

CPgeneProfiler: A lightweight R package to profile the Carbapenemase 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 Teo³, Indumathi Venkatachalam⁴, 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](https://doi.org/10.21105/joss.02473)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [William Rowe](#) ↗

Reviewers:

- [@fmaguire](#)
- [@boasvdp](#)

Submitted: 06 July 2020

Published: 11 July 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

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 co-carriage 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 resources were identified to find the antimicrobial resistance genes in nucleotide as well as amino acid sequence data (Hendriksen et al., 2019). Some of these resources include, but not limited to are ARIBA (Hunt et al., 2017), ARG-ANNOT (Gupta et al., 2014), CARD database (Jia et al., 2016), ME GARes (Lakin et al., 2017), NCBI-AMR Finder (<https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMR Finder/>), KmerResistance (<https://cge.cbs.dtu.dk/services/KmerResistance/>), SRST2 (Inouye et al., 2014) and ResFinder (Zankari et al., 2012).

Current antimicrobial resistance gene finding tools can detect the genes from either genome assemblies or sequencing reads but do not exclusively report a genetic profile for the presence of CP genes across all the genome assemblies and neither provide the ability to extract and visualize the co-carriage of multiple CP genes. Therefore, to address this need, we describe here a lightweight R package, **CPgeneProfiler** that scans bacterial genome sequences to detect the presence of CP genes using R framework.

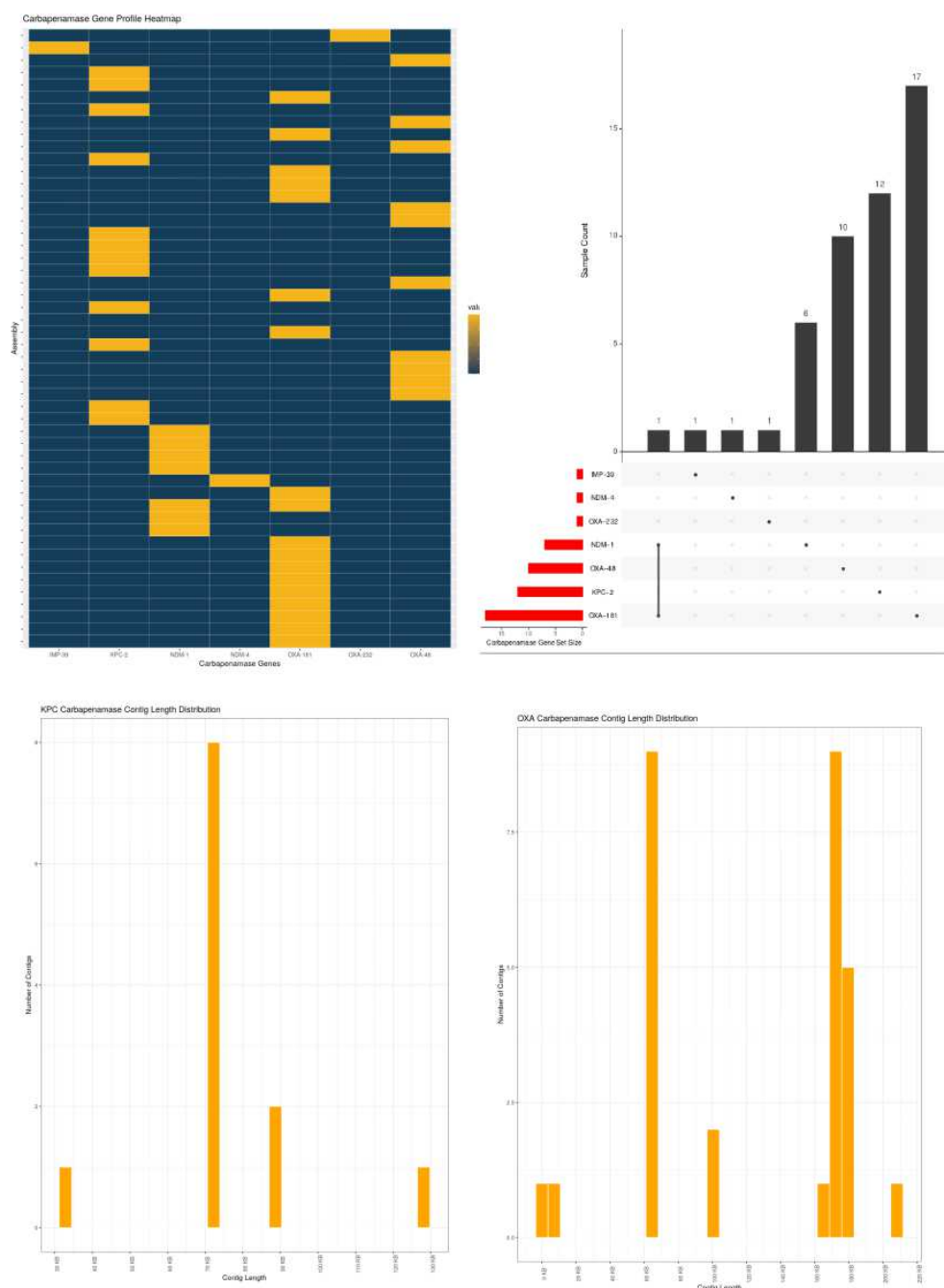


Figure 1: a) CP gene profile (Top left) b) Set intersection plot showing the number of genome assemblies for various CP genes (Top right) c) CP gene-containing contig length plots KPC (Bottom left) & OXA genes (Bottom right)

In addition to profiling the CP genes and finding co-carriage of CP genes in genome assemblies (Figure 1a), the tool also generates plots which visualizes set intersections of CP genes across all the input genome assemblies (Figure 1b). Lastly, CPgeneProfiler also generates simple

N50 & N90 statistic plots and CP gene-containing contig length distributions summary plots (Figure 1c, 1d). Currently, CPgeneProfiler can be executed under the Linux/Unix operating system.

CPgeneProfiler R Package

The R package CPgeneProfiler, currently in version 2.1.0 on Github (<https://github.com/ramadatta/CPgeneProfiler>), checks for the presence of CP genes in a list of genome assemblies provided in fasta file format. A wiki page with more detailed instructions on how to install the package on a system can be accessed at: <https://github.com/ramadatta/CPgeneProfiler/wiki/Step-by-Step-Guide>

Step 1: Download CP gene database using R

```
# Specify CP gene database URL
> url <- "https://raw.githubusercontent.com/ramadatta/CPgene-profiler
/master/ARG-annot_CPGene_DB.fasta"

# Specify destination where CP gene database file should be saved
> path <- "/home/user/db" # Can change to preferable location
> setwd(path)
> destfile <- "ARG-annot_CPGene_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")
```

After installation, the package can be used as follows:

```
> CPgeneProfiler("/path/Multiple_FastaFiles_Location/", "/home/user/db/")
```

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

- 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.
- 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.
- 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.
- 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.
- 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).
- 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.
- Jia, B., Raphenya, A. R., Alcock, B., Wagglechner, N., Guo, P., Tsang, K. K., Lago, B. A., et al. (2016). CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic acids research*, gkw1004.
- Lakin, S. M., Dean, C., Noyes, N. R., Dettenwanger, A., Ross, A. S., Doster, E., Rovira, P., et al. (2017). MEGARes: An antimicrobial resistance database for high throughput sequencing. *Nucleic acids research*, 45(D1), D574–D580.
- Nordmann, P., & Poirel, L. (2019). Epidemiology and diagnostics of carbapenem resistance in gram-negative bacteria. *Clinical Infectious Diseases*, 69(Supplement_7), S521–S528.
- 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.
- 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).
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., et al. (2012). Identification of acquired antimicrobial resistance genes. *Journal of antimicrobial chemotherapy*, 67(11), 2640–2644.