bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

[1] Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME, F-75018 Paris, France

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

*Corresponding author:

Rémi GSCHWIND, PhD

INSERM UMR1137 IAME

Faculté de Médecine Bichat

16 rue Henri Huchard

75108 Paris, France

+33(0)658543545

remi.gschwind@inserm.fr

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. However, ARGs composing the resistome of the human gut microbiota or the environment 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 database second version, ResFinderFG v2.0. Functional metagenomics studies were analyzed and DNA sequences described were retrieved, deduplicated and annotated. Sequences were curated to include only ARG sequences. ResFinderFG v2.0 was then compared to other databases for their relative sensitivity in searches for ARGs in subcatalogs from the Global Microbial Gene Catalog (GMGC). Fifty publications were considered, for a total of 23'764 ARGs identified from different environments. After deduplication, annotation and curation, 3'913 ARGs were included. New ARGs included are mainly glycopeptides/cycloserine or beta-lactams resistance genes identified mostly in human-associated samples. Results of GMGC gene subcatalogs annotation showed that ResFinderFG v2.0 detected comparable or higher ARG numbers than those detected with other databases. Most of the unigene hits obtained were database-specific and ResFinderFG v2.0 specific unigene hits included among others: glycopeptides/cycloserine, sulofnamides/trimethoprim resistance genes and beta-lactamases encoding genes. 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 ARG represents a direct risk for human health[4], genes are able to travel from one environment to another by strain dissemination or horizontal gene transfer[5]. In this way, some ARGs represent a risk as they may be transferred to pathogenic bacteria.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 and, thanks to sequences comparison with specific databases, allows ARG identification in an environment or host. Several AMR public reference databases[7] exist such as CARD[8] or ResFinder[9]. However, since the detection is matching newly obtained sequences to ARG sequences in databases, only sequences that are similar to previously-described ones will be detected with an acceptable degree of confidence. Therefore, unknown ARGs, or ARGs sharing a low identity with ARGs included in the chosen database may not be detected.

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. In order to detect new ARGs or low sequence similarity percentage ARGs, functional metagenomics has been used[10]. This technique is based on phenotypic detection by expressing exogenous DNA in an antibiotic-susceptible host. Using functional metagenomics, ARGs sharing low amino acid identity to their closest homologue in NCBI[11], or even not previously classified as ARGs[12], could be detected in human[12–22], animal[22–29], wastewater[30–37] and other environmental samples[5,11,38–60]. 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[61] and FARME DB[62]. 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 cured to include only ARGs sequences[63]. 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.

Methods

Construction of ResFinderFG v.2.0

To retrieve publications using functional metagenomics for the identification of antibiotic resistance genes, the 4 publications used to construct ResFinderFG v1.0 were first considered. Then, all the publications which were cited by these 4 publications and all the publications that cited one of these publications were collected. In addition, publications found with the following

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 to check whether functional metagenomics was actually used to study antibiotic resistance and whether insert sequences described were available. Database construction and curation was then performed as follows (Figure 1). Accession numbers describing insert DNA sequences functionally selected using antibiotics were included and DNA sequences were retrieved using Batch Entrez. CD-HIT[64] was used to remove redundant DNA sequences and annotation of the remaining was done using PROKKA v.1.14[65]. 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. Accession number of the remaining inserts were used to retrieve information on the insert DNA sequences, such as the origin of the sample and the antibiotic used for selection. Additional filtering steps were added to check the antibiotic used for selection and ARG annotation link, minimum gene size with at list the median amino acid (aa) size of antibiotic resistance determinant (ARD) from the same ARD family (260 aa for beta-lactamase, 378 for tetracycline efflux genes, 641 aa for tetracycline resistance ribosomal protection genes, 178 aa for chloramphenicol acetyltransferase, 247 aa for methyltransferase genes and 158 aa for dihydrofolate reductase genes) and the presence of one unique annotation corresponding to an ARG on the insert DNA sequence. The database also includes metadata (habitat categorization, antibiotic used for selection, ARG family) and ARO annotation for each gene for comparability with other databases using ARO ontologies[8].

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

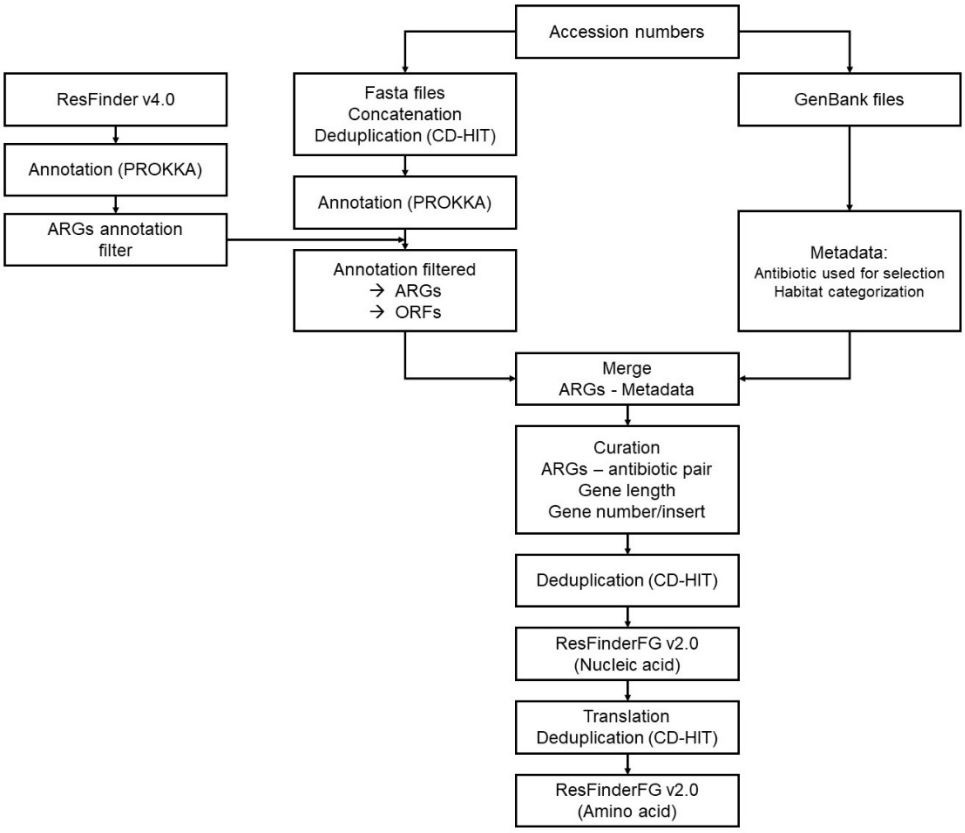

Figure 1: ResFinderFG v.2.0 construction workflow.

Description of ResFinderFG v2.0

To assess the update of the ResFinderFG v2.0 database, the database was first compared in terms of number of ARGs, ARG families and sample sources with ResFinderFG v1.0. 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, plants-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[66] was run using default parameters with different databases (ResFinderFG v2.0, ResFinder v4.0, CARD v3.0.8, ARG-ANNOT v5, NCBI v3.6).

Data and code availability

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 GitHub repository: https://github.com/RemiGSC/ResFinder_FG_Construction. The database was also deposited on the Center of Genomic Epidemiology (CGE) server, where it can be used online https://cge.food.dtu.dk/services/ResFinderFG/. Analysis processes for the description of ResFinderFG v2.0 are accessible on the following public GitHub repository: https://github.com/RemiGSC/ResFinder_FG_Analysis.

Results

Construction of ResFinderFG v2.0

A total of 50 publications using functional metagenomics to analyze ARG content were selected, resulting in 23’776 accession numbers. 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 matched with an ARG annotation of ResFinder v4.0 (228 unique ARG annotations). Another 1’165 ORFs were removed because of a discordance between the annotation and the antibiotic used for selection in the functional metagenomics experiment, 1’064 for an unexpected size relative to the ARG family, and 398 because more than one putative ARG was present in the insert. A second round of CD-HIT was used to avoid redundancy (100% sequence identity) in the ARG sequences and 3’913 ARGs remained and form the database.

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 glycopeptides/cycloserine (+906 genes) and beta-lactams (+333 genes) resistance genes. The glycopeptides/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).

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

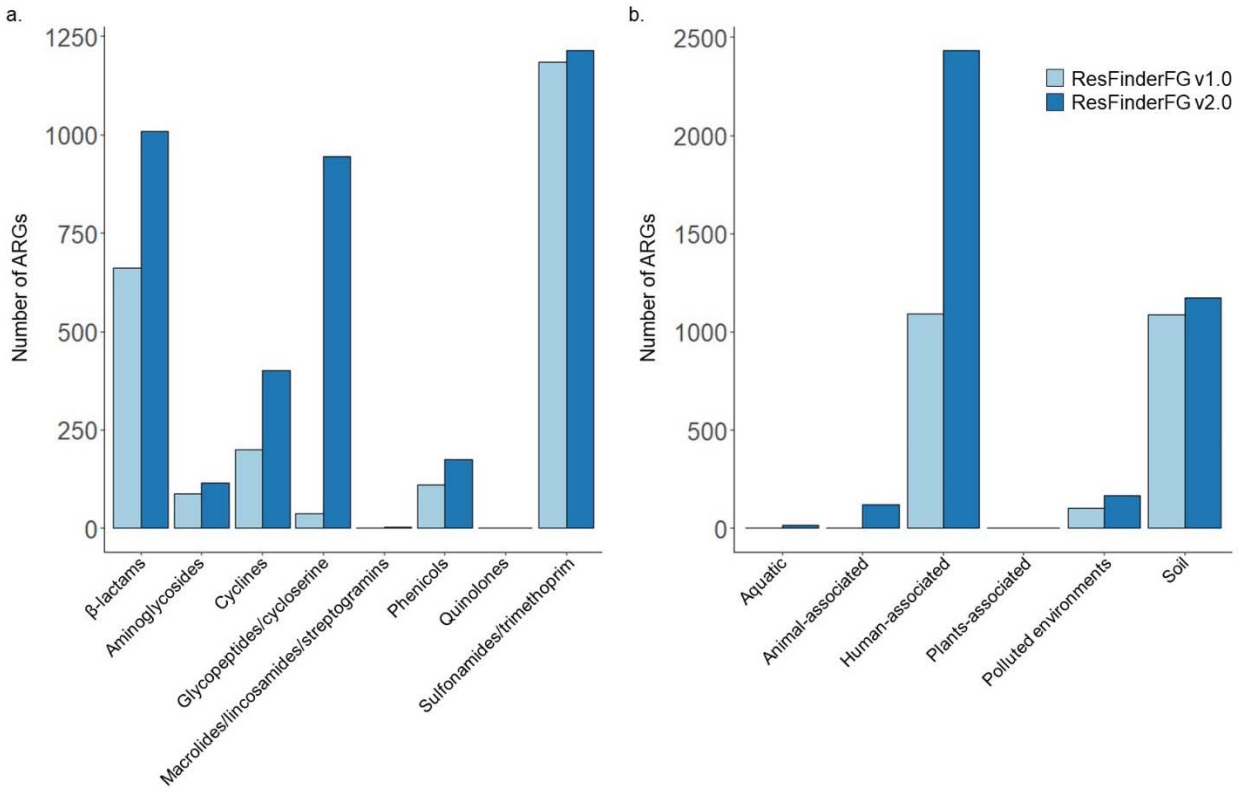

Figure 2: a. Number of ARGs in the ResFinderFG v.1.0 and v.2.0 databases depending on a. the antibiotic families involved; b. the sample sources.

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 1a.) and aquatic (marine and freshwater) subcatalogs (Supplementary Figure 1b.). Using ResFinderFG v2.0, 3'025, 211 and 129 unigene hits were obtained analyzing human gut, soil and aquatic subcatalogs respectively. The 3 most frequently detected ARG families in all gene catalogs were glycopeptides/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-lactamases 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 macrolides/ lincosamides/streptogramins 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 2).

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

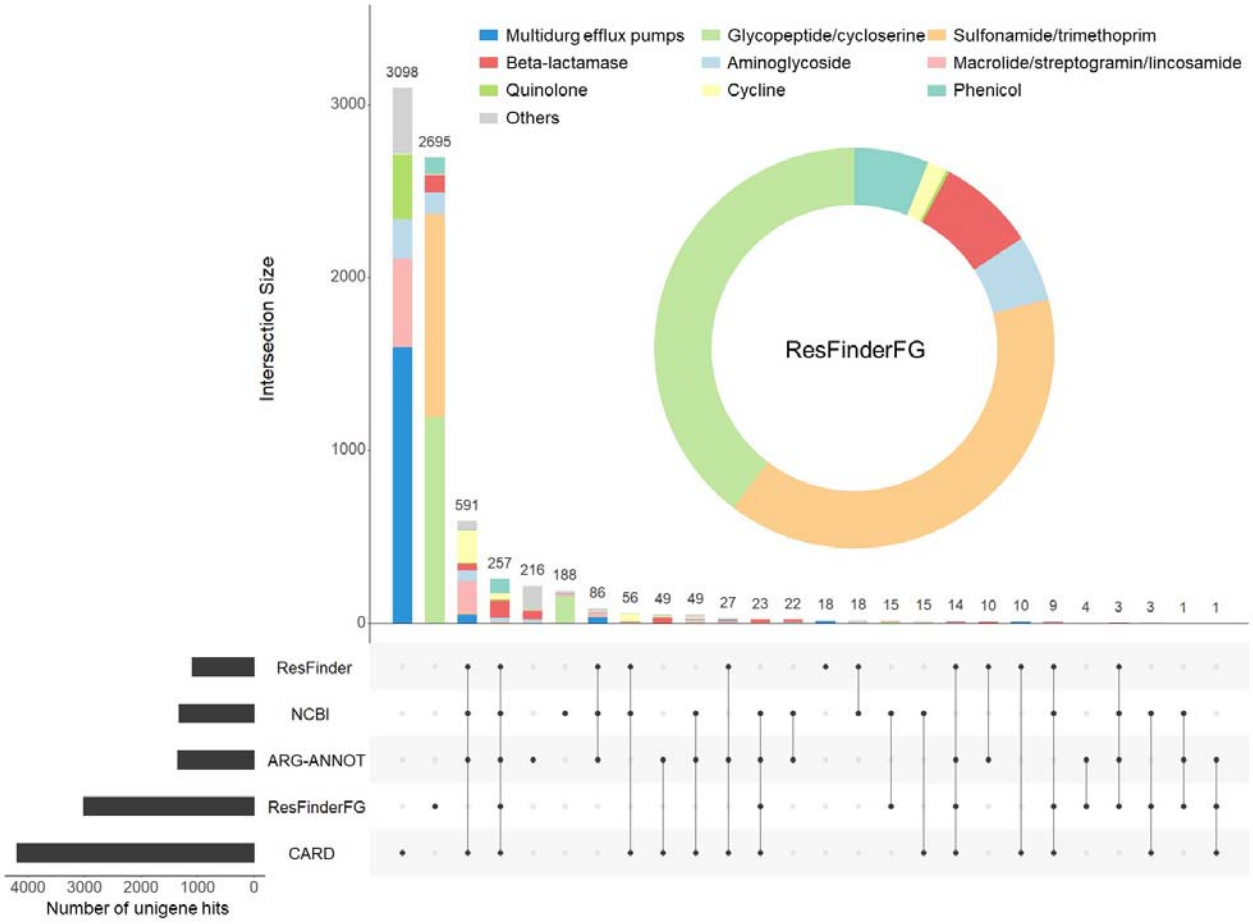

Figure 3: Number of unigene hits obtained analyzing GMGC human gut subcatalog using several databases (ResFinder v4.0, NCBI v3.6, ARG-ANNOT v5, ResFinderFG v2.0 and CARD v3.0.8) annotated by their antibiotic family. Others: bicyclomycin, beta-lactams, bleomycin, disinfectant and antiseptic agents, fosfomycin, 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, glycopeptides/cycloserine resistance gene

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 glycopeptides/cycloserine resistance genes (D-alanine-D-alanine ligase ; supplementary table 3) 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 3).

Between 2.6 and 4.2% of all unigene hits depending on the gene subcatalog analyzed were shared by all the databases used. Beta-lactamases – 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 1).

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 sulfonamides/trimethoprim, glycopeptides/cycloserine resistant genes and beta-lactamase encoding genes.

Exhaustive description of ARG content in the environment can be complicated since most ARG databases are biased towards ARGs coming from culturable and/or pathogenic bacteria. One way to detect genes that are not described in such databases, or that are too different from described genes, is to use functional metagenomics: a laborious and low throughput method that was used by only a few research groups[10] which allows phenotypic identification rather than sequence-based identification of ARGs. Yet, most of the ARGs characterized using functional metagenomics were not deposited in ARG databases until the creation of ResFinderFG v1.0 in

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2016[61]. Since then, the database has not been updated but another database called FARME DB was made including data coming from 30 publications[62]. Nevertheless, it contains all the inserts sequences selected with functional metagenomics and therefore it also contains genes that are not ARGs[63]. Therefore, we updated ResFinderFG v1.0 by including more publications using functional metagenomics to characterize ARGs and we made a curation effort to ensure that the sequences described are the unique ARGs responsible for the resistance phenotype in the initial insert sequence.

ResFinderFG v2.0 includes more ARGs coming from human-associated samples[12–22]. For example, characterization of the gut resistome with functional metagenomics showed that its ARGs were not well described in ARG databases[14]. 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 glycopeptides/cycloserine or beta-lactams resistance genes. Glycopeptides/cycloserine resistance genes were selected using cycloserine, an antibiotic used in the therapy of tuberculosis caused by multi resistant mycobacteria[67]. Beta-lactams resistant genes are of high concern because beta-lactams antibiotics are widely used against priority pathogens[68].

Using ResFinderFG v2.0, sulfonamides/trimethoprim, glycopeptides/cycloserine, beta-lactams, phenicols, cyclines, quinolones, macrolides/lincosamides/streptogramins and aminoglycosides 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, sulfonamides/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 a significant proportion of beta-lactamases 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[11,14,31,34,46,54,59] and a distant one has been evidenced recently from soil samples[39]. Other antibiotic families were even more specifically associated with ResFinderFG v2.0, such as sulfonamides/trimethoprim, phenicols, glycopeptides/cycloserine resistance genes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our study has limitations, however. To ensure that genes included are true ARGs, we selected only insert which had part of their sequence annotated as an ARG by PROKKA. Thus, ARGs not identified by PROKKA may have been missed. Moreover, we did not recheck whether described sequences were actually conferring resistance *in vitro*. Only the sequence corresponding to the ARG annotation was included and we were not able to determine if the surrounding insert DNA sequence was required to produce the resistant phenotype. Yet, since the original accession numbers are available in each ResFinderFG v2.0 ARG sequence header, researchers can easily obtain the complete insert DNA sequence to investigate.

Conclusion

ResFinderFG v2.0 is the new version of the ResFinderFG v1.0 database and includes 1’631 additional ARGs. This makes possible the detection of ARGs which would not be identified using other currently used databases. Nevertheless, other databases also contain ARGs that are absent in ResFinderFG v2.0. Therefore, to make an exhaustive description of the resistome of a sample, ResFinderFG v2.0 should be used alongside other databases.

Acknowledgements

The authors are grateful to Frank Møller Aarestrup and Maja Weiss for hosting ResFinderFG v2.0 on the Center for Genomic Epidemiology website, and to Andrew Bielski for English editing.

Conflicts of interest

All authors: none

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).

References

1. WHO. Antimicrobial resistance. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

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

10. 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).

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

14. 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).

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

16. 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).

17. 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).

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

19. 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).

20. 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).

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

22. 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).

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

25. Wichmann, F., Udikovic-Kolic, N., Andrew, S. & Handelsman, J. Diverse antibiotic resistance genes in dairy cow manure. *mBio* **5**, e01017 (2014).

26. 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).

27. 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).

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

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

30. 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).

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

32. 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).

33. 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).

34. 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).

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

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

37. 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).

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

39. 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).

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. 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).

41. 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).

42. 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).

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

44. 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**, (2017).

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

46. 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).

47. 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).

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

49. 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).

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

50. 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).

51. 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).

52. 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).

53. 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).

54. 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).

55. 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).

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

57. 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).

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58. 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).

59. 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).

60. 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).

61. 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).

62. 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).

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

64. 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).

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

66. Seemann, T. Abricate.

67. 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).

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512667; this version posted October 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

420