



Quantitative analysis of geno/pheno data - 24 June 2025

Goals

To define AMRrules for interpreting acquired genes/mutations based on quantitative analysis of matched genome/AST data

- Automate as far as possible
 - AMRgen, AMRrulemakeR packages
 - Followed by expert curation to review and encode other sources of evidence and PMIDs
- Track quantity and quality of supporting data for each rule
 - Decide logic and thresholds for setting rules, encoding evidence grades and limitations
 - Add new fields to the specification to encode quantitative evidence
- Starting point: AMRfinderplus genotype data, with disk and/or MIC assay data (CARD/ResFinder compatibility to follow later)
 - Decide minimum requirements for input data

Agenda today

Tools in development

- AMRgen package
- AMRrulemakeR package

Topics for discussion

- Rule-defining logic
- Data quality
- Thresholds for data quality / quantity and how these should map to
 - evidence grade
 - evidence limitations
- Quantitative fields to include in AMRrules specification

AMRgen R package

<https://github.com/AMRverse/AMRgen>

```
geno <- import_amrfp("amrfinderplus.tsv", sample_col = "Name")
```

Input

Name	'Gene symbol'	'Element type'	'Element subtype'	Class	Subclass	Method
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	002_S03 aac(3)-IIe	AMR	AMR	AMINOGLYCOSIDE	GENTAMICIN	EXACTX
2	002_S03 aph(3')-Ib	AMR	AMR	AMINOGLYCOSIDE	STREPTOMYCIN	EXACTX
3	002_S03 aph(6)-Id	AMR	AMR	AMINOGLYCOSIDE	STREPTOMYCIN	EXACTX
4	002_S03 aadA2	AMR	AMR	AMINOGLYCOSIDE	STREPTOMYCIN	EXACTX
5	002_S03 aadA1	AMR	AMR	AMINOGLYCOSIDE	STREPTOMYCIN	PARTI...
6	002_S03 oqx8	AMR	AMR	PHENICOL/QUINOLONE	PHENICOL/QU...	BLASTX
7	002_S03 oqxA	AMR	AMR	PHENICOL/QUINOLONE	PHENICOL/QU...	EXACTX
8	003_S08 rmtB1	AMR	AMR	AMINOGLYCOSIDE	AMINOGLYCOS...	ALLEL...
9	003_S08 aac(3)-IID	AMR	AMR	AMINOGLYCOSIDE	GENTAMICIN	EXACTX
10	003_S08 aph(3')-Ia	AMR	AMR	AMINOGLYCOSIDE	KANAMYCIN	EXACTX

Output

Name	gene	mutation	node	'variation type'	marker	marker.label	drug_agent	drug_class	'Gene symbol'
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	002_S03 ompK36	Asp135AspGlyAsp	ompK36	Protein variant detected	ompK36_D135DGD	ompK36:Asp135AspGlyAsp	NA	Carbapenems	ompK36_D135D...
2	002_S03 gyrA	Ser83Ile	gyrA	Protein variant detected	gyrA_S83I	gyrA:Ser83Ile	NA	Quinolones	gyrA_S83I
3	002_S03 parC	Ser80Ile	parC	Protein variant detected	parC_S80I	parC:Ser80Ile	NA	Quinolones	parC_S80I
4	002_S03 aac(3)-IIe	NA	aac(3)-IIe	Gene presence detected	aac(3)-IIe	aac(3)-IIe	GEN	Aminoglycosi...	aac(3)-IIe
5	002_S03 aph(3')-Ib	NA	aph(3')-Ib	Gene presence detected	aph(3')-Ib	aph(3')-Ib	STR1	Aminoglycosi...	aph(3')-Ib
6	002_S03 aph(6)-Id	NA	aph(6)-Id	Gene presence detected	aph(6)-Id	aph(6)-Id	STR1	Aminoglycosi...	aph(6)-Id
7	002_S03 aadA2	NA	aadA2	Gene presence detected	aadA2	aadA2	STR1	Aminoglycosi...	aadA2
8	002_S03 aadA1	NA	aadA1	Inactivating mutation detected	aadA1	aadA1:-	STR1	Aminoglycosi...	aadA1
9	002_S03 oqx8	NA	oqx8	Gene presence detected	oqx8	oqx8	NA	Amphenicols	oqx8
10	002_S03 oqxA	NA	oqxA	Gene presence detected	oqxA	oqxA	NA	Amphenicols	oqxA
11	002_S03 oqxB	NA	oqxB	Gene presence detected	oqxB	oqxB	NA	Quinolones	oqxB
12	002_S03 oqxA	NA	oqxA	Gene presence detected	oqxA	oqxA	NA	Quinolones	oqxA

AMRrules syntax for specifying variants

marker.label for analysis (node:mutation)

} copy for
each class

AMRgen R package

<https://github.com/AMRverse/AMRgen>

Import AST results to standard format using AMR classes

> ast	# A tibble: 455,723 × 6					
	id	drug_agent	pheno	source	mic	disk
	<chr>	<ab>	<sir>	<chr>	<mic>	<dsks>
1	SAMN02138648	TZP	R	NCBI	64	NA
2	SAMN02138670	AMP	R	NCBI	16	NA
3	SAMN02138669	AMP	R	NCBI	16	NA
4	SAMN02138668	AMP	S	NCBI	4	NA
5	SAMN02356582	AMK	S	NCBI	2	NA
6	SAMN02356581	AMK	S	NCBI	2	NA
7	SAMN02138649	FEP	I	NCBI	2	NA
8	SAMN02138647	AMK	S	NCBI	2	NA

← AMR package classes

source to track unique datasets

AMRgen R package

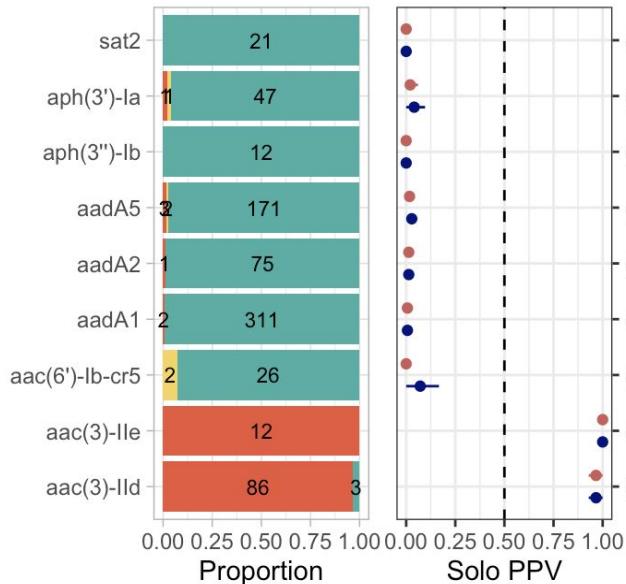
<https://github.com/AMRverse/AMRgen>

Calculate positive predictive value (PPV) of markers found ‘solo’

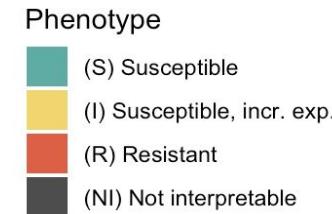
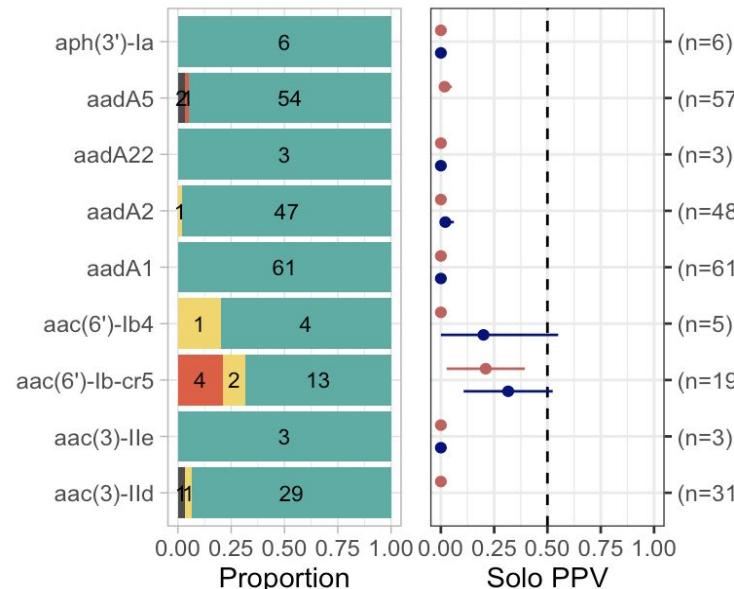
```
solo_ppv_analysis(geno, ast, drug_class_list=c("Aminoglycosides"), antibiotic="Gentamicin")
```

```
solo_ppv_analysis(geno, ast, drug_class_list=c("Aminoglycosides"), antibiotic="Amikacin")
```

Solo markers for class: Aminoglycosides
vs phenotype for drug: Gentamicin



Solo markers for class: Aminoglycosides
vs phenotype for drug: Amikacin



Category

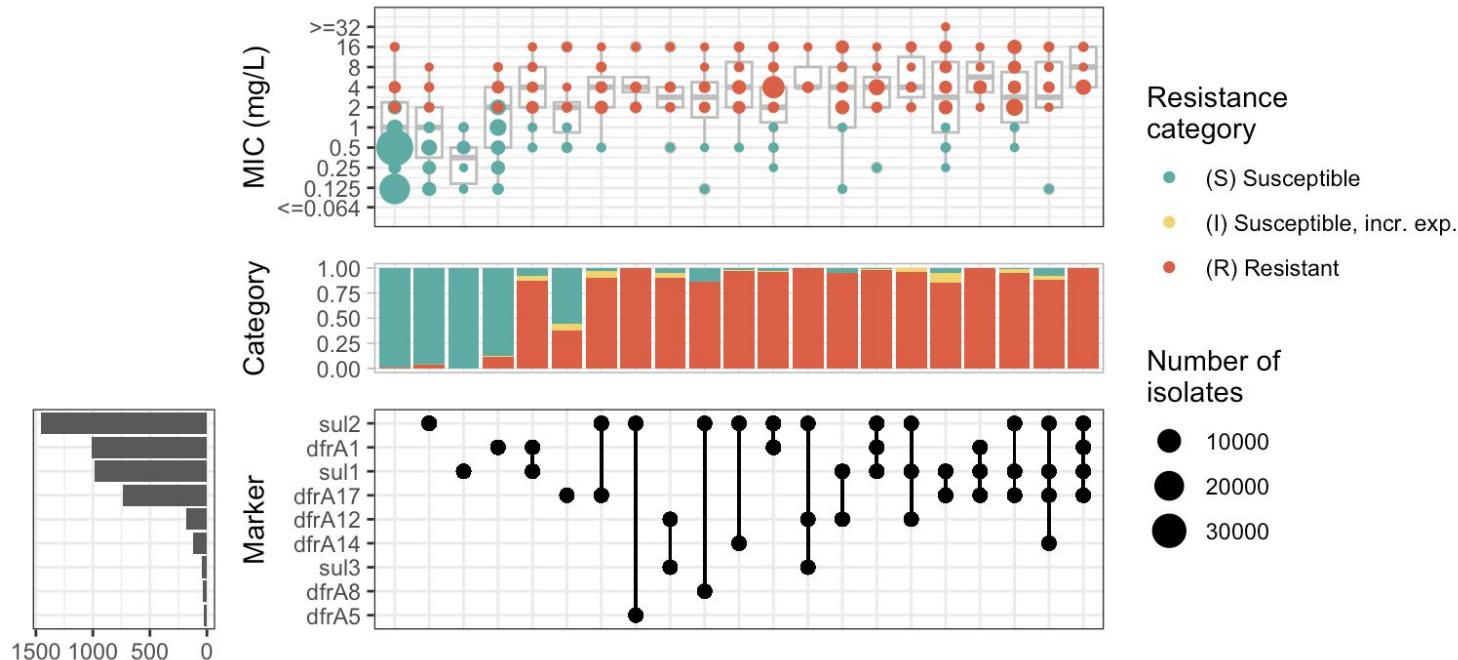
- NWT
- R

AMRgen R package

<https://github.com/AMRverse/AMRgen>

Upset plot, to explore MIC distribution associated with marker combinations

```
amr_upset(get_binary_matrix(geno, ast, antibiotic="Trimethoprim-sulfamethoxazole",
                             drug_class_list=c("Trimethoprim", "Sulfonamides"), min_set_size=20, assay="mic"))
```

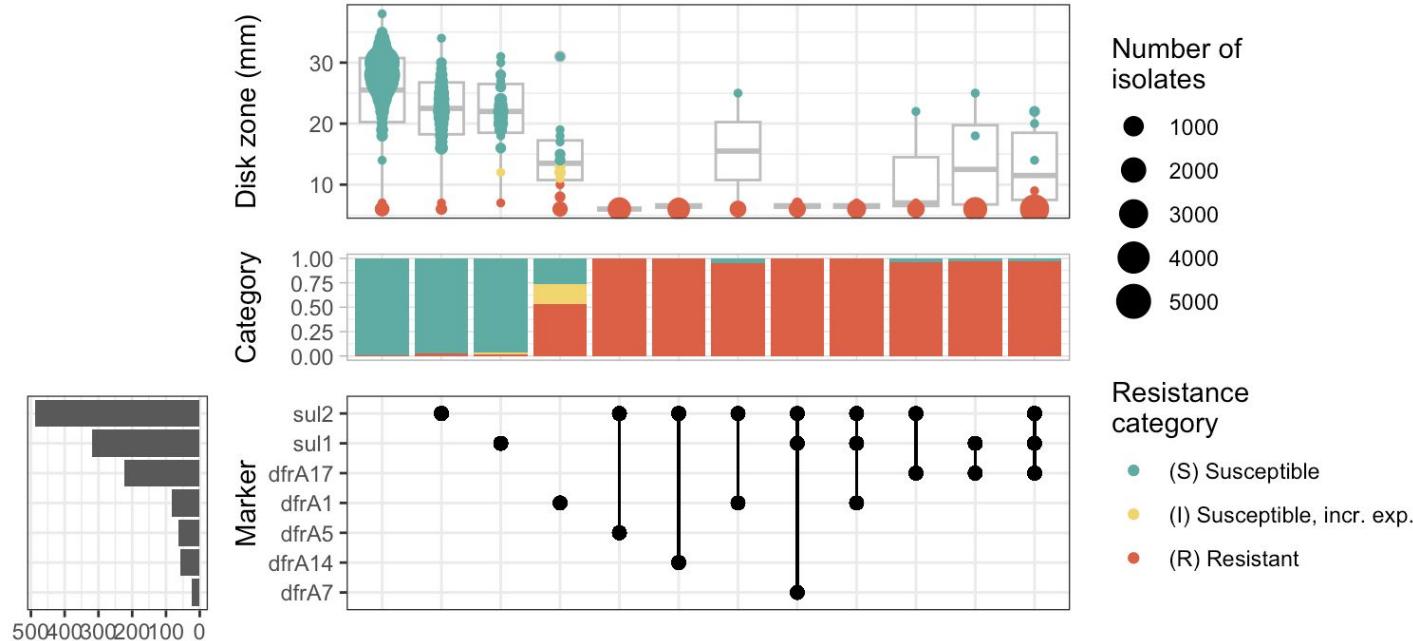


AMRgen R package

<https://github.com/AMRverse/AMRgen>

Upset plot, to explore MIC distribution associated with marker combinations

```
amr_upset(get_binary_matrix(geno, ast, antibiotic="Trimethoprim-sulfamethoxazole",
                            drug_class_list=c("Trimethoprim", "Sulfonamides"), min_set_size=20, assay="disk")
```



AMRrulemakeR package

<https://github.com/AMRverse/AMRrulemakeR>

```
summarise_data(afp, ast, "Ciprofloxacin", drug_class_list=c("Quinolones"), geno_sample_col="Name", pheno_sample_col="id")
```

```
summarise_data(  
  geno_table,  
  pheno_table,  
  antibiotic,  
  drug_class_list,  
  geno_sample_col = "Name",  
  pheno_sample_col = "id",  
  species,  
  guide = "EUCAST 2024"  
)
```

Samples with Quinolones genotypes:	120350
Samples with Ciprofloxacin phenotypes:	19325
- MIC	10013
- DISK	2150
Samples with genotypes and phenotypes:	4633
- MIC	2237
- DISK	979
EUCAST breakpoint sites:	3
- DISK / Non-meningitis	25, 22
- MIC / Non-meningitis	0.25, 0.5
- MIC / Meningitis	0.125, 0.125
EUCAST ECOFFs:	2
- DISK / NA	25, 25
- MIC / NA	0.064, 0.064

AMRrulemakeR package

<https://github.com/AMRverse/AMRrulemakeR>

```
analysis <- amrrules_analysis(afp, ast, antibiotic="Ciprofloxacin", drug_class_list=c("Quinolones"), species="E. coli")
```

	Value
amrrules_analysis(A list containing:
geno_table,	reference_mic_plot EUCAST reference MIC distribution plot
pheno_table,	reference_disk_plot EUCAST reference disk zone distribution plot
antibiotic,	summary Output of <code>summarise_data</code> showing sample and breakpoint summaries
drug_class_list,	solo_stats PPV statistics for individual markers
species,	solo_binary Binary matrix of individual marker presence by sample
sir_col = "pheno",	amr_binary Binary matrix of AMR marker presence by sample
geno_sample_col = "Name",	ppv_plot Bar plot summarizing PPV results
pheno_sample_col = "id",	logistic_mat Binary matrix used in logistic regression
marker_col = "marker.label",	logistic_plot Plot of logistic regression estimates
minPPV = 1,	ppv_logistic_plot Combined PPV/logistic plot
mafLogReg = 5,	modelR Logistic model for predicting resistance
mafUpset = 5	modelNWT Logistic model for predicting NWT (non-wild-type)
)	allstatsR Merged statistics for R category
	allstatsNWT Merged statistics for NWT category
	upset_mic_plot Upset plot for MIC data
	upset_disk_plot Upset plot for disk data
	upset_mic_summary MIC data summarised per marker or combination
	upset_disk_summary Disk diffusion data summarised per marker or combination
	combination_summary_values Summary of genotype combinations from both MIC and disk
	afp_hits List of AMR markers detected
	species The species used in the analysis
	antibiotic The antibiotic used in the analysis

AMRrulemakeR package

<https://github.com/AMRverse/AMRrulemakeR>

```
rules <- makerules(analysis, guide="EUCAST 2025", bp_site=NULL, weak_threshold=20, core_threshold=0.9)
```

```
makerules(  
  amrrules,  
  minObs = 3,  
  weak_threshold = 20,  
  core_threshold = 0.9,  
  use_mic = TRUE,  
  mic_S = NULL,  
  mic_R = NULL,  
  mic_ecoff = NULL,  
  use_disk = TRUE,  
  disk_S = NULL,  
  disk_R = NULL,  
  disk_ecoff = NULL,  
  guide = "EUCAST 2025",  
  bp_site = NULL,  
  rule_prefix = NULL,  
  ruleID_start = 1000,  
  note_prefix = "Quantitative geno-pheno analysis by ESGEM-AMR WG",  
  regression = TRUE  
)
```

Propose rulesets based on outputs
of `amrrules_analysis()`

For single markers and
combinations

AMRrulemakeR package

<https://github.com/AMRverse/AMRrulemakeR>

```
rules <- makerules(analysis, guide="EUCAST 2025", bp_site=NULL, weak_threshold=20, core_threshold=0.9)
```

A	C	D	I	J	L	N	O	P	Q	S	T	U	X
ruleID	gene	nodeID	mutation	variation type	drug	phenotype	clinical	breakpoint	breakpoint stan	evidence code	evidence gra	evidence limitations	marker
KPN5062	ompK36	ompK36	Asp135AspAsp	Protein variant detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	weak	No solo geno-pheno d	ompK36:Asp135AspAsp
KPN5046	ompK36	ompK36	Gln131Ter	Protein variant detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	ompK36:Gln131Ter
KPN5062	ompK36	ompK36	Asp135AspAsp	Protein variant detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	weak	No solo geno-pheno d	ompK36:Asp135AspAsp
KPN5045	ompK36	ompK36	Asp135AspGlyAsp	Protein variant detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Con	ompK36:Asp135AspGlyAsp
KPN5047	ompK36	ompK36	Thr136ThrAspThr	Protein variant detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Con	ompK36:Thr136ThrAspThr
KPN5044	ompK35	ompK35	Glu132Lys	Protein variant detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	moderate		ompK35:Glu132Lys
KPN5043	blaVIM-1	blaVIM-1	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaVIM-1
KPN5029	blaOXA-48	blaOXA-48	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	moderate		blaOXA-48
KPN5028	blaOXA-232	blaOXA-232	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Con	blaOXA-232
KPN5024	blaNDM-5	blaNDM-5	-	Gene presence detected	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaNDM-5
KPN5023	blaNDM-4	blaNDM-4	-	Gene presence detected	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	blaNDM-4
KPN5022	blaNDM-1	blaNDM-1	-	Gene presence detected	Imipener	nonwildtype	I	disk zone >= 19 & < 22 mm	EUCAST 2025	ECO:0001103	weak	Conflicting evidence.	blaNDM-1
KPN5020	blaKPC-3	blaKPC-3	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Conflicting evidence.	blaKPC-3
KPN5019	blaKPC-2	blaKPC-2	-	Gene presence detected	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	moderate		blaKPC-2
KPN5017	blaIMP-4	blaIMP-4	-	Gene presence detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	blaIMP-4
KPN5016	blaIMP-19	blaIMP-19	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Con	blaIMP-19
KPN5015	blaIMP-1	blaIMP-1	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Conflicting evidence.	blaIMP-1
KPN5144	KPN5027 & KPN5047		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaOXA-181, ompK36:Thr136ThrAspThr
KPN5148	KPN5020 & KPN5062		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaKPC-3, ompK36:Asp135AspAsp
KPN5149	KPN5020 & KPN5047		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	moderate	Limited MIC datasets.	blaKPC-3, ompK36:Thr136ThrAspThr
KPN5151	KPN5020 & KPN5045		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	moderate		blaKPC-3, ompK36:Asp135AspGlyAsp
KPN5153	KPN5022 & KPN5055		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	moderate		blaNDM-1, ompK35:Trp230Ter
KPN5154	KPN5010 & KPN5022		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	blaNDM-1, blaCTX-M-55
KPN5155	KPN5009 & KPN5022		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	blaNDM-1, blaCTX-M-3
KPN5156	KPN5022 & KPN5041		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaNDM-1, blaSHV:-112C>A

rule curation note

Quantitative geno-pheno analysis by ESGEM-AMR WG. ompK36:Asp135AspAsp. Category call 'S' based on logistic regression. Phenotype call 'wildtype' based on logistic regression. OR for R: 2.281 [0.687-7.575], p=1.782e-01. OR for NWT: 2.286 [0.666-7.848], p=1.891e-01.

Quantitative geno-pheno analysis by ESGEM-AMR WG. ompK35:Glu132Lys. Category call 'S' based on Solo PPV. MIC. Disk. Phenotype call 'wildtype' based on MIC. Disk. R PPV=1% (1/166). NWT PPV=1% (2/166). 14 solo datasets: 13 S, 1 I, 1 R. MIC: median 0.25 [IQR 0.25-0.5]; n=118. 16 MIC datasets: 14 S, 4 I, 9 R. Disk: median 26 [IQR 25-28]; n=48. 3 disk datasets: 3 S, 0 I, 0 R.

Agenda today

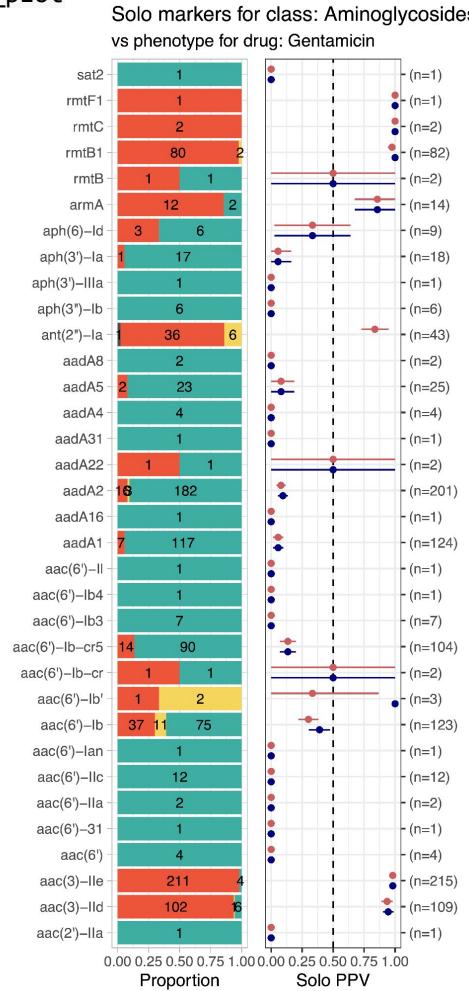
Tools in development

- AMRgen package
- AMRrulemakeR package

Topics for discussion

- Rule-defining logic
- Data quality
- Thresholds for data quality / quantity and how these should map to
 - evidence grade
 - evidence limitations
- Quantitative fields to include in AMRrules specification

analysis\$ppv_plot



Solo PPV provides strong indication of association with R or NWT

but sample size limited as rare to find a single marker

analysis\$allstatsR

analysis\$allstatsNWT

marker	x	n	ppv	ci.lower.ppv	ci.upper.ppv
aac(3)-lle	211	215	0.9814	0.96333319	0.999457505
aadA2	16	201	0.0796	0.04218164	0.11702234
aadA1	7	124	0.05645	0.01582922	0.09707401
aac(6')-lb	37	123	0.30081	0.21976385	0.381862163
aac(3)-lld	102	109	0.93578	0.88975782	0.98180181
aac(6')-lb-cr5	14	104	0.13462	0.06901727	0.200213496
rmtB1	80	82	0.97561	0.94222138	1
ant(2')-la	36	43	0.83721	0.72686405	0.947554556
aadA5	2	25	0.08	0	0.186346934
aph(3')-la	1	18	0.05556	0	0.161376534
armA	12	14	0.85714	0.67383983	1
aac(6')-llc	0	12	0	0	0
aph(6')-ld	3	9	0.33333	0.02534905	0.64131762
aac(6')-lb3	0	7	0	0	0
aph(3')-lb	0	6	0	0	0
aadA4	0	4	0	0	0
aac(6')	0	4	0	0	0
aac(6')-lb'	1	3	0.33333	0	0.866777766
rmtC	2	2	1	1	1
aadA22	1	2	0.5	0	1
rmtB	1	2	0.5	0	1
aac(6')-lb-cr	1	2	0.5	0	1
aadA8	0	2	0	0	0

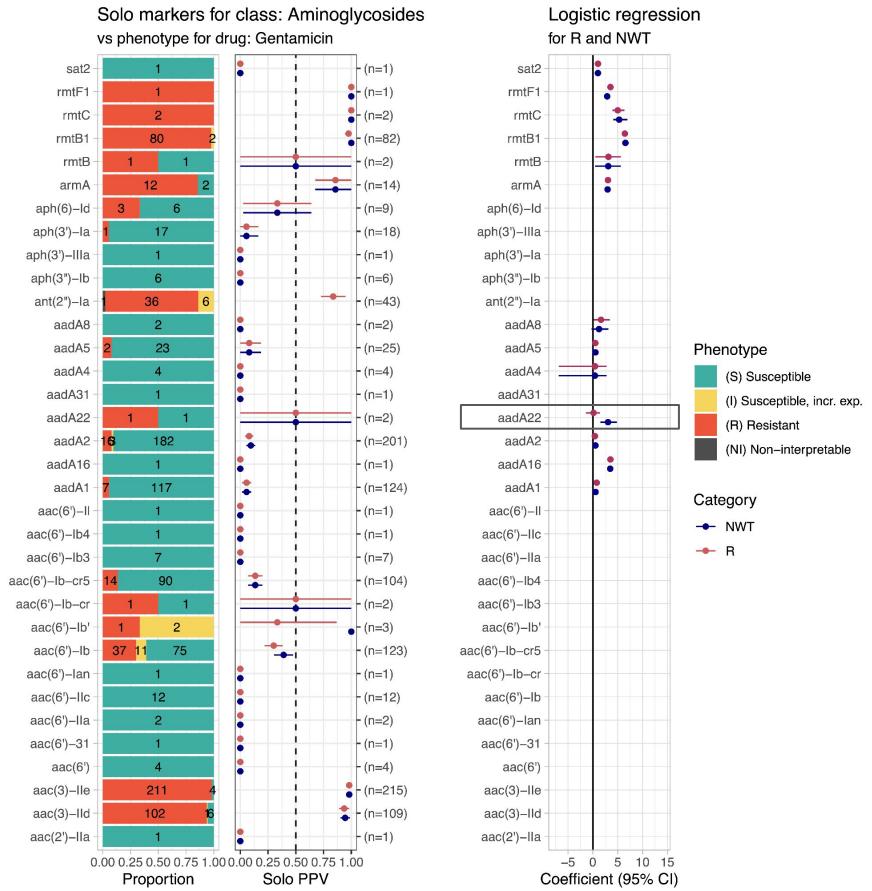
Initial logic and thresholds

`makerules(minObs=3)`

Step 1: Solo PPV (based on interpretation data, for markers found solo)

NWT PPV	R PPV	call	phenotype	if expected R	if expected I
...	>=80%	R	NWT	WT	NWT
>50%	NA	I	NWT	REVIEW: expected R	WT
>50%	<80%	I	NWT	REVIEW: expected R	WT
NA	>50%	I	NWT	REVIEW: expected R	WT
NA	<=50%	S	WT	REVIEW: expected R	WT
<=50%	...	S	WT	REVIEW: expected R	WT

analysis\$ppv_logistic_plot



Logistic regression has utility in providing evidence against R

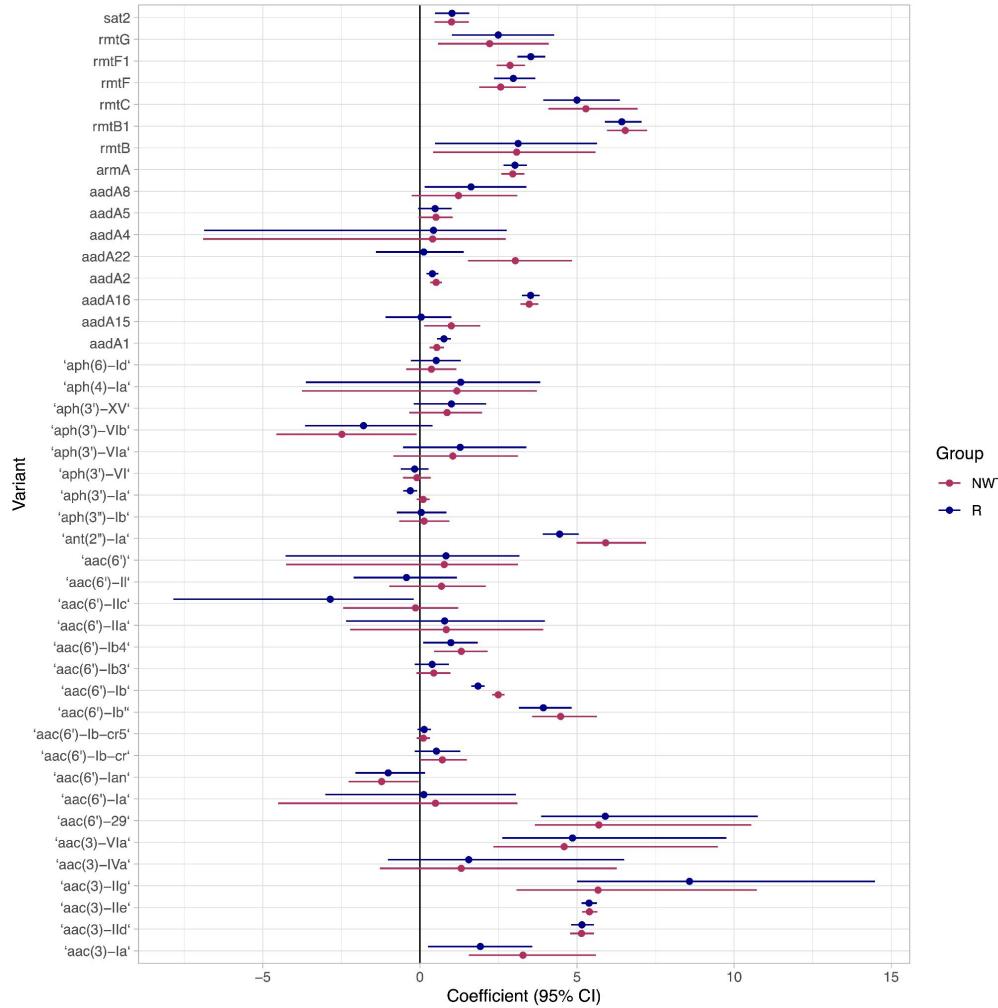
helpful when marker found rarely found solo

analysis\$upset_mic_summary

marker_list	ppv	R	n
aadA22	0	0	1
aac(6')-lb, aadA22	0	0	8
aph(3')-Ia, aadA22	0	0	1
aadA22, aac(3')-Vla	1	1	1
aph(3')-lb, aac(6')-lb, aadA22	1	1	1
aac(6')-lb, ant(2')-Ia, aadA22	1	1	1
aph(6')-Id, aac(3')-Ild, aadA22	1	1	1
aph(3')-Ia, aac(6')-Id, aac(3')-Ild, aadA22	1	1	1
aac(3)-IVa, aadA2, aac(4)-Ia, aac(6')-Ild, aac(6')-lb-cr5, aadA16, rmtB1, aadA22	1	1	1
aac(3)-IVa, aadA2, aac(4)-Ia, aac(6')-Id, aac(3)-Ild, rmtB1, aadA22	1	2	2
aph(3')-Ia, aac(6')-Id, aac(3')-Ild, aac(6')-lb-cr5, aadA16, aadA22	1	13	13
			21 31
Markers responsible for resistance			
ant(2')-Ia	85%	35	42
aac(3)-Ild	95%	91	96
aph(6)-Id	75%	3	4
rmtB1	98%	80	82

```
analysis$logistic_plot
```

**Data on all markers
(observed in at least
X genomes, default 5)**



Initial logic and thresholds

Step 2: Logistic regression

Models run:

NWT ~ marker₁ + marker₂ + marker₃ +

amrrules_analysis(mafLogReg=5)

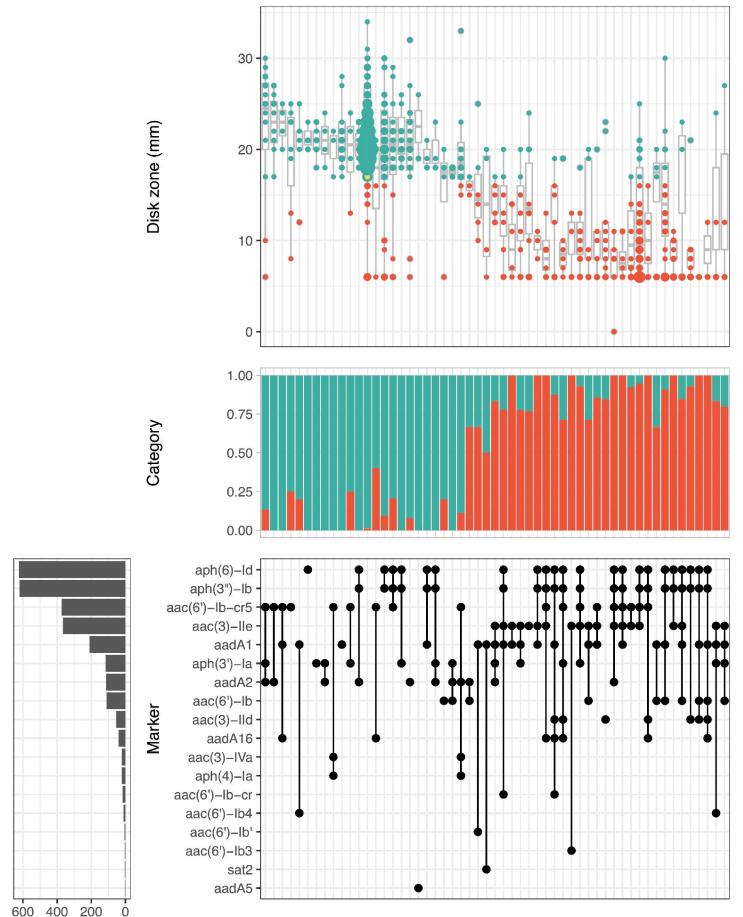
R ~ marker₁ + marker₂ + marker₃ +

(including all markers observed min [5] times)

If marker not found solo, but observed min [5] times:

NWT p-value	NWT CI	R p-value	R CI	expected	expected	phenotype
>0.05	spans 0	>0.05	spans 0	-	S	wildtype
>0.05	spans 0	>0.05	spans 0	I	I	wildtype

analysis\$upset_disk_plot



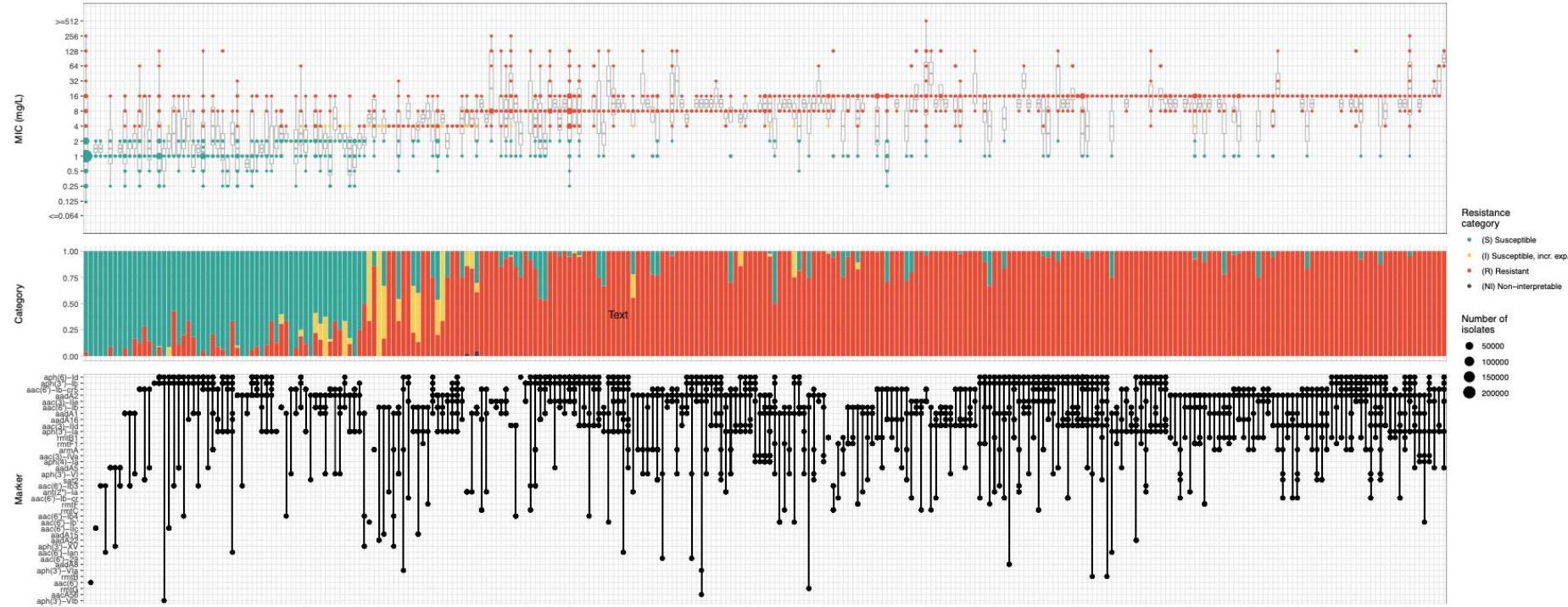
analysis\$upset_disk_summary

```
> print(analysis$upset_disk_summary, n=30)
```

	marker_list	marker_count	median	q25	q75	ppv	R	n
		<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	" "	0	20	19	22	0.0125	31	2476
2	"aph(3')-IIIa"	1	20	20	20	0	0	1
3	"aadA4"	1	21.5	20.8	22.2	0	0	2
4	"aac(2')-IIa"	1	18	18	18	0	0	1
5	"aadA8"	1	17	17	17	0	0	1
6	"aadA22"	1	16	16	16	1	1	1
7	"sat4, aph(3')-IIIa"	2	18	18	18	0	0	1
8	"ant(2')-Ia"	1	6	6	6	1	1	1
9	"aac(6')-I1"	1	20	20	20	0	0	1
10	"aac(6')-Ib3, rmtC"	2	6	6	6	1	1	1
11	"aadA5"	1	19	19	22.5	0	0	3
12	"aac(6')-Ib4"	1	18	18	18	0	0	1
13	"aac(6')-Ib"	1	18	17	18	0.2	1	5
14	"aac(6')-Ib, aadA16"	2	21	21	21	0	0	1
15	"aadA1"	1	20.5	19.2	21.8	0	0	34
16	"aadA1, aac(6')-Ib"	2	14.5	12.5	17.2	0.667	4	6
17	"aadA1, ant(2')-Ia"	2	13	13	13	1	1	1
18	"aadA1, aac(6')-I1"	2	17	17	17	0	0	1
19	"aadA1, sat2"	2	14	8.25	19.2	0.5	2	4
20	"aadA1, aac(3)-Ia"	2	13	13	13	1	1	1
21	"aadA1, aac(6')-Ib3"	2	27.5	27.2	27.8	0	0	2
22	"aadA1, aac(6')-Ib4"	2	21	20	22	0.2	1	5
23	"aadA1, aac(6')-Ib"	2	6	6	17	0.667	6	9
24	"aac(6')-Ib-cr5"	1	21.5	17.5	23.2	0.25	2	8
25	"aac(6')-Ib-cr5, aadA16"	2	20	16	20	0.4	2	5
26	"aac(6')-Ib-cr5, aadA1"	2	12	9	15	0.5	1	2
27	"aac(6')-Ib-cr5, aadA1, aadA16"	3	23	21.5	24.2	0	0	4
28	"aac(6')-Ib-cr5, aadA1, aadA16, sat2"	4	20	20	20	0	0	1
29	"aac(3)-IIe"	1	10	6	14	0.769	10	13
30	"aac(3)-IIe, ant(2')-Ia"	2	8	7	9	1	2	2

i 137 more rows

```
analysis$upset_mic_plot
```



Initial logic and thresholds

`makerules(minObs=3)`

Step 3: MIC data

MIC.median	call
> mic_R	R
> mic_S & <= mic_R	I
<= mic_S	S

Step 4: disk data

disk.median	call
< mic_R	R
< mic_S & >= mic_R	I
>= mic_S	S

(calculated if marker observed solo min [5] times)

MIC.median	phenotype
> mic_ecoff	NWT
<= mic_ecoff	WT

disk.median	phenotype
< disk_ecoff	NWT
>= mic_ecoff	WT

Initial logic

```
amrrules_analysis(mafLogReg=5)  
makerules(minObs=3)
```

Step 5: Compare available data on each individual marker or combination

Solo PPV call	MIC call	Disk call	LogReg	Call	Note
available	none	none	-	Solo PPV	Solo PPV.
available	agree	none	-	Solo PPV	Solo PPV. MIC.
available	disagree	none	-	Solo PPV	Solo PPV. MIC disagrees
available	agree	agree	-	Solo PPV	Solo PPV. MIC. Disk.
available	disagree	disagree	-	Solo PPV	Solo PPV. MIC disagrees. Disk disagrees.
none	available	none	-	MIC	MIC.
none	available	agree	-	MIC	MIC. Disk.
none	available	disagree	-	MIC	MIC. Disk disagrees.
none	none	available	-	Disk	Disk.
none	none	none	WT S/I	Logistic regression	Logistic regression.

Initial logic and thresholds

`makerules(minObs=3)`

Step 6: Assess marker combinations

Determine interpretation for combination by comparing MIC/disk vs breakpoints/ECOFF

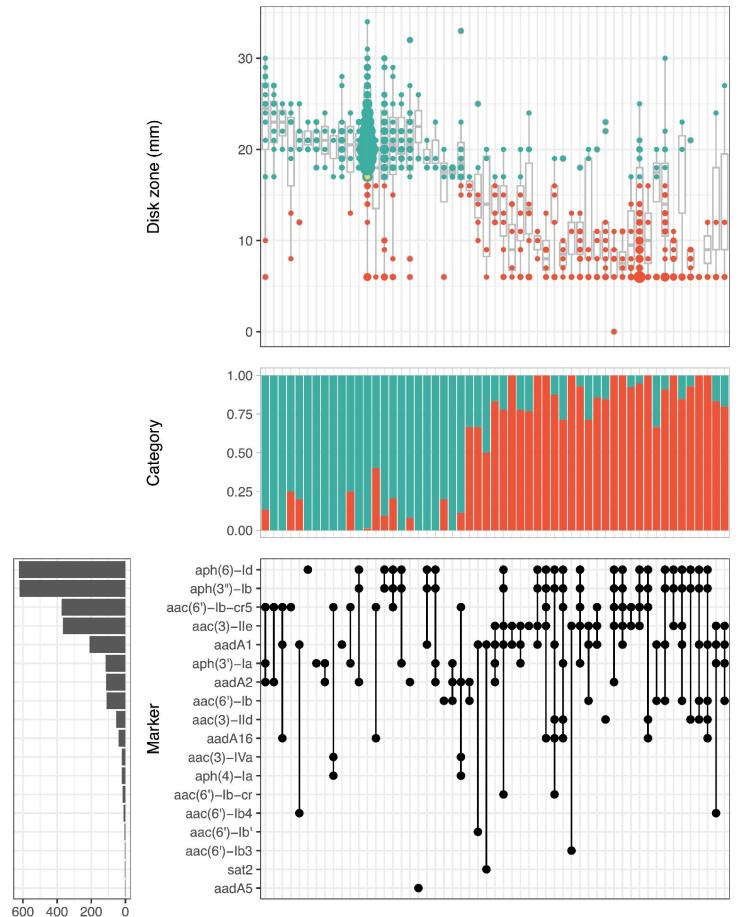
Find rules recorded for any subset (individual or combinations)

- Get highest call (R>I>S) for any subset
- Check if any subset interpreted as NWT

Compare calls for each marker combination vs components

- If combination call ‘higher’ than ALL subsets (R>I>S, NWT>WT)
 - Keep this combination rule
- Otherwise
 - Delete this combination rule

analysis\$upset_disk_plot



analysis\$upset_disk_summary

```
> print(analysis$upset_disk_summary, n=30)
```

A tibble: 167 × 8

	marker_list	marker_count	median	q25	q75	ppv	R	n
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	" "	0	20	19	22	0.0125	31	2476
2	"aph(3')-IIla"	1	20	20	20	0	0	1
3	"aadA4"	1	21.5	20.8	22.2	0	0	2
4	"aac(2')-IIa"	1	18	18	18	0	0	1
5	"aadA8"	1	17	17	17	0	0	1
6	"aadA22"	1	16	16	16	1	1	1
7	"sat4, aph(3')-IIIA"	2	18	18	18	0	0	1
8	"ant(2')-Ia"	1	6	6	6	1	1	1
9	"aac(6')-I1"	1	20	20	20	0	0	1
10	"aac(6')-Ib3, rmtC"	2	6	6	6	1	1	1
11	"aadA5"	1	19	19	22.5	0	0	3
12	"aac(6')-Ib4"	1	18	18	18	0	0	1
13	"aac(6')-Ib"	1	18	17	18	0.2	1	5
14	"aac(6')-Ib, aadA16"	2	21	21	21	0	0	1
15	"aadA1"	1	20.5	19.2	21.8	0	0	34
16	"aadA1, aac(6')-Ib"	2	14.5	12.5	17.2	0.667	4	6
17	"aadA1, ant(2')-Ia"	2	13	13	13	1	1	1
18	"aadA1, aac(6')-I1"	2	17	17	17	0	0	1
19	"aadA1, sat2"	2	14	8.25	19.2	0.5	2	4
20	"aadA1, aac(3)-Ia"	2	13	13	13	1	1	1
21	"aadA1, aac(6')-Ib3"	2	27.5	27.2	27.8	0	0	2
22	"aadA1, aac(6')-Ib4"	2	21	20	22	0.2	1	5
23	"aadA1, aac(6')-Ib"	2	6	6	17	0.667	6	9
24	"aac(6')-Ib-cr5"	1	21.5	17.5	23.2	0.25	2	8
25	"aac(6')-Ib-cr5, aadA16"	2	20	16	20	0.4	2	5
26	"aac(6')-Ib-cr5, aadA1"	2	12	9	15	0.5	1	2
27	"aac(6')-Ib-cr5, aadA1, aadA16"	3	23	21.5	24.2	0	0	4
28	"aac(6')-Ib-cr5, aadA1, aadA16, sat2"	4	20	20	20	0	0	1
29	"aac(3)-IIe"	1	10	6	14	0.769	10	13
30	"aac(3)-IIe, ant(2')-Ia"	2	8	7	9	1	2	2

i 137 more rows

Initial logic and thresholds

`makerules(weak_threshold=20)`

Step 7: Assign evidence

All calls assigned

- Evidence code: ECO:0001103 natural variation mutant evidence
- Evidence grade: moderate

Exceptions:

Criteria	Grade	Limitations
Call based on solo PPV, R.ppv.n < <code>weak_threshold</code>	weak	Limited samples
Call based on MIC, MIC.n < <code>weak_threshold</code>	weak	Limited samples
Call based on disk, Disk.n < <code>weak_threshold</code>	weak	Limited samples
Disagreement between PPV/MIC/Disk data	weak	Conflicting evidence
Marker not found solo	-	No solo geno-pheno data.
Call based on marker detected in <3 datasets	-	Limited datasets.

Tracking data sources

Currently this does not affect the evidence grade,
just recorded in 'evidence limitations'

- Could consider in evidence grade

Could do similar to track STs/phylogroups

> ast						
# A tibble: 455,723 × 6						
id	drug_agent	pheno	source	mic	disk	
<chr>	<ab>	<sir>	<chr>	<mic>	<dsk>	
1	SAMN02138648	TZP	R	NCBI	64	NA
2	SAMN02138670	AMP	R	NCBI	16	NA
3	SAMN02138669	AMP	R	NCBI	16	NA
4	SAMN02138668	AMP	S	NCBI	4	NA
5	SAMN02356582	AMK	S	NCBI	2	NA
6	SAMN02356581	AMK	S	NCBI	2	NA
7	SAMN02138649	FEP	I	NCBI	2	NA
8	SAMN02138647	AMK	S	NCBI	2	NA

↑
source to track unique datasets

Step 8: Assess number of source datasets contributing to each rule

marker	solo.sources	solo.sources.SIR	mic.sources	mic.sources.SIR	disk.sources	disk.sources.SIR
ompK36:Asp135AspAsp	1	0 S, 0 I, 1 R	6	1 S, 1 I, 5 R	1	0 S, 0 I, 1 R
ompK36:Gln313Ter	1	1 S, 0 I, 0 R	6	4 S, 1 I, 3 R	0	0 S, 0 I, 0 R
ompK36:Asp135AspAsp	1	0 S, 0 I, 1 R	6	1 S, 1 I, 5 R	1	0 S, 0 I, 1 R
ompK36:Asp135AspGlyAsp	3	2 S, 0 I, 2 R	15	11 S, 8 I, 14 R	1	0 S, 1 I, 0 R

AMRrulemakeR package

<https://github.com/AMRverse/AMRrulemakeR>

```
rules <- makerules(analysis, guide="EUCAST 2025", bp_site=NULL, weak_threshold=20, core_threshold=0.9)
```

A	C	D	I	J	L	N	O	P	Q	S	T	U	X
ruleID	gene	nodeID	mutation	variation type	drug	phenotype	clinical	breakpoint	breakpoint stan	evidence code	evidence gra	evidence limitations	marker
KPN5062	ompK36	ompK36	Asp135AspAsp	Protein variant detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	weak	No solo geno-pheno d	ompK36:Asp135AspAsp
KPN5046	ompK36	ompK36	Gln131Ter	Protein variant detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	ompK36:Gln131Ter
KPN5062	ompK36	ompK36	Asp135AspAsp	Protein variant detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	weak	No solo geno-pheno d	ompK36:Asp135AspAsp
KPN5045	ompK36	ompK36	Asp135AspGlyAsp	Protein variant detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Con	ompK36:Asp135AspGlyAsp
KPN5047	ompK36	ompK36	Thr136ThrAspThr	Protein variant detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Con	ompK36:Thr136ThrAspThr
KPN5044	ompK35	ompK35	Glu132Lys	Protein variant detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	moderate		ompK35:Glu132Lys
KPN5043	blaVIM-1	blaVIM-1	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaVIM-1
KPN5029	blaOXA-48	blaOXA-48	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	moderate		blaOXA-48
KPN5028	blaOXA-232	blaOXA-232	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Con	blaOXA-232
KPN5024	blaNDM-5	blaNDM-5	-	Gene presence detected	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaNDM-5
KPN5023	blaNDM-4	blaNDM-4	-	Gene presence detected	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	blaNDM-4
KPN5022	blaNDM-1	blaNDM-1	-	Gene presence detected	Imipener	nonwildtype	I	disk zone >= 19 & < 22 mm	EUCAST 2025	ECO:0001103	weak	Conflicting evidence.	blaNDM-1
KPN5020	blaKPC-3	blaKPC-3	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Conflicting evidence.	blaKPC-3
KPN5019	blaKPC-2	blaKPC-2	-	Gene presence detected	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	moderate		blaKPC-2
KPN5017	blaIMP-4	blaIMP-4	-	Gene presence detected	Imipener	wildtype	S	MIC <= 2 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	blaIMP-4
KPN5016	blaIMP-19	blaIMP-19	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Con	blaIMP-19
KPN5015	blaIMP-1	blaIMP-1	-	Gene presence detected	Imipener	nonwildtype	I	MIC > 2 & <= 4 mg/L	EUCAST 2025	ECO:0001103	weak	Conflicting evidence.	blaIMP-1
KPN5144	KPN5027 & KPN5047		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaOXA-181, ompK36:Thr136ThrAspThr
KPN5148	KPN5020 & KPN5062		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaKPC-3, ompK36:Asp135AspAsp
KPN5149	KPN5020 & KPN5047		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	moderate	Limited MIC datasets.	blaKPC-3, ompK36:Thr136ThrAspThr
KPN5151	KPN5020 & KPN5045		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	moderate		blaKPC-3, ompK36:Asp135AspGlyAsp
KPN5153	KPN5022 & KPN5055		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	moderate		blaNDM-1, ompK35:Trp230Ter
KPN5154	KPN5010 & KPN5022		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	blaNDM-1, blaCTX-M-55
KPN5155	KPN5009 & KPN5022		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples. Limi	blaNDM-1, blaCTX-M-3
KPN5156	KPN5022 & KPN5041		-	-	Imipener	nonwildtype	R	MIC > 4 mg/L	EUCAST 2025	ECO:0001103	weak	Limited samples.	blaNDM-1, blaSHV:-112C>A

rule curation note

Quantitative geno-pheno analysis by ESGEM-AMR WG. ompK36:Asp135AspAsp. Category call 'S' based on logistic regression. Phenotype call 'wildtype' based on logistic regression. OR for R: 2.281 [0.687-7.575], p=1.782e-01. OR for NWT: 2.286 [0.666-7.848], p=1.891e-01.

Quantitative geno-pheno analysis by ESGEM-AMR WG. ompK35:Glu132Lys. Category call 'S' based on Solo PPV. MIC. Disk. Phenotype call 'wildtype' based on MIC. Disk. R PPV=1% (1/166). NWT PPV=1% (2/166). 14 solo datasets: 13 S, 1 I, 1 R. MIC: median 0.25 [IQR 0.25-0.5]; n=118. 16 MIC datasets: 14 S, 4 I, 9 R. Disk: median 26 [IQR 25-28]; n=48. 3 disk datasets: 3 S, 0 I, 0 R.

Agenda today

Tools in development

- AMRgen package
- AMRrulemakeR package

Topics for discussion

- Rule-defining logic
- Data quality
- Thresholds for data quality / quantity and how these should map to
 - evidence grade
 - evidence limitations
- Quantitative fields to include in AMRrules specification

Data quality

Genome quality - assessed how?

- Assembly QC
- Contamination or mixed culture

Genotype calls

- Unit of analysis
- Dealing with uncertain/imprecise matches

AST quality

Which drugs?

Genome data quality - examples

	GenomeTrackr		KlebNET	Global Pneumo Seq
	<i>Salmonella</i>	<i>E. coli</i>	<i>K. pneumoniae</i>	<i>S. pneumoniae</i>
Read quality	≥30	≥30	-	-
Read depth	≥30x	≥40x	≥40x	≥20x
Genome size	4.3–5.2	4.5–5.9	4.9–6.1	1.9–2.3
Contig count	≤300	≤500	≤500	≤500
G+C content	-	-	56.35–57.98	-
Reads matching another genus	-	-	<5%	<2%

<https://onehealthoutlook.biomedcentral.com/articles/10.1186/s42522-020-00026-3>

<https://bigsdb.pasteur.fr/klebsiella/genome-quality-check/>

<https://www.biorxiv.org/content/10.1101/2024.11.27.625679v1>

Genome data quality - examples

Klebsiella pneumoniae	all	Checkm	Completeness	\geq	95.57	
Klebsiella pneumoniae	all	Checkm	Contamination	\leq	4.14	
Klebsiella pneumoniae	all	Checkm	GC	\leq	58.0	
Klebsiella pneumoniae	all	Checkm	GC	\geq	56.0	
Klebsiella pneumoniae	all	Checkm	Genome size (bp)	\leq	6400000	
Klebsiella pneumoniae	all	Checkm	Genome size (bp)	\geq	4900000	
Klebsiella pneumoniae	all	Checkm	Marker lineage	regex	<code>^Klebsiella pneumoniae</code>	species_field
Klebsiella pneumoniae	all	Quast	GC (%)	\leq	58.0	
Klebsiella pneumoniae	all	Quast	GC (%)	\geq	56.0	
Klebsiella pneumoniae	all	Quast	Ns per 100 kbp	\leq	1000	
Klebsiella pneumoniae	all	Quast	Total length (≥ 0 bp)	\leq	6400000	
Klebsiella pneumoniae	all	Quast	Total length (≥ 0 bp)	\geq	4900000	

<https://github.com/happykhan/speccheck/blob/main/criteria.csv>

Genotype calls

AMRfinderplus reports different kinds of hits, need to decide how to treat these in analyses

Unit of analysis?

- Group inexact matches separately?
- How to treat mutations in hits with divergent matches?

ARTICLES · Volume 5, Issue 11, 100913, November 2024 · Open Access

 Download Full Issue

Exploring uncatalogued genetic variation in antimicrobial resistance gene families in *Escherichia coli*: an observational analysis

Samuel Lipworth, DPhil   · Prof Derrick Crook, FRCP ^{a,b,c,d} · Prof A Sarah Walker, PhD ^{a,c,d} · Prof Tim Peto, FRCP ^{a,b,c,*} · Prof Nicole Stoesser, DPhil ^{a,b,c,d,*}

AST quality

S/I/R vs assay data

When to re-interpret, vs use S/I/R calls

When to use data for defining rules, vs testing/validating

EUCAST vs CLSI vs mix

MIC vs disk

- Default is to prioritise MIC where available
- Could modify this

Automated phenotyping platforms and calibration

- Ranges and censoring upper/lower values
- NOTE: MIC median/IQR calculations ignore values expressed as ranges

Which drugs? Which genotypes?

Which drugs?

- To start with, focus on those with EUCAST breakpoints
- If no breakpoints but many genotypes, consider using ECOFF or CLSI breakpoints

Which genotypes/markers?

- If core genes are included in AMRfinderplus, may need to suppress these in order to exclude them from determining which other markers are found ‘solo’ against this core background (e.g. oqxAB in *K. pneumoniae*)
- Note that for carbapenem drugs, need to test against markers assigned to either Cephalosporins or Carbapenems, otherwise will miss when combinations of enzymes/mutations contribute to resistance

Agenda today

Tools in development

- AMRgen package
- AMRrulemakeR package

Topics for discussion

- Rule-defining logic
- Data quality
- Thresholds for data quality / quantity and how these should map to
 - evidence grade
 - evidence limitations
- Quantitative fields to include in AMRrules specification

Thresholds

General principle: it is better to encode a rule based on limited evidence, and note the sample size and its limitations, than encode no rule at all

- Manual curation can add experimental evidence, boosting ‘moderate’ to ‘high’
- Current algorithm has these **default parameters**:
 - Marker included in logistic regression if observed in $n \geq 5$ genomes
 - PPV and MIC/disk range assessed for each individual marker/combination observed $n \geq 3$ times
 - Evidence grade ‘weak’ if this number is $n \leq 20$ genomes
 - Evidence grade ‘weak’ if any disagreement between PPV/MIC/disk signals
- Data diversity? Currently track # datasets, could track # STs/lineages

When can data-only rules be graded as high?

Currently:

- Statistical evidence only: moderate (if $N > 20$) or weak ($N \leq 20$)
- Experimental evidence required for a grade of 'high'

Sample size >X?

Independent datasets >Y?

Diversity >Z?

- Would need to have an additional input to assess this (e.g. ST or similar), to be defined for each organism

Agenda today

Tools in development

- AMRgen package
- AMRrulemakeR package

Topics for discussion

- Rule-defining logic
- Data quality
- Thresholds for data quality / quantity and how these should map to
 - evidence grade
 - evidence limitations
- Quantitative fields to include in AMRrules specification

Values to record in spec?

R.ppv	R.ppv.n	R.ppv.x	NWT.ppv	NWT.ppv.n	NWT.ppv.x	solo.sources	solo.sources	MIC.n	MIC.median	MIC.q25	MIC.q75	mic.sources	mic.sources.SIR	Disk.n	Disk.median	Disk.q25	Disk.q75	disk.sources	disk.sources.SIR
-	-	-	-	-	-	1 0S,0I,1R	-	-	-	-	-	6 1S,1I,5R	-	-	-	-	1 0S,0I,1R		
0	3	0	0	3	0	1 1S,0I,0R	3	1	1	1	6 4S,1I,3R	-	-	-	-	0 0S,0I,0R			
-	-	-	-	-	-	1 0S,0I,1R	-	-	-	-	6 1S,1I,5R	-	-	-	-	1 0S,0I,1R			
0.583	12	7	0.58333	12	7	3 2S,0I,2R	12	6	1	10	15 11S,8I,14R	-	-	-	-	1 0S,1I,0R			
0.538	13	7	0.53846	13	7	2 1S,0I,1R	13	16	0.25	16	12 7S,5I,10R	-	-	-	-	1 0S,1I,1R			
0.006	166	1	0.01205	166	2	14 13S,1I,1R	118	0.25	0.25	0.5	16 14S,4I,9R	48	26	25	28	3 3S,0I,0R			
0.5	16	8	0.8125	16	13	5 1S,4I,5R	15	4	4	8	5 1S,3I,5R	-	-	-	-	1 1S,1I,0R			
0.429	21	9	0.7619	21	16	6 4S,4I,3R	20	4	3.5	4	11 7S,6I,9R	-	-	-	-	1 1S,0I,1R			
0.556	9	5	0.77778	9	7	2 2S,1I,1R	9	8	2	16	7 3S,5I,7R	-	-	-	-	1 0S,1I,1R			
0.875	8	7	0.875	8	7	3 1S,0I,3R	8	16	8	16	11 2S,2I,11R	-	-	-	-	1 0S,0I,1R			
1	10	10	1	10	10	1 0S,0I,1R	10	16	16	16	4 0S,0I,4R	-	-	-	-	0 0S,0I,0R			
0.739	23	17	0.91304	23	21	8 2S,3I,6R	20	8	4	16	14 8S,6I,14R	3	19	18	21	2 1S,1I,2R			
0.776	299	232	0.8796	299	263	9 5S,6I,9R	261	8	4	16	8 4S,6I,8R	38	22	14.5	25	2 2S,2I,2R			
0.885	235	208	0.91489	235	215	11 5S,4I,10R	235	8	8	8	12 8S,4I,12R	-	-	-	-	0 0S,0I,0R			
0.111	9	1	0.11111	9	1	2 2S,0I,1R	9	1	1	1	LogRegR.est	LogRegR.ci.lc	LogRegR.ci.u	LogRegR.pva	LogRegNWT.i	LogRegNWT.j	LogRegNWT.k	LogRegNWT.l	freq
0.2	5	1	1	5	5	1 0S,1I,1R	5	2	2	2	0.82448173	-0.3758406	2.02480407	0.17821639	0.82659292	-0.4070918	2.06027763	0.18911076	0.00408359
0.227	44	10	0.70455	44	31	1 1S,1I,1R	44	2	2	2	0.01559981	-2.4653807	2.49658028	0.99016729	1.15188714	-0.573058	2.87683232	0.19059247	0.00156137
											0.82448173	-0.3758406	2.02480407	0.17821639	0.82659292	-0.4070918	2.06027763	0.18911076	0.00408359
											1.40360729	1.16241279	1.64480179	0	1.34272001	1.07946556	1.60597445	0	0.13415806
											2.15453488	1.76588803	2.54318173	0	2.17352523	1.73155861	2.61549185	0	0.10929618
											-0.3135716	-0.9472164	0.32007324	0.33208306	-0.227452	-0.7872999	0.33239583	0.42586708	0.03026663
											3.16419774	2.47320846	3.85518701	0	3.8724004	3.10356285	4.64123794	0	0.00480423
											2.53091544	2.21183328	2.8499976	0	3.04908477	2.71995247	3.37821707	0	0.03459044
											2.38269321	2.027273825	2.69264816	0	3.25956339	2.90926239	3.60986438	0	0.07578669
											4.50420213	3.95737711	5.05102715	0	4.69614622	4.02313636	5.36915607	0	0.05020418
											7.94671627	5.19434761	10.6990849	1.523789527	7.73940734	4.97701413	10.5018005	3.991251906	0.04479942
											4.54398046	4.23868536	4.84927556	0	4.940689	4.56539937	5.31597862	0	0.10148931
											4.56571635	4.31428163	4.81715106	0	4.91160114	4.6178273	5.20537499	0	0.08887821
											5.15956798	4.89096189	5.42817407	0	5.16233482	4.87356732	5.45110232	0	0.16298343
											1.77380985	0.39831361	3.14930609	0.01148691	1.45326128	0.07941364	2.82710892	0.03814833	0.00156137
											2.046989	0.63021113	3.46376686	0.00462875	5.06832942	3.33379685	6.80286198	1.021960838	0.00132116
											1.99956785	1.37412491	2.6250108	3.702295137	3.61236495	3.04315556	4.18157433	0	0.00696613

TB catalog

<https://www.who.int/publications/i/item/9789240082410>

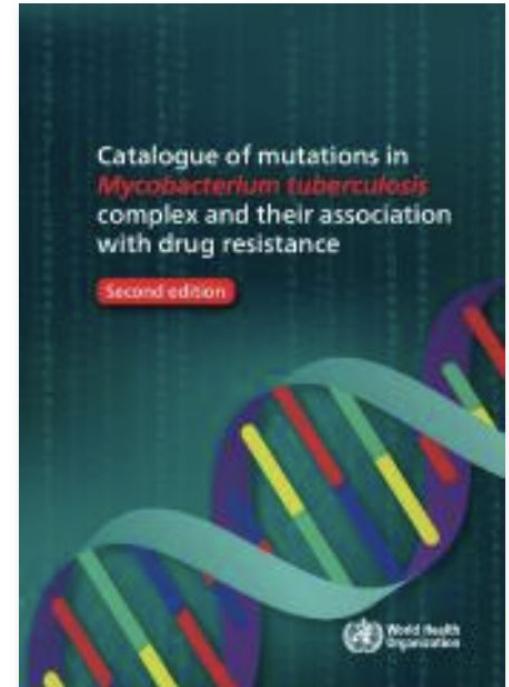
<https://github.com/GTB-tbsequencing/mutation-catalogue-2023/tree/main/Final%20Result%20Files>

drug	gene	mutation	variant	tier	effect	genomic position	algorithm_pass	Present_SOLO_SR	Present_SOLO_R	Present_SOLO_S	Present_R	Present_S	Absent_R	Absent_S	Sens	Sens_ib	Sens_ub	Spec	Spec_ib	Spec_ub	PPV	PPV_lb	PPV_ub	PPV_SOLO	PPV_SOLO_lb	PPV_SOLO_ub		
Isoniazid	inhA	p.Val203Ala	inhA_p.Val203Ala	1	missense_variant	(see "Genomic")	1	2	0	2	0	2	20846	27466	0.00%	0.00%	0.02%	99.99%	99.97%	100.00%	0.00%	0.00%	84.19%	0.00%	0.00%	84.19%		
Isoniazid	inhA	p.Val28Leu	inhA_p.Val28Leu	1	missense_variant	(see "Genomic")	0	0	0	0	1	0	20845	27468	0.00%	0.00%	0.03%	100.00%	99.99%	100.00%	100.00%	100.00%	2.50%	100.00%	NA	NA	NA	
Isoniazid	inhA	p.Val65Ala	inhA_p.Val65Ala	1	missense_variant	(see "Genomic")	0	0	0	0	1	0	20845	27468	0.00%	0.00%	0.03%	100.00%	99.99%	100.00%	100.00%	100.00%	2.50%	100.00%	NA	NA	NA	
Isoniazid	inhA	p.Val78Ala	inhA_p.Val78Ala	1	missense_variant	(see "Genomic")	0	0	0	0	8	62	20838	27406	0.04%	0.02%	0.08%	99.77%	99.71%	99.83%	11.43%	5.07%	21.28%	NA	NA	NA		
Isoniazid	inhA	p.Val91Ala	inhA_p.Val91Ala	1	missense_variant	(see "Genomic")	1	1	0	1	0	1	20846	27467	0.00%	0.00%	0.02%	100.00%	99.98%	100.00%	0.00%	0.00%	97.50%	0.00%	0.00%	97.50%		
Isoniazid	inhA	p.Val92Met	inhA_p.Val92Met	1	missense_variant	(see "Genomic")	0	0	0	0	1	0	20845	27468	0.00%	0.00%	0.03%	100.00%	99.99%	100.00%	100.00%	100.00%	2.50%	100.00%	NA	NA	NA	
Isoniazid	katG	c.1008G>T	katG_c.1008G>T	1	synonymous_variant	(see "Genomic")	NA	NA	NA	NA	NA	1	0	20845	27468	0.00%	0.00%	0.03%	100.00%	99.99%	100.00%	100.00%	100.00%	2.50%	100.00%	NA	NA	NA
Isoniazid	katG	c.-100G>A	katG_c.-100G>A	1	upstream_gene_variant	2156211	1	2	0	2	0	2	20846	27466	0.00%	0.00%	0.02%	99.99%	99.97%	100.00%	0.00%	0.00%	84.19%	0.00%	0.00%	84.19%		
Isoniazid	katG	c.1014C>G	katG_c.1014C>G	1	synonymous_variant	(see "Genomic")	1	NA	NA	NA	0	1	20846	27467	0.00%	0.00%	0.02%	100.00%	99.98%	100.00%	0.00%	0.00%	97.50%	NA	NA	NA		
Isoniazid	katG	c.1014C>T	katG_c.1014C>T	1	synonymous_variant	(see "Genomic")	1	NA	NA	NA	1	4	20845	27464	0.00%	0.00%	0.03%	99.99%	99.96%	100.00%	20.00%	0.51%	71.64%	NA	NA	NA		
Isoniazid	katG	c.1029G>A	katG_c.1029G>A	1	synonymous_variant	(see "Genomic")	1	NA	NA	NA	0	1	20846	27467	0.00%	0.00%	0.02%	100.00%	99.98%	100.00%	0.00%	0.00%	97.50%	NA	NA	NA		
Isoniazid	katG	c.-102A>G	katG_c.-102A>G	1	upstream_gene_variant	2156213	1	0	0	0	0	1	20846	27467	0.00%	0.00%	0.02%	100.00%	99.98%	100.00%	0.00%	0.00%	97.50%	NA	NA	NA		
Isoniazid	katG	c.1032G>A	katG_c.1032G>A	1	synonymous_variant	(see "Genomic")	1	NA	NA	NA	2	8	20844	27460	0.01%	0.00%	0.03%	99.97%	99.94%	99.99%	20.00%	2.52%	55.61%	NA	NA	NA		
Isoniazid	katG	c.1035G>A	katG_c.1035G>A	1	synonymous_variant	(see "Genomic")	1	NA	NA	NA	0	2	20846	27466	0.00%	0.00%	0.02%	99.99%	99.97%	100.00%	0.00%	0.00%	84.19%	NA	NA	NA		
Isoniazid	katG	c.1047C>A	katG_c.1047C>A	1	synonymous_variant	(see "Genomic")	1	NA	NA	NA	1	3	20845	27465	0.00%	0.00%	0.03%	99.99%	99.97%	100.00%	25.00%	0.63%	80.59%	NA	NA	NA		

TB catalog

Present_SOLO_SR
Present_SOLO_R
Present_SOLO_S
Present_R
Present_S
Absent_R
Absent_S
Sens
Sens_lb
Sens_ub
Spec
Spec_lb
Spec_ub
PPV
PPV_lb
PPV_ub
PPV_SOLO
PPV_SOLO_lb
PPV_SOLO_ub
PPV_conditional_SOLO
PPV_conditional_SOLO_lb
PPV_conditional_SOLO_ub

OR_SOLO
OR_SOLO_exact_lb
OR_SOLO_exact_ub
OR_SOLO_pvalue
OR_SOLO_pval_rank
k
OR_SOLO_FE_sig
Neutral_masked
OR
OR_exact_lb
OR_exact_ub
Sens_SOLO
Sens_SOLO_lb
Sens_SOLO_ub
Spec_SOLO
Spec_SOLO_lb
Spec_SOLO_ub
Initial confidence grading ALL dataset
INITIAL CONFIDENCE GRADING
DATASET(S)
Initial confidence grading WHO dataset
algorithm_pass





Timeline for quantitative rule generation

July-Aug

- Quant group: develop protocol and thresholds
- Within organism subgroups:
 - discuss & test AMRrulemakeR package
 - feedback issues and refinements needed for criteria / logic / specification

Sep-Dec: Generate and curate initial quantitative rules within organism subgroups

Jan-Feb: Aggregate rules into AMRrules v2.0 (Python package) release

Mar-Apr: Publish and present at Wellcome AMR Big Data & ESCMID Global