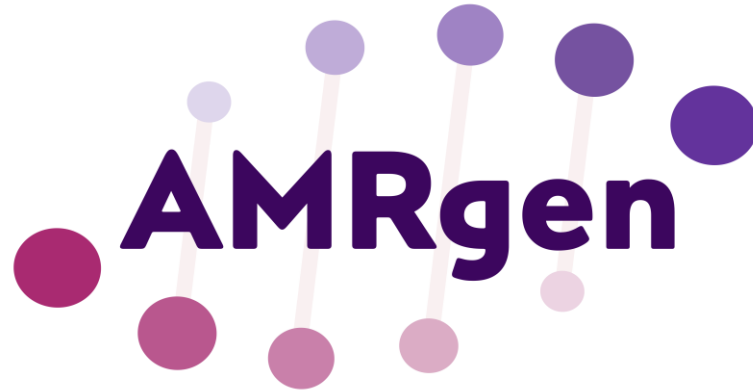
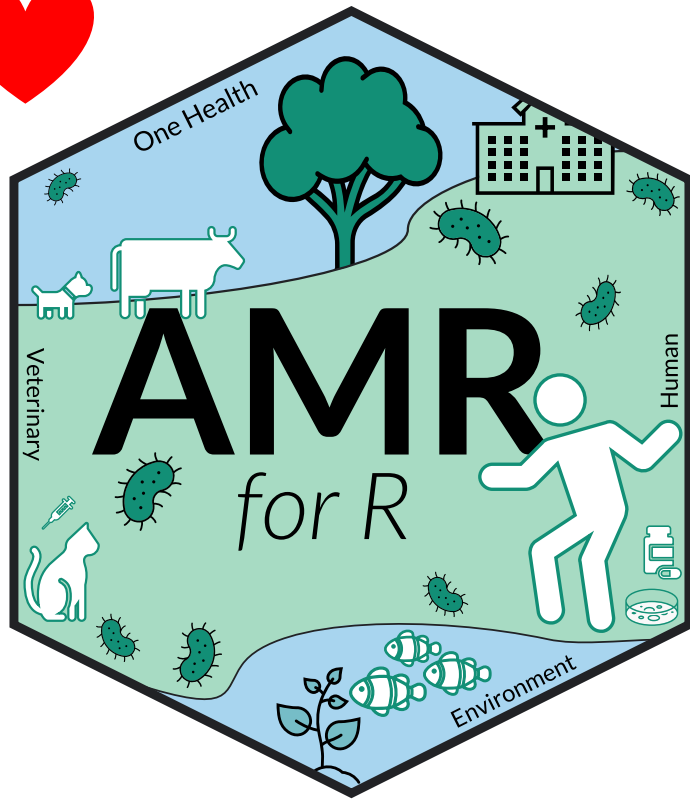
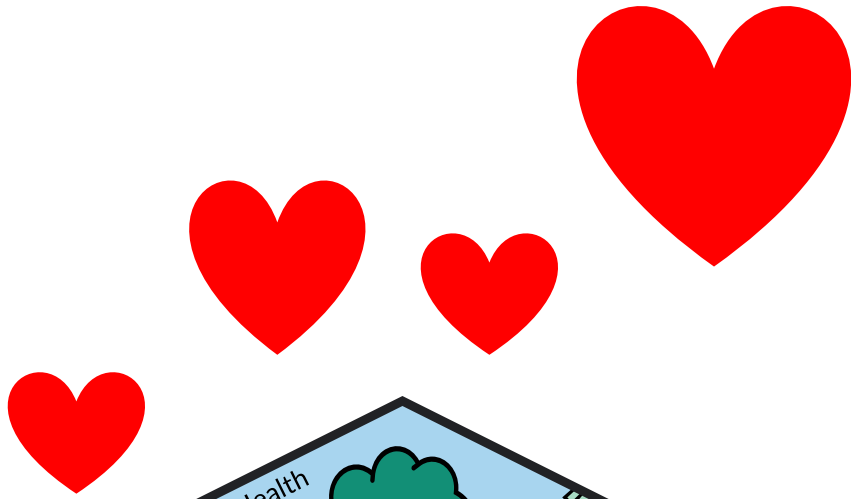




# An R package for antimicrobial resistance genotype-phenotype analysis

**LSHTM:** Kat Holt, Gwen Knight, Ebenezer Foster-Nyarko, Zoe Dyson, Kara Tsang

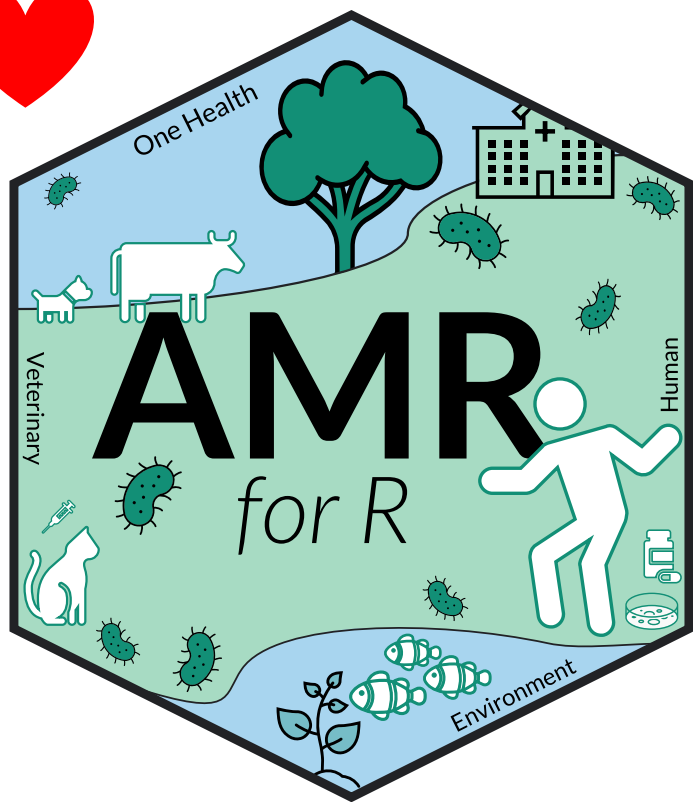
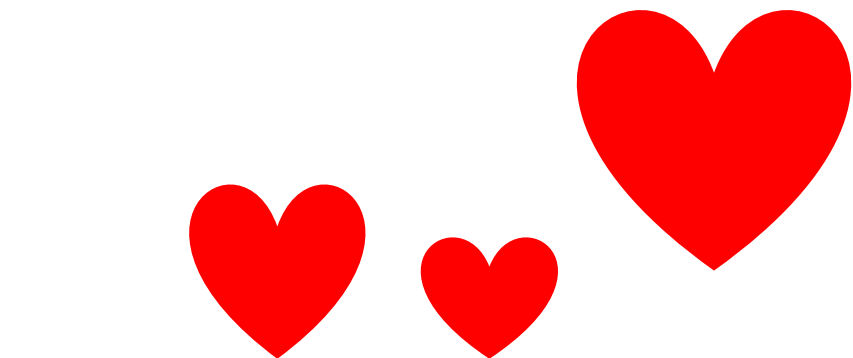
**Collaborators:** Silvia Argimón, Dominique Chaput, Natacha Couto, Jane Hawkey, Richard Goodman, Dorottya Nagy, Arjun Prasad, Leonor Sánchez-Busó, Matthijs Berends



## R package for antimicrobial phenotype data

Smart software that *understands* what AMR data is  
...and what users want to do with it

- Organisms (species, genus)
- Drugs and drug classes
- Types of assay data (MIC, disk zones)
- Breakpoints to interpret assay measurements into **S / I / R**
- Common analyses and plots

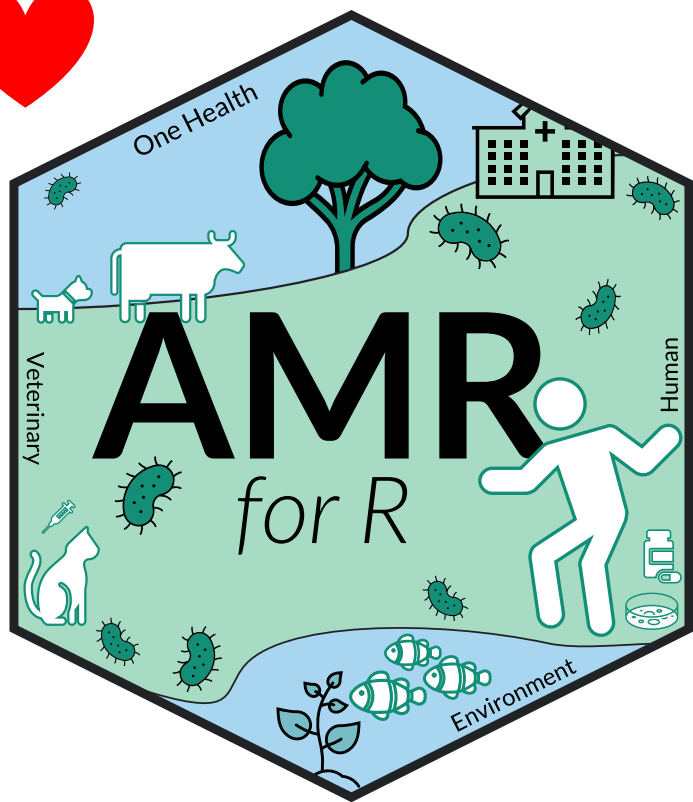
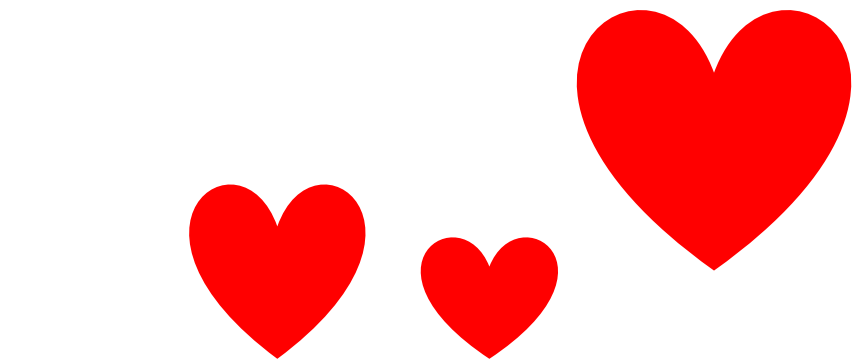


+



=

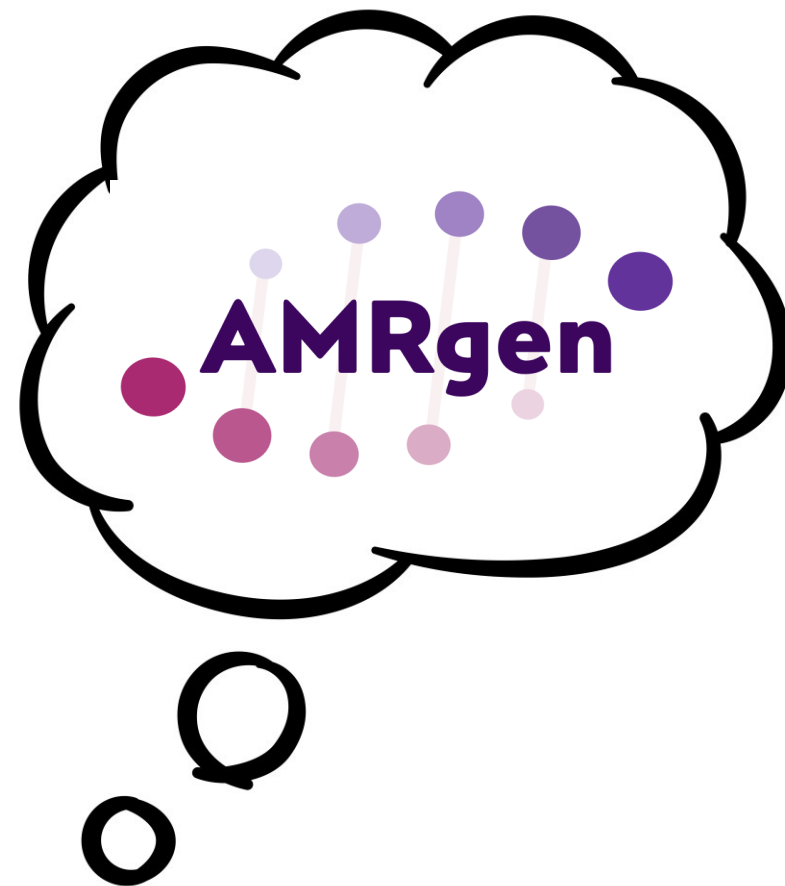


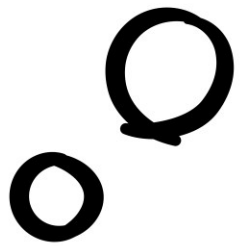


+



=





ANTIMICROBIAL  
RESISTANCE  
CENTRE



**ESCMID** | ESGEM-AMR

## Hackathon, Jan 2025





[github.com/AMRverse/AMRgen](https://github.com/AMRverse/AMRgen)



[AMRgen.org](https://AMRgen.org)

## Vignettes

[Analysing Geno-Pheno Data](#)

[Downloading geno-pheno data from NCBI and EBI](#)

[Assessing geno-pheno concordance](#)

[Example using large-scale regional/national surveillance data](#)

[Example with custom stratification by isolate source](#)

[Example with custom classification of genotype hits](#)

[Analysing the impact of deletion variants on susceptibility](#)

## Developers

Kathryn E. Holt

Author 

Silvia Argimón

Author 

Dominique L. Chaput

Author 

Natacha Couto

Author 

Zoe A. Dyson

Author 

Jane Hawkey

Author 

Ebenezer Foster-Nyarko

Author 

Richard N. Goodman

Author 

Gwen M. Knight

Author 

Dorottya Nagy

Author 

Arjun B. Prasad

Author 

Leonor Sánchez-Busó

Author 

Kara K. Tsang

Author 

Matthijs S. Berends

Author, maintainer 



THE PREPRINT SERVER FOR BIOLOGY

## **AMRgen: an R package for antimicrobial resistance genotype-phenotype analysis**

 KE Holt,  S Argimón,  DL Chaput,  N Couto,  ZA Dyson,  
 E Foster-Nyarko,  RN Goodman,  J Hawkey,  GM Knight,  D Nagy,  
 AB Prasad,  L Sánchez-Busó,  KK Tsang,  MS Berends

**doi:** <https://doi.org/10.64898/2026.05.01.722195>

# Supporting AMR geno/pheno data sharing

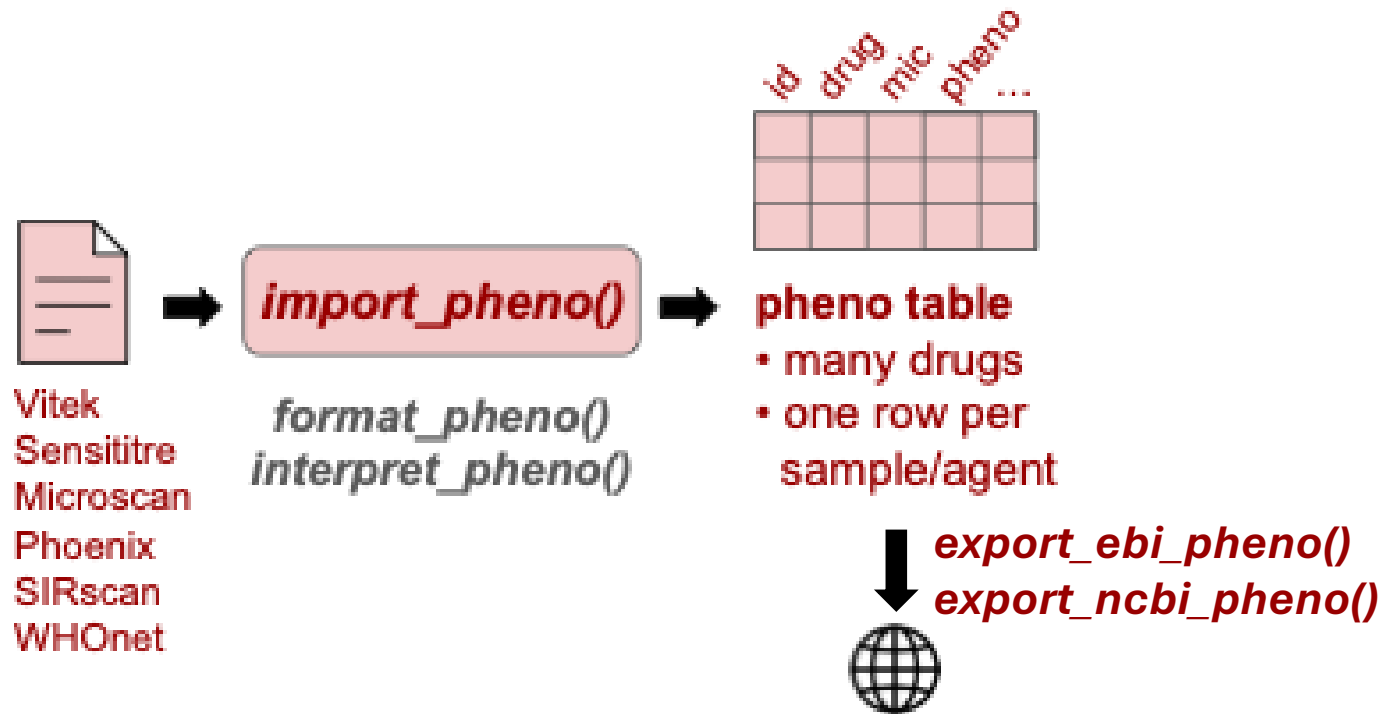
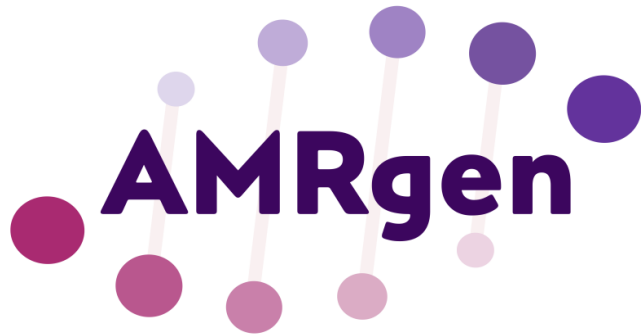


EMBL-EBI



- Standard practice to deposit bacterial genome sequence data
- DBs also accept AMR phenotype data, linked to deposited genomes
- ...but <3% of public genomes also have AMR phenotypes!

# Functions to convert data to submission-ready formats



# Web app for non-R users to convert data for submission



Upload antibiogram file

**Browse...** No file selected

Input format

Sensititre

InstructionsImported dataSummary**Exports**

**NCBI**EBI tableEBI JSON

Export a TSV file suitable for submission to NCBI. See [NCBI BioSample Antibigrams](#) for details.

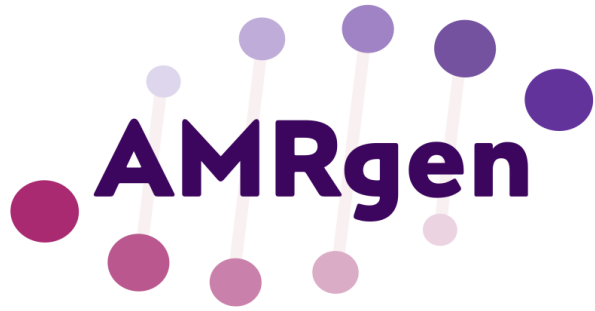
Import data to choose a phenotype column for export.

**Download NCBI TSV**

Guideline

EUCAST

Vendor (optional)



## Open from start to finish

- R ❤️ Build on existing packages, contribute to the open ecosystem
- Open development, through hackathon and subsequent calls
- Inclusive development
  - Volunteers to test and develop examples, not just code
  - Input from academia, clinical labs, public health labs, WHO, NCBI
- Accessible outputs
  - Example data, step-by-step tutorials, web apps for key functions



# R package for AMR geno/pheno analysis

bioRxiv  
THE PREPRINT SERVER FOR BIOLOGY



## AMRgen: an R package for antimicrobial resistance genotype-phenotype analysis

 KE Holt,  S Argimón,  DL Chaput,  N Couto,  ZA Dyson,  
 E Foster-Nyarko,  RN Goodman,  J Hawkey,  GM Knight,  D Nagy,  
 AB Prasad,  L Sánchez-Busó,  KK Tsang,  MS Berends

doi: <https://doi.org/10.64898/2026.05.01.722195>

### Vignettes

[Analysing Geno-Pheno Data](#)

[Downloading geno-pheno data from NCBI and EBI](#)

[Assessing geno-pheno concordance](#)

[Example using large-scale regional/national surveillance data](#)

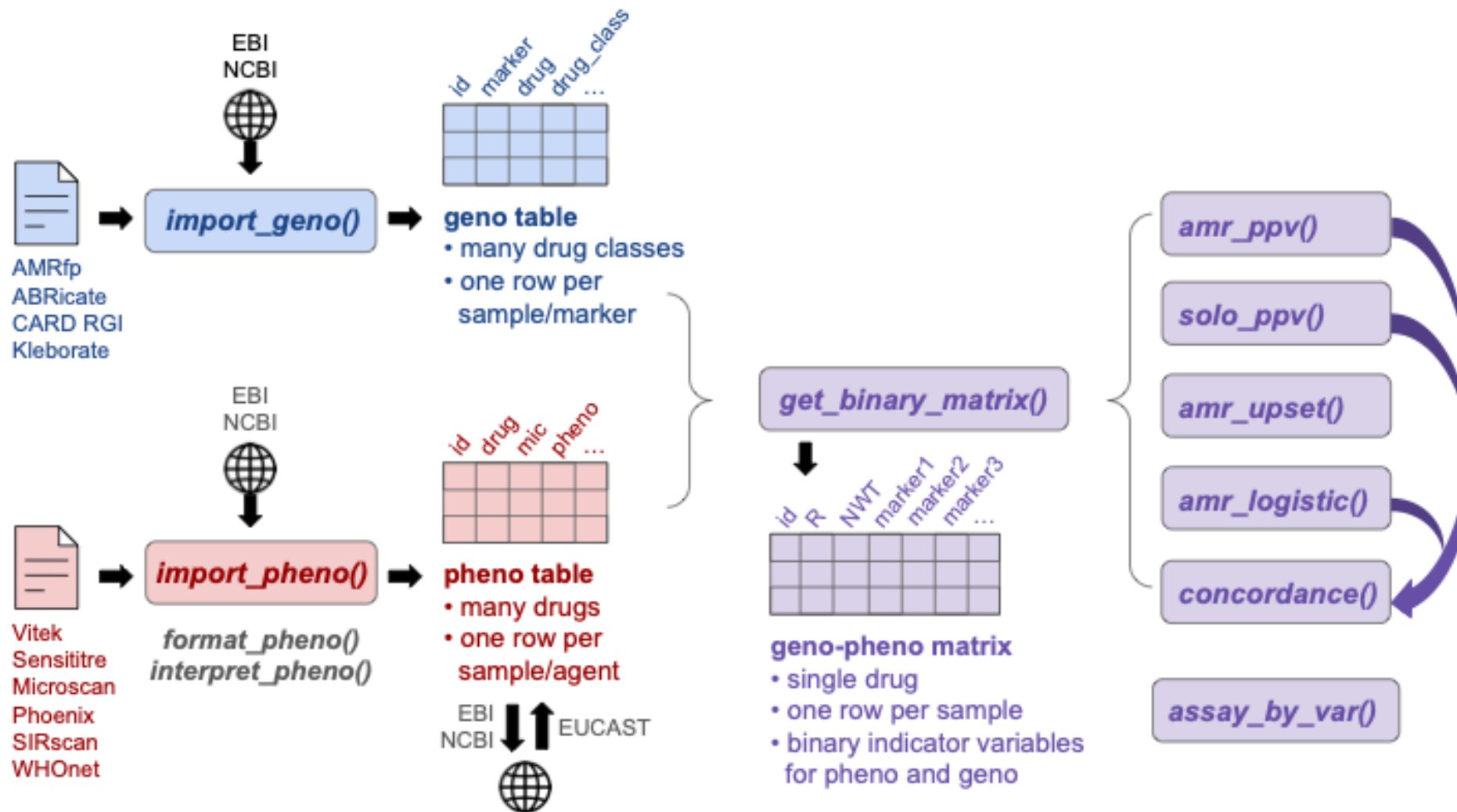
[Example with custom stratification by isolate source](#)

[Example with custom classification of genotype hits](#)

[Analysing the impact of deletion variants on susceptibility](#)



# R package for AMR geno/pheno analysis



## Developers

Matthijs S. Berends

Author, maintainer

Kathryn E. Holt

Author

Jane Hawkey

Author

Natacha Couto

Author

Ebenezer Foster-Nyarko

Author

Gwen M. Knight

Author

Zoe A. Dyson

Author

Kara K. Tsang

Author

Arjun B. Prasad

Author

Leonor Sánchez-Busó

Author

Silvia Argimón

Author

Dorottya Nagy

Author

Dominique L. Chaput

Author

Richard N. Goodman

Author



# Import genotypes

```
> import_geno(ecoli_geno_raw, format="amrfp")
```

```
# A tibble: 50,720 × 37
```

|   | id      | marker   | gene    | mutation   | drug | drug_class    | `variation type`   | node    | marker.label   |
|---|---------|----------|---------|------------|------|---------------|--------------------|---------|----------------|
|   | <chr>   | <chr>    | <chr>   | <chr>      | <ab> | <chr>         | <chr>              | <chr>   | <chr>          |
| 1 | SAMN... | blaEC    | blaEC   | NA         | NA   | Beta-lacta... | Gene presence d... | blaEC   | blaEC          |
| 2 | SAMN... | acrF     | acrF    | NA         | NA   | Efflux        | Gene presence d... | acrF    | acrF           |
| 3 | SAMN... | glpT_... | glpT    | Glu448L... | FOS  | Phosphonics   | Protein variant... | glpT    | glpT:Glu448... |
| 4 | SAMN... | floR     | floR    | NA         | CHL  | Phenicol      | Gene presence d... | floR    | floR           |
| 5 | SAMN... | floR     | floR    | NA         | FLR  | Phenicol      | Gene presence d... | floR    | floR           |
| 6 | SAMN... | mdtM     | mdtM    | NA         | NA   | Efflux        | Gene presence d... | mdtM    | mdtM           |
| 7 | SAMN... | blaTE... | blaT... | NA         | NA   | Beta-lacta... | Gene presence d... | blaT... | blaTEM-1       |

*Genotype importers for AMRFinderPlus, CARD RGI, ABRicate, Kleborate*



# Import phenotypes

```
> import_pheno(ecoli_pheno_raw, format="ncbi")
```

```
# A tibble: 10 x 29
```

|   | id           | drug | mic    | disk  | guideline | method         | platform | pheno_provided | spp_pheno    |
|---|--------------|------|--------|-------|-----------|----------------|----------|----------------|--------------|
|   | <chr>        | <ab> | <mic>  | <dsk> | <chr>     | <chr>          | <chr>    | <sir>          | <mo>         |
| 1 | SAMN11638310 | CIP  | 256.00 | NA    | CLSI      | broth dilution | NA       | R              | B_ESCHR_COLI |
| 2 | SAMN05729964 | CIP  | 64.00  | NA    | CLSI      | Etest          | Etest    | R              | B_ESCHR_COLI |
| 3 | SAMN10620111 | CIP  | >=4.00 | NA    | CLSI      | broth dilution | NA       | R              | B_ESCHR_COLI |
| 4 | SAMN10620168 | CIP  | >=4.00 | NA    | CLSI      | broth dilution | NA       | R              | B_ESCHR_COLI |
| 5 | SAMN10620104 | CIP  | <=0.25 | NA    | CLSI      | broth dilution | NA       | S              | B_ESCHR_COLI |
| 6 | SAMN10620102 | CIP  | >=4.00 | NA    | CLSI      | broth dilution | NA       | R              | B_ESCHR_COLI |
| 7 | SAMN10620129 | CIP  | >=4.00 | NA    | CLSI      | broth dilution | NA       | R              | B_ESCHR_COLI |

*Phenotype importers for Vitek, BD Phoenix, Sensititre, Microscan, WHOnet, NCBI, EBI*

*Re-call MIC or disk values against latest EUCAST or CLSI breakpoints (via AMR R package)*



# Retrieve geno/pheno data from EBI or NCBI

```
> download_ebi(data = "phenotype", species = "Escherichia coli", pheno_drug = "Cipro")
```

```
# A tibble: 20,279 × 37
```

|    | id       | drug | mic   | disk  | pheno_provided | guideline | method    | platform | source   | spp_pheno    |
|----|----------|------|-------|-------|----------------|-----------|-----------|----------|----------|--------------|
|    | <chr>    | <ab> | <mic> | <dsk> | <sir>          | <chr>     | <chr>     | <chr>    | <chr>    | <mo>         |
| 1  | SAMEA... | CIP  | NA    | NA    | R              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 2  | SAMEA... | CIP  | NA    | NA    | R              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 3  | SAMEA... | CIP  | NA    | NA    | S              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 4  | SAMEA... | CIP  | NA    | NA    | R              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 5  | SAMEA... | CIP  | NA    | NA    | R              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 6  | SAMEA... | CIP  | NA    | NA    | S              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 7  | SAMEA... | CIP  | NA    | NA    | R              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 8  | SAMEA... | CIP  | NA    | NA    | S              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 9  | SAMEA... | CIP  | NA    | NA    | S              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |
| 10 | SAMEA... | CIP  | NA    | NA    | S              | EUCAST    | broth ... | NA       | 30550... | B_ESCHR_COLI |

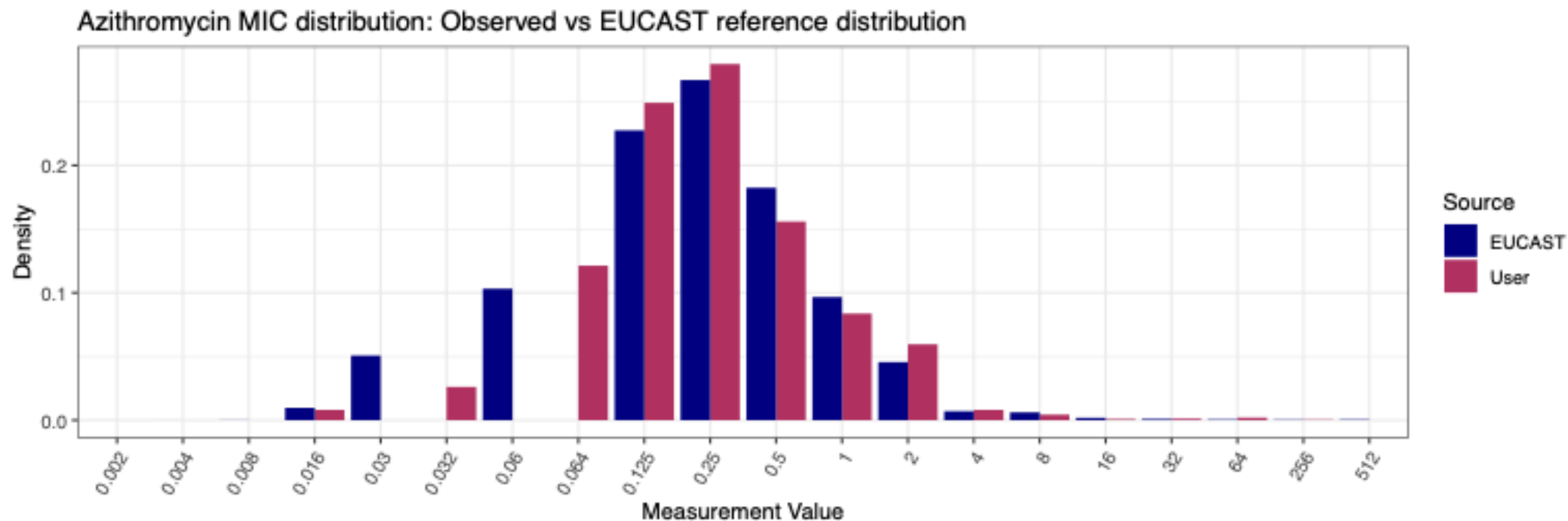
*Download functions for EBI (ftp), NCBI (cloud or web)*

*Phenotypes and genotypes*



# Retrieve reference distributions from EUCAST

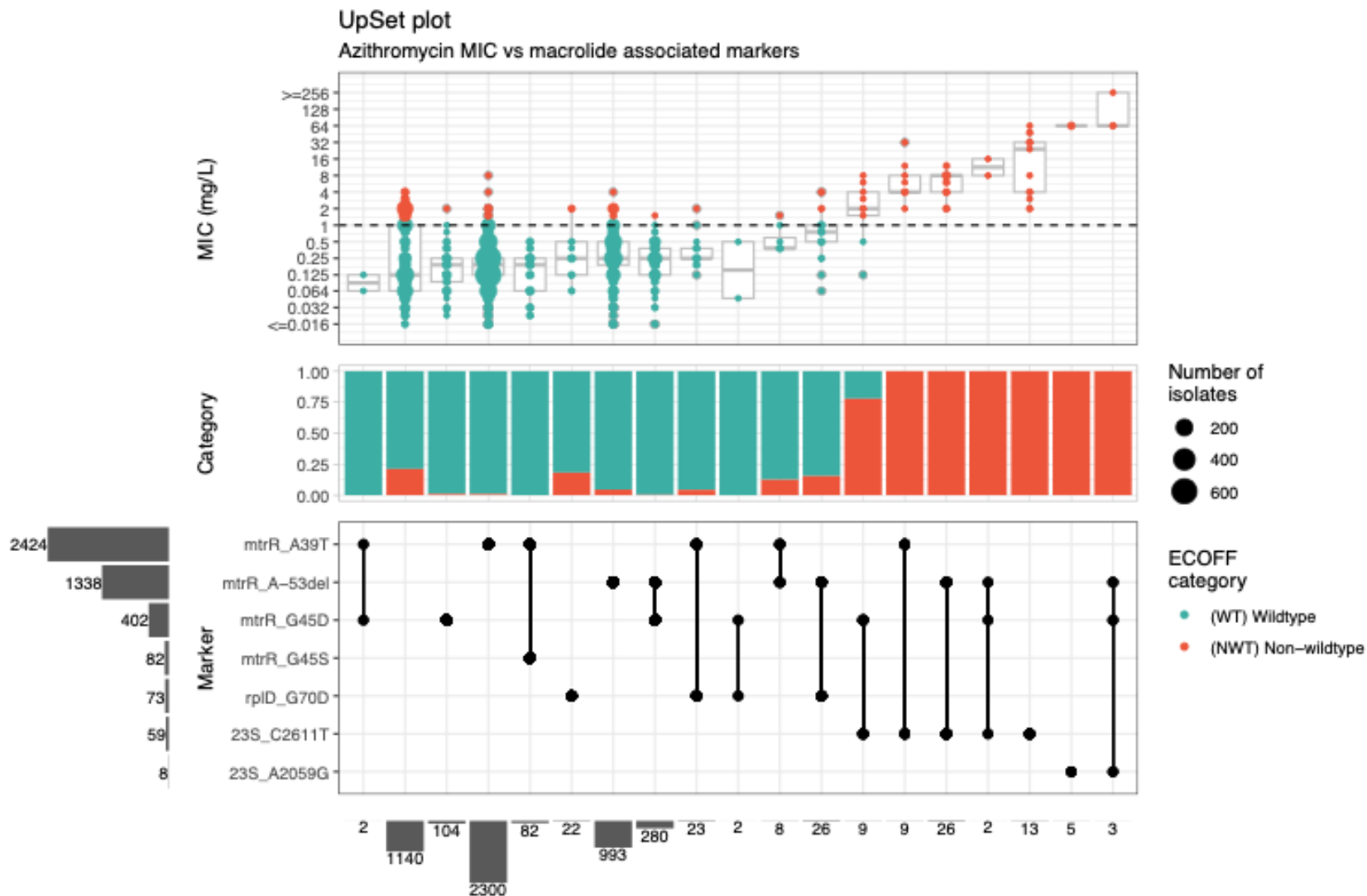
```
autoplot(compare_mic_with_eucast(mics = azm_data, ab = "Azithromycin", mo = "Neisseria gonorrhoeae"))
```





# Upset plots

```
> amr_upset(geno_table, pheno_table, pheno_drug = "Azithromycin", assay = "mic")
```



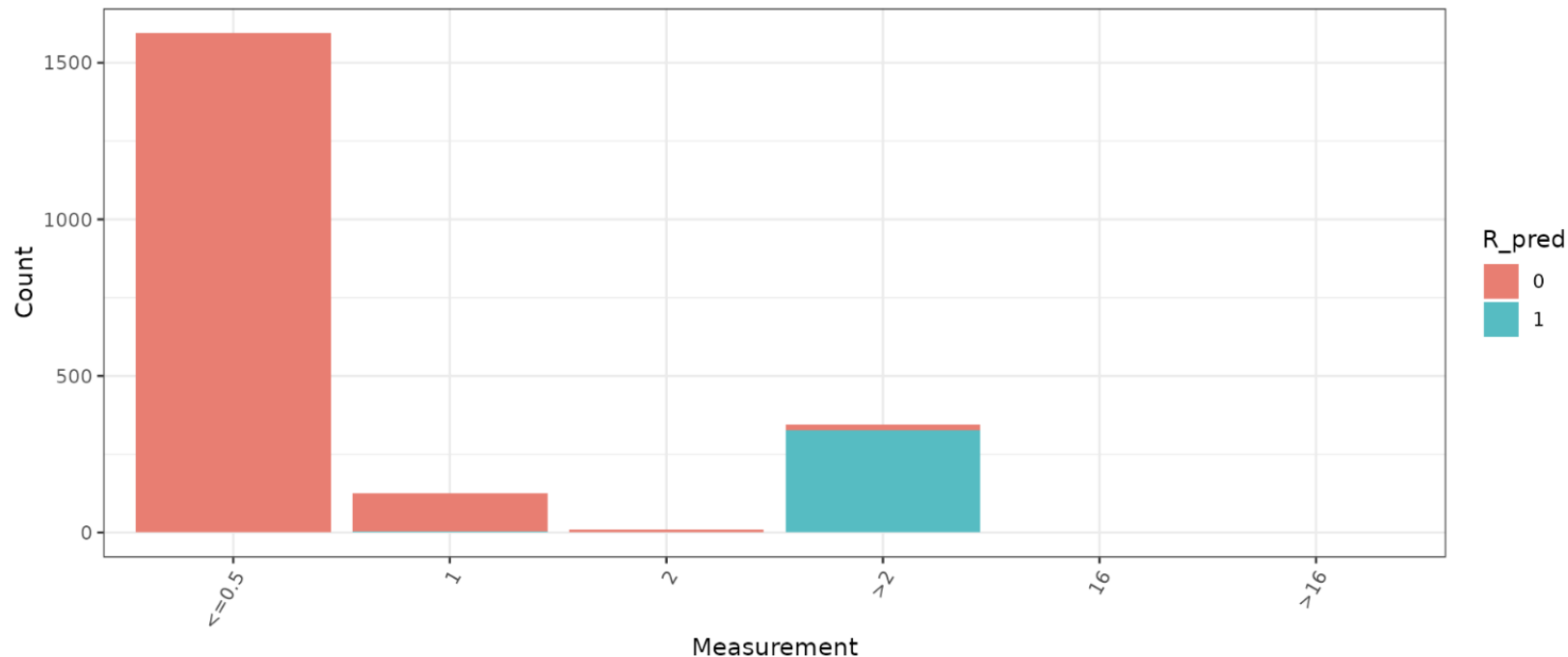


# Concordance analysis

```
model <- amr_logistic(binary_matrix = ecoli_cip)
metrics <- concordance(ecoli_cip, truth = "R", prediction_rule = "logistic", logreg_results = model)
assay_by_var(metrics$data, colour_by = "R_pred")
```

MIC for Ciprofloxacin in *E. coli*

coloured by genotypic prediction based on logistic regression



#> Metrics:

#> Sensitivity : 0.9481

#> Specificity : 0.9954

#> PPV : 0.9763

#> NPV : 0.9896

#> Accuracy : 0.9875

#> Kappa : 0.9545

#> F-measure : 0.9620

#> VME : 0.0519

#> ME : 0.0046