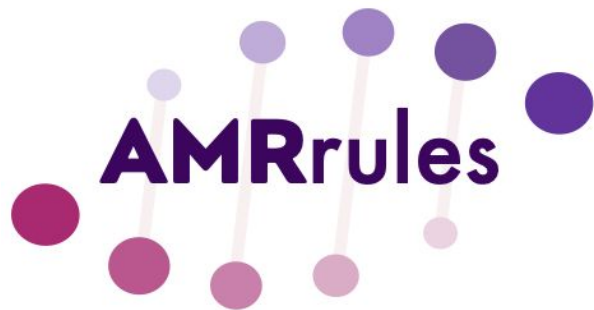# Agenda

1. Resource Updates

    a. Validation code

    b. AMRgen R package

    c. AMRrules R package

2. Preliminary AMRrules submitted by subgroups

3. Future Planning

    a. First release AMRrules

    b. Paper/s

    c. New subgroups

4. Funding Opportunities

    a. ESCMID SG grants?

    b. Other funding?

Rule validation

python validate_rules.py Pseudomonas_aeruginosa_v0.1.txt

Validating rules file: ~/drafts/Pseudomonas/Pseudomonas_aeruginosa_v0.1.txt

Checking that all required columns for spec v0.5 are present…

Continuting to validate values in each column…

Checking ruleID column...
All rule IDs have passed auto validation
✅ All values are valid
Rule prefix: PSA

Checking organism column...
✅ All organism names passed auto validation

Unique organism names: s__Pseudomonas aeruginosa

Checking gene column...
✅ All gene values are valid

Now checking for combinatorial rules in gene column...
✅ All gene combinatorial rule IDs are valid

Checking nodeID, refseq accession, GenBank accession and HMM accession columns...
✅ All rows contain at least one value in one of these columns.
❌ One or more accessions aren't present in either the NCBI Reference Gene Catalog (for nodeID, refseq accession and genbank accession) or the NCBI Reference HMM Catalog (for HMM accession). Empty cells must be specified by '-'.

Summary of checks:
✅ Passed: 17
 - ruleID
 - organism
 - gene
 - ARO accession
 - mutation
 - variation type
 - variation type mutation concordance
 - context
 - drug and drug class
 - phenotype
 - clinical category
 - breakpoint
 - clinical category and breakpoint concordance
 - breakpoint standard
 - PMID
 - evidence code
 - evidence grade and limitations
❌ Failed: 1
 - gene accessions

AMRgen R package

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

Import AMRfinderplus genotyping results to standard format

```
> geno
# A tibble: 3,544,510 × 6
   Name          marker      drug_agent drug_class                Class        Subclass
   <chr>         <gene>      <ab>       <chr>                     <chr>        <chr>
 1 SAMN19594082  glpT_E448K  FOS        Fosfomycin                FOSFOMYCIN   FOSFOMYCIN
 2 SAMN19594082  acrF        NA         Efflux                    EFFLUX       EFFLUX
 3 SAMN19594082  pmrB_Y358N  COL        Polymyxins                COLISTIN     COLISTIN
 4 SAMN19594082  blaEC       NA         Beta-lactams/penicillins  BETA-LACTAM  BETA-LACTAM
 5 SAMN19594082  mdtM        NA         Efflux                    EFFLUX       EFFLUX
 6 SAMN19594082  tet(B)      NA         Tetracyclines             TETRACYCLINE TETRACYCLINE
 7 SAMEA8798279  glpT_E448K  FOS        Fosfomycin                FOSFOMYCIN   FOSFOMYCIN
 8 SAMEA8798279  pmrB_Y358N  COL        Polymyxins                COLISTIN     COLISTIN
```

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> <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
```
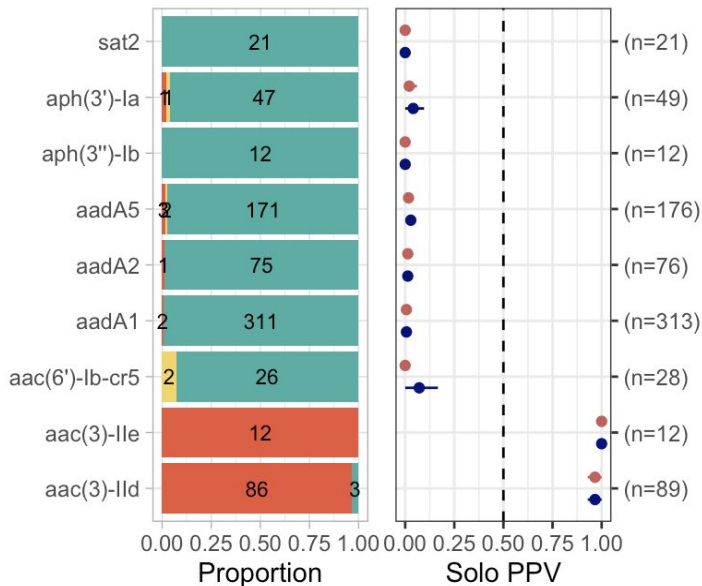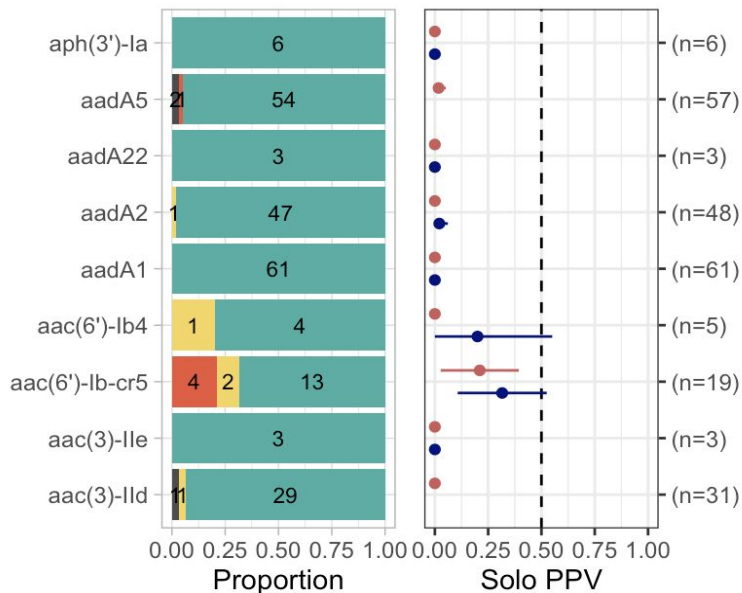
AMRgen R package

Calculate positive predictive value (PPV) of markers found 'solo'

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

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

Phenotype
- (S) Susceptible
- (I) Susceptible, incr. exp.
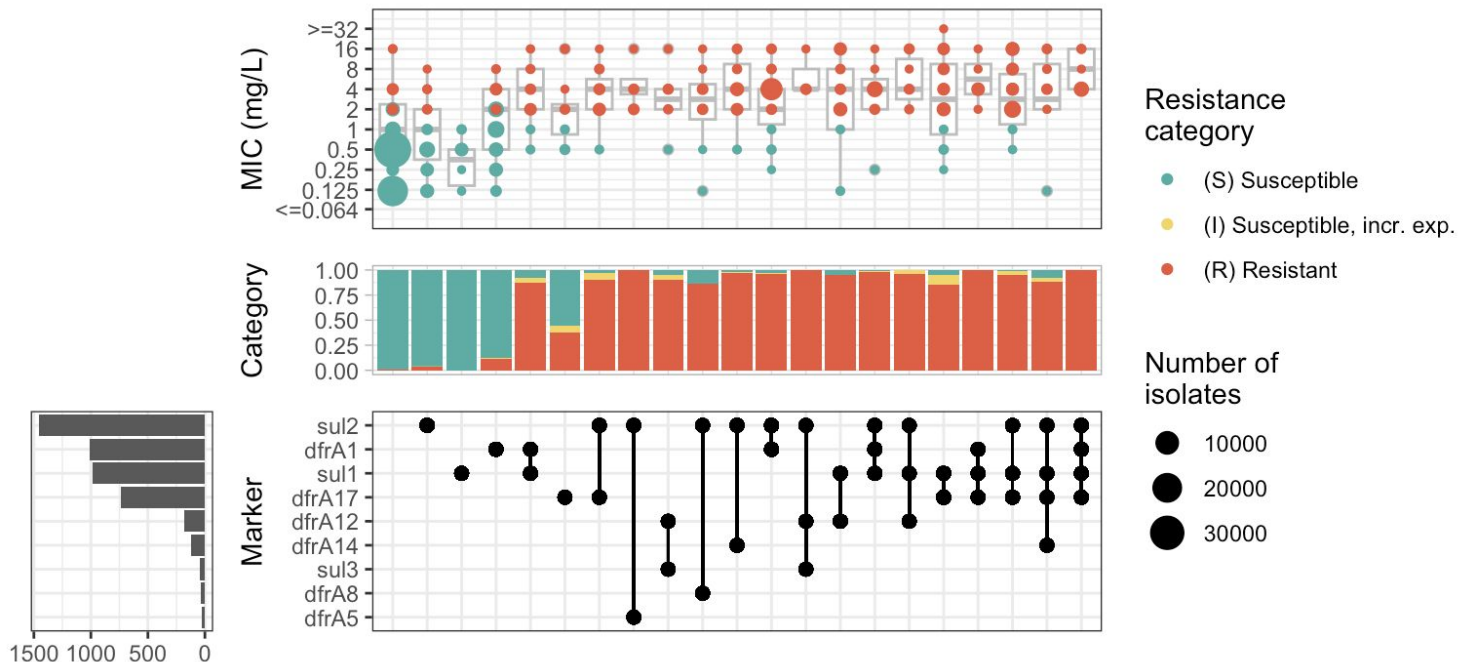- (R) Resistant
- (NI) Not interpretable

Category
- NWT
- R

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

AMRgen R package

Upset plot, to explore MIC distribution associated with marker combinations


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

AMRgen R package

Upset plot, to explore disk zone distribution associated with marker combinations
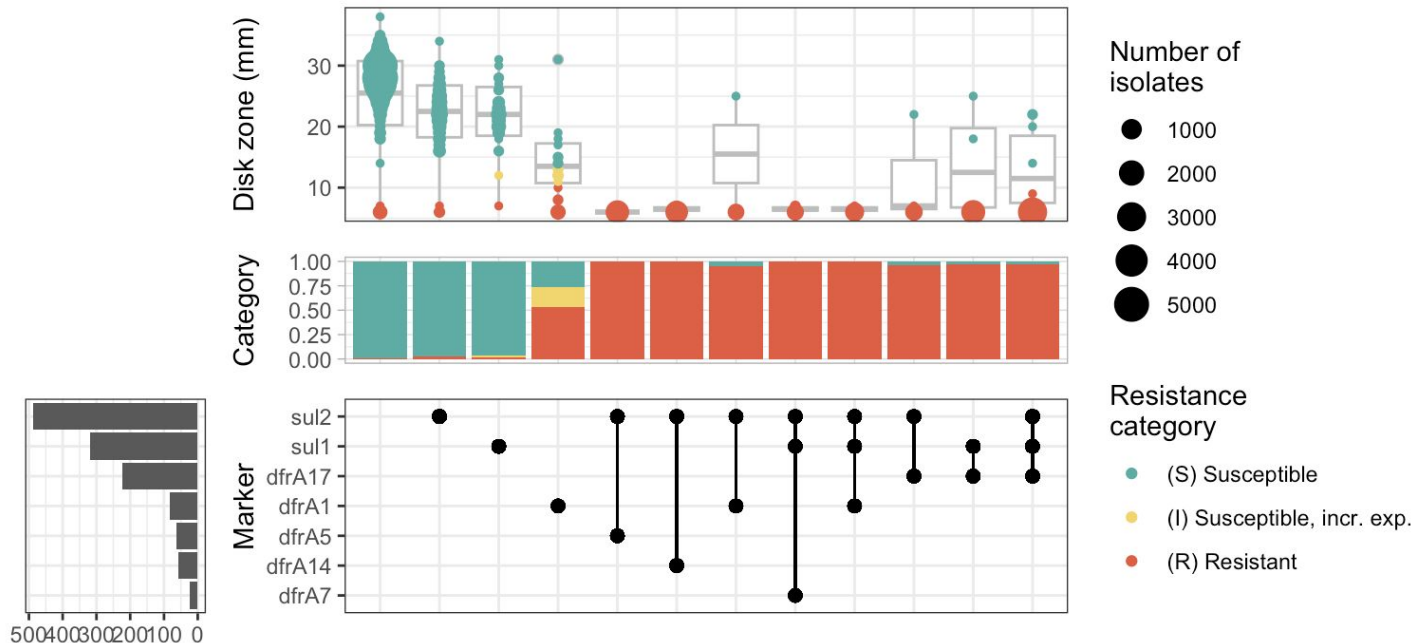


```
amr_upset(get_binary_matrix(geno, ast, antibiotic="Trimethoprim-sulfamethoxazole",
          drug_class_list=c("Trimethoprims", "Sulfonamides"), min_set_size=20, assay="disk")
```

AMRrules R package   *(work in progress)*

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

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

AMRrules R package   *(work in progress)*

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

- Download and plot EUCAST reference distributions for MIC and disk diffusion (using `get_eucast_mic_distribution`)
- Summarise the available data using `summarise_data`
- Summarise the frequency of each marker in the matched geno/pheno dataset
- Run a solo PPV analysis using `solo_ppv_analysis` (for all markers meeting a minimum count, default n=1 i.e. all)
- Run a logistic regression using `amr_logistic` (for all markers meeting a minimum count, default n=5)
- Generate upset plots and summary tables reporting the categorical counts and median/IQR values for individual markers and combinations, for MIC and disk data, using `amr_upset`

```
rules <- makerules(analysis, guide="EUCAST 2024", bp_site=NULL, weak_threshold=20, core_threshold=0.9)
```

- Draft rules, based on
  a. Solo PPV for single markers
  b. MIC/disk vs breakpoint and ECOFF for single markers
  c. MIC/disk vs breakpoint and ECOFF for combinations (where phenotype exceeds the single markers)

AMRrules R package   *(work in progress)*

```
rules <- makerules(analysis, guide="EUCAST 2024", bp_site=NULL, weak_threshold=20, core_threshold=0.9)
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ruleID	organism	gene	nodeID	mutation	variation typ	context	drug	phenotype	clinical cate	breakpoint	breakpoint st	evidence co	evidence gra	evidence limitations	rule curation
2	ECO1001	s_Escherichi	qnrS13	qnrS13	-	Gene presen	core	Ciprofloxacir	nonwildtype	I	MIC > 0.25 &	EUCAST 2024	ECO:000110	weak	limited geno-pheno data wit	Quantitative
3	ECO1002	s_Escherichi	gyrA	gyrA	p.Asp87Gly	Protein varia	accessory	Ciprofloxacir	nonwildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	limited geno-pheno data wit	Quantitative
4	ECO1003	s_Escherichi	qnrS2	qnrS2	-	Gene presen	core	Ciprofloxacir	nonwildtype	I	MIC > 0.25 &	EUCAST 2024	ECO:000110	weak	limited geno-pheno data wit	Quantitative
5	ECO1004	s_Escherichi	qnrB4	qnrB4	-	Gene presen	core	Ciprofloxacir	nonwildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	limited geno-pheno data wit	Quantitative
6	ECO1005	s_Escherichi	parC	parC	p.Ala56Thr	Protein varia	accessory	Ciprofloxacir	nonwildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	limited geno-pheno data wit	Quantitative
7	ECO1006	s_Escherichi	parE	parE	p.Asp475Glu	Protein varia	accessory	Ciprofloxacir	nonwildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	moderate		Quantitative
8	ECO1007	s_Escherichi	qnrS1	qnrS1	-	Gene presen	core	Ciprofloxacir	nonwildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	moderate		Quantitative
9	ECO1008	s_Escherichi	qnrB19	qnrB19	-	Gene presen	core	Ciprofloxacir	nonwildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	moderate		Quantitative
10	ECO1024	s_Escherichi	parE	parE	p.Ser458Thr	Protein varia	accessory	Ciprofloxacir	wildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	no geno-pheno data with this	Quantitative
11	ECO1025	s_Escherichi	parE	parE	p.Leu416Phe	Protein varia	accessory	Ciprofloxacir	wildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	no geno-pheno data with this	Quantitative
12	ECO1026	s_Escherichi	parC	parC	p.Glu84Gly	Protein varia	accessory	Ciprofloxacir	wildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	no geno-pheno data with this	Quantitative
13	ECO1027	s_Escherichi	gyrA	gyrA	p.Ser83Val	Protein varia	accessory	Ciprofloxacir	-	-	-	EUCAST 2024	-	-	-	Quantitative
14	ECO1028	s_Escherichi	qepA4	qepA4	-	Gene presen	core	Ciprofloxacir	wildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	no geno-pheno data with this	Quantitative
15	ECO1029	s_Escherichi	qepA9	qepA9	-	Gene presen	core	Ciprofloxacir	wildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	no geno-pheno data with this	Quantitative
16	ECO1030	s_Escherichi	parC	parC	p.Glu84Lys	Protein varia	accessory	Ciprofloxacir	wildtype	S	MIC <= 0.25 r	EUCAST 2024	ECO:000110	weak	no geno-pheno data with this	Quantitative
17	ECO1062	s_Escherichi	ECO1033 & E	-	-	-	-	Ciprofloxacir	nonwildtype	R	MIC > 0.5 mg	EUCAST 2024	ECO:000110	weak	limited geno-pheno MIC data	Quantitative
18	ECO1063	s_Escherichi	ECO1004 & E	-	-	-	-	Ciprofloxacir	nonwildtype	I	MIC > 0.25 &	EUCAST 2024	ECO:000110	weak	limited geno-pheno MIC data	Quantitative
19	ECO1064	s_Escherichi	ECO1007 & E	-	-	-	-	Ciprofloxacir	nonwildtype	I	MIC > 0.25 &	EUCAST 2024	ECO:000110	weak	limited geno-pheno MIC data	Quantitative
20	ECO1065	s_Escherichi	ECO1007 & E	-	-	-	-	Ciprofloxacir	nonwildtype	R	MIC > 0.5 mg	EUCAST 2024	ECO:000110	weak	limited geno-pheno MIC data	Quantitative
21	ECO1066	s_Escherichi	ECO1007 & E	-	-	-	-	Ciprofloxacir	nonwildtype	R	MIC > 0.5 mg	EUCAST 2024	ECO:000110	weak	limited geno-pheno MIC data	Quantitative
22	ECO1084	s_Escherichi	ECO1007 & E	-	-	-	-	Ciprofloxacir	REVIEW	REVIEW	-	EUCAST 2024	ECO:000110	weak	limited geno-pheno disk data	Quantitative

gyrA_D87Y, qnrS1. Disk: median 16 [IQR 16-16]; n=1. R PPV=100% (1/1). 1 disk datasets (0 S, 0 I, 1 R).
REVIEW: MIC data conflicts with disk data. MIC: median 0.5 [IQR 0.5-0.5]; n=5. R PPV=20% (1/5). 2 MIC datasets (0 S, 2 I, 1 R).

Agenda

1. Resource Updates

 a. Validation code

 b. AMRgen R package

 c. AMRrules R package

2. Preliminary AMRrules submitted by subgroups

3. Future Planning

 a. First release AMRrules

 b. Paper/s

 c. New subgroups

4. Funding Opportunities

 a. ESCMID SG grants?

 b. Other funding?

AMRrulesCuration / **draftrules** /

katholt draft rules for 4 spp   4544db1 · now   History

Name	Last commit message	Last commit date
..		
.DS_Store	draft rules for 4 spp	now
Acinetobacter_baumannii.tsv	Initial draft rules for Acinetobacter baumannii	4 months ago
Bordetella_v0.1.txt	draft rules for 4 spp	now
Bordetella_v0.1_check.txt	draft rules for 4 spp	now
Klebsiella_pneumoniae_v0.1.txt	draft rules for 4 spp	now
Klebsiella_pneumoniae_v0.1_check.txt	draft rules for 4 spp	now
Pseudomonas_aeruginosa_v0.1.txt	draft rules for 4 spp	now
Pseudomonas_aeruginosa_v0.1_check.txt	draft rules for 4 spp	now
Yersinia_v0.1.txt	draft rules for 4 spp	now
Yersinia_v0.1_check.txt	draft rules for 4 spp	now

<u>Core gene</u> rules submitted and validated

Pseudomonas aeruginosa
- 97 rules, for 14 genes and 28 drugs
- *Refining PMID/evidence codes*

Yersinia enterocolitica and **pseudotuberculosis**
- *Y. enterocolitica:* 39 rules, for 24 genes and 19 drugs
- *Y. pseudotuberculosis:* 13 rules, 12 genes and 7 drugs
- *Minor fixes needed to nodeIDs etc*

Klebsiella pneumoniae
- 7 rules, for 7 genes and 3 drugs

Bordetella
- 4 core gene rules, for core *bla* genes and 1 expected resistance (4 species)
- +11 acquired gene rules, for 6 acquired resistances (3 species)
- *Minor refinements needed, clarify appropriate accession for 23S rRNA*

Pseudomonas aeruginosa

gene	nodeID	refseq accession	GenBank accession	HMM accession	ARO accession
aph(3')-IIb	aph(3')-IIb	WP_003113011.1	AGY72832.1	-	ARO:3002645
blaOXA-50_fam	blaOXA-50_fam	-	-	NF000432.2	ARO:3001796
blaPDC_gen	blaPDC_gen	-	-	-	ARO:3000098
catB7	catB7	WP_003112709.1	AAD02068.1	-	ARO:3002679
crpP	crpP	-	AEQ93536.1	-	ARO:3004467
fosA_PA1129	fosA_PA1129	-	-	NF000094.1	ARO:3000149
mexA	mexA	-	AAA74436.1	NF033834.1	ARO:3000377
mexB	-	-	AAA74437.1	-	ARO:3000378
mexE	mexE	-	AAG05881.1	NF045904.1	ARO:3000803
mexF	-	-	AAG05882.1	-	ARO:3000804
mexX	mexX	-	BAA34299.1	-	ARO:3003034
mexY	-	-	BAA34300.1	-	ARO:3003033
oprM	-	WP_003084633.1	AAG03816.1	-	ARO:3000379
oprN	-	-	AAG05883.1	-	ARO:3000805

Pseudomonas aeruginosa

drug	phenotype	category
amikacin	wildtype	S
amoxicillin	wildtype	R
amoxicillin-clavulanic acid	wildtype	R
ampicillin	wildtype	R
ampicillin-sulbactam	wildtype	R
aztreonam	wildtype	I
cefepime	wildtype	I
cefotaxime	wildtype	R
ceftazidime	wildtype	I
ceftazidime-avibactam	wildtype	S
ceftolozane-tazobactam	wildtype	S
ceftriaxone	wildtype	R
chloramphenicol	wildtype	R
ciprofloxacin	wildtype	I

drug	phenotype	category
ertapenem	wildtype	R
fosfomycin	wildtype	S
kanamycin	wildtype	R
levofloxacin	wildtype	I
meropenem	wildtype	S
neomycin	wildtype	R
norfloxacin	wildtype	R
piperacillin	wildtype	I
piperacillin-tazobactam	wildtype	I
sulfamethoxazole	wildtype	R
tetracycline	wildtype	R
tigecycline	wildtype	R
tobramycin	wildtype	S
trimethoprim	wildtype	R

Yersinia enterocolitica

gene	nodeID	refseq accession	ARO accession
aac(3)-IVa	aac(3)-IVa	-	-
ampC	-	-	-
aph(3')-IIa	aph(3')-IIa	-	-
aph(4)-Ia	aph(4)-Ia	-	-
blaA	blaA_Yent	-	-
blaOXA-48	blaOXA-48	-	-
blaTEM-181	blaTEM-181	-	-
dfrA1	dfrA1	-	-
dfrA12	dfrA12	-	ARO: 39292
dfrA14	dfrA14	-	-
dfrB1	dfrB1	-	-
floR	floR	-	-
gyrA	-	WP_013650157.1	-
qnrB19	qnrB19	-	-
sat2	sat2_fam	-	ARO:3002895
sul1	sul1	-	-
sul2	sul2	-	-
sul3	sul3	-	ARO: 3000413
tet(A)	tet(A)	WP_000804064.1	-
tet(B)	tet(B)	WP_012881330.1	-
tet(C)	tet(C)	WP_001297013.1	-
tet(H)	tet(H)	WP_006248867.1	-
tet(M)	tet(M)	-	ARO:3000186
vat(F)	vat(F)	WP_005164767.1	-

Yersinia pseudotuberculosis

gene	nodeID	refseq accession	GenBank accession	drug	phenotype	clinical category
ant(3")-Ia	ant(3")-Ia	-	-	Streptomycin	nonwildtype	R
aph(3")-Ib	aph(3")-Ib	-	-	Streptomycin	nonwildtype	R
aph(6)-Id	aph(6)-Id	-	-	Streptomycin	nonwildtype	R
blaOXA-2	blaOXA-2	-	-	Amoxicillin	nonwildtype	R
blaTEM-1	blaTEM-1	AAR25033.1	-	Amoxicillin	nonwildtype	R
blaTEM-1	blaTEM-1	AAR25033.1	-	Ticarcillin	nonwildtype	R
catA2	catA2	-	-	Chloramphenicol	nonwildtype	R
dfrA1	dfrA1	-	-	Trimethoprim	nonwildtype	R
dfrA14	dfrA14	-	-	Trimethoprim	nonwildtype	R
sul1	sul1	-	-	Sulfamethoxazole	nonwildtype	R
sul2	sul2	-	-	Sulfamethoxazole	nonwildtype	R
tetA(A)	-	-	CAD57192.1	Tetracycline	nonwildtype	R
tetA(D)	-	-	AAL75563.1	Tetracycline	nonwildtype	R

Klebsiella pneumoniae

gene	nodeID	refseq accession	ARO accession	drug	drug class	phenotype	category
blaSHV	blaSHV	WP_419672855.1	ARO:3000015	-	penams	wildtype	R
oqxA	oqxA	WP_419141451.1	ARO:3003922	-	fluoroquinolone	wildtype	S
oqxB	oqxB	WP_419388593.1	ARO:3003923	-	fluoroquinolone	wildtype	S
fosA5_fam	fosA5_fam	WP_417645943.1	-	fosfomycin	-	wildtype	S
fosA5	fosA5	WP_012579083.1	ARO:3003209	fosfomycin	-	wildtype	S
fosA6	fosA6	WP_069174570.1	ARO:3004111	fosfomycin	-	wildtype	S
fosA10	fosA10	WP_004214174.1	-	fosfomycin	-	wildtype	S

Bordetella - core genes & expected resistance

organism	gene	nodeID	refseq accession	drug class	phenotype	clinical category
s__Bordetella bronchiseptica	blaBOR	blaBOR	WP_010926363.1	penams	wildtype	R
s__Bordetella parapertussis	blaBOR	blaBOR	WP_010926363.1	penams	wildtype	R
s__Bordetella holmesii	blaHBL	blaHBL	WP_080700357.1	penams	wildtype	R
s__Bordetella hinzii	blaHBL	blaHBL	WP_080700357.1	penams	wildtype	R

Bordetella - acquired genes & resistances

organism	gene	nodeID	refseq accession	ARO accession	mutation	drug	drug class	phenotype	category
s__Bordetella pertussis	**23s rDNA**	**23s rDNA**	**NC_002929.2**	ARO:3004125	c.[2047A>G]3	-	macrolides	nonwildtype	R
s__Bordetella bronchiseptica	blaTEM	blaTEM	WP_080699425.1	-	-	Ceftiofur	-	wildtype	R
s__Bordetella bronchiseptica	floR2	floR2	WP_000214125.1	-	-	-	phenicols	nonwildtype	R
s__Bordetella bronchiseptica	sul1	sul1	WP_000259031.1	-	-	-	sulfonamide	nonwildtype	R
s__Bordetella bronchiseptica	sul2	sul2	WP_001043260.1	-	-	-	sulfonamide	nonwildtype	R
s__Bordetella bronchiseptica	aph(3")-Ib	aph(3")-Ib	WP_001082319.1	-	-	streptomycin	-	nonwildtype	R
s__Bordetella bronchiseptica	aph(6)-Id	aph(6)-Id	WP_000480968.1	-	-	streptomycin	-	nonwildtype	R
s__Bordetella bronchiseptica	tet(G)	tet(G)	-	-	-	-		nonwildtype	R
s__Bordetella bronchiseptica	aph(3")-Ib	aph(3")-Ib	WP_001082319.1	-	-	neomycin	-	nonwildtype	R
s__Bordetella bronchiseptica	aph(6)-Id	aph(6)-Id	WP_000480968.1	-	-	neomycin	-	nonwildtype	R
s__Bordetella bronchiseptica	blaBOR	blaBOR	WP_010926363.1	-	-	-	penams	wildtype	R
s__Bordetella parapertussis	blaBOR	blaBOR	WP_010926363.1	-	-	-	penams	wildtype	R
s__Bordetella holmesii	blaHBL	blaHBL	WP_080700357.1	-	-	-	penams	wildtype	R
s__Bordetella hinzii	blaHBL	blaHBL	WP_080700357.1	-	-	-	penams	wildtype	R

Agenda

1. Resource Updates

   a.   Validation code
   b.   AMRgen R package
   c.   AMRrules R package

2. Preliminary AMRrules submitted by subgroups

3. Future Planning

   a.   First release AMRrules
   b.   Paper/s
   c.   New subgroups

4. Funding Opportunities

   a.   ESCMID SG grants?
   b.   Other funding?

Future Planning

First release of AMRrules (focus on core genes & expected R)

- Aiming for ESKAPEE + *Salmonella, Yersinia, Bordetella* in time for ABPHM conference (21 May, 2025)

First release of AMRrules interpretation engine (for AMRfinderplus output) - June?

- Expansion beyond AMRfinderplus inputs

Quantitative approach (using AMRgen to analyse geno/pheno)

Papers

- Concept, specification, and rules for core genes/expected R
- AMRgen package
- Quantitative approach for acquired genes (& AMRrules package)
- Analysis of mobile gene rules across organisms/groups
- Individual subgroup papers?
- Ideas? From the group - learnings from work so far:
    - Gaps in knowledge about intrinsic resistance mechanisms
    - Gaps in upstream AMR tools & databases

Funding opportunities

ESCMID Research grants

- Study Group Grant: EUR30k, open 21 May–16 July
- Study Group Collaboration Grant: EUR180k, open 20 Aug–29 Oct

https://www.escmid.org/science-research/grants-awards/research-grants/

Other funding?

Questions? / Any other business?

ESGEM-AMR

ESCMID

AMRrules