we also have **ancientbiotic**

**Brief Communication: Mass Spectroscopic Characterization of Tetracycline in the Skeletal Remains of an Ancient Population From Sudanese Nubia 350–550 CE**

Mark L. Nelson,<sup>1\*</sup> Andrew Dinardo,<sup>2</sup> Jeffery Hochberg,<sup>3</sup> and George J. Armelagos<sup>3\*</sup>

<sup>1</sup>Paratek Pharmaceuticals, Inc., Boston, MA 02111

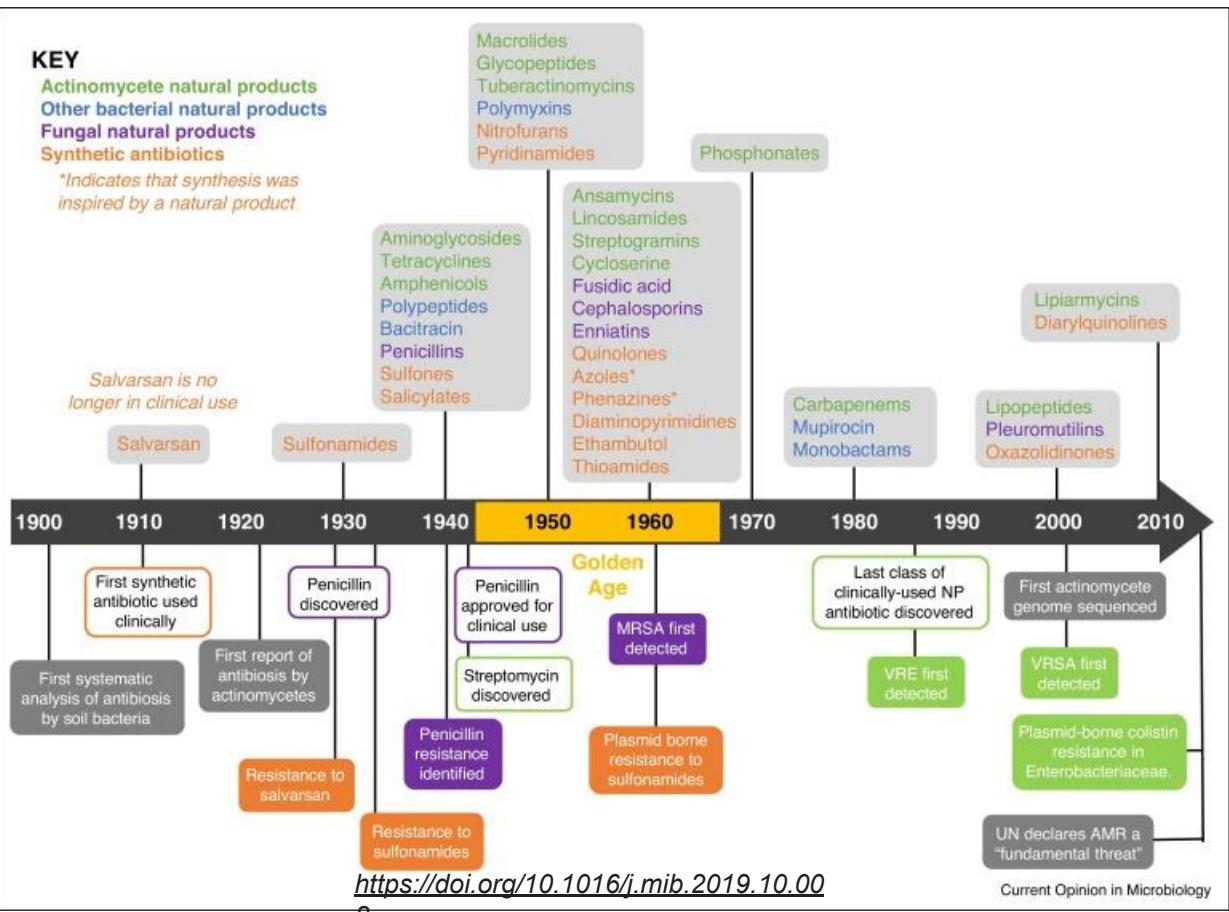
<sup>2</sup>Hospital of University of Pennsylvania, Philadelphia, PA 19104

<sup>3</sup>Department of Anthropology, Emory University, Atlanta, GA 30322

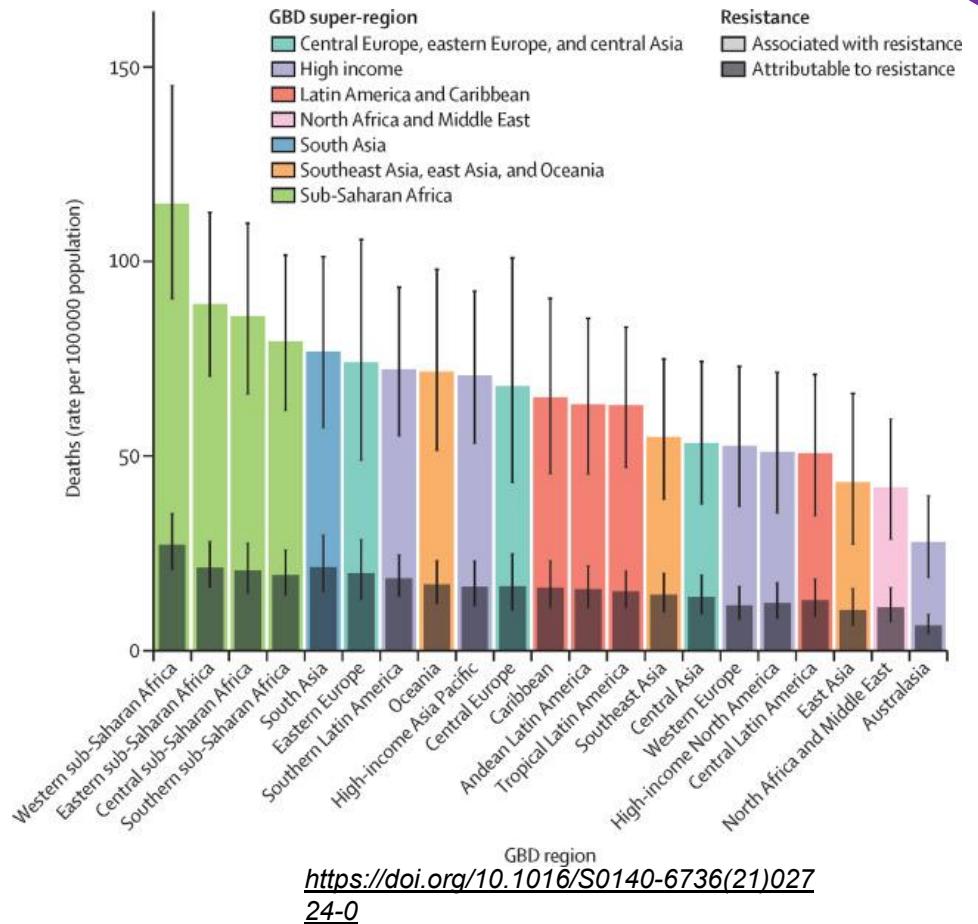
The source ? It was beer !!!  
And of course not our modern beer.



# Antibiotic resistance



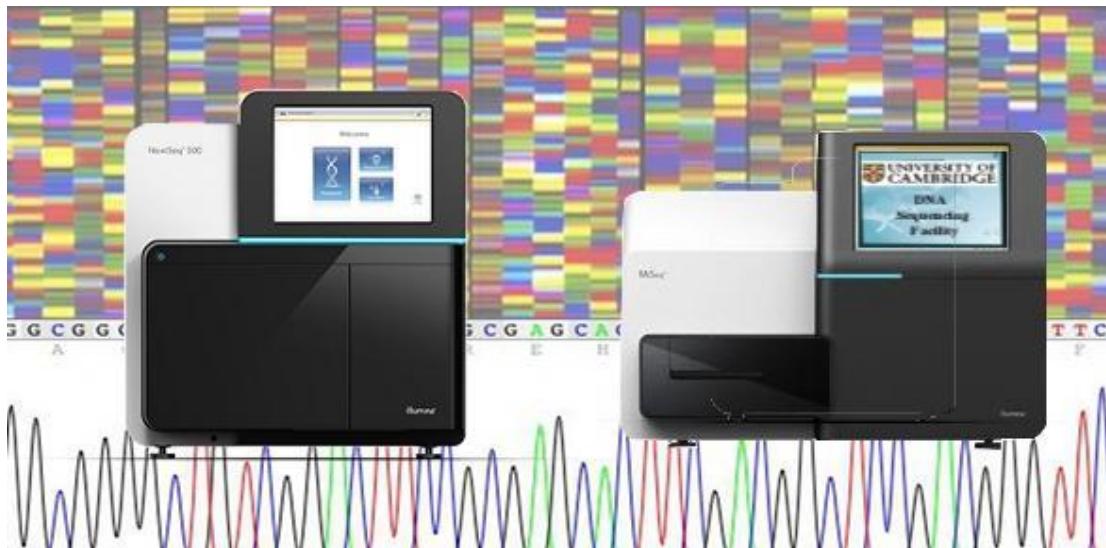
8



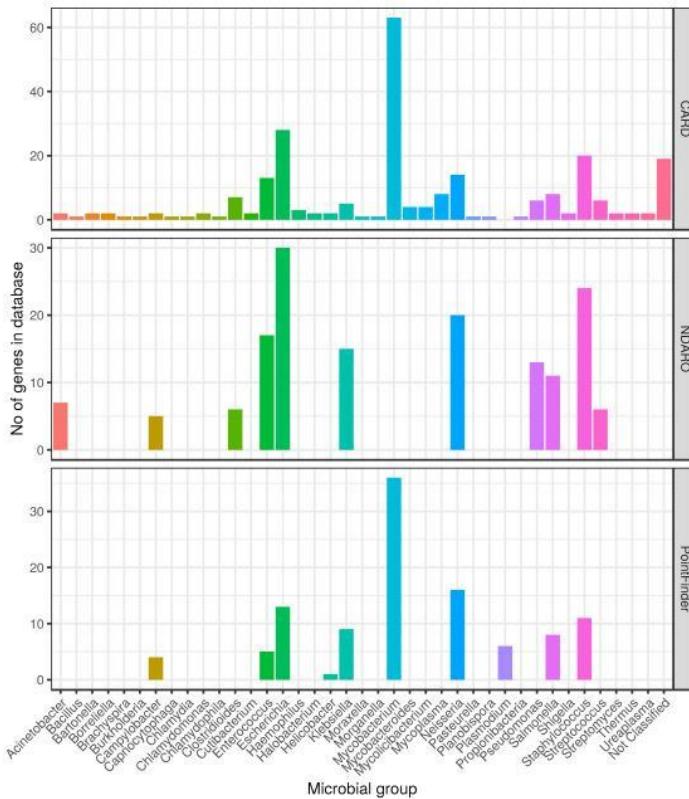
# How we detect AMR

AMRs was discovered and been detecting by its phenotypes (lab intensively, time consuming and limited targets).

*In silico* typing since 2010s.  
We **BLAST** every possible genes.



# AMR gene database



[doi:10.3390/antibiotics11030339](https://doi.org/10.3390/antibiotics11030339)



National Database of Antibiotic  
Resistant Organisms (**NDARO**)



The Comprehensive Antibiotic  
Resistance Database (**CARD**)



# Know your species

# Know your AMR gene

The database consists of various information, make sure you know what you looking for



## Mechanism

- Antibiotic inactivation
- Antibiotic efflux
- Antibiotic target alteration
- Reduced permeability to antibiotic
- Antibiotic target protection
- Antibiotic target replacement
- Resistance by absence



## Drug Class

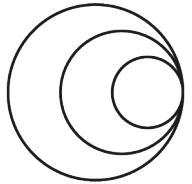
- Carbapenem
- Fluoroquinolone
- Aminoglycoside
- Cephalosporin
- Macrolide
- Glycylcycline
- Tetracycline
- Etc.



## Species

- *Escherichia coli*
- *Klebsiella pneumoniae*
- *Staphylococcus aureus*
- *Acinetobacter baumannii*
- *Pseudomonas aeruginosa*
- *Etc.*

How ? Literature review or <https://card.mcmaster.ca/home> (totally a cheat sheet)



wellcome  
connecting  
science

# ACORN.



# Again, any Question ?