

Crushing Antimicrobial Resistance using Explainable Artificial intelligence

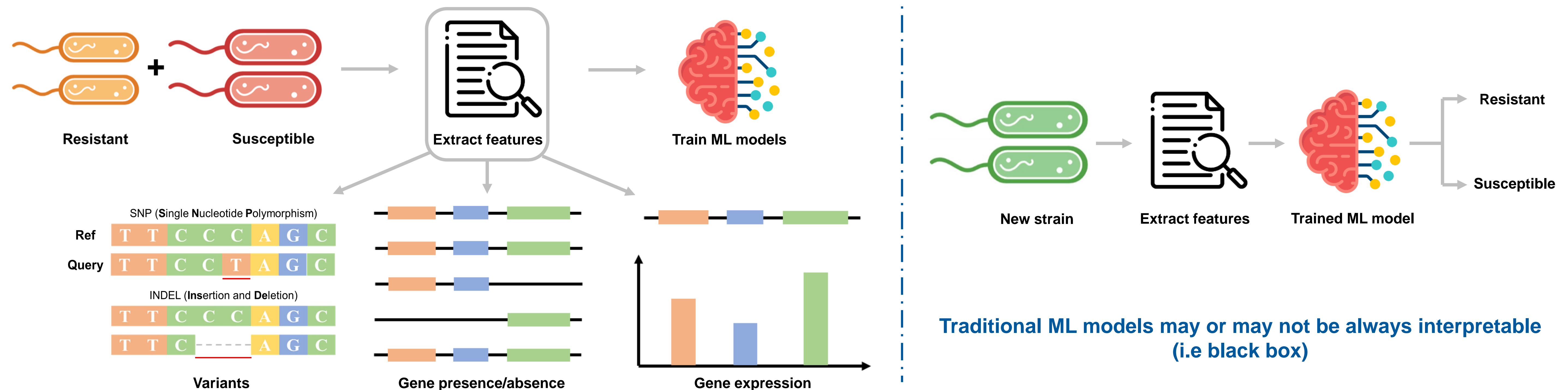
Amay Ajaykumar Agrawal ^{1,2}, Nils Walter ⁴, Alper Yurtseven ^{1,2}, Jilles Vreeken ⁴ and Olga V. Kalinina ^{1,2,3}

¹ Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarbrücken, Germany; ² Center for Bioinformatics, Saarland University, Saarbrücken, Germany; ³ Faculty of Medicine, Saarland University, Homburg, Germany; ⁴ CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

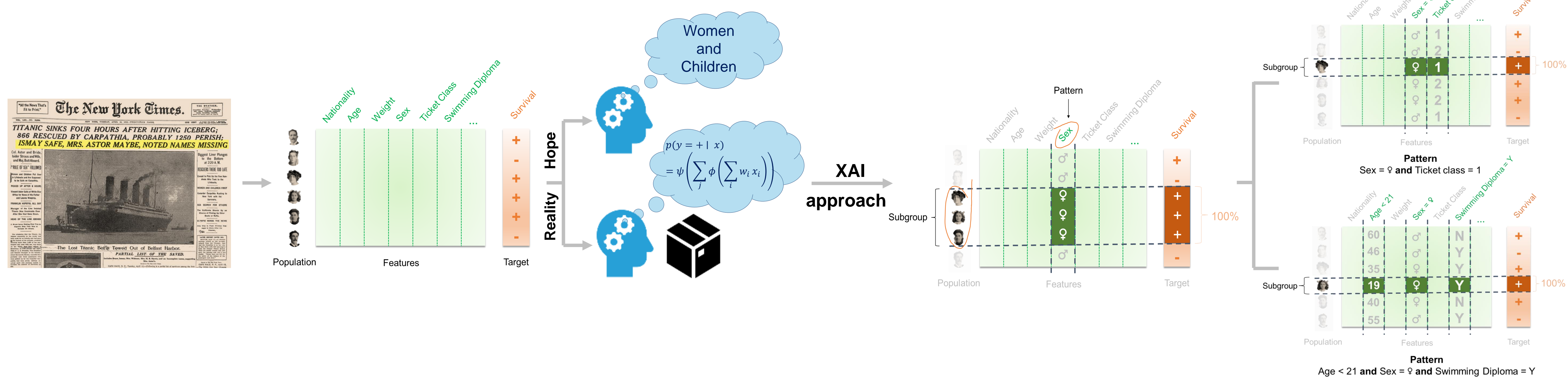
Introduction

Antimicrobial Resistance (AMR) is an urgent threat to human health worldwide as microbes have developed resistance to even the most advanced drugs. Whole-genome sequencing followed by rule-based analysis has accelerated the AMR diagnosis, but it cannot detect resistance caused by unknown AMR markers. Prediction approaches based on machine learning (ML) provide superior performance over rule-based approaches, but they offer limited interpretability. In this work, we propose a novel computational approach to develop inherently interpretable models based on pattern set mining. These models discover patterns of different lengths comprising features (mutations and genes) that are associated with resistance. We applied our models on publicly available AMR data from the PATRIC database for *Mycobacterium tuberculosis* against commonly prescribed antibiotics such as fluoroquinolones and aminoglycosides. Using a basic pattern enumerator model, we rediscovered most known common resistance mechanisms. Additionally, we observe novel sets of genes and mutations that can be related to resistance in rare cases.

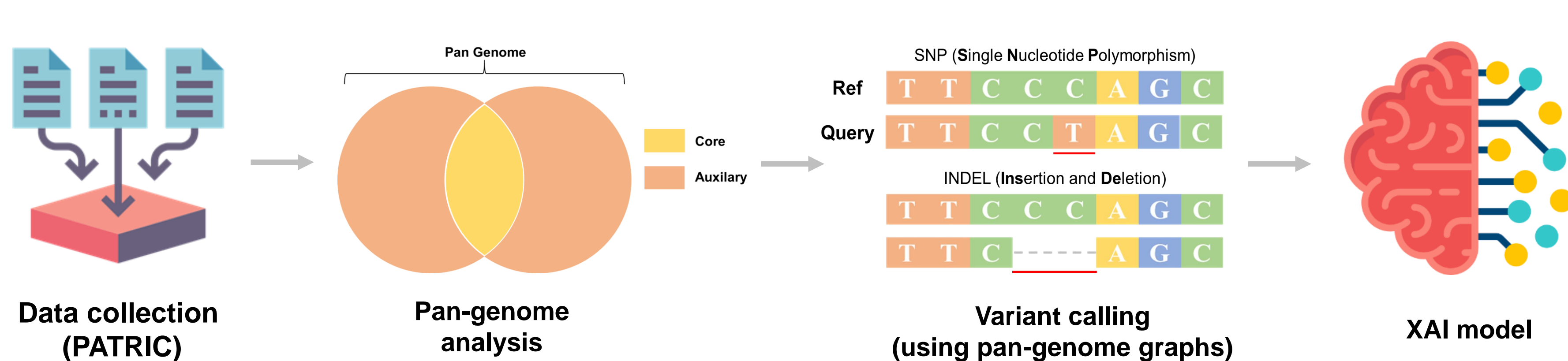
Machine learning (ML) approach for phenotype detection



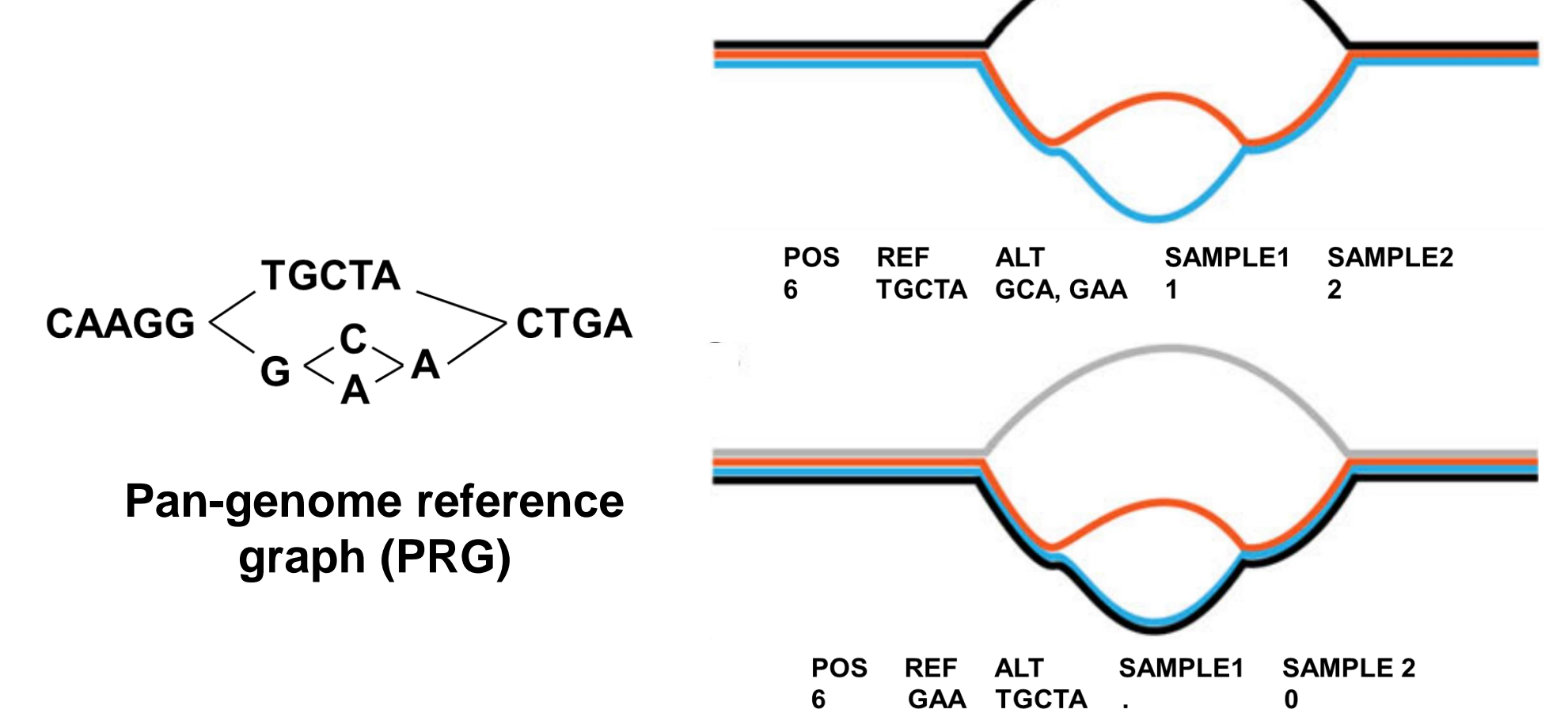
Explainable Artificial Intelligence (XAI) approach to discover patterns of subgroup



Our pipeline



Variant calling using PRG



Results

Ofloxacin

Pattern 1	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 40 Support among susceptible strains (%): 1 P(resistant pattern): 96.89%	(7570, 'C', 'T', 'snp')	gyrA	Fluoroquinolone
Pattern 2	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 20 Support among susceptible strains (%): 0 P(resistant pattern): 98.36%	(206339, 'T', 'C', 'snp')	mecIE	Fluoroquinolone
	(761110, 'A', 'G', 'snp')	rpoB	Rifampicin
	(839129, 'C', 'G', 'snp')	PE_PGRS10	
	(1025106, 'T', 'C', 'snp')		
	(1967237, 'C', 'A', 'snp')		
	(3798095, 'A', 'C', 'snp')	idbB	
Pattern 3	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 20 Support among susceptible strains (%): 0 P(resistant pattern): 96.83%	(261869, 'T', 'C', 'snp')	rpoB	Rifampicin
	(761110, 'A', 'G', 'snp')	phoR	
	(852910, 'C', 'T', 'snp')	hspP	
	(1074558, 'G', 'A', 'snp')	kdpD	
	(1473246, 'A', 'G', 'snp')	rrs	Aminoglycoside
	(1676290, 'C', 'A', 'snp')		
	(1901493, 'T', 'C', 'snp')	blaI	
	(2096186, 'A', 'G', 'snp')		
	(2065760, 'A', 'G', 'snp')		
	(3228143, 'G', 'T', 'snp')		
	(3486977, 'A', 'G', 'snp')	cyp14I	

Mutation (A90V) in gyrA gene known in literature for fluoroquinolone resistance

Amikacin

Pattern 1	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 83 Support among susceptible strains (%): 1 P(resistant pattern): 97.18%	(1473246, 'A', 'G', 'snp')	rrs	Aminoglycoside
Pattern 2	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 44 Support among susceptible strains (%): 0 P(resistant pattern): 98.91%	(265772, 'A', 'C', 'snp')	atfD	Aminoglycoside
	(840496, 'A', 'G', 'snp')	PE_PGRS10	
	(927385, 'A', 'G', 'snp')	PE_PGRS13	
	(979704, 'G', 'C', 'snp')		
	(2155168, 'C', 'G', 'snp')	katG	Isoniazid
	(2984740, 'A', 'G', 'snp')		
	(3080795, 'A', 'G', 'snp')		
	(3746409, 'A', 'G', 'snp')	PPE55	
	(4222073, 'A', 'G', 'snp')		
	(4306155, 'C', 'T', 'snp')		
Pattern 3	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 66 Support among susceptible strains (%): 0 P(resistant pattern): 98.56%	(1473246, 'A', 'G', 'snp')	rrs	Aminoglycoside
	(1620181, 'C', 'G', 'snp')	bisC	
	(2440926, 'G', 'T', 'snp')	aroG	
	(2601576, 'T', 'G', 'snp')	PE23	
	(3036826, 'C', 'T', 'snp')	fadE20	
	(4285120, 'A', 'G', 'snp')	proX	
	(2074754, 'C', 'T', 'snp')		
	(3310626, 'T', 'C', 'snp')		

Mutation in rrs gene known in literature for aminoglycoside resistance

Kanamycin

Pattern 1	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 72 Support among susceptible strains (%): 1 P(resistant pattern): 96.38%	(1473246, 'A', 'G', 'snp')	rrs	Aminoglycoside
Pattern 2	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 60 Support among susceptible strains (%): 0 P(resistant pattern): 97.81%	(420008, 'A', 'G', 'snp')	dnaK	Aminoglycoside
	(836658, 'A', 'G', 'snp')	PE_PGRS9	
	(2155168, 'C', 'G', 'snp')	katG	Isoniazid
Pattern 3	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 19 Support among susceptible strains (%): 0 P(resistant pattern): 98.21%	(1473246, 'A', 'G', 'snp')	rrs	Aminoglycoside
	(773809, 'G', 'A', 'snp')	echA4	
	(812008, 'G', 'A', 'snp')	rpsN1	
	(839949, 'C', 'G', 'snp')	PE_PGRS10	
	(1040459, 'C', 'T', 'snp')	psaS2	
	(1087279, 'T', 'C', 'snp')	accD2	
	(1142266, 'A', 'C', 'snp')	mfd	
	(1144664, 'G', 'A', 'snp')	eno	
	(1446923, 'T', 'G', 'snp')	argS	
	(1452071, 'C', 'A', 'snp')	thrB	
	(162581, 'G', 'A', 'snp')	ephF	
	(178941, 'G', 'T', 'snp')	PE1	
	(180025, 'C', 'T', 'snp')	PE2	
	(196522, 'C', 'T', 'snp')	fadD5	
	(203269, 'C', 'T', 'snp')	mceI1D	

Mutation in rrs gene known in literature for aminoglycoside resistance

Streptomycin

Pattern 1	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 49 Support among susceptible strains (%): 1 P(resistant pattern): 95.39%	(781087, 'A', 'G', 'snp')	rpsL	Streptomycin
Pattern 2	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 15 Support among susceptible strains (%): 0 P(resistant pattern): 99.45%	(311889, 'C', 'T', 'snp')	rpsL	Streptomycin
	(484005, 'G', 'T', 'snp')	fadD30	
	(1151304, 'C', 'A', 'snp')	kdpD	
	(1364706, 'G', 'A', 'snp')	sigE	Isoniazid
	(1446733, 'G', 'A', 'snp')	argS	
	(2095716, 'C', 'T', 'snp')	blaR	
	(2155168, 'C', 'G', 'snp')	katG	Isoniazid
	(2472610, 'C', 'T', 'snp')	colS	
	(2668467, 'G', 'A', 'snp')	mhtD	
	(2774555, 'T', 'C', 'snp')	aglA	
	(2838897, 'C', 'G', 'snp')	fas	
	(2844573, 'C', 'T', 'snp')	PE_PGRS44	
	(2922591, 'C', 'A', 'snp')	altA	Ethambutol
	(4238675, 'C', 'T', 'snp')		
Pattern 3	Variant (POS, REF, ALT, TYPE)	Gene	Resistance against
Support among resistant strains (%): 23 Support among susceptible strains (%): 0 P(resistant pattern): 98.55%	(781087, 'A', 'G', 'snp')	rpsL	Streptomycin
	(781155, 'C', 'T', 'snp')	rpoB	Rifampicin
	(2155168, 'C', 'G', 'snp')	katG	Isoniazid
	(4053050, 'A', 'G', 'snp')		

Mutation (K43R) in rpsL gene known in literature for streptomycin resistance

Conclusion

- Performed pan-genome analysis to split the genome into core & auxiliary genome
- Performed variant calling using PRG's as reference [1]
- Developed inherently interpretable models based on pattern set mining & applied it on *Mycobacterium tuberculosis*
- Rediscovered most known common resistance markers. Additionally, we discovered novel markers that can be related to resistance in rare cases

Outlook

- Models can be further tuned to take cross resistances into account
- Function of novel markers can be determined & further used in AMR diagnostics

To know more scan me!



References

- Colquhoun et al. (2021) *Genome biology* 22
- Wattam et al. (2014) *Nucleic acids research* 42
- Kim JI et al. (2022) *Clinical Microbiology Reviews* 35.3
- McArthur Andrew G. et al. (2013) *Antimicrobial agents and chemotherapy* 57.7
- Loh HW et al. (2022) *Computer Methods and Programs in Biomedicine*