The Journal of Open Source Software

CPgeneProfiler: A lightweight R package to profile the Carbapenamase genes from genome assemblies

Prakki Sai Rama Sridatta[1,2], Natascha M Thevasagayam[1,2], Weizhen Xu[1,2], Kalisvar Marimuthu[1,2], Jeanette W P Teo[3], Indumathi Venkatachalam[4], and Ng Oon Tek[1,2,5]

1 National Centre for Infectious Diseases, Singapore **2** Tan Tock Seng Hospital, Singapore **3** National University Hospital, Singapore **4** Singapore General Hospital, Singapore **5** Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

DOI: 10.21105/joss.02473

Software

- Review
- Repository
- Archive

Editor: William Rowe

Reviewers:

- @fmaguire
- @boasvdp

Submitted: 06 July 2020
Published: 12 October 2020

Summary

"Carbapenems" are a specific subset of antibiotics considered to possess a higher spectrum of antimicrobial activity (Papp-Wallace, Endimiani, Taracila, & Bonomo, 2011) against Gram-positive and Gram-negative bacteria. Even so, there are pathogens which are resistant to carbapenems due to the presence of carbapenemase genes (CP genes) which have the ability to hydrolyze carbapenems.

Studies show that those cases infected by carbapenem-resistant pathogens have a higher morbidity and mortality rate compared with those who are infected by non-carbapenem-resistant pathogens (Cai et al., 2017; Duin, Kaye, Neuner, & Bonomo, 2013). Therefore, early discerning of the CP genes and their resistance mechanisms are considered crucial to aid in infection control as well as lessen the likelihood of mortality, duration of hospitalization stay, and related medical costs (Duin et al., 2013; Nordmann & Poirel, 2019). Further, it is understood that the cocarriage of genes encoding different classes of carbapenemases could confer higher resistance to carbapenem antibiotics, which may promote further spread of the disease (Wang et al., 2019).

The detection of the resistance genes from various bacterial strains using techniques such as polymerase chain reaction (PCR) and microarrays in real-time are very time consuming and costly. With the advancement in whole-genome sequencing (WGS) technologies the costs are more accessible and WGS provides an alternative method for detection of resistance genes, given that the relevant analysis tools are available.

To this end, several freely available bioinformatics tools such as `ABRicate` (https://github.com/tseemann/abricate), `AMRPlusPlus` (Doster et al., 2020), `ARG-ANNOT` (Gupta et al., 2014), `ARIBA` (Hunt et al., 2017), `Comprehensive Antibiotic Resistance Database - Resistance Gene Identifier` (CARD-RGI) (Alcock et al., 2020), `NCBI AMRFinderPlus` (https://ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/), `KmerResistance` (Clausen, Aarestrup, & Lund, 2018; Clausen, Zankari, Aarestrup, & Lund, 2016), `PointFinder` (Zankari et al., 2017), `Resfinder` (Bortolaia et al., 2020), `sraX` (Panunzi, 2020) and `SRST2` (Inouye et al., 2014) assist in finding of the antimicrobial resistance genes from the sequence data (Hendriksen et al., 2019).

Statement of Need

Undeniably, all the above-mentioned tools are focused around the antimicrobial-resistant genes and tools such as ABRicate and CARD-RGI can even generate comparative tables across

genomes and sraX can help in visualization of comprehensive AMR gene complement. Nevertheless, they do not readily generate a genetic profile for the presence of CP genes, extract and visualize the set intersections of cocarriage of CP genes. Achieving this currently necessitates a restructuring and transformation of the output from these tools. Furthermore, in the research settings where it is crucial to quickly examine the transmission of CP genes, it is useful to have a tool that is catered to CP gene dataset that provides easily interpretable visualizations and statistics. Therefore, to address this need, we describe here a lightweight R package, **CPgeneProfiler** that scans multiple bacterial genome assemblies to detect and visualize the presence of CP genes and their cocarriage using the R framework. Additionally, this package also allows to assess the size of CP contigs to check if the CP genes are distributed on the particular sequence size by generating the contig length distribution plots.

Implementation

In order to detect CP genes from the genome assemblies, `NCBI Bacterial Antimicrobial Resistance Reference Gene Database` (2020-07-16.2) (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047) was used for generating a CP gene database. Only those resistant genes whose subclass is categorized as "CARBAPENEM" in the reference gene catalog were considered for database preparation. This excluded the possibility of having resistant gene variants which are beta-lactamases but do not show carbapenem-resistant activity. For example, although OXA-48 is a carbapenemase gene, other OXA variants such as OXA-163 and OXA-405 have been concluded to be devoid of any carbapenemase activity (Dortet & Naas, 2017) and therefore their subclass is not categorized as "CARBAPENEM" in the NCBI Bacterial Antimicrobial Resistance Reference Gene Database. Therefore, both OXA variants OXA-163, OXA-405 were not included in the CP gene database.

The tool first uses the `cpblast` command, by which each fasta file is searched against the CP gene database using NCBI BLAST+ (Camacho et al., 2009) (version 2.9.0+) which is pre-installed in the local system as a dependency. The presence of a CP gene in an assembled genome is confirmed if the CP gene meets the identity and coverage thresholds (default: 100%) when aligned with the genome sequence. The genome sequences which meet the thresholds are extracted from the BLAST results using the `filt_blast` command.

Visualizing the presence of CP genes and its corresponding counts across all the genome assemblies in a simple heatmap enables one to find CP gene variants that are found across the samples and aids in exploring the pattern of CP gene presence with reference to species or sequence type (ST). In order to facilitate this, the `cpprofile` command generates a profile of CP genes (Figure 1A) across the genome assemblies, while the `cocarriage` command finds cocarriage of CP genes in the genome assemblies. In addition to this, the tool also generates plots to visualize set intersections of CP genes across all the input genome assemblies using the command `upsetR_plot` (Figure 1B). It is understood that isolates which harbour multiple carbapenemase genes are considered to produce high resistant phenotypes and running the commands `cocarriage` and `upsetR_plot` provide an overview of the CP genes as well as their cocarriages present in all the genomes.

Given a set of bacterial genomes that are of same species, it would be useful to explore if the CP genes are found on specific plasmids or scattered across multiple plasmids/chromosomes of different sequence lengths. This can be achieved by plotting the number of contigs across the contig length by using the `plot_conlen` command (Figure 2).

Lastly, `CPgeneProfiler` can also generate the N50, N90, Assembly size statistics for each of the genome assemblies and also plots the assembly size against N50 and N90 using the `assembly_stat` command (Figure 3A, 3B). This would help in quickly assessing and comparing the quality of the assembled genomes provided as an input. All the generated output files from various commands of the package are arranged accordingly into respective folders using the `cp_summarize` command.

Availability and Implementation

The R package `CPgeneProfiler` (version 2.1.1) is supported on UNIX/Linux machines. The source code, guide and datasets are currently available on Github repository (https://github.com/ramadatta/CPgeneProfiler).

Step 1: Download CP gene database using R console

```
# Specify CP gene database URL
url <- "https://raw.githubusercontent.com/ramadatta/CPgeneProfiler/
master/testData/db/NCBI_BARRGD_CPG_DB.fasta"

# Specify destination where CP gene database file should be saved
path <- "/home/user/db" # Can be changed to preferred location
setwd(path)
destfile <- "NCBI_BARRGD_CPG_DB.fasta"

# Download the CP gene database file to the folder set in "path"
download.file(url, destfile)
```

Step 2: Install CPgeneProfiler package

The R package `CPgeneProfiler` can be installed by typing the following in R:

```
devtools::install_github("ramadatta/CPgeneProfiler")
```

Figures

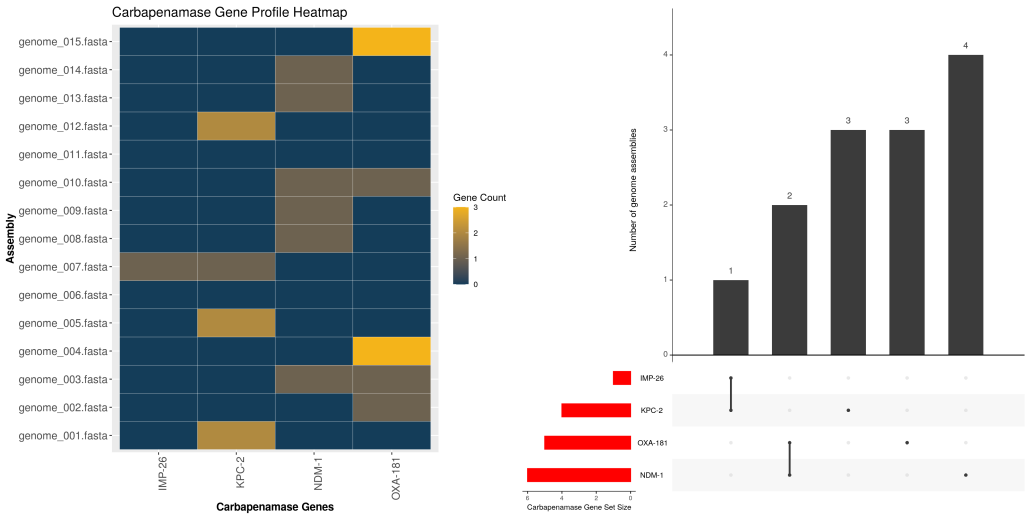

Figure 1. **(A)** CP gene profile obtained by 'cpprofile' command **(B)** Set intersection plot of the available CP genes across genome assemblies, obtained by the 'upsetR_plot' command

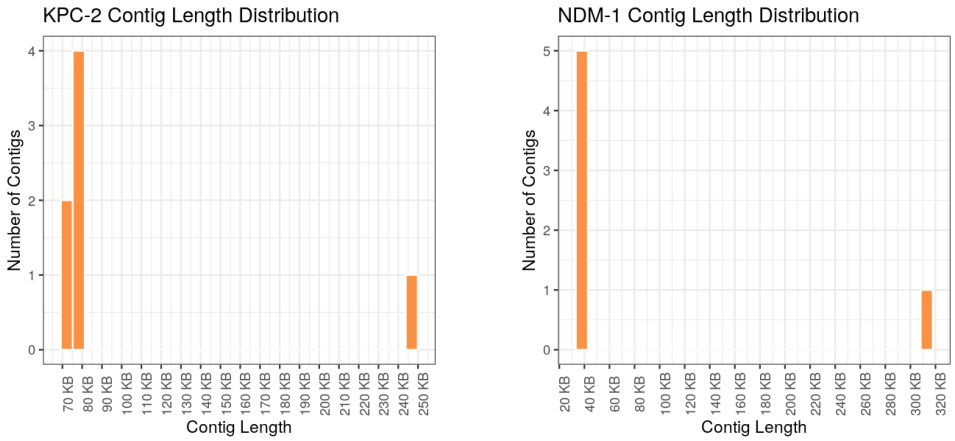

Figure 2. CP gene contig length distribution plots obtained by the 'plot_conlen' command.

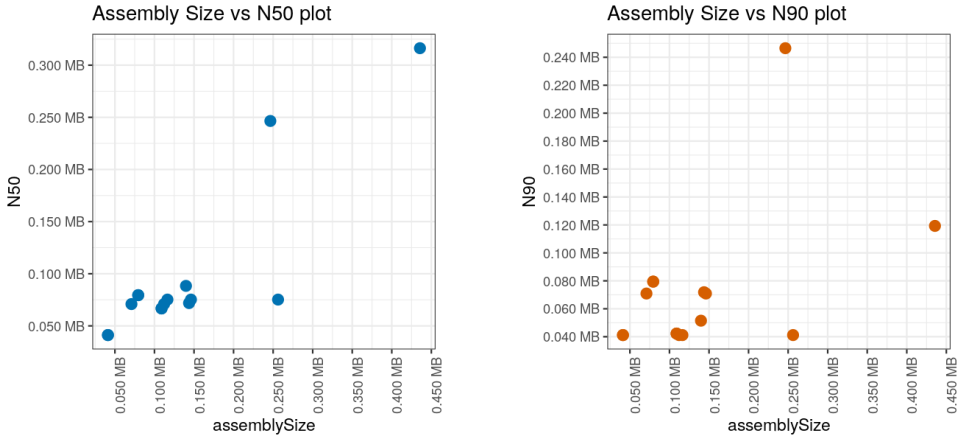

Figure 3. Plots generated by the 'assembly_stat' command **(A)** Assembly size vs N50 **(B)** Assembly size vs N90

Conclusion

`CPgeneProfiler` is useful to understand the CP gene profile of a set of bacterial genome assemblies. It generates a simple heatmap for visualization of the CP gene profile and reports details on cocarriage of CP genes within the genomes. The capability to identify and visualize the presence of CP genes across multiple genomes would have useful applications, for example, in a dataset of outbreak samples, and the CPgeneProfiler could aid researchers in obtaining an overview of the samples and their CP gene carriage.

Acknowledgement

The authors would like to thank Victor Ong and Wang Liang De for generating the sequence data that was used for developing and testing the tool.

Funding

This work is supported by the Singapore Ministry of Health's National Medical Research Council under its NMRC Collaborative Grant: Collaborative Solutions Targeting Antimicrobial Resistance Threats in Health Systems (CoSTAR-HS) (NMRC CGAug16C005) and NMRC Clinician Scientist Award (MOH-000276). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of MOH/NMRC.

References

Alcock, B. P., Raphenya, A. R., Lau, T. T., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., et al. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research*, *48*(D1), D517–D525. doi:10.1093/nar/gkz935

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A., et al. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy*. doi:10.1093/jac/dkaa345

Cai, B., Echols, R., Magee, G., Arjona Ferreira, J. C., Morgan, G., Ariyasu, M., Sawada, T., et al. (2017). Prevalence of carbapenem-resistant gram-negative infections in the united states predominated by acinetobacter baumannii and pseudomonas aeruginosa. In *Open forum infectious diseases* (Vol. 4). Oxford University Press. doi:10.1093/ofid/ofx176

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC bioinformatics*, *10*(1), 421. doi:10.1186/1471-2105-10-421

Clausen, P. T., Aarestrup, F. M., & Lund, O. (2018). Rapid and precise alignment of raw reads against redundant databases with kma. *BMC bioinformatics*, *19*(1), 1–8. doi:10.1186/s12859-018-2336-6

Clausen, P. T., Zankari, E., Aarestrup, F. M., & Lund, O. (2016). Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *Journal of Antimicrobial Chemotherapy*, *71*(9), 2484–2488. doi:10.1093/jac/dkw184

Dortet, L., & Naas, T. (2017). Noncarbapenemase oxa-48 variants (oxa-163 and oxa-405) falsely detected as carbapenemases by the β carba test. *Journal of Clinical Microbiology*, *55*(2), 654–655. doi:10.1128/JCM.02086-16

Doster, E., Lakin, S. M., Dean, C. J., Wolfe, C., Young, J. G., Boucher, C., Belk, K. E., et al. (2020). MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic acids research*, *48*(D1), D561–D569. doi:10.1093/nar/gkz1010

Duin, D. van, Kaye, K. S., Neuner, E. A., & Bonomo, R. A. (2013). Carbapenem-resistant enterobacteriaceae: A review of treatment and outcomes. *Diagnostic microbiology and infectious disease*, *75*(2), 115–120. doi:10.1016/j.diagmicrobio.2012.11.009

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., & Rolain, J.-M. (2014). ARG-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrobial agents and chemotherapy*, *58*(1), 212–220. doi:10.1128/aac.01310-13

Hendriksen, R. S., Bortolaia, V., Tate, H., Tyson, G., Aarestrup, F. M., & McDermott, P. (2019). Using genomics to track global antimicrobial resistance. *Frontiers in public health*, *7*, 242. doi:10.3389/fpubh.2019.00242

Hunt, M., Mather, A. E., Sánchez-Busó, L., Page, A. J., Parkhill, J., Keane, J. A., & Harris, S. R. (2017). ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial genomics*, *3*(10). doi:10.1099/mgen.0.000131

Inouye, M., Dashnow, H., Raven, L.-A., Schultz, M. B., Pope, B. J., Tomita, T., Zobel, J., et al. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome medicine*, *6*(11), 90. doi:10.1186/s13073-014-0090-6

Nordmann, P., & Poirel, L. (2019). Epidemiology and diagnostics of carbapenem resistance in gram-negative bacteria. *Clinical Infectious Diseases*, *69*(Supplement_7), S521–S528. doi:10.1093/cid/ciz824

Panunzi, L. G. (2020). SraX: A novel comprehensive resistome analysis tool. *Frontiers in microbiology*, *11*, 52. doi:10.3389/fmicb.2020.00052

Papp-Wallace, K. M., Endimiani, A., Taracila, M. A., & Bonomo, R. A. (2011). Carbapenems: Past, present, and future. *Antimicrobial agents and chemotherapy*, *55*(11), 4943–4960. doi:10.1128/AAC.00296-11

Wang, Y.-C., Tang, H.-L., Liao, Y.-C., Chiou, C.-S., Chen, Y.-T., Chiang, M.-K., Lu, M.-C., et al. (2019). Cocarriage of distinct blaKPC-2 and blaOXA-48 plasmids in a single sequence type 11 carbapenem-resistant klebsiella pneumoniae isolate. *Antimicrobial agents and chemotherapy*, *63*(6). doi:10.1128/aac.02282-18

Zankari, E., Allesøe, R., Joensen, K. G., Cavaco, L. M., Lund, O., & Aarestrup, F. M. (2017). PointFinder: A novel web tool for wgs-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *Journal of Antimicrobial Chemotherapy*, *72*(10), 2764–2768. doi:10.1093/jac/dkx217