ResFinderFG v2.0: a database of antibiotic resistance genes obtained by functional metagenomics

AUTHORS

Rémi Gschwind[1]*, Svetlana Ugarcina Perovic[2], Maja Weiss[3], Marie Petitjean[1], Julie Lao[1], Luis Pedro Coelho[2], Etienne Ruppé[1]

[1] University of Paris Cité, INSERM UMR 1137 IAME, F-75018 Paris, France

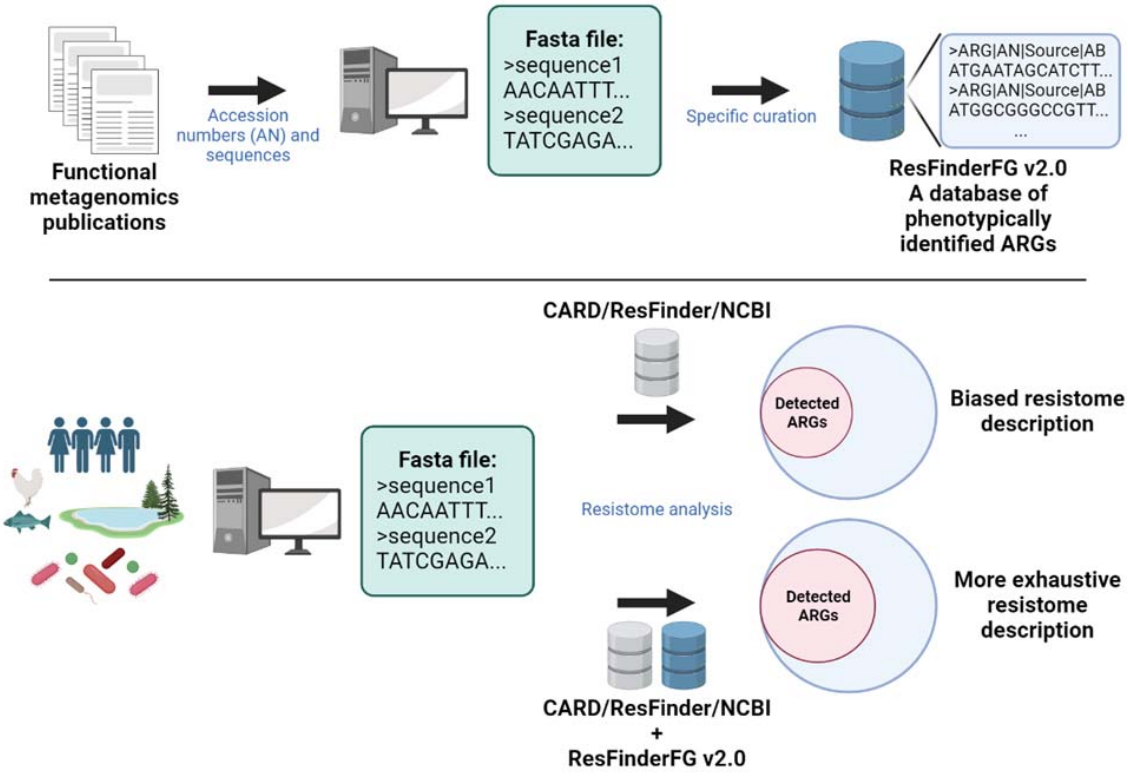
[2] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, 200433, China

[3] Research Group for Genomic Epidemiology, Technical University of Denmark, Kgs, Lyngby, 2800, Denmark

* To whom correspondence should be addressed. Tel: +336658543545; Fax: -; Email: remi.gschwind@inserm.fr

Present Address: Rémi GSCHWIND, University of Paris Cité, INSERM UMR 1137 IAME, 16 rue Henri Huchard, 75018 Paris, France

GRAPHICAL ABSTRACT



Caption: Additional use of ResFinderFG v2.0 database (composed of antibiotic resistance genes obtained with functional metagenomics) on the Center of Genomic Epidemiology webserver (https://cge.food.dtu.dk/services/ResFinderFG/), allows for more exhaustive resistome descriptions.

ABSTRACT

Metagenomics can be used to monitor the spread of antibiotic resistance genes (ARGs). ARGs found in databases such as ResFinder and CARD primarily originate from culturable and pathogenic bacteria, while ARGs from non-culturable and non-pathogenic bacteria remain understudied. Functional metagenomics is based on phenotypic gene selection and can identify ARGs from non-culturable bacteria with a potentially low identity shared with known ARGs. In 2016, the ResFinderFG v1.0 database was created to collect ARGs from functional metagenomics studies. Here, we present the second version of the database, ResFinderFG v2.0, which is available on the Center of Genomic Epidemiology web server (https://cge.food.dtu.dk/services/ResFinderFG/). It comprises 3,913 ARGs identified by functional metagenomics from 50 carefully curated datasets. We assessed its potential to detect ARGs in comparison to other popular databases in gut, soil and water (marine + freshwater) Global Microbial Gene Catalogues (https://gmgc.embl.de). ResFinderFG v2.0 allowed for the detection of ARGs that were not detected using other databases. These included ARGs conferring resistance to beta-lactams, cycline, phenicol, glycopeptide/cycloserine and trimethoprim/sulfonamide. Thus, ResFinderFG v2.0 can be used to identify ARGs differing from those found in conventional databases and therefore improve the description of resistomes.

INTRODUCTION

Antimicrobial resistance (AMR) is recognized as a global threat possibly leading to the lack of efficient treatment against deadly infections[1]. From a genetic perspective, AMR is driven by mutational events (e.g. fluoroquinolone resistance is driven by mutations in the topoisomerase-encoding genes) and the expression of antibiotic resistance genes (ARGs). ARGs are widespread in human- and animal-associated microbiomes, and in the environment[2]. Hence, these microbial niches are now considered in a One Health manner[3]. Although not every ARGs represents a direct risk for human health[4], genes are able to travel

from one environment to another by strain dissemination and then to pathogenic bacteria by horizontal gene transfer[5].

Identifying ARGs and assessing this risk is essential to better understand and putatively find means to prevent their dissemination in pathogenic bacteria. To identify ARGs, culture-based methods, PCR, qPCR[6], genomic and metagenomic sequencing have been used. Metagenomics makes it possible to sequence all the DNA from a sample. Then with the help of alignment (e.g. BLAST[7]) or Hidden Markov Models (e.g. HMMER[8]) – based tools, ARGs can be identified based on their sequence identity with known ARGs. To this end, several ARG databases[9] exist such as CARD[10] or ResFinder[11]. However, their sensitivity and specificity are highly dependent on which database is used and on the parameters used for the search. Moreover, methods using databases of known ARG cannot detect truly novel ARG i.e. not described to date. Pairwise Comparative Modelling (PCM), combining sequence and structure homology to detect ARGs, showed good performances to detect ARGs distantly related to known ARGs in the human gut microbiota, yet some predictions were not all functionally validated[12].

Although culturable and/or pathogenic bacteria only represent a small fraction of microbial diversity, their genes make up the vast majority of the ARGs present in existing databases. Thus, the description of the resistome can be biased in environments where non-culturable or non-pathogenic bacteria dominate. In order to detect ARGs distantly-related to known ARGs or ARGs from novel ARG families, functional metagenomics has been used[13]. This method is based on phenotypic detection by expressing exogenous DNA in an antibiotic-susceptible host. Metagenomic DNA is sheared and cloned into an expression vector used to transform a host strain susceptible to antibiotics. Transformant strains are then selected with culture media supplemented with antibiotics. If growth is observed, it implies that the DNA cloned into the expression vector is responsible for the resistance phenotype observed. DNA inserts can then be amplified, sequenced and annotated to identify ARGs[14]. Using functional metagenomics, ARGs sharing low amino acid identity to their closest homologue in NCBI[15], or even not previously classified as ARGs[16], could be detected in human[16–26], animal[22,27–33], wastewater[34–41] and other environmental samples[5,15,42–64]. Despite being a laborious technique, genes described by functional metagenomics are mainly absent in classical ARG databases. Two databases listing specifically functionally identified ARGs were created:

ResFinderFG v1.0[65] and FARME DB[66]. ResFinderFG v1.0 (https://cge.food.dtu.dk/services/ResFinderFG-1.0/) was based on the data coming from 4 publications, while FARME DB includes data from 30 publications, mainly reporting environmental genes which were not necessarily curated to include only ARG sequences[67]. Here, we report a new version of the ResFinderFG database, ResFinderFG v2.0, providing well-curated data from functional metagenomics publications available until 2021 that include environmental and host-associated samples.

MATERIAL AND METHODS

Construction of ResFinderFG v.2.0

To retrieve publications using functional metagenomics for the identification of antibiotic resistance genes, the four publications used to construct ResFinderFG v1.0 were first considered. Then, all the publications which were cited by these four publications and all the publications that cited one of these publications were collected. In addition, publications found with the following terms on PubMed: "functional metagenomics" AND "antibiotic resistance", were added to this pool. After filtering out all the reviews, publications were screened one by one. First, the use of functional metagenomics to detect ARGs, meaning the expression of exogenous DNA thanks to an expression vector into a host selected on an antibiotic containing medium, was checked. Second, the accessibility to the insert sequences was also checked. Database construction and curation was then performed as follows (Figure 1). The accession numbers describing insert DNA sequences functionally selected using antibiotics were included and DNA sequences were retrieved using Batch Entrez. CD-HIT[68] was used to remove redundant DNA sequences and annotation of the remaining was done using PROKKA v.1.14[69]. To specifically select insert DNA sequences with ARG annotations, a representative pool of ARG annotations was obtained by applying the PROKKA annotation process to the ResFinder v4.0 database. Resulting annotations were used as a reference to specifically select insert DNA sequences containing an ARG annotation. Metadata (sample origin and antibiotic used for selection) associated with each insert were collected. Then, several curation steps were added. First, a curation based on the consistency between the ARG and the antibiotic family used for selection (e.g. beta-lactamase-encoding gene selected with a beta-lactam antibiotic). Second, we apply a filter on the size in amino acids of the

predicted protein: 260 for beta-lactamase, 378 for tetracycline efflux genes, 641 for tetracycline resistance ribosomal protection genes, 178 for chloramphenicol acetyltransferase, 247 for methyltransferase genes and 158 aa for dihydrofolate reductase genes). Finally, insert sequences containing more than one ARG annotation (consistent with the antibiotic used for selection) were discarded as we could not know which one was responsible for the observed phenotype. The database also includes metadata retrieved in the GenBank metadata retrieval process and ARO annotation for each gene for comparability with other databases using ARO ontologies[10]. Besides, we provide in the supplementary material the access numbers of inserts (n=5'256) containing a single ORF among which potential ARG from new families can be characterized (Supplementary Table 1).

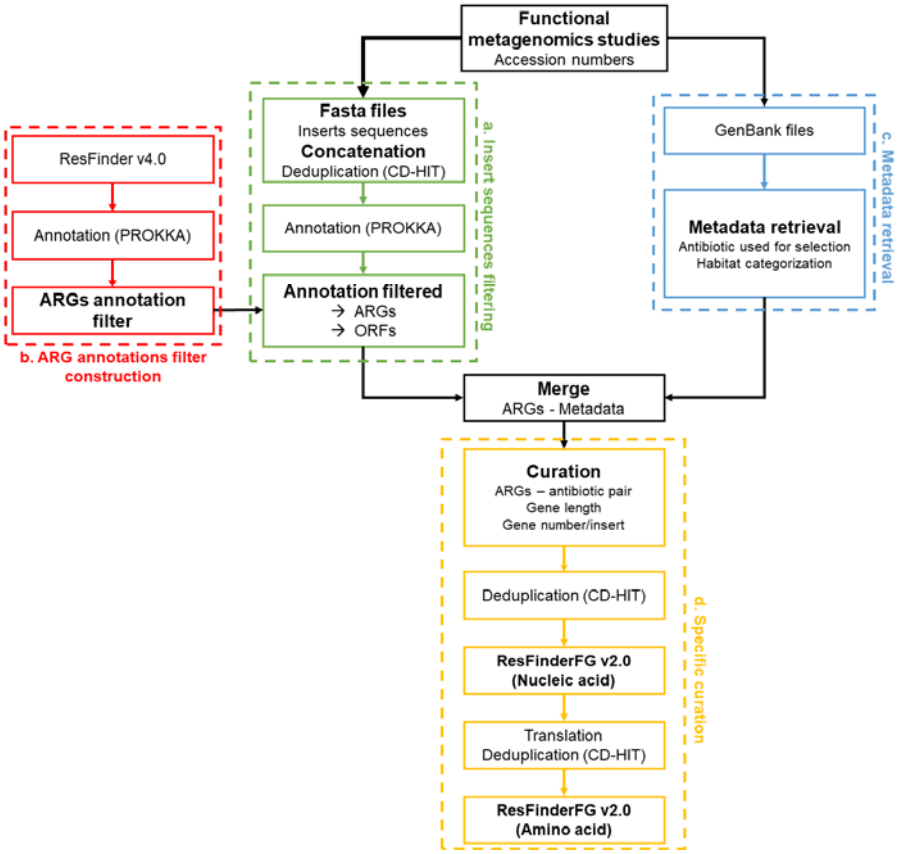


Figure 1. ResFinderFG v.2.0 construction workflow. a. The insert sequences filtering step consists of concatenating all the insert sequences and selecting only those containing an ARG annotation. b. ARG annotations filter construction was done by annotating the ResFinder v4.0 database using PROKKA to have a pool of ARG annotations. c. Metadata was retrieved using GenBank files to get the origin of the samples and the antibiotic used for

selection of each insert sequence. d. Specific curation was done to discard false positives and include only ARGs.

Comparison between ResFinderFG v1.0 and ResFinderFG v2.0

To assess the update of the ResFinderFG database and compare v2.0 to v1.0, the number of ARGs in each ARG family and regarding the sample sources was evaluated in both versions. ARG families were categorized according to the antibiotic families they conferred resistance to: glycopeptides/cycloserine, sulfonamides/trimethoprim, beta-lactams, aminoglycosides, macrolides/lincosamides/streptogramins, tetracyclines, phenicols and quinolones. Sample sources were categorized as follows: aquatic, animal-associated, human-associated, plant-associated, polluted environment and soil. Then, to detect the presence of ARGs in several gene subcatalogs (human gut, soil and marine+freshwater) coming from the Global Microbial Gene Catalog (GMGC, https://gmgc.embl.de/download.cgi[2]), ABRicate[70] was run using default parameters with different databases (ResFinderFG v2.0, ResFinder v4.0, CARD v3.0.8, ARG-ANNOT v5, NCBI v3.6).

RESULTS

Construction of ResFinderFG v2.0

A total of 50 publications using functional metagenomics to analyse ARG content were selected, resulting in 23,776 accession numbers (Supplementary Table 2). CD-HIT identified 2,629 perfectly redundant insert sequences (100% sequence identity). PROKKA identified 41,977 open reading frames (ORFs). Among them, 7,787 ORFs had an annotation matching with an ARG annotation of ResFinder v4.0 (228 unique ARG annotations). Then, we did not consider the following ARGs: (i) inconsistency between the predicted ARG and the antibiotic used for selection (n=1,165), (ii) ARGs with excessively small size with respect to the ARG family (n=1,064) and ARGs colocating on an insert (n=398). A second round of CD-HIT was used to remove redundancy (100% nucleic acid sequence identity), which resulted in 3,913 unique ARGs which were finally included in the database which can be used on the Center of Genomic Epidemiology website (https://cge.food.dtu.dk/services/ResFinderFG/; supplementary figure 1).

Comparison with ResFinderFG v1.0

First, the ARGs present in ResFinderFG v.2.0 were compared to the ones present in ResFinderFG v1.0 (Figure 2). A total of 1,631 new ARGs were present in ResFinderFG v.2.0, mainly due to new glycopeptide/cycloserine (+906 genes) and beta-lactam (+333 genes) resistance genes. The glycopeptide/cycloserine resistance genes were mostly annotated as homologues of D-Ala-D-X ligase. New beta-lactams antibiotics used for functional selection compared to v1.0 were cefepime, meropenem and tazobactam. Regarding the sources of ARGs, new ARGs mostly originated from human-associated samples (+1,333 genes).

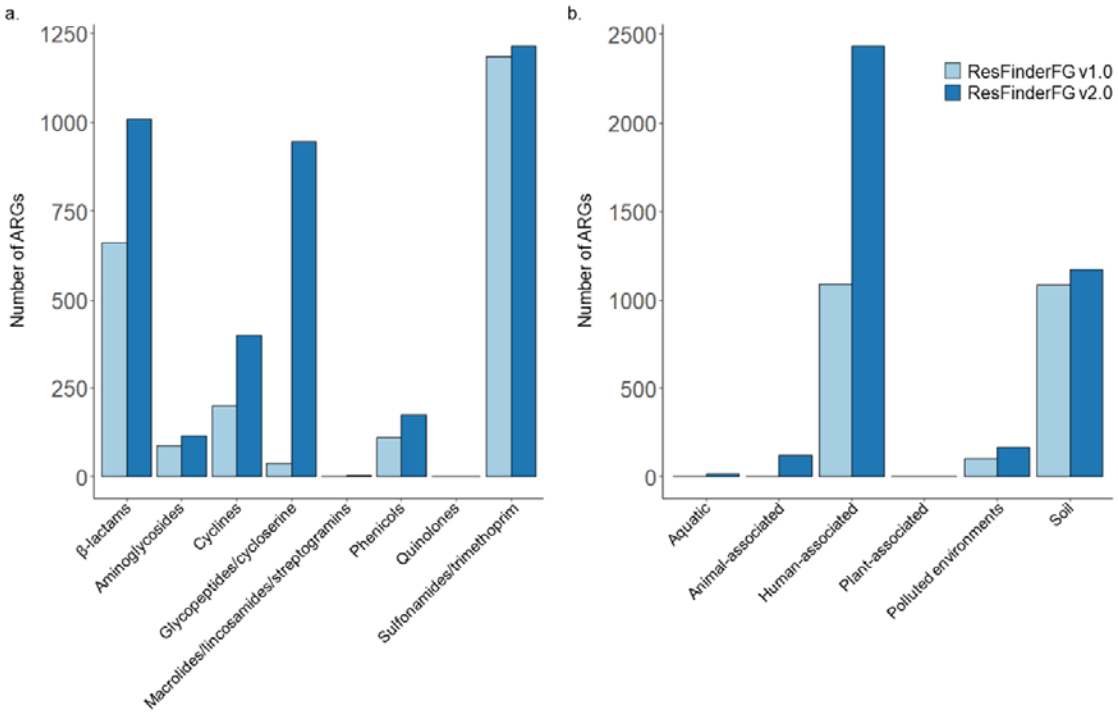


Figure 2. a. Number of ARGs in the ResFinderFG v.1.0 and v.2.0 databases depending on a. the antibiotic families involved; b. the sample sources.

ARG detection in several GMGC gene subcatalogs using ResFinderFG v.2.0 and other databases

ABRicate (default parameters) was used to detect ARGs in GMGC human gut (Figure 3), soil (Supplementary Figure 2a.) and aquatic (marine and freshwater) subcatalogs (Supplementary Figure 2b.). Using ResFinderFG v2.0, 3,025, 211 and 129 unigene hits were obtained analyzing human gut, soil and aquatic subcatalogs respectively. The three most frequently detected

ARG families in all gene catalogs were glycopeptide/cycloserine resistance genes (20.9 to 39.7% of detected ARGs), sulfonamides/trimethoprim resistance genes (21.8 to 58.1% of detected ARGs) and beta-lactamase encoding genes (7.9 to 25.6% of detected ARGs). Phenicols (up to 6.0% of detected ARGs), aminoglycosides (up to 5.3%), cyclines (up to 6.2%) and macrolide/ lincosamide/streptogramin resistance genes (up to 0.03%) were also detected. Also, ResFinderFG v2.0 provides habitat information on where a given ARG was first identified by functional metagenomics. A majority of ARGs identified in the gut subcatalog (90.2%) were indeed initially identified in the human gut by functional metagenomics (supplementary table 3). In the soil gene subcatalog, 62.6% of ARGs detected were also genes identified initially in soil with functional metagenomics. However, ARGs detected in the aquatic gene subcatalog were primarily first identified by functional metagenomics in soil.

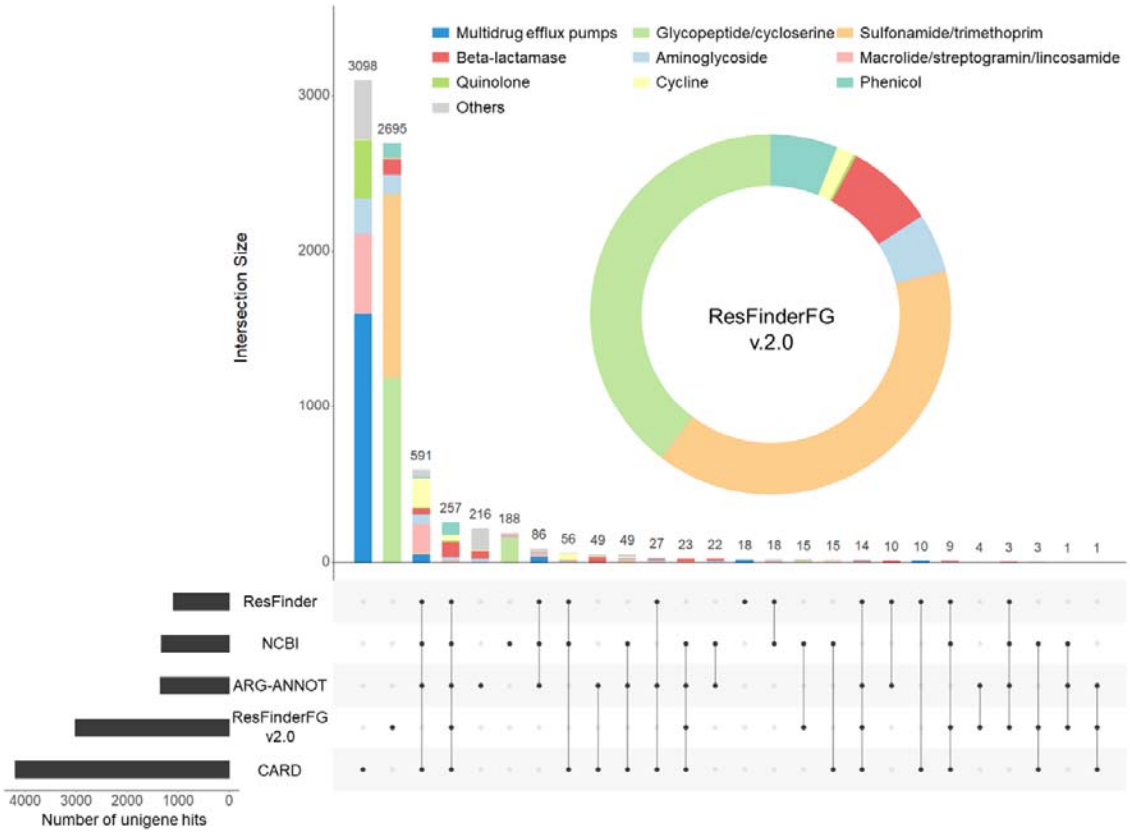


Figure 3. Proportion of antibiotic family unigene hits obtained analyzing GMGC human gut subcatalog using ResFinderFG v2.0 and number of unigene hits obtained analyzing GMGC human gut subcatalog using several databases (ResFinder v4.0, NCBI v3.6, ARG-ANNOT v5,

ResFinderFG v2.0 and CARD v3.0.8) annotated by their antibiotic family. Others: bicyclomycin, beta-lactams, bleomycin, disinfectant and antiseptic agents, fosfomycin, fusidic acid, multidrug, mupirocin, nitroimidazole, nucleoside, peptide, rifampicin, streptothricin.

To compare ResFinderFG v2.0 to other databases, we ran the same ABRicate analysis of GMGC gene subcatalogs using ResFinder v.4.0, CARD v3.0.8, ARG-ANNOT v5 and NCBI v3.6. ResFinderFG v2.0 identified a comparable or even greater number of ARGs compared to other databases. We observed that the most frequently observed ARG family depended on the database used. In the human gut gene subcatalog, glycopeptide/cycloserine resistance gene was the most frequent ARG family found by ResFinderFG v2.0 (39.7% of all unigene hits obtained with ResFinderFG v2.0). In contrast, the beta-lactamase family was the top ARG family with ARG-ANNOT (21.2%). NCBI and ResFinder detected mostly tetracycline resistance genes (20.4 and 23.8% respectively). Finally, multidrug efflux pump unigene hits were the most frequent using CARD (39.4%).

ResFinderFG v2.0 was the database with the highest fraction of database-specific hits, with 89.1% of specific unigene hits composed mainly by glycopeptide/cycloserine resistance genes (D-alanine-D-alanine ligase; supplementary table 4) and sulfonamides/trimethoprim resistance genes (dihydrofolate reductase). By comparison, CARD had 73.7% of specific unigene hits, mostly composed by gene encoding multidrug efflux pumps. Of note, 16.2% of unique CARD specific multidrug efflux pump unigene hits found in the human gut were regulatory genes (Supplementary Table 4).

Between 2.6 and 4.2% of all unigene hits, depending on the gene subcatalog analyzed, were shared by all the databases used. Beta-lactamase encoding genes were the most prevalent among them (ranging from 38.1 to 51.3% of the shared unigene hits), followed by, phenicols, aminoglycosides and tetracyclines resistance genes. However, 25.1, 23.2 and 46.3% of beta-lactamases, aminoglycosides and phenicols resistance genes respectively, were only detected using ResFinderFG v2.0 (Figure 3; supplementary Figure 2).

DISCUSSION

ResFinderFG v2.0 contains 3,913 ARGs which were described with functional metagenomics in 50 publications. Here, we showed that using ResFinderFG v2.0 enabled us to describe the

resistome with ARGs that were not detected by other databases. Notably, ResFinderFG v2.0 permitted a better description of sulfonamide/trimethoprim, glycopeptide/cycloserine resistant genes and beta-lactamase encoding genes.

ResFinderFG v2.0 includes more ARGs coming from human-associated samples[16–26]. For example, characterization of the gut resistome with functional metagenomics showed that its ARGs were not well described in ARG databases[18]. Inclusion of these ARGs is therefore important for future metagenomic characterization of resistomes. Regarding the ARG family concerned, most of the new ARGs included compared to ResFinderFG v1.0 are glycopeptide/cycloserine or beta-lactam resistance genes. Glycopeptide/cycloserine resistance genes were selected using cycloserine, an antibiotic used in the therapy of tuberculosis caused by multi resistant mycobacteria[71]. Beta-lactam resistant genes are of high concern because beta-lactam antibiotics are widely used against priority pathogens[72].

Using ResFinderFG v2.0, sulfonamide/trimethoprim, glycopeptide/cycloserine, beta-lactam, phenicol, cycline, quinolone, macrolide/lincosamide/streptogramin and aminoglycoside resistance genes were evidenced studying three GMGC gene subcatalogs (human gut, soil and aquatic). As expected, regarding their representation in the database, the most frequent unigene hits were glycopeptide/cycloserine and sulfonamide/trimethoprim resistance genes. Analogous analyses performed with other databases showed that ResFinderFG v2.0 detected a comparable or higher number of ARGs depending on the other database used. Beta-lactamase encoding genes were the most represented ARGs in unigene hits shared by all databases. Yet, ResFinderFG v2.0 allowed the detection of beta-lactamase encoding genes which were not detected with other databases. It was expected since many publications using functional metagenomics reported beta-lactamase encoding genes distant from the ones described in ARG databases[15,18,35,38,50,58,63] and a distant one has been evidenced recently from soil samples[43]. Other antibiotic families were even more specifically associated with ResFinderFG v2.0, such as sulfonamide/trimethoprim, phenicol and glycopeptide/cycloserine resistance genes.

Our study has limitations, though. The strength of functional metagenomics is its ability to identify ARGs without relying on their DNA sequence and to provide phenotypical data associated with ARGs. This makes it possible to identify new resistance mechanisms from

new families of ARGs. However, to properly identify the latter and avoid false positives, substantial laboratory work is required and is not done for most inserts in functional metagenomics experiments. In order to characterise an ARG, further cloning is required to study which part of the insert is important for the observation of the phenotype and also to what extent this may modify the associated MICs. Consequently, ResFinderFG v2.0 includes ARGs belonging to known families as opposed to truly novel ARGs from new families which would require further experiments to be identified. We are also aware that some genes included in ResFinderFG v2.0 require overexpression to confer resistance to an antibiotic. This is the case for example for D-alanine-D-alanine ligase or dihydrofolate reductase encoding genes which confer resistance to cycloserine or trimethoprim respectively when they are overexpressed. Yet, these genes are also described in other ARG databases and they fit the operational definition of an ARG[4]. Finally, since the original accession numbers are available in each ARG sequence header, researchers can easily obtain the complete insert DNA sequence to investigate.

DATA AVAILABILITY

All the computational steps and data used in the construction of the ResFinderFG v2.0 database and the database itself are available on the following public repository: https://zenodo.org/badge/latestdoi/470536952. The database was also deposited on the Center of Genomic Epidemiology (CGE) web server, where it can be used online https://cge.food.dtu.dk/services/ResFinderFG/. Supplementary figure 1 describes how to use the web server. Analysis processes for the description of ResFinderFG v2.0 are accessible on the following public repository: https://zenodo.org/badge/latestdoi/507027650.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

AUTHOR CONTRIBUTIONS

RG made the database update, the comparison analysis of ABRicate results with other databases and wrote the manuscript. SUP made the initial ABRicate analysis on GMGC subcatalogs and reviewed the manuscript. JL and MP helped building the figures and the scripts for database construction and analysis. MW made the update available on the CGE

webserver. LPC reviewed the manuscript. ER designed the study and reviewed the manuscript.

ACKNOWLEDGEMENTS

The authors are grateful to Frank Møller Aarestrup for hosting ResFinderFG v2.0 on the Center for Genomic Epidemiology website, and to Andrew Bielski for English editing. Graphical abstract was created using BioRender.

FUNDING

This work was funded by the Joint Program Initiative for Antimicrobial Resistance (JPIAMR) EMBARK (Establishing a Monitoring Baseline for Antimicrobial Resistance in Key environments) project (International Development Research Centre, IDRC, grant 109304-001 to LPC, Agence Nationale de la Recherche, ANR, grant ANR-19-JAMR-0004 to ER).

CONFLICT OF INTEREST

All authors: none

REFERENCES

1. WHO. Antimicrobial resistance. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.
2. Coelho, L. P. *et al.* Towards the biogeography of prokaryotic genes. *Nature* **601**, 252–256 (2022).
3. Mackenzie, J. S. & Jeggo, M. The One Health Approach—Why Is It So Important? *Trop. Med. Infect. Dis.* **4**, 88 (2019).
4. Martínez, J. L., Coque, T. M. & Baquero, F. What is a resistance gene? Ranking risk in resistomes. *Nat. Rev. Microbiol.* **13**, 116–123 (2015).
5. Forsberg, K. J. *et al.* The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
6. Waseem, H. *et al.* Contributions and Challenges of High Throughput qPCR for Determining Antimicrobial Resistance in the Environment: A Critical Review. *Molecules* **24**, 163 (2019).

7. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

8. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29-37 (2011).

9. Boolchandani, M., D'Souza, A. W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat. Rev. Genet.* **20**, 356–370 (2019).

10. Alcock, B. P. *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, D517–D525 (2020).

11. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).

12. Ruppé, E. *et al.* Prediction of the intestinal resistome by a three-dimensional structure-based method. *Nat. Microbiol.* **4**, 112–123 (2019).

13. Dos Santos, D. F. K., Istvan, P., Quirino, B. F. & Kruger, R. H. Functional Metagenomics as a Tool for Identification of New Antibiotic Resistance Genes from Natural Environments. *Microb. Ecol.* **73**, 479–491 (2017).

14. Moore, A. M., Munck, C., Sommer, M. O. A. & Dantas, G. Functional metagenomic investigations of the human intestinal microbiota. *Front. Microbiol.* **2**, 188 (2011).

15. Forsberg, K. J. *et al.* Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612–616 (2014).

16. Gibson, M. K. *et al.* Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol.* **1**, 16024 (2016).

17. Moore, A. M. *et al.* Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PloS One* **8**, e78822 (2013).

18. Sommer, M. O. A., Dantas, G. & Church, G. M. Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science* **325**, 1128–1131 (2009).

19. Pehrsson, E. C. *et al.* Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).

20. Reynolds, L. J., Anjum, M. F. & Roberts, A. P. Detection of a Novel, and Likely Ancestral, Tn916-Like Element from a Human Saliva Metagenomic Library. *Genes* **11**, E548 (2020).

21. Kintses, B. *et al.* Phylogenetic barriers to horizontal transfer of antimicrobial peptide resistance genes in the human gut microbiota. *Nat. Microbiol.* **4**, 447–458 (2019).

22. Clemente, J. C. *et al.* The microbiome of uncontacted Amerindians. *Sci. Adv.* **1**, e1500183 (2015).

23. Card, R. M. *et al.* Application of microarray and functional-based screening methods for the detection of antimicrobial resistance genes in the microbiomes of healthy humans. *PloS One* **9**, e86428 (2014).

24. Cheng, G. *et al.* Functional screening of antibiotic resistance genes from human gut microbiota reveals a novel gene fusion. *FEMS Microbiol. Lett.* **336**, 11–16 (2012).

25. Moore, A. M. *et al.* Gut resistome development in healthy twin pairs in the first year of life. *Microbiome* **3**, 27 (2015).

26. Campbell, T. P. *et al.* The microbiome and resistome of chimpanzees, gorillas, and humans across host lifestyle and geography. *ISME J.* **14**, 1584–1599 (2020).

27. Flórez, A. B., Vázquez, L. & Mayo, B. A Functional Metagenomic Analysis of Tetracycline Resistance in Cheese Bacteria. *Front. Microbiol.* **8**, 907 (2017).

28. Versluis, D. *et al.* Sponge Microbiota Are a Reservoir of Functional Antibiotic Resistance Genes. *Front. Microbiol.* **7**, 1848 (2016).

29. Wichmann, F., Udikovic-Kolic, N., Andrew, S. & Handelsman, J. Diverse Antibiotic Resistance Genes in Dairy Cow Manure. *mBio* **5**, (2014).

30. Tian, B., Fadhil, N. H., Powell, J. E., Kwong, W. K. & Moran, N. A. Long-term exposure to antibiotics has caused accumulation of resistance determinants in the gut microbiota of honeybees. *mBio* **3**, (2012).

31. Martiny, A. C., Martiny, J. B. H., Weihe, C., Field, A. & Ellis, J. C. Functional Metagenomics Reveals Previously Unrecognized Diversity of Antibiotic Resistance Genes in Gulls. *Front. Microbiol.* **2**, (2011).

32. Allen, H. K. *et al.* Resident microbiota of the gypsy moth midgut harbors antibiotic resistance determinants. *DNA Cell Biol.* **28**, 109–117 (2009).

33. Wang, X.-R. *et al.* Duck wastes as a potential reservoir of novel antibiotic resistance genes. *Sci. Total Environ.* **771**, 144828 (2021).

34. McGivern, B. B., McDonell, R. K., Morris, S. K., LaPara, T. M. & Donato, J. J. Novel class 1 integron harboring antibiotic resistance genes in wastewater-derived bacteria as revealed by functional metagenomics. *Plasmid* **114**, 102563 (2021).

35. Marathe, N. P. *et al.* Sewage effluent from an Indian hospital harbors novel carbapenemases and integron-borne antibiotic resistance genes. *Microbiome* **7**, 97 (2019).

36. Cameron, A. *et al.* Functional screening for triclosan resistance in a wastewater metagenome and isolates of Escherichia coli and Enterococcus spp. from a large Canadian healthcare region. *PloS One* **14**, e0211144 (2019).

37. Zhang, L. *et al.* Novel clinically relevant antibiotic resistance genes associated with sewage sludge and industrial waste streams revealed by functional metagenomic screening. *Environ. Int.* **132**, 105120 (2019).

38. Marathe, N. P. *et al.* Functional metagenomics reveals a novel carbapenem-hydrolyzing mobile beta-lactamase from Indian river sediments contaminated with antibiotic production waste. *Environ. Int.* **112**, 279–286 (2018).

39. González-Plaza, J. J. *et al.* Functional Repertoire of Antibiotic Resistance Genes in Antibiotic Manufacturing Effluents and Receiving Freshwater Sediments. *Front. Microbiol.* **8**, (2018).

40. Munck, C. *et al.* Limited dissemination of the wastewater treatment plant core resistome. *Nat. Commun.* **6**, 8452 (2015).

41. Uyaguari, M. I., Fichot, E. B., Scott, G. I. & Norman, R. S. Characterization and quantitation of a novel β-lactamase gene found in a wastewater treatment facility and the surrounding coastal ecosystem. *Appl. Environ. Microbiol.* **77**, 8226–8233 (2011).

42. Willms, I. M. *et al.* Discovery of Novel Antibiotic Resistance Determinants in Forest and Grassland Soil Metagenomes. *Front. Microbiol.* **10**, 460 (2019).

43. Willms, I. M. *et al.* Novel Soil-Derived Beta-Lactam, Chloramphenicol, Fosfomycin and Trimethoprim Resistance Genes Revealed by Functional Metagenomics. *Antibiot. Basel Switz.* **10**, 378 (2021).

44. Obermeier, M. M. *et al.* Plant resistome profiling in evolutionary old bog vegetation provides new clues to understand emergence of multi-resistance. *ISME J.* **15**, 921–937 (2021).

45. Böhm, M.-E., Razavi, M., Flach, C.-F. & Larsson, D. G. J. A Novel, Integron-Regulated, Class C β-Lactamase. *Antibiot. Basel Switz.* **9**, E123 (2020).

46. Park, K. S. *et al.* The novel metallo-β-lactamase PNGM-1 from a deep-sea sediment metagenome: crystallization and X-ray crystallographic analysis. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **74**, 644–649 (2018).

47. Wang, S. *et al.* Tetracycline Resistance Genes Identified from Distinct Soil Environments in China by Functional Metagenomics. *Front. Microbiol.* **8**, 1406 (2017).

48. Lau, C. H.-F., van Engelen, K., Gordon, S., Renaud, J. & Topp, E. Novel Antibiotic Resistance Determinants from Agricultural Soil Exposed to Antibiotics Widely Used in Human Medicine and Animal Farming. *Appl. Environ. Microbiol.* **83**, e00989-17 (2017).

49. Pawlowski, A. C. *et al.* A diverse intrinsic antibiotic resistome from a cave bacterium. *Nat. Commun.* **7**, 13803 (2016).

50. Gudeta, D. D. *et al.* Expanding the Repertoire of Carbapenem-Hydrolyzing Metallo-ß-Lactamases by Functional Metagenomic Analysis of Soil Microbiota. *Front. Microbiol.* **7**, 1985 (2016).

51. Im, H., Kim, K. M., Lee, S.-H. & Ryu, C.-M. Functional Metagenome Mining of Soil for a Novel Gentamicin Resistance Gene. *J. Microbiol. Biotechnol.* **26**, 521–529 (2016).

52. Hatosy, S. M. & Martiny, A. C. The ocean as a global reservoir of antibiotic resistance genes. *Appl. Environ. Microbiol.* **81**, 7593–7599 (2015).

53. Perron, G. G. *et al.* Functional Characterization of Bacteria Isolated from Ancient Arctic Soil Exposes Diverse Resistance Mechanisms to Modern Antibiotics. *PLoS ONE* **10**, (2015).

54. Udikovic-Kolic, N., Wichmann, F., Broderick, N. A. & Handelsman, J. Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15202–15207 (2014).

55. Su, J. Q., Wei, B., Xu, C. Y., Qiao, M. & Zhu, Y. G. Functional metagenomic characterization of antibiotic resistance genes in agricultural soils from China. *Environ. Int.* **65**, 9–15 (2014).

56. Cheng, G. *et al.* Identification of a novel fosfomycin-resistant UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) from a soil metagenome. *Biotechnol. Lett.* **35**, 273 (2013).

57. López-Pérez, M., Mirete, S., Jardón-Valadez, E. & González-Pastor, J. E. Identification and modeling of a novel chloramphenicol resistance protein detected by functional metagenomics in a wetland of Lerma, Mexico. *Int. Microbiol. Off. J. Span. Soc. Microbiol.* **16**, 103–111 (2013).

58. Vercammen, K. *et al.* Identification of a metagenomic gene cluster containing a new class A beta-lactamase and toxin-antitoxin systems. *MicrobiologyOpen* **2**, 674–683 (2013).

59. McGarvey, K. M., Queitsch, K. & Fields, S. Wide variation in antibiotic resistance proteins identified by functional metagenomic screening of a soil DNA library. *Appl. Environ. Microbiol.* **78**, 1708–1714 (2012).

60. Torres-Cortés, G. *et al.* Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. *Environ. Microbiol.* **13**, 1101–1114 (2011).

61. Lang, K. S. *et al.* Novel florfenicol and chloramphenicol resistance gene discovered in Alaskan soil by using functional metagenomics. *Appl. Environ. Microbiol.* **76**, 5321–5326 (2010).

62. Donato, J. J. *et al.* Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. *Appl. Environ. Microbiol.* **76**, 4396–4401 (2010).

63. Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A. & Handelsman, J. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J.* **3**, 243–251 (2009).

64. Riesenfeld, C. S., Goodman, R. M. & Handelsman, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* **6**, 981–989 (2004).

65. Munk, P. *et al.* Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nat. Microbiol.* **3**, 898–908 (2018).

66. Wallace, J. C., Port, J. A., Smith, M. N. & Faustman, E. M. FARME DB: a functional antibiotic resistance element database. *Database J. Biol. Databases Curation* **2017**, (2017).

67. Lal Gupta, C., Kumar Tiwari, R. & Cytryn, E. Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes. *Environ. Int.* **138**, 105667 (2020).

68. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinforma. Oxf. Engl.* **22**, 1658–1659 (2006).

69. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* **30**, 2068–2069 (2014).

70. Seemann, T. Abricate.

71. Di Perri, G. & Bonora, S. Which agents should we use for the treatment of multidrug-resistant Mycobacterium tuberculosis? *J. Antimicrob. Chemother.* **54**, 593–602 (2004).

72. WHO. Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report: 2021. https://www.who.int/publications-detail-redirect/9789240027336.

TABLE AND FIGURES LEGENDS

Figure 1: ResFinderFG v.2.0 construction workflow. a. The insert sequences filtering step consists of concatenating all the insert sequences and selecting only those containing an ARG annotation. b. ARG annotations filter construction was done by annotating the ResFinder v4.0 database using PROKKA to have a pool of ARG annotations. c. Metadata was retrieved using GenBank files to get the origin of the samples and the antibiotic used for

selection of each insert sequence. d. Specific curation was done to discard false positives and include only ARGs.

Figure 2: a. Number of ARGs in the ResFinderFG v.1.0 and v.2.0 databases depending on a. the antibiotic families involved; b. the sample sources.

Figure 3. Proportion of antibiotic family unigene hits obtained analyzing GMGC human gut subcatalog using ResFinderFG v2.0 and number of unigene hits obtained analyzing GMGC human gut subcatalog using several databases (ResFinder v4.0, NCBI v3.6, ARG-ANNOT v5, ResFinderFG v2.0 and CARD v3.0.8) annotated by their antibiotic family. Others: bicyclomycin, beta-lactams, bleomycin, disinfectant and antiseptic agents, fosfomycin, fusidic acid, multidrug, mupirocin, nitroimidazole, nucleoside, peptide, rifampicin, streptothricin.