ESGEM-AMR Working Group: Interpretive Standards for AMR Genotypes

Overview

ESGEM-AMR is a working group coordinated by the ESCMID Study Group for Epidemiological Markers (ESGEM). It is led by Prof Kat Holt (LSHTM) and Dr Natacha Couto (ESGEM Chair), and is open to ESGEM members and others with relevant expertise.

Purpose

The overall purpose of the group is to **capture expert knowledge** on the relationship between antimicrobial resistance (AMR) genotypes and antimicrobial susceptibility testing (AST) phenotypes in bacterial pathogens, in a manner that:

i. Recognises and accounts for differences between species (is organism-specific);
ii. Connects with EUCAST Expert Rules, Expected Phenotypes, and other standards as far as practicable;
iii. Uses standardised data structures to capture expert knowledge, which are interoperable with a range of informatics tools and databases (i.e. not platform-specific).

The initial focus of the ESGEM-AMR working group will be expert curation of interpretive standards for AMR genotypes in the AMRrules format, but may expand in future to include other activities/projects. ESGEM-AMR will work with EUCAST to ensure alignment of the AMRrules approach with the EUCAST Subcommittee on WGS and Phenotypic AST, including their initial report (2017) and ongoing updates. We will also seek to engage with other relevant groups within ESCMID and externally.

Membership

Anyone with relevant expertise can apply to join the ESGEM-AMR working group. Prospective members will be asked to register their interest via an online form, to be reviewed by the ESGEM-AMR leads and ESGEM Executive Committee, and will be selected on the basis of expertise and ensuring a spread across organisms.

Working group members will be required to sign a Memorandum of Understanding outlining the roles and responsibilities of the working, and a Code of Conduct, to formalize their involvement with the working group. Active working group members will be encouraged to note their membership on their CV, and will be invited to co-author outputs.

As a working group of ESGEM (the ESCMID Study Group for Epidemiological Markers), ESGEM-AMR members will be encouraged to also join ESGEM, which requires current membership of ESCMID, however this is not a requirement.

Background

Introduction: AMRrules

Detecting and interpreting genetic determinants of antimicrobial resistance (AMR) in bacterial genomes is a fundamental task in microbial genomics, with applications across the clinical, public health and research domains. National or international WGS-based surveillance programs require the correct interpretation of genotypic data to avoid overestimation of resistance. Additionally, genotypes are also used as a proxy for treatment options in recent clinical studies using metagenomics (direct sequencing) directly from clinical samples, where a phenotypic test is often not available, and as we move towards such diagnostic scenarios more confident interpretations will lead to improved outcomes.

Progress has been made in the development of standardized bioinformatics tools and resources to support this activity, principally the development of databases of AMR determinants (NCBI *refgene* being the primary sequence database; CARD *Antibiotic Resistance Ontology* being the primary ontology); tools for finding AMR determinants in query genomes to generate an AMR genotype profile (including NCBI *AMRfinderplus*, CARD *Resistance Gene Identifier*, and *ResFinder*); and tools for harmonizing AMR genotype profiles generated using different tools (*hAMRonization*). However there remains a lack of standards for **interpreting** AMR genotype profiles, in terms of the likely functional implications of detecting a given genetic determinant in a given organism.

An analogy can be drawn to the interpretive standards that exist for interpreting antimicrobial susceptibility testing (AST) assay data in terms of clinical categories (S/I/R), which capture expectations about the likely clinical response to treatment for a given organism with a given minimum inhibitory concentration (MIC) or disk diffusion (DD) zone size measured by AST. Clinical and laboratory data are reviewed regularly by expert committees (e.g. EUCAST, CLSI) to update published guidance on threshold values ('clinical breakpoints') with which to interpret MIC or DD measures for specific bug-drug combinations into S (susceptible), I (susceptible with increased exposure), or R (resistant) categories. In addition, EUCAST aggregates distributions of MIC and DD zone size values for bug-drug combinations, to define epidemiological cutoffs (ECOFF) that distinguish wildtype (WT) from nonwildtype (NWT) parts of the distribution.

The AMRrules Initiative proposes that a similar strategy is needed to develop and update published standards for interpreting the presence of AMR determinants as WT or NWT, and associated with MIC phenotypes that would be categorized as S, I, or R (resulting in six genotype categories, S^{WT}, S^{NWT}, I^{WT}, I^{NWT}, R^{WT} and R^{NWT}, as recommended by the EUCAST Subcommittee on WGS and Phenotypic AST). These standards, or interpretive rules, should be based on expert review of available data on resistance mechanisms, and matched data linking genotypes to AST phenotype profiles (MIC as gold standard), in an organism-specific manner. The genotype interpretation rules should connect as far as practicable with EUCAST Expert Rules, Expected Phenotypes, and other standards; and be captured in a standardised data structure that is interoperable with a range of informatics tools and databases (i.e. not platform-specific).

Specifically, we proposes to develop three related resources (purple in **Figure 1**):

1) **AMRrules**, a set of organism-specific rules for interpreting AMR genotypes
2) **interpretAMR**, a code base for applying the AMRrules to interpret a genotype profile
3) **A genome-phenotype database** to support systematically developing and updating rules

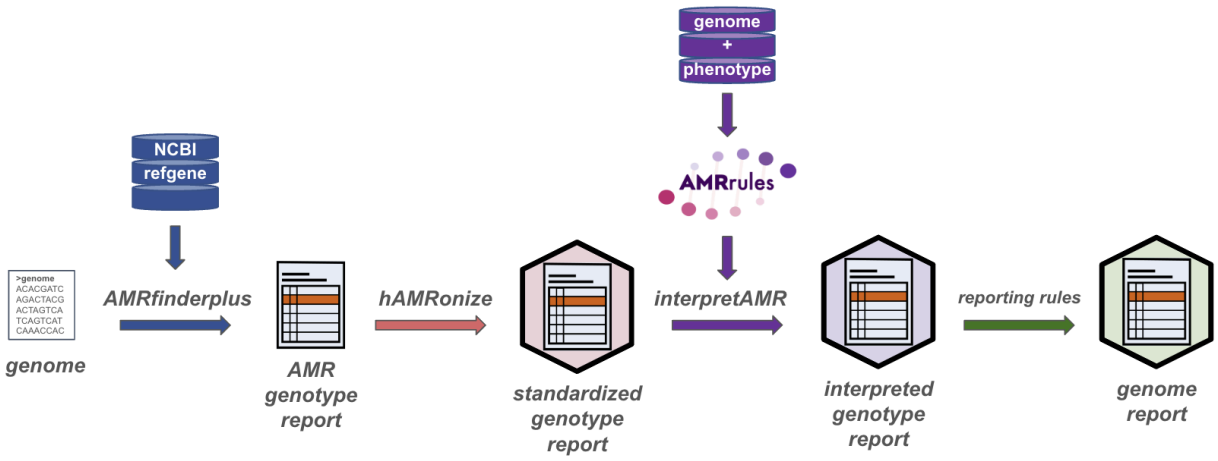

Figure 1: Schematic of current resources, and the proposed new interpretive layer (*AMRrules* and *interpretAMR*). The common task of detecting and interpreting genetic determinants of AMR in a bacterial genome involves first selecting a suitable reference database of known AMR determinants (e.g. 'NCBI refgene'), and then searching the genome sequence against this database (e.g. using '*AMRfinderplus*') to generate a genotype report. Reports from different tools and databases can be '*hAMRonize*-d' into a standardized format. In the absence of clear rules for interpreting these reports, users typically assume that the presence of a determinant labelled as linked to a particular drug or class implies that the sequenced isolate is resistant to that drug. We propose to develop interpretive rules ('*AMRrules*') that can be used to annotate a genotype report (via '*interpretAMR*' code) with organism-specific interpretation (categorized as wt/nwt, S/I/R). The interpretive rules should be populated and updated based on large-scale systematic genome-phenotype datasets. (Note that 'genotype reports' refer to one-row-per gene text file formats intended for bioinformaticians and analysis pipelines. Additional reporting logic is required to format genome reports (green in figure) to summarize relevant information for different types of end-users such as clinicians or public health officials; e.g. specifying which drugs to report, how results should be formatted such as in the style of an antibiogram, etc.)

Genotype vs phenotype

It is common to speak about the desire to 'predict resistance' from genomes or genotypes. However it is important to define what is meant by 'resistance', and indeed what is meant by 'predict'. EUCAST defines the **clinical categorization** 'R' or Resistant as meaning there is a high likelihood of therapeutic failure even when there is increased exposure. Bacterial isolates can be categorized as R on the basis of a measured **laboratory phenotype** (e.g. MIC, or disk diffusion diameter), using clinical breakpoints to interpret the measurement (e.g. if the MIC exceeds the R threshold for the organism and drug in question, the isolate is categorized as Resistant). Most studies attempting to model the relationship between 'resistance' and genetic data use 'categorized as R' (or, equivalently, 'MIC exceeding the R breakpoint') as the outcome to model. This is essentially modeling the value of an established predictor (the clinical categorization based on laboratory phenotype) using a new type of data (genetic). It is important to note that this is different from actually modeling the likelihood of clinical failure based on genetic data. To do that, one would need sufficient systematic data matching genome sequence to clinical response, in order to train a suitable model or classifier; it is not sufficient to predict the intermediate laboratory phenotype and treat that as a clinical prediction.

In 2017, the EUCAST Subcommittee on the Role of Whole Genome Sequencing (WGS) in Antimicrobial Susceptibility Testing (AST) of Bacteria recommended that the primary phenotype comparator for prediction from genetic data should be the epidemiological cut-off value (ECOFF), which aims to distinguish the wildtype MIC distribution from nonwildtype. Clinical breakpoints must not divide wildtype MIC distributions, thus the ECOFF is the lowest possible susceptible breakpoint, and the most relevant 'cut-off' to screen for acquired resistance. The Subcommittee also noted that a wildtype organism can be categorized as 'S', 'I' or 'R' and a nonwildtype organism may still be categorized as 'S'. This means that if one wants to encompass both wildtype/nonwildtype status and clinical 'S', 'I' and 'R', the possible categories are S^{WT}, S^{NWT}, I^{WT}, I^{NWT}, R^{WT} and R^{NWT}. The report gives the example of *Pseudomonas aeruginosa* and tigecycline to illustrate R^{WT}; another would be *Klebsiella pneumoniae* and ampicillin, or indeed most combinations noted in the EUCAST Expected Resistance tables (not all, because the definition of expected is >90% should be resistant, which does not necessarily imply universality across the species and may occur if e.g. only resistant lineages of a species are associated with disease and dominate the MIC distribution). The well-understood scenario of an otherwise carbapenem-susceptible organism acquiring a carbapenemase such as KPC-2, would be characterized as changing the phenotype from S^{WT} to R^{NWT}.

AMR genotype profiles

Most genomic analyses of AMR involve screening WGS data against a database of known AMR determinants (genes and mutations), to generate an AMR genotype profile. The results are typically reported in a way that links each determinant with associated drugs/classes. Users tend to assume that the presence of an AMR determinant implies the sequenced isolate (a) is nonwildtype; and (b) would likely be categorized as 'R' to the associated drugs/classes. Although not often expressed this way, the assumption tends to be that the presence of a determinant linked to drug X implies 'R^{NWT}' or 'I^{NWT}' to drug X. In many cases concerning mobile

AMR genes, the simple interpretation of 'presence of gene X' implies 'R^{NWT} to drug Y' is accurate. However these assumptions do not always hold, and can differ by species. For example, wildtype isolates of *Klebsiella pneumoniae* carry chromosomal core genes *bla*SHV (ampicillin R^{WT}), *oqxAB* (ciprofloxacin S^{WT}) and *fosA* (fosfomycin S^{WT}). In contrast, wildtype *Escherichia coli* lack these genes, and their presence in *E. coli* often implies different phenotypes: *bla*SHV (ampicillin R^{NWT}), *oqxAB* (ciprofloxacin R^{NWT}) and *fosA* (fosfomycin R^{NWT}). To complicate matters, certain alleles of *bla*SHV are also associated with R^{NWT} to other drugs, e.g. *bla*SHV-12 is an extended-spectrum beta-lactamase associated with ceftriaxone R^{NWT} in both *Klebsiella pneumoniae* and *E. coli*. Therefore to interpret AMR determinants, it is important to consider both the gene/allele and the species, and often the genetic context.

Sometimes the simple interpretation of AMR genotypes is made explicit and AMR genotype profiles are presented to mimic an antibiogram, where the presence of determinants are interpreted in terms of a clinical categorization ('S', 'I', 'R'). In many cases these antibiograms are constructed without explanation of (i) how the clinical categorizations are inferred from the genotype profiles, (ii) the evidence supporting these inferences, or (iii) the confidence in the inference. In essence the antibiogram structure is used to organize results (grouping determinants by drug/class), and the underlying logic is simply the general assumption that presence of a gene linked to drug X in the database implies 'R' for drug X. Sometimes these implicit assumptions will be more or less correct, but often they will not be. In any case, the presentation as an antibiogram can be potentially misleading, as its use in reporting AST implies the data have been interpreted against a particular standard in order to assign a clinical category. Indeed it is expected that antibiograms be accompanied by a statement as to which standard was used to interpret the laboratory data.

Interpretive standards

We propose that standards are needed for interpreting AMR genotype profiles, similar to those established for interpreting phenotypic AST data. This is particularly important where results are organized in the form of an antibiogram, which is arguably a desirable format in terms of clarity of communicating information to key end users. As outlined above, genes can have different activity in different contexts, and it is therefore essential that interpretive standards for AMR genotypes be organism-specific in the same way as those for AST data.

Importantly, genetic data offers the opportunity to distinguish formally between wildtype and nonwildtype states, because it is possible to definitively determine whether a gene is 'core' or 'intrinsic' to the species (and thus associated with wildtype phenotypes) vs 'acquired' or 'accessory' (and thus may be associated with nonwildtype phenotypes). We therefore propose to use genotype profiles to assign isolates to six '**genotype categories**' comprising S^{WT}, S^{NWT}, I^{WT}, I^{NWT}, R^{WT} and R^{NWT}. As outlined above, these are not direct predictions of clinical phenotype, but rather represent categorizations of genotypes and thus should be referred to as 'genotype categories'. As they are based on observed associations between genotypes and laboratory phenotypes, which are themselves interpreted as S/I/R according to breakpoints applied to observed MIC measurements, the 'genotype category' can be interpreted as both (i) an indication of whether the genotype is core or acquired in this species, and (ii) a prediction of

whether the isolate would exceed the clinical breakpoints for 'I' or 'R' if MIC were measured. The former is useful in itself, especially when characterizing an emerging clone or outbreak.

While users will inevitably want to interpret the S/I/R part as a prediction of clinical treatment response, it is important to recognize and reiterate that genotypes cannot be interpreted that way unless and until there is data directly linking genotype to clinical response. Notably, we propose to use above/below ECOFF as a secondary element of the interpretation, rather than primary. This is because the combination of wildtype/nonwildtype and S/I/R is more complete and informative than ECOFF, and the relevance of ECOFF to interpreting the effect of wildtype or core genes (a key concern for interpreting genotype profiles) is not clear.

Objectives and Scope

The AMRrules interpretive standards will aim to capture all exceptions to the generalized interpretation of 'presence of gene X' implies 'R^{NWT} or I^{NWT} to drug Y'. Ultimately, they should also differentiate R^{NWT}' from I^{NWT}', and interpret combinations of genes.

However, the initial focus of the ESGEM-AMR Working Group will be on clearly delineating **'wildtype' (core) genotypes underlying 'wildtype' (intrinsic/expected) phenotypes** for clinically relevant bacteria. This is analogous to EUCAST's Expected Resistance and Expert Rules for susceptibility testing, which capture expert knowledge on interpretive rules for AST, and ideally will capture genetic mechanisms behind all expected resistances.

Initial development will work species-by-species, generating **one rule set for each species**, which can be combined together into a single resource, via the AMRrules project.

Subsequently, we envisage moving on to address rules concerning the interpretation of **acquired determinants**, including genes, mutations and **combinations** thereof, and how these should be interpreted in terms of expected **deviation from wildtype phenotypes** in each organism. This will require high-volume high-quality genome-phenotype data, which is not yet available for most organisms and will need to be generated and aggregated.

Organisms

The scope of organisms to be included will depend on the availability of expertise to contribute to the working group, but we will aim to first develop rule sets for:

1. ESKAPEE pathogens:
   Enterococcus faecium
   Staphylococcus aureus
   Klebsiella pneumoniae species complex*
   Acinetobacter baumannii
   Pseudomonas aeruginosa
   Enterobacter cloacae complex*
   Escherichia coli

(*Note the ESKAPE list was established prior to WGS-informed understanding of the structure and diversity of the *Klebsiella pneumoniae* or *Enterobacter cloacae* species complexes, so referred to *Klebsiella pneumoniae* and *Enterobacter* spp.)

2. Other organisms on the WHO Priority Pathogens list:
   Salmonella spp., *Shigella* spp., and other *Enterobacteriaceae*
   Neisseria gonorrhoeae
   Streptococcus pneumoniae
   Haemophilus influenzae
   Helicobacter pylori
   Campylobacter spp.

   (Note the current list dates from 2017 and an update is due Q2 2024, priorities will be reviewed once it is released)

3. Other organisms of clinical relevance where sufficient expertise and data is available, prioritizing those with Expected Resistant phenotypes

Out of scope: WGS prediction

There is intense academic and commercial interest in training classifiers to predict AMR phenotypes from 'whole genomes' and/or transcriptomes, using different kinds of genomic features (e.g. kmers, genes, single nucleotide variants). However the ESGEM-AMR Working Group will focus on the specific issue of **interpreting genotype profiles**, generated by screening WGS data against a database of known AMR determinants as discussed above (e.g. screening against the refgenes AMR database using the AMRfinderplus tool).

Genotype profiling is the most widely used and accessible approach to investigating AMR from WGS data, is based on decades of accumulated knowledge of molecular mechanisms of resistance, and the resulting profiles hold epidemiological value in understanding the mechanisms and spread of resistance in addition to their value in predicting phenotypes. We therefore expect that genotype interpretation will remain relevant and complementary to current and future whole-genome, whole-transcriptome, and other phenotype prediction/profiling approaches.

Both approaches (genotype profiling and interpretation, and whole-genome prediction) share a common need for high-volume high-quality matched genome-phenotype datasets, and as such the working group will engage with the broader WGS prediction community to further our common goals in developing open and accessible data resources.

Mode of working

Individual members or subgroups will be **allocated to work on specific organism/s** that match their expertise. The main task will be to **propose rule sets** for the assigned organism/s, in the AMRrules format (see Table 1 below and Technical Guidance). Preparation of publications describing the rationale and testing for a given organism or set thereof will also be encouraged.

Ideally, the interpretive standards should capture all exceptions to the generalized interpretation of 'presence of gene X' implies nonwildtype resistant (R^{NWT}) to drug Y'. Ultimately, they should also differentiate R^{NWT} from I^{NWT}, and interpret combinations of genes. However, the initial priority will be creating rule sets that clearly delineate **core genes** associated with **'wildtype' phenotypes** for each species (see AMRrules overview for the reasons for this approach).

Working group members will be encouraged to identify public matched genome/AST data to use in validation and testing, and to seek or share additional unpublished data if possible. If unpublished data is shared for this purpose, it must be treated in confidence under the terms of the working group's Memorandum of Understanding and Code of Conduct, which all members must sign before starting. The publishing and public deposition of data is strongly encouraged, either independently (so it can be cited by the working group publication) or as part of the working group (if this suits the needs and timelines of the data owners). However using or sharing unpublished data for the purpose of setting or testing rules by the working group does not obligate anyone to publicly release data, this will always be the decision of the data owners.

Once a proposed rule set for a given organism has been drafted, it should be submitted for review by the ESGEM-AMR chairs and lead bioinformatician (Jane Hawkey, Monash), according to agreed criteria. Approved rule sets will be added to the master scheme and included in the next release.

Technical guidance

See: ESGEM-AMR Technical Guidance

Planned Outputs

Rule sets

The key output of the ESGEM-AMR working group will be organism-specific rule sets with which to interpret AMR genotypes. We will adhere to FAIR principles to ensure the outputs are findable, interoperable, accessible, and reusable. Rule sets will follow a **standard format** (see Table 1 below and Technical Details). Details of the rule specification will be further developed and refined by the working group during the course of their activities. Rule sets will be made **publicly available** via open-access repositories under a permissive license (GNU General Public License v3.0). They will be **versioned** via numbered releases, and issued with stable document object identifiers (DOIs).

Table 1. Proposed format for organism-specific rules, populated with example rules for *Klebsiella pneumoniae*.

species	allele	context	drug	category	PMID	note
Klebsiella pneumoniae	blaSHV	core	ampicillin	wt R	32284385	Specific alleles can be ESBL, these are mostly mobile
Klebsiella pneumoniae	oqxA	core	ciprofloxacin	wt S	30834112	Wildtype core gene, not expected to confer multiple drug resistance unless mobilised under strong promoter
Klebsiella pneumoniae	oqxB	core	ciprofloxacin	wt S	30834112	Wildtype core gene, not expected to confer multiple drug resistance unless mobilised under strong promoter
Klebsiella pneumoniae	fosA5_fam	core	fosfomycin	wt S	27261267	Wildtype core gene, not expected to confer multiple drug resistance unless mobilised under strong promoter

Rule sets and associated tools should be **compatible with existing resources** for AMR genotype analysis as far as possible, including use of common sequence identifiers, standard gene nomenclature, and data formats. Preliminary code has been developed for annotating the gene-level reports output by the AMRFinderPlus tool, which uses the NCBI refgene database of AMR determinants. Working group outputs should be interoperable with the hAMRonization format, to facilitate compatibility with the outputs of CARD RGI, ResFinder, and >12 other AMR genotyping tools whose outputs can be readily converted to hAMRonization format.

Table 2. Example AMRFinderPlus genotype report for a *K. pneumoniae* (ERR257656), annotated with additional columns using the rule set shown in **Table 1**. Note a column reporting the source of the organism-specific interpretation is also annotated ("Klebsiella pneumoniae; v1.1") but is not shown here for brevity.

Gene symbol	Context	Org interpretation	Drug	Class	Subclass	% Coverage	% Identity
oqxB19	core	wt S	ciprofloxacin	PHENICOL/QUINOLONE	PHENICOL/QUINOLONE	100	100
oqxA3	core	wt S	ciprofloxacin	PHENICOL/QUINOLONE	PHENICOL/QUINOLONE	100	100
blaSHV-33	core	wt R	ampicillin	BETA-LACTAM	BETA-LACTAM	100	100
fosA	core	wt S	fosfomycin	FOSFOMYCIN	FOSFOMYCIN	100	100
dfrA1			trimethoprim	TRIMETHOPRIM	TRIMETHOPRIM	100	100

Publications

The overall motivations, goals, approach and rule sets will be described in an article co-authored by all contributors to the working group.

Additional articles describing the details for specific organisms or sets thereof, or describing associated tools or protocols, may also be developed by the working group, with authorship to be determined by contribution.

Wherever possible, publications should be accompanied by accessions for publicly available genome sequence and AST data supporting the rule development and testing. However the timing of this remains the decision of the data owners as outlined above, and in the Memorandum of Understanding.

Conferences

The working group will aim to present their activities at relevant conferences, upcoming ones that we plan to target include:

ASM-NGS (Oct 13-16, 2024, Washington DC; late-breaker abstracts Aug 14 2024)
ECCMID (April 12-15, 2025, Vienna; abstracts due 27 Nov 2024)
ABPHM (May 21-23, 2025, Hinxton; abstracts due ~Jan 2025)
IMMEM (Sep 17-20, 2025, Porto; abstracts due ~Mar 2025)

Work Plan and Timeline

April 2024 — Working group launch and call for interest - ESGEM session @ ECCMID

May 2024 — Introductory webinars

- 2x, to accommodate different time zones
- Advertised to members of ESGEM and other ESCMID study groups; SEDRIC, PHA4GE, twitter etc
- Introduce project goals, mode of working, commitment needed, benefits
- Attendees invited to ask questions and provide feedback
- Attendees invited to complete online form to register EOI

2 June 2024 — EOI form due

Mid June — Chairs Review EOI submissions

- Identify submissions with suitable expertise
- Review spread of expertise and decide individual organisms / groups of organisms and identify potential subgroup leads
- Invite selected submissions to join working group (by late June)
- Members sign MOU and Code of Ethics

July — Initial meeting of working group

- Review members, goals, modes of working
- Data formats, protocols, support tools
- Discuss standards and process for review of rules
- Set goals and timelines for next 6 months
- Establish communications model, to connect within subgroups, but also discuss issues around formats, coding, analysis, etc across organisms

September — Monthly progress meetings

- Issues emerging related to protocols, rule specification, etc
- Updates on each organism, learning across subgroups
- Ideas for outputs, conferences, publications, funding

April 2025 (?) — Manuscript submission & presentation at ECCMID