VRprofile2: detection of antibiotic resistance-associated mobilome in bacterial pathogens

Meng Wang1, Ying-Xian Goh1, Cui Tai1, Hui Wang2, Zixin Deng1 and Hong-Yu Ou1,∗

1State Key Laboratory of Microbial Metabolism, Joint International Laboratory on Metabolic & Developmental Sciences, School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, China and 2State Key Laboratory of Pathogens and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China

Received March 12, 2022; Revised April 10, 2022; Editorial Decision April 19, 2022; Accepted April 21, 2022

ABSTRACT

VRprofile2 is an updated pipeline that rapidly identifies diverse mobile genetic elements in bacterial genome sequences. Compared with the previous version, three major improvements were made. First, the user-friendly visualization could aid users in investigating the antibiotic resistance gene cassettes in conjunction with various mobile elements in the multiple resistance region with mosaic structure. VRprofile2 could compare the predicted mobile elements to the collected known mobile elements with similar architecture. A new mobilome indicator was proposed to give an overall estimation of the mobilome size in individual bacterial genomes. Second, the relationship between antibiotic resistance genes, mobile elements, and host strains would be efficiently examined with the aid of predicted strain’s sequence typing, the incompatibility group and the transferability of plasmids. Finally, the updated back-end database, MobilomeDB2, now collected nearly a thousand active mobile elements retrieved from literature or based on prediction. The pre-computed results of the antibiotic resistance gene-carrying mobile elements of >5500 ESKAPEE genomes were also provided. We expect that VRprofile2 will provide better support for researchers interested in bacterial mobile elements and the dissemination of antibiotic resistance. VRprofile2 is freely available to all users without any login requirement at https://tool2-mml.sjtu.edu.cn/VRprofile.

GRAPHICAL ABSTRACT

INTRODUCTION

The genome sequence of an individual bacterium can be divided into the chromosome backbone, common to the entire species, and a long list of the accessory regions. The latter is termed the ‘mobilome’ (mobile genome) (1,2), which includes a myriad of short IS elements and longer spans of episomal plasmids, transposons, integrons, gene cassettes, prophages, integrative and conjugative elements (ICEs), and genomic islands (GIs). These mobile genetic elements (MGEs) frequently carry antibiotic resistance genes (ARGs). It is a key avenue for the broad distribution of antibiotic resistance in the pathogens with clinical significance, especially for the ESKAPEE bacteria (3). The ESKAPEE group encompasses Gram-positive and Gram-negative species, made up of Enterococcus faecium, Staphy-
lococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, Enterobacter species and Escherichia coli. The ARG-carrying plasmids, ICEs and other large MGEs often display a mosaic structure composed of multiple modules acquired by complex recombination, insertion, and/or reversion events (4,5). For example, a multiple resistance region contains several antibiotic resistance cassettes in conjunction with various ISs, transposons, as well as integrons and their degenerate elements (6). A recent study revealed that the interaction between the conjugative plasmids and IS elements mediated the ARG transfers (7).

There are many publicly available sources for predicting bacterial MGEs, such as the PLSDB database for plasmids (8), PHASTER tool for prophages (9), ICEberg database for ICEs (10), PAIDB database for pathogenicity islands (11), IslandViewer tool for GIs (12), TnCentral database for transposons (13) and ISfinder database for IS elements (14). They focus on the individual types of MGEs. Several prediction pipelines covering diverse MGE types were recently proposed, such as the MobileElementFinder (15) and MGEfinder (16). MobileElementFinder employs simple BLAST searches against the collected known IS, transposon, and ICE datasets. MGEfinder identifies integrative MGEs from short-read sequencing data but requires bioinformatics to operate its local version. In 2017, we developed the open-access web server VRprofile that allows rapid detection of virulence and antibiotic resistance genes and their flanking MGEs (17). It uses the homologous searches with the combination of the hit co-localization approach to detect gene clusters related to MGEs. However, these tools cannot meet the needs of visualizing and comparing the various mobile elements and their flanking ARGs. To preferably predict the transferability of putative MGEs and to investigate the transmission of antibiotic resistance and virulence factors carried by the bacterial MGEs, it is necessary to obtain the whole picture of mobilome and interpret their interactions, especially in multiple resistance regions, which have a widespread distribution in the clinical isolates of ESKAPE.

Here, we report the release of VRprofile version 2.0, which offers three major enhancements: (i) graphical representation of the multiple resistance regions with mosaic structures and comparison to the known MGEs with similar architecture; (ii) to aid in exploring the relationships of antibiotic resistance genes, mobile elements, and host strains; (iii) pre-computed mobilome of >5500 ESKAPE bacterial genomes. We expect that VRprofile2 will provide better support for researchers interested in bacterial diverse mobile elements and the dissemination of antibiotic resistance.

MATERIALS AND METHODS
Detection and visualization of the MGE regions
The VRprofile2 provides a rapid identification pipeline for various MGEs in the chromosomal or plasmid sequences, including the prophage, ICE, island-like region, IS element, integron and SCCmec (staphylococcal chromosomal cassettes mec) (Supplementary Figure S1). Compared to the original version, the update pipeline newly integrated the following modules: the prediction of integron using IntegronFinder (18), the detection of SCCmec using SCCmecFinder (19), and the BLASTp searches of the transposase (TnPA) and resolvase (TnPR) for the transposons retrieved from the TnCentral (13) and TnRegistry (20) databases. The cluster of ISs/transposons (15) and the flanking ARGs were also detected with an interval <5 kb, which is often present in the multiple resistance region with the complex mosaic structure in a resistance plasmid (7,13,21).

As shown in Supplementary Figure S1, VRprofile2 has different preprocessing modules for the different uploaded sequences. (i) For the chromosomal sequence, the strain typing is done using the script mlst (https://github.com/tseemann/mlst) that employs the PubMLST typing schemes (22). (ii) For the plasmid genome sequence, the plasmid incompatibility group is classified using the tool ABRicate (https://github.com/tseemann/abricate) with an inbuilt database of PlasmidFinder (23). The transferability of the plasmid is predicted as conjugative or mobilizable by using the tool oriTfinder (24). (iii) For the assembled contigs of a bacterial genome or the metagenomic contigs assembled from long reads, the tool viralVerify was integrated into the VRprofile2 pipeline to classify contigs as the fragments of chromosome or plasmid (25). Kraken2 was also employed to determine the taxonomic classification of metagenomic contigs (26).

The VRprofile2 output web page dynamically presents the chromosome or plasmid map with the aid of the AngularPlasmid HTML markup (http://angularplasmid.vixis.com/). And then, an expanded view of the identified MGE region was visualized using SVGene (https://github.com/krakenmlst) that employs the PubMLST typing schemes. These interactive presentations allow the users to conveniently check the genomic coordinates and features of the identified MGE regions and their harboring or flanking ARGs.

Estimation of the mobilome size
To provide an overview of the mobilome of individual strains, we proposed a new indicator, mV (mobilome value), which is estimated by dividing the mobilome size by its whole genome size. The mobilome size is the sum of MGE lengths identified by VRprofile2, irrespective of their types. There are two points to be considered in the mV estimation: (i) for two completely overlapping MGE regions, the length of the larger one is counted while that of the smaller one is omitted. (ii) for two partially overlapping MGE regions, the length of the merged region is counted. In addition, another value, mAV, is employed to examine the mobile elements that harbor ARGs or the ISs/transposons flanked by ARGs with an interval <5 kb. It is calculated by dividing the size of the ARG-associated mobilome by its whole genome size.

Mobilome of ESKAPEE bacteria
The VRprofile2 web server provided the pre-calculated results of the mobilome of the sequenced ESKAPEE bacterial
A total of 5,536 completely sequenced genomes of ESKAPE bacteria were retrieved from GenBank in October 2021, including 235 E. faecium, 1002 S. aureus, 1063 K. pneumoniae, 300 A. baumannii, 359 P. aeruginosa, 207 Enterobacter spp. and 2370 E. coli. The pre-calculated results were listed and visualized on the web pages ‘Browse’, ‘Search’ and ‘Statistics’. On the ‘Browse’ web page, the users can fetch the strain information and the details of identified MGEs. On the ‘Search’ web page, by entering an ARG name, the distribution of the queried ARG in the MGEs of host bacteria can be retrieved (Supplementary Figure S2).

A total of 294 461 MGEs of ESKAPE were detected, of which 15 650 carry ARGs (Supplementary Table S1). The sequences of these ARG-carrying MGEs were pairwise aligned using Mash (27) and then grouped using NetworkX with the Mash distances <0.005 (28). The distribution of these obtained ARG-carrying MGE groups among ESKAPE was examined, and 106 groups were widespread among different MLST of the individual species. These groups were thus labeled as ‘active’.

MobilomeDB2, a collection of active ARG-associated MGEs

The back-end database MobilomeDB2 newly added three datasets related to the active MGEs that are prone to transfer ARGs (Supplementary Table S2). The first dataset, Dataset I, consists of 586 transferable MGEs with experimental supports. It contains 165 conjugative plasmids, 180 ICEs and 241 transposons collected from the literature retrieved from PubMed. The second dataset, Dataset II, was taken from the VRprofile2-predicted active ESKAPE’s MGEs as defined above. In addition, the third dataset, Dataset III, which contains the publicly available MGE resources, was also integrated into MobilomeDB (Supplementary Table S2).

RESULTS AND DISCUSSION

Case Study 1: detection and comparison of a multiple resistance region in the resistance plasmid

VRprofile2 provides a flexible and biologist-friendly web interface consisting of an input page and a result page. In the example shown in Figure 1, the genome sequence of an E. coli M505 plasmid pM505-NDM5 was uploaded with the file in FASTA format (GenBank accession number: AP023220.2) (29). The result page listed the prediction of the incompatibility group (IncFII) and transferability (conjugative) of pM505-NDM5 (Figure 1A). The identified MGE regions were listed and could also be downloaded as an individual text file, while the MGE regions carrying ARGs were highlighted in the table (Figure 1B). Notably, the result page also offered the functional tabs for visualization and further analysis of the ARG-associated MGEs (Figure 1C and D). A carbapenemase gene NDM-5 and eight other ARGs were clustered in a 27-kb region (MGE1 region in Figure 1C), carrying four IS26 elements. Interestingly, an integron that contained three resistance gene cassettes (MGE2 regions in Figure 1C) was located within this multiple resistance region.

In addition, the VRprofile2-identified MGE regions can be compared to the known MGEs with similar architecture by using Mash. MobilomeDB2 Database II contains 106 active ARG-associated MGE groups taken from ESKAPE. Using the whole plasmid pM505-NDM5 as the query to search against Database II, it showed that the skeleton of pM505-NDM5 was highly similar to an active ARG-associated MGE group, which consists of 14 plasmids of similar structure but carrying different ARGs in E. coli and K. pneumoniae (Figure 1E). Remarkably, there was no match for the MGE1 region (an IS and ARG cluster) but massive hits for the MGE2 region (an integron with resistance cassettes) (Figure 1F), indicating that the NDM-5-carrying resistance region in pM505-NDM5 exhibited a unique combination, i.e., a conserved integron located within a new cluster of ISs.

Three other examples with the input of E. coli chromosomal genome sequence, K. pneumoniae plasmid genome sequence (30), and the MinION sequencing metagenomic contigs (31) are available in Supplementary Figures S3, S4 and S5, respectively.

Case Study 2: examination of the relationship between ARGs, plasmids and host strains

VRprofile2 can facilitate an in-depth analysis of the relationships between the host strains with diverse MLSTs, different mobile elements, and various ARGs. The ‘Statistics’ web page of the VRprofile2 web server provides the estimated mobilome sizes of >5500 ESKAPE bacterial genomes (Figure 2A). The mv values of these ESKAPE bacteria ranged from 0.01 to 0.36, while the values vary across different species and even within a species.

There are 3525 plasmids in 1063 K. pneumoniae genomes, representing about 30% of the total mobilome in an individual strain (Figure 2B and Supplementary Figure S6). The relationships between ARGs, plasmids, and host strains are graphically shown in Figure 2C and Supplementary Figure...
Figure 1. An overview of VRprofile2 outputs using the *E. coli* M505 plasmid pM505-NDM52 as an example. (A) The predicted incompatibility group and transferability of pM505-NDM5. (B) A table of the identified MGE regions. (C) A circular representation of pM505-NDM52 showing the locations of the identified MGE1 and MGE2 regions (filled with blue) and ARGs (highlighted in red). Four modules related to plasmid conjugation are also marked, including an origin region of transfer (*oriT*), relaxase gene, type IV coupling protein (T4CP) gene and gene cluster for the type IV secretion system (T4SS) apparatus. (D) An expanded view of the MGE1 region contains four IS26 elements, one integron (MGE2 region), and nine ARGs. (E) A table containing the Mash-facilitated sequence comparison of pM505-NDM52 against the MobilomeDB Dataset II, which contained the active ARG-associated MGE groups of ESKAPEE bacteria. (F) A table containing the sequence comparison of the MGE2 region against the Dataset II.

S7 and Table S3. For example, out of all the 3525 plasmids under analysis, 314 plasmids carried 326 *bla*KPC genes. Notably, 190 *bla*KPC genes were harbored in 182 IncFII-related plasmids, including 64.3% (117/182) IncFII plasmids in the ST11 strains and 11.5% (21/182) in the ST258 strains.

CONCLUSION

Here, we reported a major update of VRprofile and focused on ARG-carrying mobilome, especially in the ESKAPEE bacteria. The rapid detection, user-friendly visualization and comparison could aid users in investigating the multiple resistance regions and other mosaic regions. With the prediction of the strain MLST, plasmid incompatibility group, and transferability, VRprofile2 could facilitate examining the relationships between ARGs, MGEs and host strains. The newly proposed indicator *mV* also gives an overall estimation of the mobilome size in an individual bacterial genome. The pre-computed results of the mobilome of >5500 ESKAPEE bacterial genomes are provided. They will be updated and improved regularly from the rapidly growing bacterial genomic data. Ultimately, we propose an updated mobilome detection pipeline to efficiently explore large numbers of mobile elements that play key roles in the rapid dissemination of antibiotic resistance.
Figure 2. VRprofile2-estimated size of the mobilome borne by individual strains of the ESKAPE group. (A) Bubble chart of the estimated mobilome size. The complete genome sequences of 5,536 ESKAPE bacteria were retrieved from GenBank. The abscissa axis represents the $m^V$ value of individual strains, calculated by dividing the mobilome size by its whole genome size. The bubble size represents the magnitude of the individual strains under analysis. (B) The distribution of MGE types includes the GI, ICE, integron (In), IS, prophage, SCCmec and plasmid. (C) The relationship between the strains ($n = 1063$), plasmids ($n = 3525$), and plasmid-carrying ARGs ($n = 8924$) of Klebsiella pneumoniae. The heat map presents the profile of the plasmids belonging to different incompatibility groups (Inc). The number (log value) of all the identified Inc groups in the single- or multiple-replicon plasmids (Supplementary Table S3) is plotted alongside the map diagonal. The number of the Inc pairs in the individual multiple-replicon plasmids is depicted in the cell. The left bar chart indicates the plasmid number and the proportion in different sequence typing (MLST) strains. For the plasmid containing two replicon regions, the plasmid number of each Inc group was counted separately. The right bar chart denotes the plasmid transferability. The top bar chart presents the number of $bla_{KPC}$ and other carbapenemase genes (see Supplementary Figure S7 for details). The down bar chart shows the number of ARGs and virulence genes.
