Big Data Analytics and Molecular Diagnostic Methods for Treatment Pathways of Priority Pathogens

Webinar at ICGEB-DBT
Apex, BTIC, New Delhi

Anshu Bhardwaj, Ph.D.

Senior Scientist & Assistant Professor
Bioinformatics Center
CSIR-Institute of Microbial Technology
Chandigarh, India

WHO Priority Pathogens List (PPL) - 2017

CRITICAL	HIGH	MEDIUM
Acinetobacter baumannii, carbapenem-resistant	*Enterococcus faecium,* vancomycin-resistant	*Streptococcus pneumoniae,* penicillin-non-susceptible
Pseudomonas aeruginosa, carbapenem-resistant	*Staphylococcus aureus,* methicillin-resistant, vancomycin intermediate resistant	*Haemophilus influenzae,* ampicillin-resistant
Enterobacteriaceae,* carbapenem-resistant, 3rd generation cephalosporin-resistant	*Neisseria gonorrhoeae, 3rd generation cephalosporin-resistant,* fluoroquinolone-resistant	*Shigella,* fluoroquinolone-resistant
* Enterobacteriaceae include: *Klebsiella pneumonia*, *Escherichia coli*, *Enterobacter spp.*, *Serratia spp.*, *Proteus spp.*, and *Providencia spp*, *Morganella spp*	*Campylobacter,* fluoroquinolone-resistant	*ACCORDING TO THE STUDY, MTB IS ON THE PRIORITY FOR NEW DRUG DISCOVERY AND DEVELOPMENT*
	Salmonella, fluoroquinolone-resistant	
	Helicobacter pylori, clarithromycin-resistant	

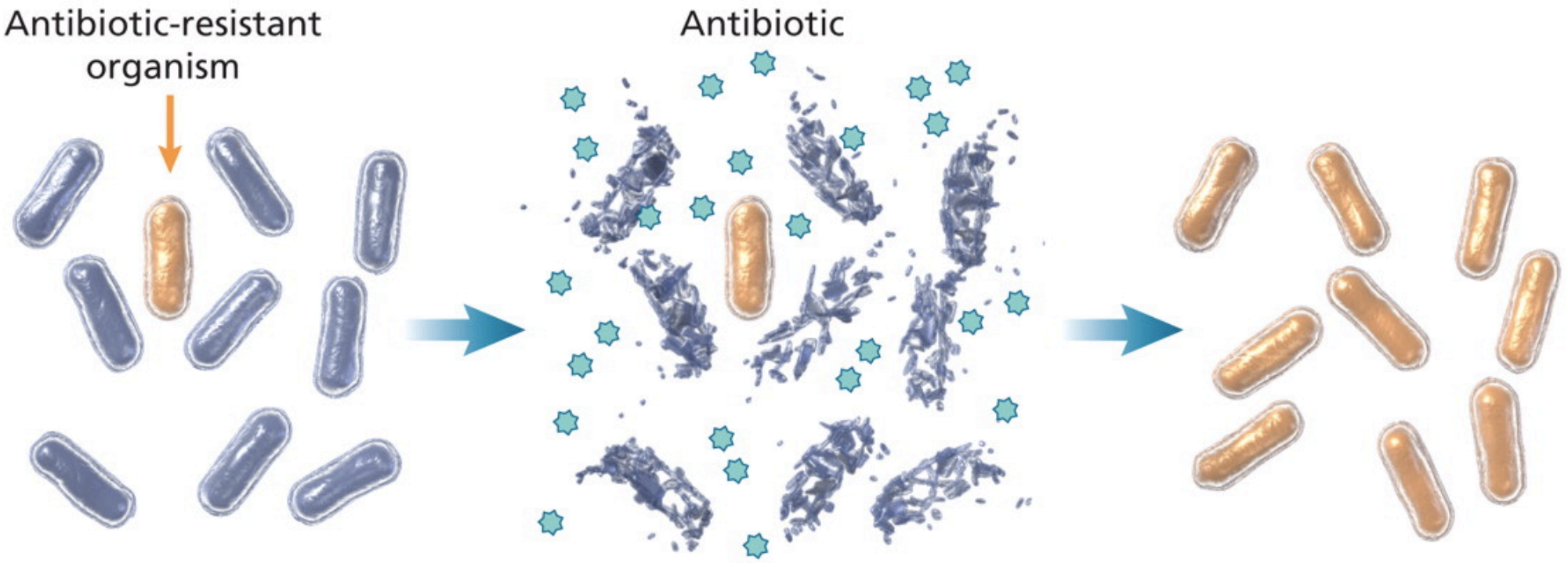
http://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf?ua=1

Antimicrobial resistance occurs when microorganisms such as bacteria, viruses, fungi and parasites change in ways that render the **medications** used to cure the infections they cause **ineffective**.

When the microorganisms become resistant to most antimicrobials they are often referred to as "**superbugs**".

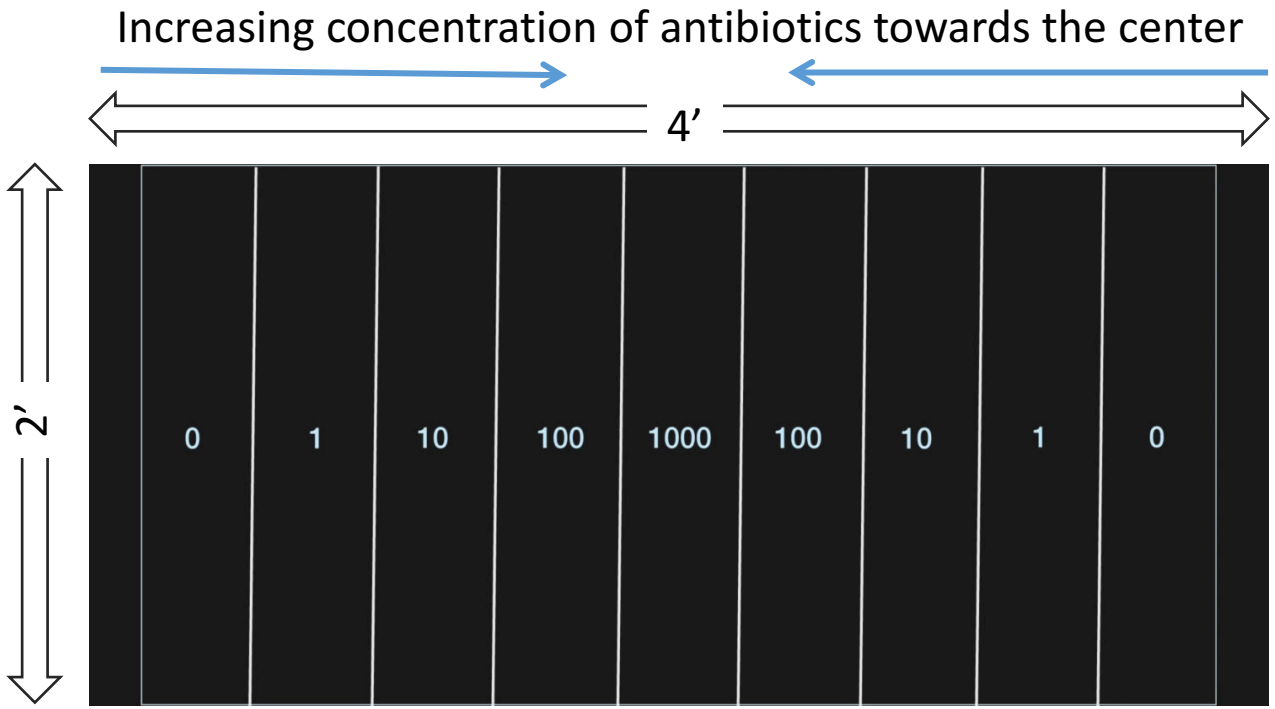
World Health Organization

Antibiotic Resistance



https://www.google.com/url?sa=i&source=imgres&cd=&cad=rja&uact=8&ved=2ahUKEwifiOiPqMDqAhXJxTgGHanwAPMQjRx6BAgBEAQ&url=https%3A%2F%2Fwww.cmaj.ca%2Fcontent%2F180%2F4%2F408%2Ftab-figures-data&psig=AOvVaw36qc4E5LvmsYs2UMdUTib2&ust=1594388842796509

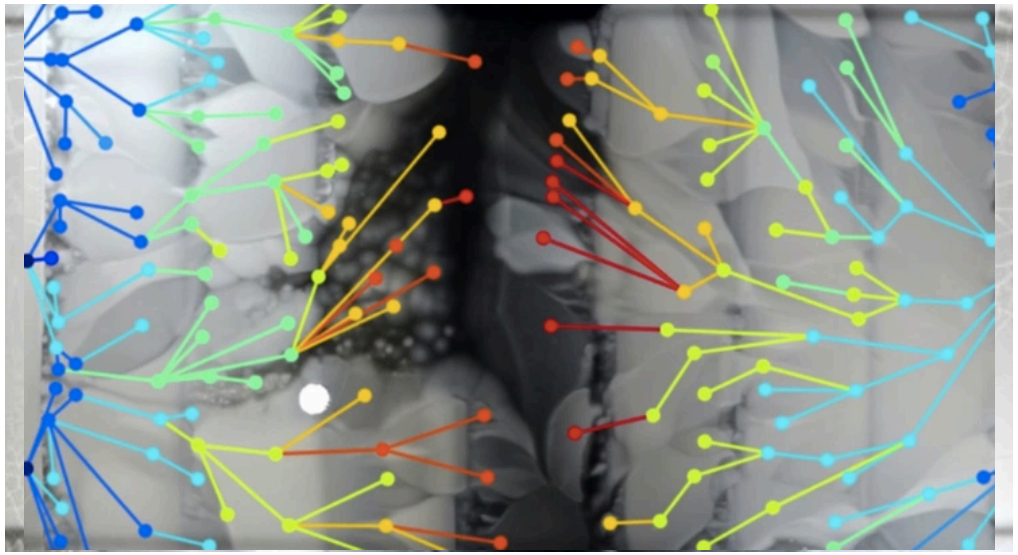
A cinematic approach to drug resistance – Harvard Gazette



An experimental device for studying microbial evolution in a spatially structured environment

Microbial Evolution and Growth Arena (MEGA) plate

E. coli appear as white on the black background

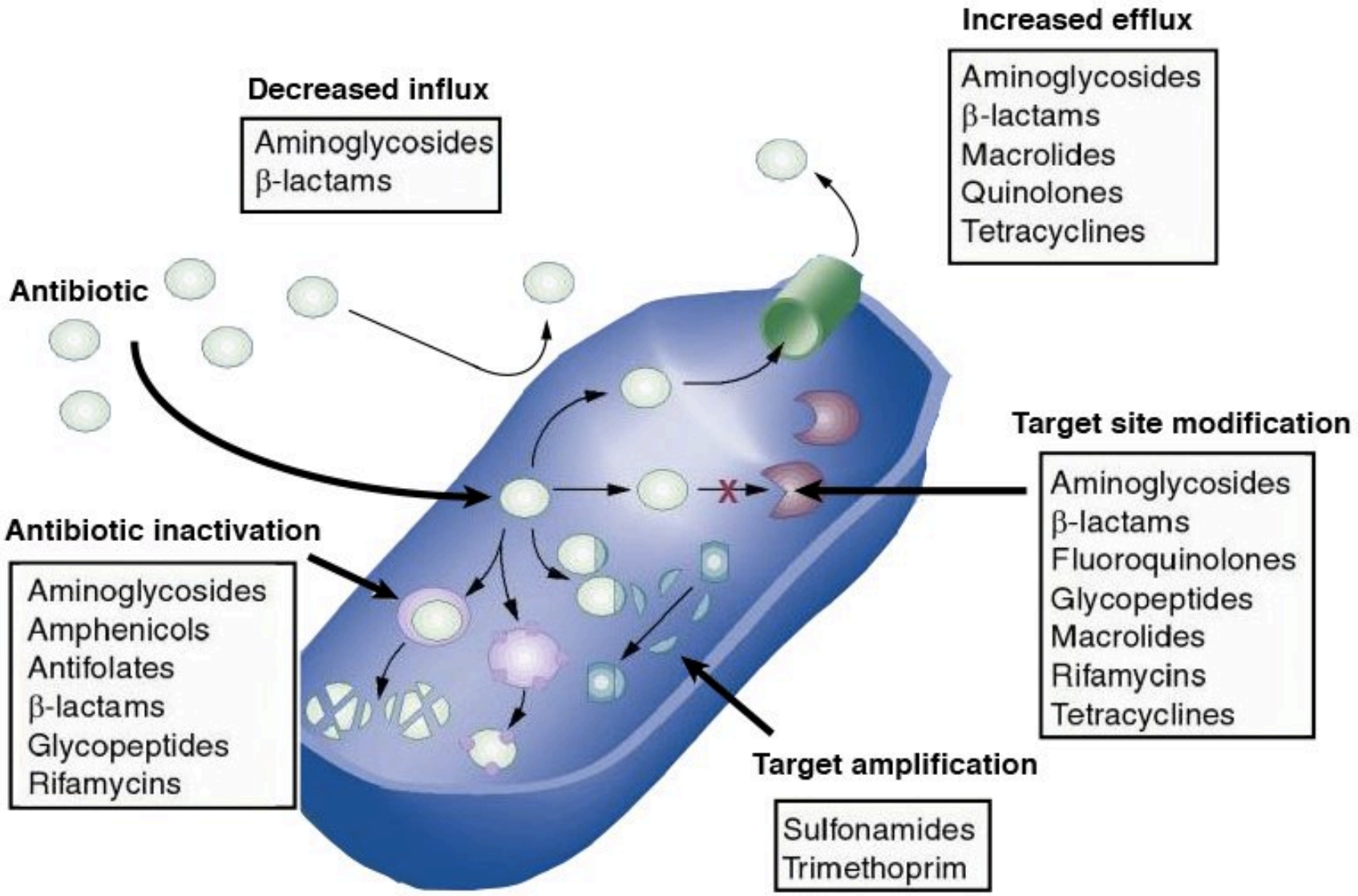


Bacteria spread as they evolve

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5534434/

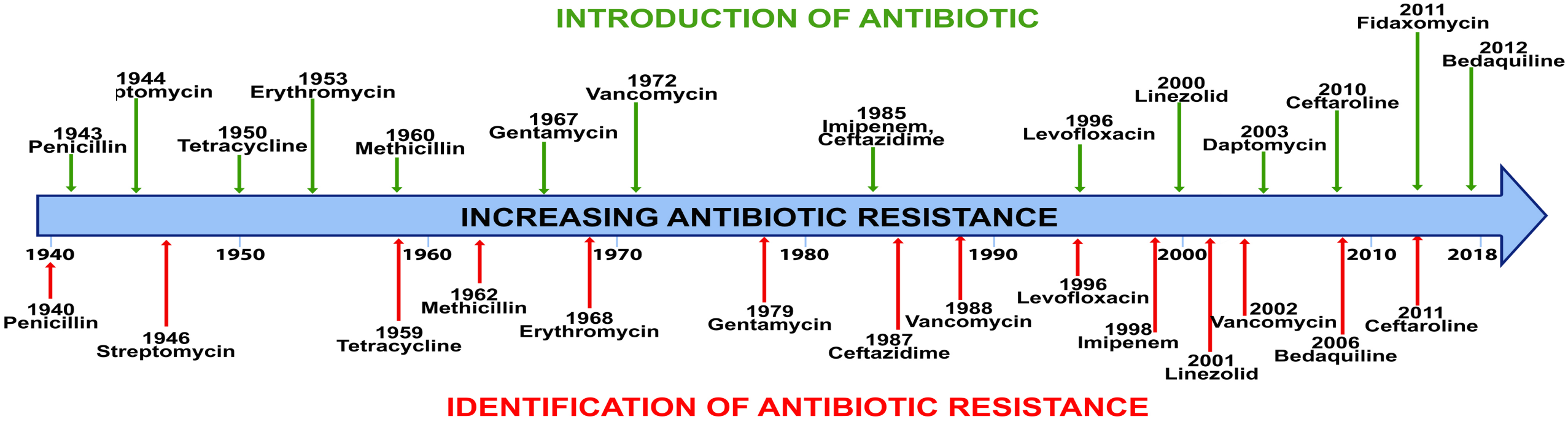
https://news.harvard.edu/gazette/story/2016/09/a-cinematic-approach-to-drug-resistance/

Drug Resistance Mechanisms



https://adailydeed.wordpress.com/tag/antibiotic-resistance/

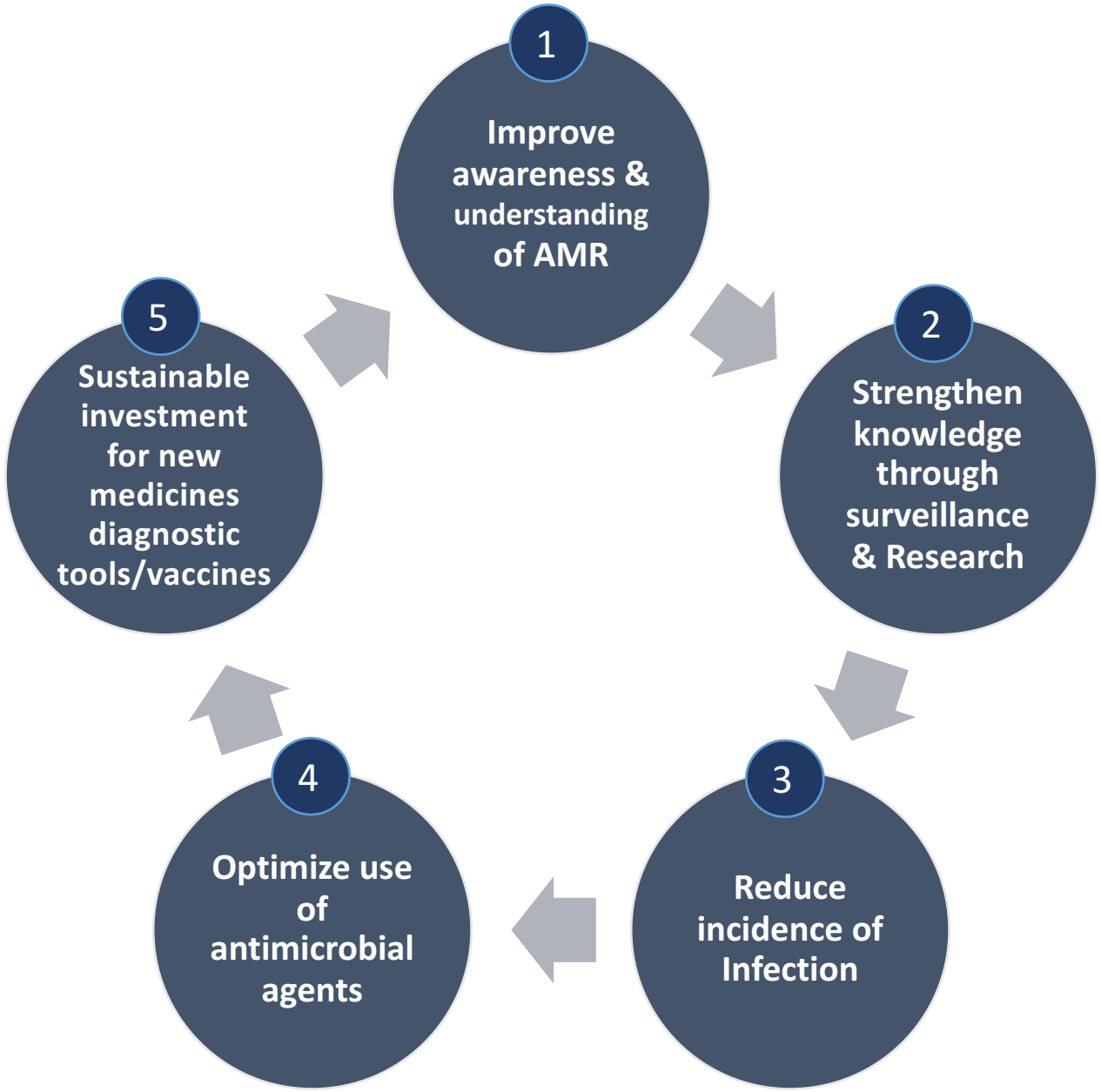
Timeline showing discovery of antibiotics, year of introduction and observation of resistance



- Discovery of antibiotics is rapidly followed by emergence of resistance against them.
- Continuing improvements are being made for discovery of new antimicrobial agents but are outpaced by emerging and re-emerging drug resistant infectious disease agents.

PMC3911012, PMC4378521

Sixty-eight World Health Assembly in May 2015 endorsed a global action plan to tackle antimicrobial resistance



Culture remains the gold standard for diagnosis of the Pathogen
So what is wrong?

- As culture takes time – empirical based approach is used for diagnosis and prescription (broad-spectrum antibiotics are prescribed to cover as many pathogens)
- Leads to overtreatment or not the right treatment
- Impacts the emergence of antimicrobial resistance
- <60% sensitive; biological sample has low number of pathogens
- Complex etiology of conditions like sepsis makes it even more challenging - 90–95% of blood cultures remain negative in immunocompromised patients suspected of infection; even in the cases where bacterial or fungal sepsis is likely

Sites of Infection

	No.	% Total
Lung	1016	37.2
Intraabdominal	801	29.3
Bowel perforation/peritonitis	226	8.3
Postoperative bowel perforation/anastomotic dehiscence	65	2.4
Spontaneous bacterial peritonitis	50	1.8
Other peritonitis	18	0.7
Intraabdominal abscess	44	1.6
Cholecystitis	40	1.5
Ascending cholangitis	43	1.6
Ischemic bowel/bowel infarction	166	6.1
Clostridium difficile enterocolitis/toxic megacolon	47	1.7
Genitourinary	293	10.7
Skin and soft tissue	197	7.2
Necrotizing soft tissue infections	74	2.7
Cellulitis	46	1.7
Operative wound infection	22	0.8
Soft tissue abscess	20	0.7
Decubitus ulcer	16	0.6
Diabetic lower extremity ulcer/cellulitis	13	0.5
Surgical site infection	31	1.1
Central nervous system infection (meningitis/abscess)	20	0.7
Intravascular catheter infection	100	3.7
Primary bloodstream infection (bacteremia without identifiable source)	120	4.4
Systemically disseminated infection (including yeast and tuberculosis)	58	2.1
Septic arthritis	21	0.8
Mediastinitis	15	0.5
Other	59	2.1

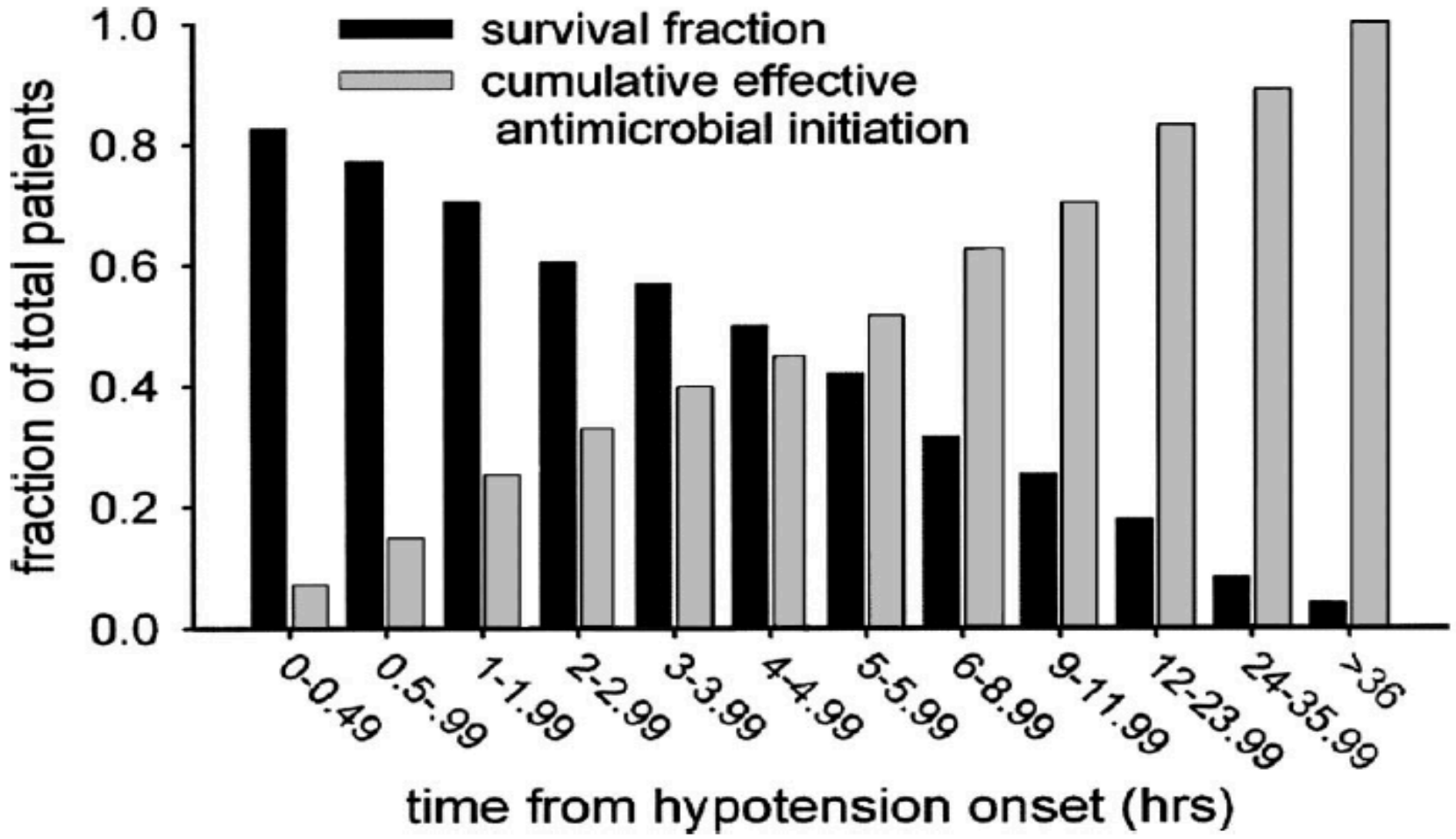
Sepsis is a systematic inflammatory response caused by infection

DeNIS Study

Pathogen	No. of Patients	% Total
Gram-negative organisms	930	47.9
Escherichia coli	435	22.4
Klebsiella species	131	6.7
Pseudomonas aeruginosa	115	5.9
Enterobacter species	80	4.1
Haemophilus influenzae	44	2.2
Proteus species	25	1.2
Acinetobacter species	21	1.1
Serratia species	20	1.0
Stenotrophomonas maltophila	16	0.8
Morganella morganii	14	0.7
Citrobacter species	13	0.7
Neisseria meningitidis	6	0.3
Burkholderia cepacia	3	0.2
Haemophilus parainfluenzae	3	0.2
Other Gram-negative bacilli	8	0.4
Gram-positive organisms	731	38.3
Staphylococcus aureus	302	15.6
Streptococcus pneumoniae	170	8.8
Streptococcus faecalis	77	4.0
Group A *Streptococcus* species	69	3.6
Other β-hemolytic streptococci	43	2.2
Viridans streptococci	37	1.9
Streptococcus faecium	29	1.5
Bacillus species	5	0.3
Corynebacterium jeijkeium	5	0.3
Staphylococcus lugdunensis	1	0.3
Yeast/fungi	160	8.2
Candida albicans	91	4.7
Candida glabrata	18	0.9
Aspergillus/Mucor species	14	0.7
Blastomyces species	10	0.5
Candida tropicalis	4	0.2
Candida parapsilosis	4	0.2
Candida krusei	3	0.2
Cryptococcus neoformans	1	0.1
Histoplasma species	1	0.1
Other unidentified yeast	13	0.6
Anaerobes	69	3.6
Clostridium difficile	46	2.4
Bacteroides fragilis	15	0.8
Other clostridia	8	0.4
Legionella species	8	0.4
Mycobacterium tuberculosis	11	0.6

Suspected primary pathogens in septic shock

The Need for Rapid Diagnosis



https://pubmed.ncbi.nlm.nih.gov/16625125/

Need a method that:

- Non culturable pathogen detection
- Can detect an entire range of pathogens
- Comprehensive
- Sensitive
- Specific
- Rapid
- Works with dead or live pathogen
- Differentiate between pathogen and human cells

Do Sequencing-based methods offer a solution?

What sequencing based methods can offer?

- Identify unculturable pathogens
- Distinguish between different pathogen strains
- Detect new drug resistance determinants, e.g. detection of mecC in *S. aureus**
- Rapid with extreme accuracy, -> better-informed treatment decisions
- Sequence data can be used to develop specific primers for PCR-based assays
    - to provide more complete sequence information and
    - to develop other specific field tests – specific genotype based

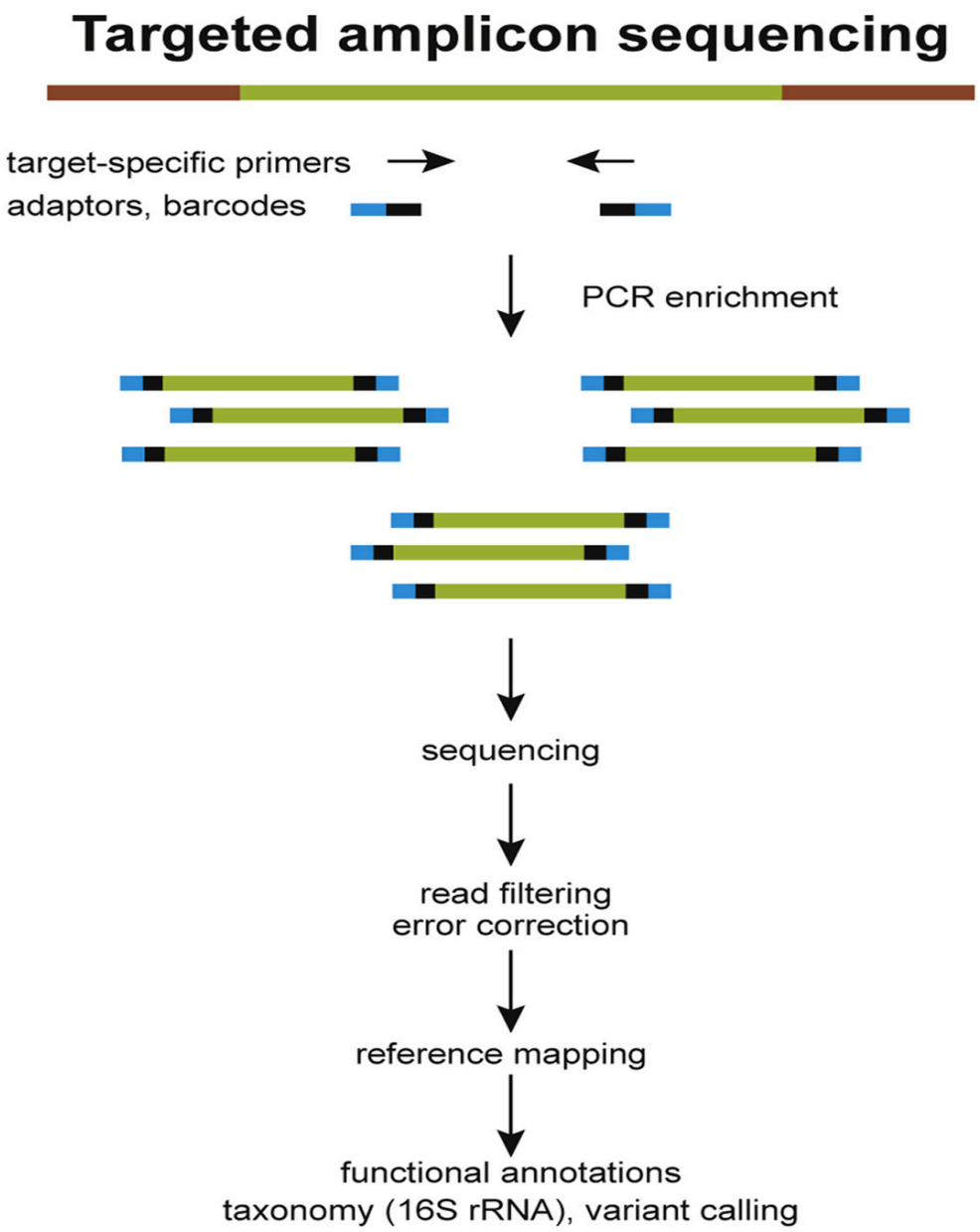
*https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3989053/

Traditional vs Next Generation Sequencing Platforms in Clinical Microbiology

- The Sanger method has limited sensitivity for minor variants when present at <15% to 20% of the viral population, whereas NGS methods can detect drug-resistant mutations (DRMs) present at approximately 1%

- At least half of the DRMs identified by NGS are missed by Sanger sequencing

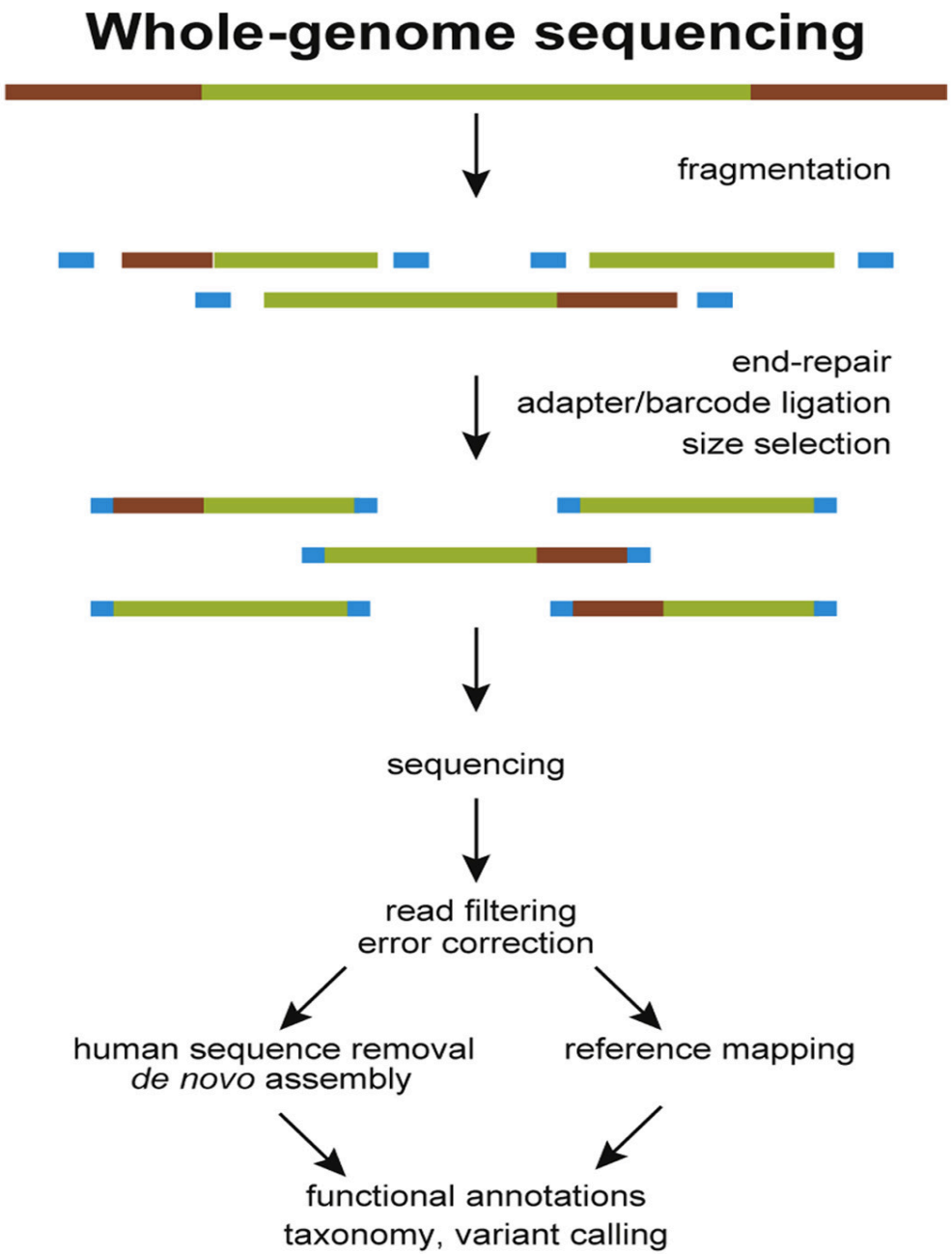
- The presence of such variants has been shown to predict an increased risk for therapy failure



Cannot determine new pathogens or

New antimicrobial resistance genes

Not comprehensive



Comprehensive

Can help in detecting novel AMR genes

Nanopore - MinIon

From bench to bedside – Point of care sequencing - promise

Companies which are pushing the boundaries of PoC with NGS in IDD

COMPANY	REGION	PRODUCT TYPE(S)	DESCRIPTION	FUNDING
Illumina and bioMérieux	U.S.	▪ NGS Service	Illumina, the dominant player in the NGS market, and bioMérieux, a leader in *in vitro* diagnostics, partnered together to launch an NGS service dedicated to infectious disease diagnostics.[14]	▪ Through a partnership between Illumina and bioMérieux, both global players in healthcare.[15]
Oxford Nanopore Technologies	U.K	▪ NGS Sequencers ▪ NGS Assays ▪ Sample Prep Kits ▪ Library Prep Kits	Founded in 2005, the company is developing portable NGS sequencing products for the analysis of single molecules.[22, 27]	▪ Raised a total of $386 million through VC funding. ▪ Closed a $109 million round in July 2015 from new and existing investors in the U.K., U.S., and mainland Europe for product development, manufacturing and commercialization.[20]
DNAe	U.K.	▪ NGS Sequencers ▪ NGS Assays	Founded in 2003, DNAe aims to revolutionize NGS diagnostic technology by bringing the entire NGS workflow from sample preparation, sequencing, and analysis into one semiconductor chip.[4]	▪ Secured a $38 million bank facility from Citibank in November 2015 for development of an NGS *in vitro* diagnostic test for Serious Blood Infections.[5] ▪ Completed a Series A fundraising round in April 2014 for an undisclosed amount.[5]
MRIGlobal	U.S.	▪ NGS Sequencers ▪ NGS Assays	Founded in 1977, MRIGlobal is an independent organization that performs contract research for the government and the industry.	▪ Awarded $14.8 million in February 2015 from the Defense Threat Reduction Agency of the U.S. Department of Defense to develop a comprehensive NGS platform for infectious diseases diagnosis.[8]
Pathoquest	France	▪ NGS Assays ▪ Bioinformatics	Pathoquest is developing a comprehensive NGS assay for diagnosing infectious diseases, including bioinformatics.	▪ Raised $5 million in July 2013 in Series B funding by IDInvest Partners, Aurinvest and Kurma Partners.[23]

A few databases and tools for DR determinant

Species identification

From clonal sample
- NCBI BLAST
- GenBank
- Other genome databases in Table 1

From nonclonal sample

Meta-assembly
- AMOS
- MIRA
- MetaVelvet

Clustering and species annotation
- MEGAN
- MG-RAST

Maximum likelihood phylogeny trees
- BEAST
- RAxML
- ClonalFrame
- ClonalOrigin

Whole-genome alignment

For SNP calling
- Mummer
- Mugsy

For structural variant calling
- Mauve

Gene annotation

Bacterial
- GLIMMER
- RAST

Drug resistance in bacteria
- ResFinder
- ARG-ANNOT

Examples of Public Bioinformatics Databases That May Be Leveraged for Multiscale Analysis of Infectious Disease[a]

Database Focus	For General Research	For Infectious Disease	
		Multipathogen	Pathogen-Specific
Genomes	• NCBI Nucleotide (GenBank/RefSeq) • ENA/EMBL • DDBJ	• ViPR • NMPDR • PATRIC • EuPathDB	• Influenza Research Database • Tuberculosis Database • LANL: Databases for HIV, HCV, and HFV
Gene products and functionality	• UniProt • KEGG	• Pathogen-Host Interaction Database • Antibiotic Resistance Genes Database • Comprehensive Antibiotic Resistance Database	
Expression and immune profiles	• GEO • ArrayExpress	• ImmPort	

List is not comprehensive

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 OCTOBER 11, 2018 VOL. 379 NO. 15

Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing

The CRyPTIC Consortium and the 100,000 Genomes Project

ABSTRACT

BACKGROUND

The World Health Organization recommends drug-susceptibility testing of *Mycobacterium tuberculosis* complex for all patients with tuberculosis to guide treatment deci-

The members of the writing group (Timothy M. Walker, D.Phil., A. Sarah Walker,

Lets know more about this study:

1. A total of 10,209 isolates were analyzed.
2. Resistance to isoniazid, rifampin, ethambutol, and pyrazinamide was correctly predicted with 97.1%, 97.5%, 94.6%, and 91.3% sensitivity, respectively
3. Susceptibility to these drugs was correctly predicted with 99.0%, 98.8%, 93.6%, and 96.8% specificity.
1. Among the 4037 phenotypic profiles that were predicted to be pansusceptible, 3952 (97.9%) were correctly predicted.

Susceptibility is predicted with higher accuracy

Clinical correlation of resistance phenotype is a limiting factor!!

Modes of action of first-and second-line anti-TB drugs, mechanisms of drug resistance and mutation frequency for each gene in clinical Mtb isolates

Group	Drug	Drug action	Drug resistance-associated gene(s)	Mutation frequency in clinical isolates (%)[a]
First-line anti-TB drugs	Rifampicin	Binding to the β-subunit of the RNA polymerase, inhibition of the elongation of messenger RNA	*rpo*B encoding for β-subunit of RNA polymerase	90–100
	Isoniazid	Activation by a catalase-peroxidase enzyme	*kat*G encoding for catalase-peroxidase	40–97
		Inhibition of the synthesis of mycolic acids through binding to NADH-ACP-reductase	*inh*A encoding for fatty acid enoyl acyl carrier protein reductase A (InhA)	8–64
	Ethambutol	Inhibition of an arabinosyl transferase involved in cell wall synthesis	*emb*B encoding for arabinosyl transferase	47–89
	Pyrazinamide	- Activation by the pyrazinamidase - Disruption of membrane energetics that inhibits membrane transport	*pnc*A encoding for pyrazinamidase	44–97
	Streptomycin	Inhibition of protein synthesis by interaction with the 16S rRNA and the S12 ribosomal protein	*rrs* encoding for 16S rRNA subunit	12–26
			*rps*L encoding for S12 ribosomal protein	40–68
		Inhibition of methylation of 16S rRNA	*gid*B encoding for 7-methylguanosine methyltransferase	5–13
Second-line anti-TB drugs	Amikacin, kanamycin, capreomycin	Inhibition of protein synthesis by interaction with the 16S rRNA	*rrs* encoding for 16S rRNA	40–90
	Kanamycin	Inhibition of acetyltransferase	*eis* encoding for aminoglycoside acetyltransferase	28–80
	Capreomycin	Inhibition of methylation of 16S rRNA & 23S rRNA	*tly*A encoding for 2′-O-methyltransferase	4–13
	Ofloxacin, levofloxacin, moxifloxacin, gatifloxacin	Inhibition of the topoisomerase II (DNA gyrase) lead to the inhibition of DNA supercoiling	*gyr*A encoding for DNA gyrase subunit A and	70–90
			*gyr*B encoding for DNA gyrase subunit B	0–11
	Ethionamide	Inhibition of the synthesis of mycolic acids by interaction with NAD that inhibits the enoyl-ACP reductase	*inh*A encoding for fatty acid enoyl acyl carrier protein reductase A (InhA)	33–62
		Inhibition of metabolic activation by interaction with the transcriptional repressor of the Monooxygenase (EthA)	*eth*A encoding for EthA	46–72
			*eth*R encoding for transcriptional repressor EthR, NADH-ACP	0–4

In Greek, pan means ‘whole’

This Article | Info for Authors | Subscribe | About

PNAS

Proceedings of the National Academy of Sciences of the United States of America

Proc Natl Acad Sci U S A. 2005 Sep 27; 102(39): 13950–13955.

PMCID: PMC1216834

Published online 2005 Sep 19. doi: 10.1073/pnas.0506758102

PMID: 16172379

Genetics

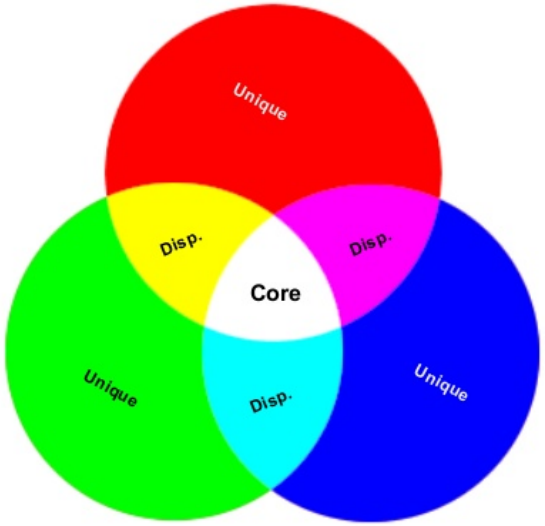
Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”

Hervé Tettelin,[a,b] Vega Masignani,[b,c] Michael J. Cieslewicz,[b,d,e] Claudio Donati,[c] Duccio Medini,[c] Naomi L. Ward,[a,f] Samuel V. Angiuoli,[a] Jonathan Crabtree,[a] Amanda L. Jones,[g] A. Scott Durkin,[a] Robert T. DeBoy,[a] Tanja M. Davidsen,[a] Marirosa Mora,[c] Maria Scarselli,[c] Immaculada Margarit y Ros,[c] Jeremy D. Peterson,[a] Christopher R. Hauser,[a] Jaideep P. Sundaram,[a] William C. Nelson,[a] Ramana Madupu,[a] Lauren M. Brinkac,[a] Robert J. Dodson,[a] Mary J. Rosovitz,[a] Steven A. Sullivan,[a] Sean C. Daugherty,[a] Daniel H. Haft,[a] Jeremy Selengut,[a] Michelle L. Gwinn,[a] Liwei Zhou,[a] Nikhat Zafar,[a] Hoda Khouri,[a] Diana Radune,[a] George Dimitrov,[a] Kisha Watkins,[a] Kevin J. B. O'Connor,[h] Shannon Smith,[i] Teresa R. Utterback,[i] Owen White,[a] Craig E. Rubens,[g] Guido Grandi,[c] Lawrence C. Madoff,[e,j] Dennis L. Kasper,[e,j] John L. Telford,[c] Michael R. Wessels,[d,e] Rino Rappuoli,[c,k,l] and Claire M. Fraser[a,b,k,m]

‣ Author information ‣ Copyright and License information Disclaimer

Strains of a bacterial species might differ substantially in their gene content and total gene pool of a species might be orders of magnitudes larger than the gene content of any single strain

Three genomes

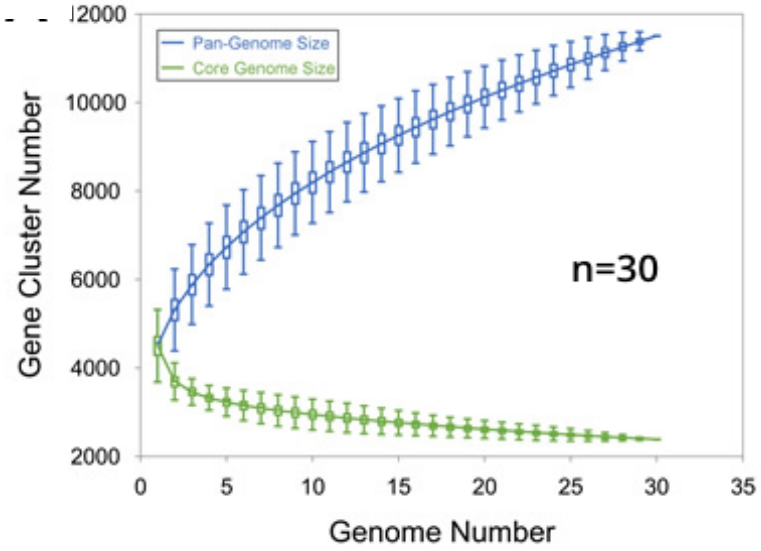


Core
In all 3 strains
(∈ *N* strains)

Dispensable
In 2 strains
(∈ [2,*N*-1] strains)

Unique
In only 1 strain
(∈ 1 strain)

This image cannot currently be displayed.



The core genes are responsible for basic aspects of the biology of the species and its major phenotypic traits;

The accessory genes and singletons usually pertain to supplementary biochemical pathways and functions that may confer selective advantages such as ecological adaptation, virulence mechanisms, antibiotic resistance, or colonization of a new host.

nature COMMUNICATIONS

Article | OPEN | Published: 17 October 2018

Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance

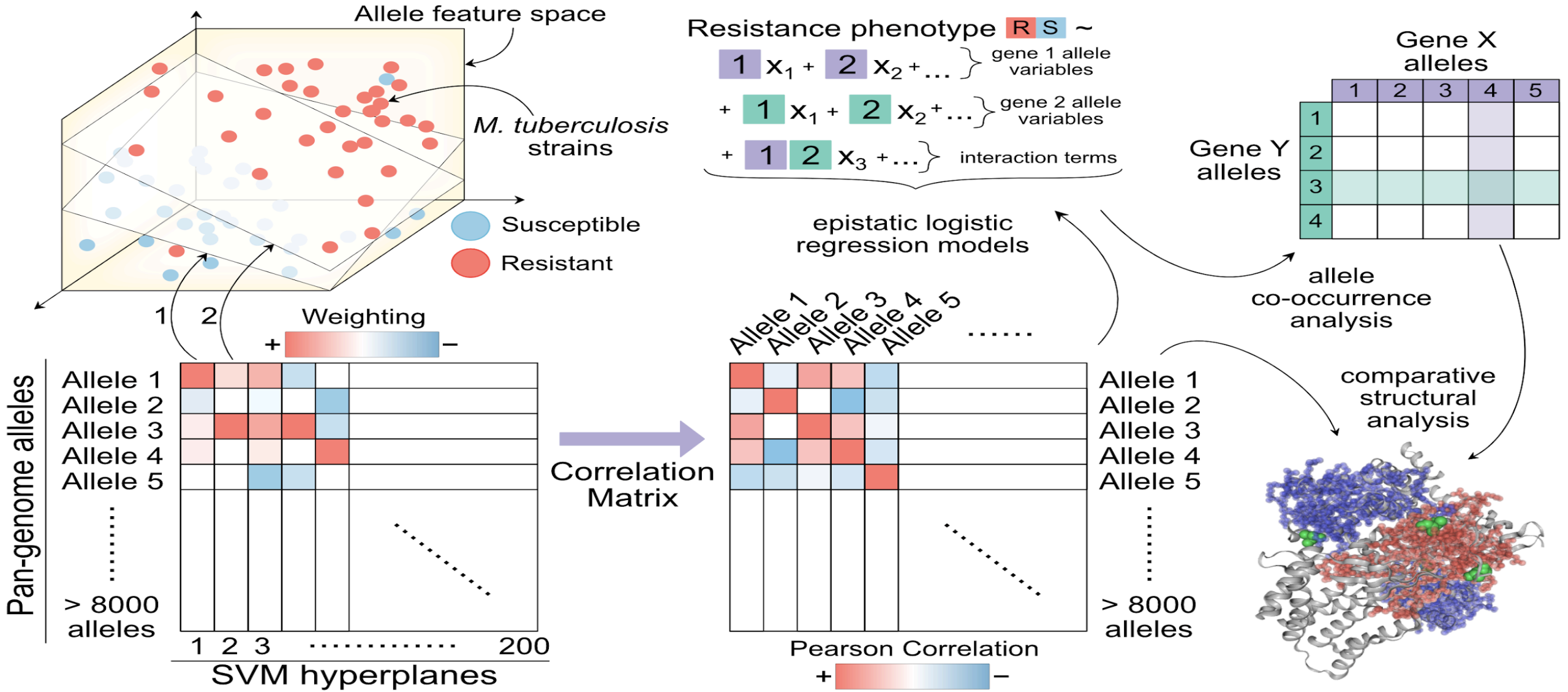
Erol S. Kavvas, Edward Catoiu, Nathan Mih, James T. Yurkovich, Yara Seif, Nicholas Dillon, David Heckmann, Amitesh Anand, Laurence Yang, Victor Nizet, Jonathan M. Monk ✉ & Bernhard O. Palsson ✉

Nature Communications **9**, Article number: 4306 (2018) | Download Citation ↓

The dataset – Genome IDs with Drug Sensitivity Profiling

genome_id	isoniazid	rifampicin	ethambutol	amikacin	ciprofloxacin	clofazimine	cycloserine	ethionamide	kanamycin	moxifloxacin	nicotinamide
1295764_3	R	R	R			S		R	R		R
1423468_3	R	R	S								
1448744_3	R	R	S								
1448649_3	S	S	S								
1773_371	R	R	R								
1421929_3	R	R									
1448748_3	S	S	S								
1455273_3	R	R	R								
1455275_3	R	R	R								
1427320_3	R	R	R	R			R	R	R		
1427186_3	R	R	S	S			S	S	S		
1448825_3	S	S	S								
1447514_3	R	R	S								
1408948_4	S	R	S	S			S	S	S	S	
1421939_3	R	R									
1447458_3	R	S	S								
1773_369	S	S	S								
1773_368	S	S	S								
1773_365	R	R	R	R			S				

Illustration of multi-layered analysis workflow



Known AMR genes uncovered by machine learning

Antibiotics	Known AMR genes
Isoniazid	*katG*[43], *inhA*[a20], *fabG1*[44]
Rifampicin	*rpoB*[45], *rpoC*[a46] *Rv3239c*[47]
Ethambutol	*embB*[48], *embC*[17], *ubiA*[a6], *embR*[a17]
Pyrazinamide	*pncA*[49]
Streptomycin	*rpsL*[50], *gidB*[51]
Ofloxacin	*gyrA*[52]
4-Aminosalicylic acid	*folC*[a7], *thyA*[a53]
Ethionamide	*ethA*[54], *inhA*[a20]
Known AMR genes associated with other antibiotics	*dprE1*[55], *ald*[5], *alr*[56], *murA*[57], *pks2*[58], *pks12*[59], *ppsA*[60], *ppsD*[60], *drrB*[61], *drrC*[61], *moeW*[55], *Rv0687*[62], *mshD*[63], *gyrB*[52], *Rv1877*[64], *Rv0194*[65]

Gene	Drug
Rv3848	EMB, XDR
embR	EMB
Rv3129	EMB
proC	EMB
kdpC	EMB
oxcA	INH
chp2	ETA
lipD	ETA
Rv3471c	ETA, XDR, SM
mmpL11	PAS
Rv0044c	PAS
Rv0954	PAS
Rv2560	PZA
Rv2090	RIF, INH
lpqZ	RIF
Rv1597	RIF, MDR, INH
Rv1543	RIF, MDR
nuoL	MDR, PAS
dnaA	SM
yajC	SM
accD5	OFX, MDR
Rv3041c	RIF, OFX, SM, MDR
VapC21	XDR

Newly proposed AMR genes

github.com/aldro61/kover

Paused

README.md

KOVER 2.0

DOI 10.5281/zenodo.2630879 build passing

Kover is an *out-of-core* implementation of rule-based machine learning algorithms that has been tailored for genomic biomarker discovery. It produces highly interpretable models, based on k-mers, that explicitly highlight genotype-to-phenotype associations.

Introduction

Understanding the relationship between the genome of a cell and its phenotype is a central problem in precision medicine. Nonetheless, genotype-to-phenotype prediction comes with great challenges for machine learning algorithms that limit their use in this setting. The high dimensionality of the data tends to hinder generalization and challenges the scalability of most learning algorithms. Additionally, most algorithms produce models that are complex and difficult to interpret. We alleviate these limitations by proposing strong performance guarantees, based on sample compression theory, for rule-based learning algorithms that produce highly interpretable models. We show that these guarantees can be leveraged to accelerate learning and improve model interpretability. Our approach

2.0.3 - Squirrel Monkey Latest
on Apr 6, 2019

+ 10 releases

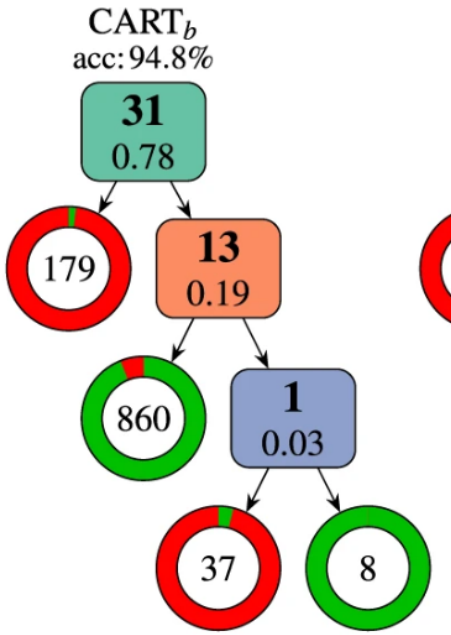
Contributors 2

aldro61 aldro61

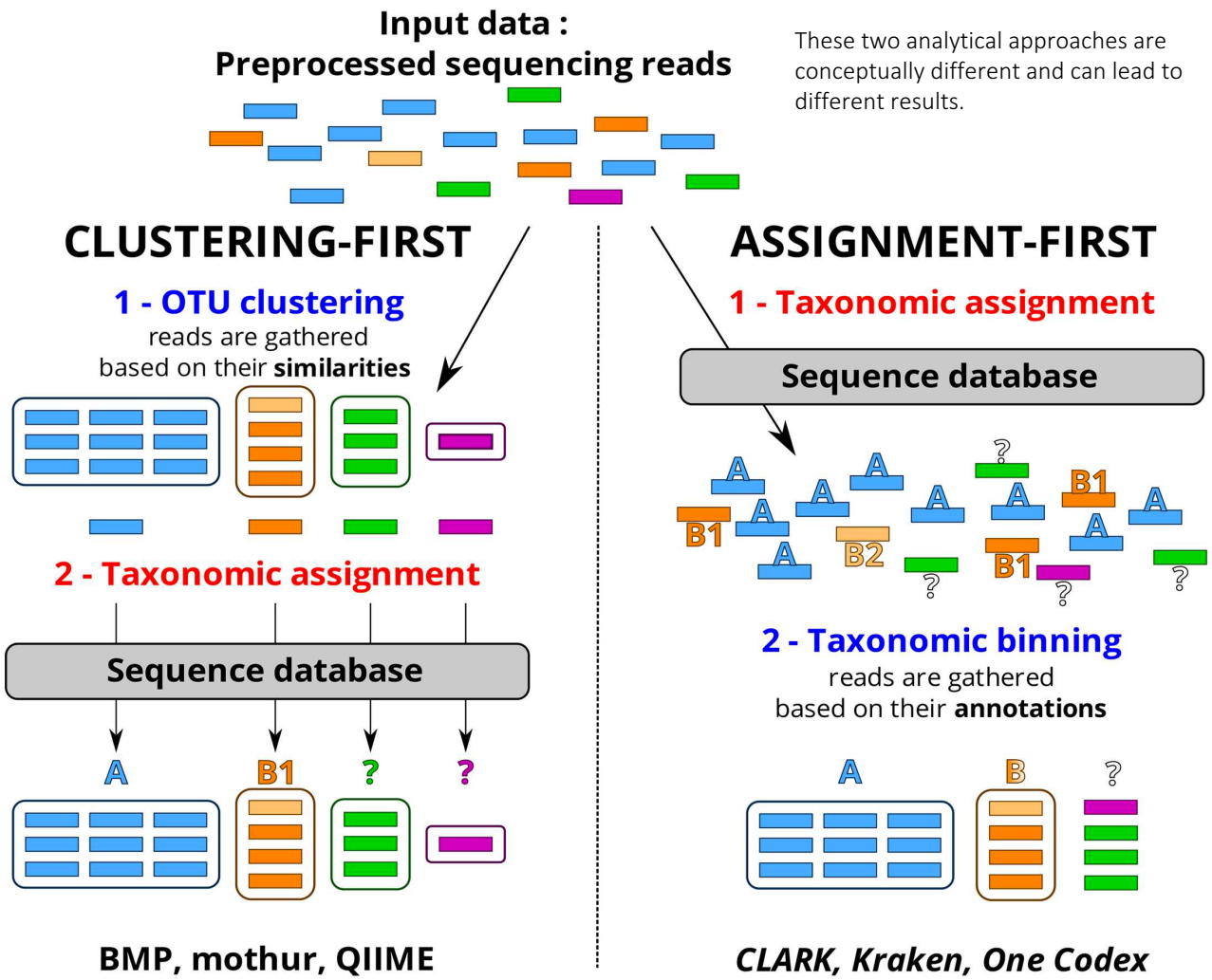
gletarte gletarte

Languages

- **Python** 71.5%
- **C++**
- **CMake** 1.6%
- **Other**



Identification of strains from metagenome sequencing



These two analytical approaches are conceptually different and can lead to different results.

PLoS One. 2017 Jan 4;12(1):e0169563

mothur does not include taxonomic annotations at the species level.
For the QIIME integrated databases, less than 7% of Greengenes sequences and
less than 45% of SILVA sequences are annotated at the species level.

KrakenUniq - NCBI Taxonomy IDs - hundreds of strains share the same taxonomy ID

ONE CODEX

~83K complete microbial genomes, including 50K distinct bacterial genomes, 27K viral genomes, and hundreds of archaeal and eukaryotic genomes.

- **Clinical metagenomics (CMg) refers to the concept of sequencing the DNA of a clinical sample (without any prior culturing step) with the purpose of recovering clinical information**

- **"Salvage microbiology" is a term used for the application of molecular diagnostic techniques in the detection of bacterial DNA directly from clinical specimens**

- **Enable the detection of pathogens and of some antibiotic resistance genes (ARGs)**

Knowledge of the full drug susceptibility profile would enable tailored treatment to improve efficacy and reduce exposure to ineffective toxic drugs

WIMP methods

Useful things to know:

- Reads can only be assigned to organisms that are included in the database and taxonomic subtree e.g. reads will not be assigned as dog (Canis canis familiaris) as this is not included in the database however, many dog reads will likely to classified as "mammalia" as they will have homology to the human reference which is included in the database.
- The WIMP application is optimised for Oxford Nanopore Technologies DNA sequence reads (both length and error profile)
- The application works best with genomic DNA sequences from whole genome or metagenome library preparation. Amplicon sequences such as the 16S rRNA sequence do not contain sufficient taxonomic context for this WIMP method to perform accurate sequence classification.
- The normalised qscore presented in the results is calibrated against the Zymo microbial mock community collection.

Generating WIMP database:

References / data source

WIMP classifies DNA sequence reads against a pre-generated centrifuge index. This index has been prepared from a DNA sequence collection that includes both the RefSeq microbial genome reference sequences and the primary chromosome assemblies from the human genome [1] . The reference genome sequences are downloaded from the NCBI primary sequence database. Dustmasker is used to filter sequence regions of low complexity, in addition to this we mask adapter sequences. The adapters masked are provided here. The WIMP database is generated using the `centrifuge-build` script and is run on an amazon x1e.2xlarge instance. The `centrifuge-inspect` m ethod is used to collate the information for the construction of the taxonomic subtree for the genomic sequences included in the centrifuge database.

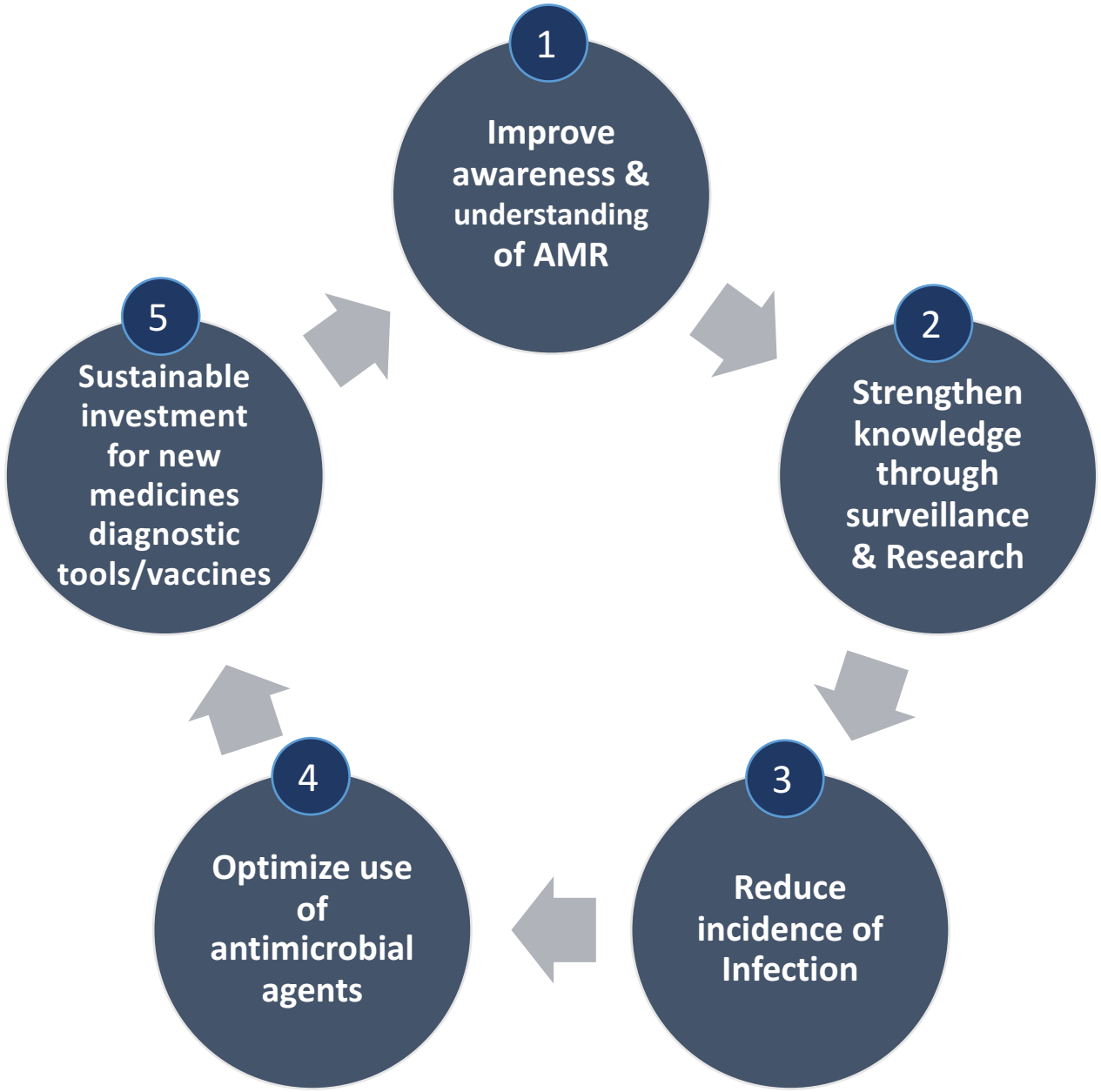
Database details:

DNA sequencing as the new microscope

- It has been about 400 years since the invention of light microscopy, a technology which continues to be used and to evolve.
- It has been only 40 years since the invention of DNA sequencing; the technologies and likely to also continue to develop in the coming decades and centuries.
- It has transformed biomedical research, and is beginning to transform clinical medicine
- DNA sequencing will have a longevity and impact on par with or exceeding that of the microscope.

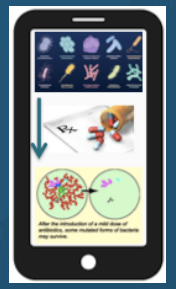
doi:10.1038/nature24286

Sixty-eight World Health Assembly in May 2015 endorsed a global action plan to tackle antimicrobial resistance



Open Antimicrobial Resistance Awareness Drive

Abot – knowledge repository for education & awareness on AMR



A Chatbot will guide the Players on the most viable treatment options

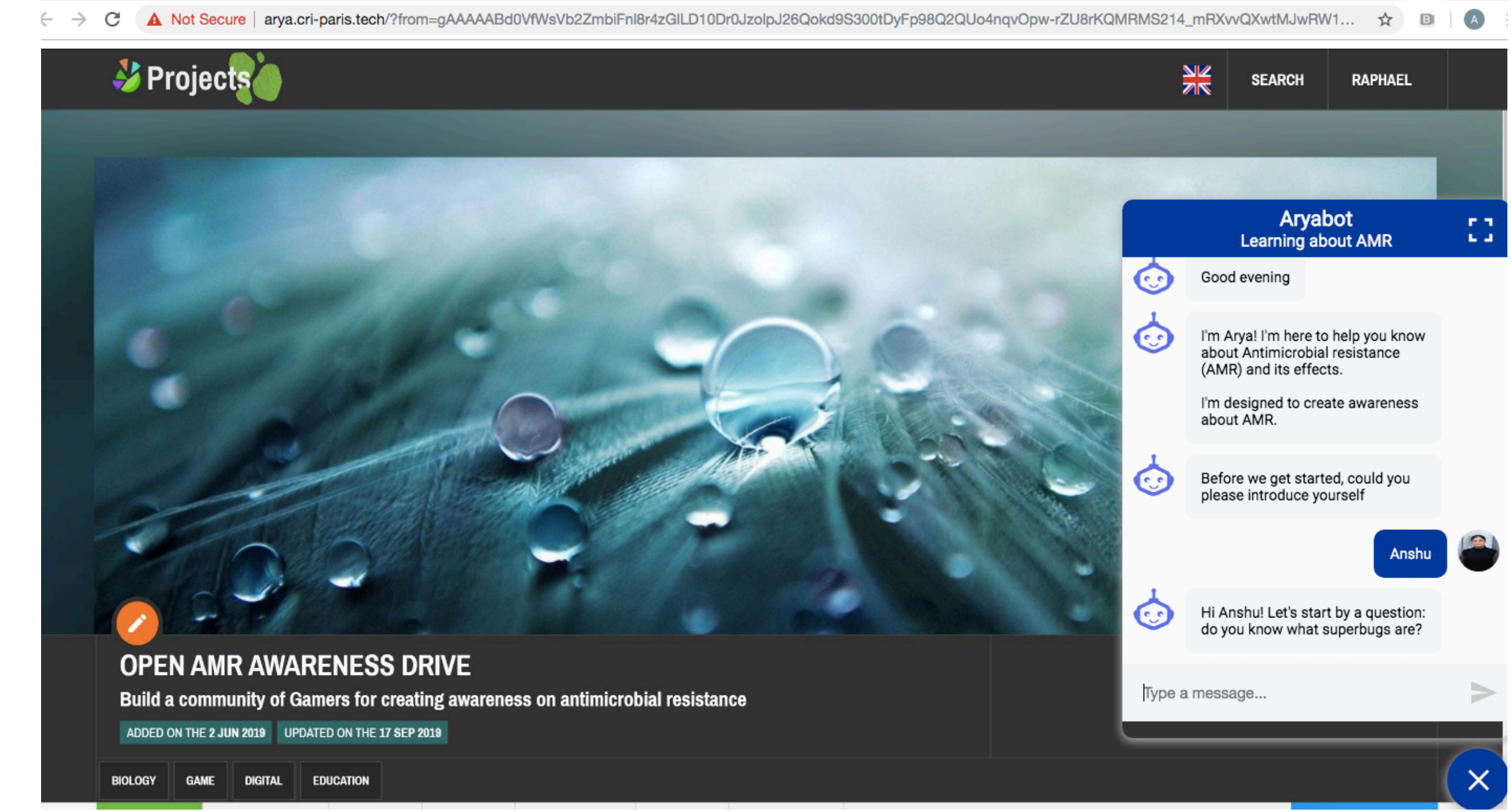
Basic – Treat drug susceptible cases
Advanced – Treat drug resistance cases
Faster cure is the winning bet

- **Bot learning**
- **Structured knowledgebase**
- **Outreach**

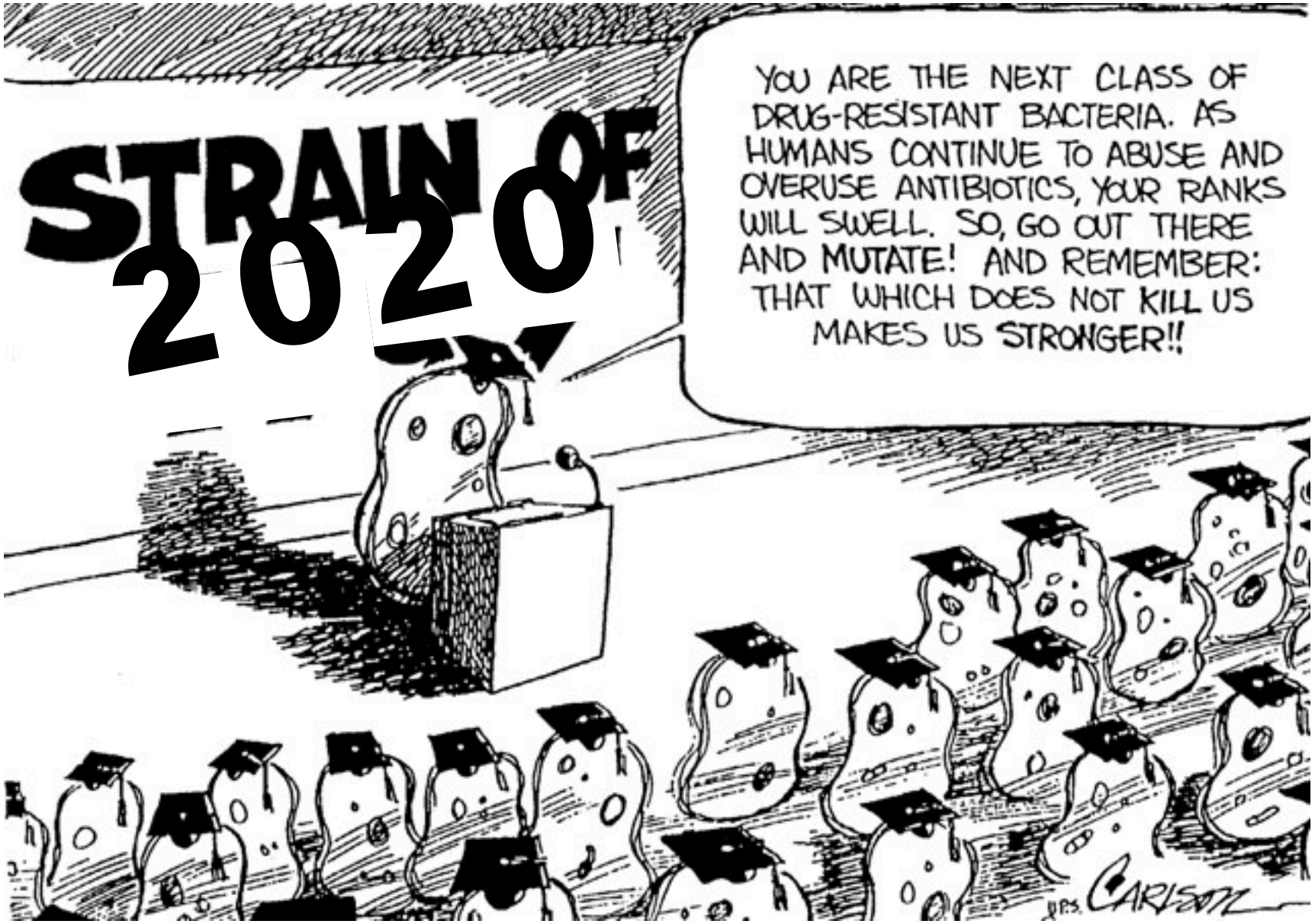
A Games on education & awareness on AMR

- **Game Designer / creative skills**
- **Gamers**
- **Coders**

https://freshdesk.com/help-desk-software/chatbot-conversations-dos-donts-blog/


Not Secure
arya.cri-paris.tech/?from=gAAAAABd0VfWsVb2ZmbiFnl8r4zGlLD10Dr0JzolpJ26Qokd9S300tDyFp98Q2QUo4nqvOpw-rZU8rKQMRMS214_mRXvvQXwtMJwRW1...
Projects
SEARCH
RAPHAEL
OPEN AMR AWARENESS DRIVE
Build a community of Gamers for creating awareness on antimicrobial resistance
ADDED ON THE 2 JUN 2019
UPDATED ON THE 17 SEP 2019
BIOLOGY
GAME
DIGITAL
EDUCATION
Aryabot
Learning about AMR
Good evening
I'm Arya! I'm here to help you know about Antimicrobial resistance (AMR) and its effects.
I'm designed to create awareness about AMR.
Before we get started, could you please introduce yourself
Anshu
Hi Anshu! Let's start by a question: do you know what superbugs are?
Type a message...

Playtest and Feedback Sessions with Aryabot



How do we convey the same in a video game?

https://slideplayer.com/slide/4685976/


00:00
-01:24

Game specifics are based on AWaRe

Initial game levels cannot unlock antibiotics in Watch / Reserve categories

In order to unlock these, the players should accumulate points in earlier layers by assigning right antibiotic for the pathogens

Community Collaborative Projects – Crowdsourcing @ work

TBrowse H37Rv
Integrative Genomics Map of *Mycobacterium tuberculosis*
PMID: 19683474

SysBorg 2.0
PMID: 21782516

OSDD Chem

Mycobacterium IPW Interactome/ Pathway Annotation tuberculosis
PMID: 22808064
PMID: 23629487
PMID: 25304862

BioPhytMol
PMID: 25360160

Cloning, Expression and Purification of Proteins

OSDDChemDesign

Prediction of Potential Inhibitors of Mtb P2I

Drug Repurposing through Network-Based Inference

Systems level modeling for identification of drug targets for gram negatives

dPABBs
Design Peptides Against Bacterial Biofilms
PMID: 26912180

doi.org/10.1007/s12039-017-1268-4
MPDS^{TB}

RepTB
PMID: 29785561

2009 2010 2011 2012 2013 2014 2015 2016 2017----2019/20

MtSNPscore
PMID: 19758471

E-MiDAS

mitolsdb
Human Mitochondrial Locus Specific Variation Database
INSTITUTE OF GENOMICS & INTEGRATIVE BIOLOGY
PMID: 23585830

SSCUE
PMID: 24434286

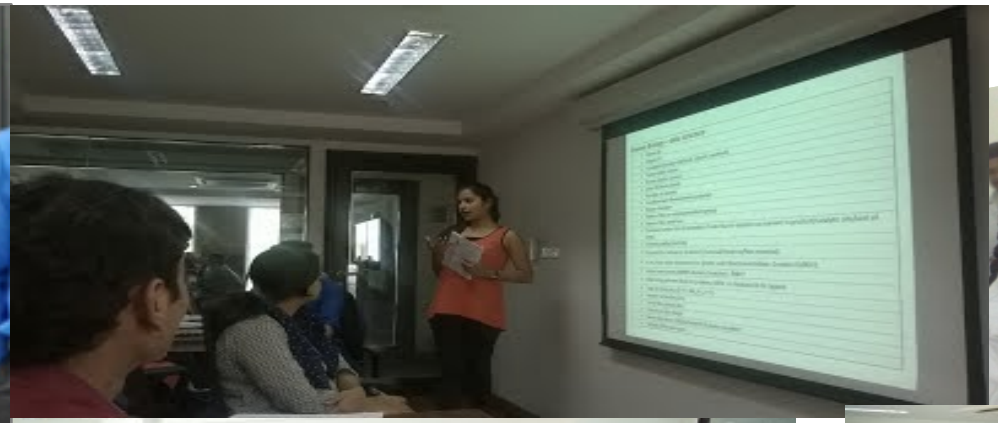
FROG
PMID: 26244889

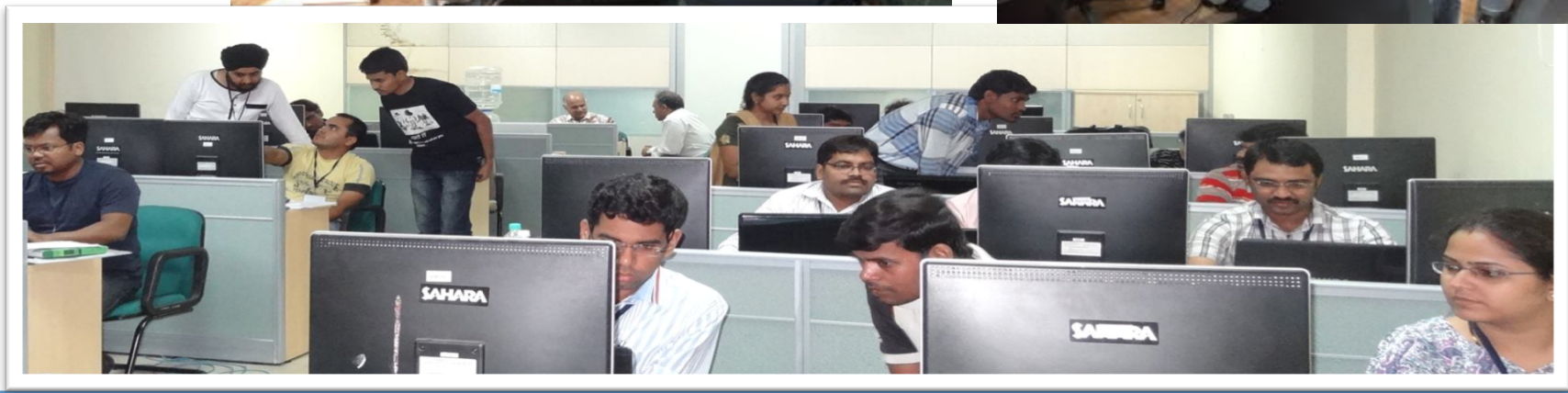
Identification of Clinically actionable variants in case of mt diseases – applicable to complex diseases

PMID: 26180633

MitoLINK
Community Collaborative Systems Level Platform

Worked with over 500 students on several genomics and drug discovery projects





dPABBs: A Novel *in silico* Approach for Predicting and Designing Anti-biofilm Peptides

SCIENTIFIC REPORTS

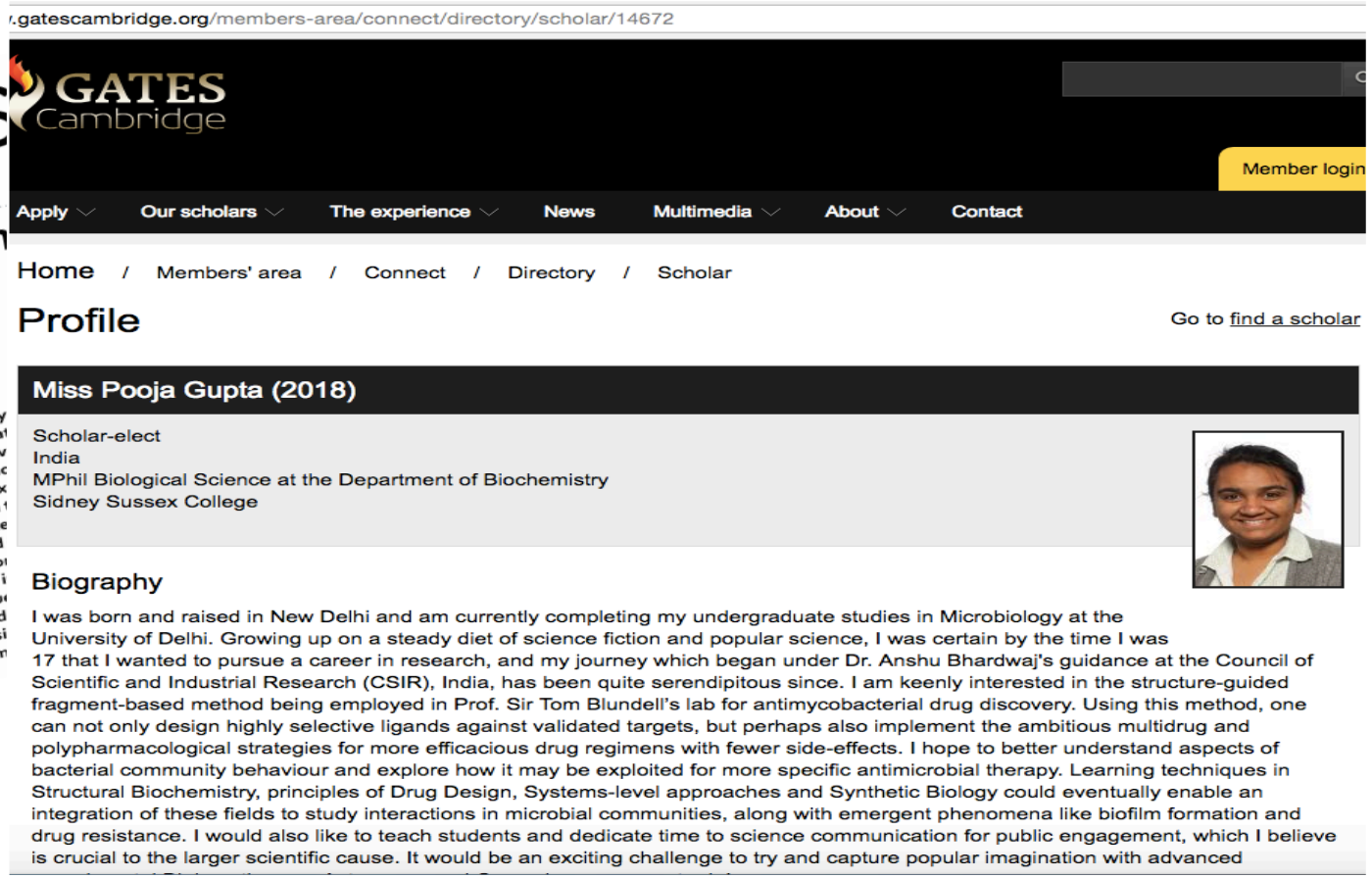
OPEN

dPABBs: A Novel *in silico* Approach for Predicting and Designing Anti-biofilm Peptides

Arun Sharma[1,2,*], Pooja Gupta[1,3,*], Rakesh Kumar[1,2] & Anshu Bhardwaj[1,2]

Received: 03 July 2015
Accepted: 27 January 2016
Published: 25 February 2016

Increasingly, biofilms are being recognised for their causative role in persistent infections (like cy
fibrosis, otitis media, diabetic foot ulcers) and nosocomial diseases (biofilm-infected vascular ca
implants and prosthetics). Given the clinical relevance of biofilms and their recalcitrance to conv
antibiotics, it is imperative that alternative therapeutics are proactively sought. We have develc
dPABBs, a web server that facilitates the prediction and design of anti-biofilm peptides. The six
and Weka models implemented on dPABBs were observed to identify anti-biofilm peptides on
of their whole amino acid composition, selected residue features and the positional preference
residues (maximum accuracy, sensitivity, specificity and MCC of 95.24%, 92.50%, 97.73% and
respectively, on the training datasets). On the N-terminus, it was seen that either of the catio
residues, R and K, is present at all five positions in case of the anti-biofilm peptides, whereas i
peptides, the uncharged polar residue S is preponderant at the first (also anionic polar residu
third and fifth positions. Positive predictions were also obtained for 29 FDA-approved peptid
and ten antimicrobial peptides in clinical development, indicating at their possible repurposi
biofilm therapy. dPABBs is freely accessible on: http://ab-openlab.csir.res.in/abp/antibiofilm



.gatescambridge.org/members-area/connect/directory/scholar/14672

GATES Cambridge

Member login

Apply | Our scholars | The experience | News | Multimedia | About | Contact

Home / Members' area / Connect / Directory / Scholar

Profile

Go to find a scholar

Miss Pooja Gupta (2018)

Scholar-elect
India
MPhil Biological Science at the Department of Biochemistry
Sidney Sussex College

Biography

I was born and raised in New Delhi and am currently completing my undergraduate studies in Microbiology at the University of Delhi. Growing up on a steady diet of science fiction and popular science, I was certain by the time I was 17 that I wanted to pursue a career in research, and my journey which began under Dr. Anshu Bhardwaj's guidance at the Council of Scientific and Industrial Research (CSIR), India, has been quite serendipitous since. I am keenly interested in the structure-guided fragment-based method being employed in Prof. Sir Tom Blundell's lab for antimycobacterial drug discovery. Using this method, one can not only design highly selective ligands against validated targets, but perhaps also implement the ambitious multidrug and polypharmacological strategies for more efficacious drug regimens with fewer side-effects. I hope to better understand aspects of bacterial community behaviour and explore how it may be exploited for more specific antimicrobial therapy. Learning techniques in Structural Biochemistry, principles of Drug Design, Systems-level approaches and Synthetic Biology could eventually enable an integration of these fields to study interactions in microbial communities, along with emergent phenomena like biofilm formation and drug resistance. I would also like to teach students and dedicate time to science communication for public engagement, which I believe is crucial to the larger scientific cause. It would be an exciting challenge to try and capture popular imagination with advanced

Individual Competence to Team Delivery

Journal of Cheminformatics

JOURNAL OF TRANSLATIONAL MEDICINE

nature

SCIENTIFIC REPORTS

CURRENT SCIENCE

PLOS ONE

BMC Bioinformatics

Science Translational Medicine
AAAS

F1000Research
AN OPEN SCIENCE JOURNAL FOR LIFE SCIENTISTS

And more…

AB-OpenLab – India & France

- OSDD Community
- JOGL
- LWB

anshu@imtech.res.in
anshu.bhardwaj
@anshub