AMRScan: A hybrid R and Nextflow toolkit for rapid antimicrobial resistance gene detection from sequencing data

Kaitao Lai[1]

2025-07-08

[1] University of Sydney

1 Summary

AMRScan is a hybrid bioinformatics toolkit implemented in both R and Nextflow for the rapid and reproducible detection of antimicrobial resistance (AMR) genes from next-generation sequencing (NGS) data. The toolkit enables users to identify AMR gene hits in sequencing reads by aligning them against reference databases such as CARD using BLAST (Altschul et al. 1990).

The R implementation provides a concise, script-based approach suitable for single-sample analysis, teaching, and rapid prototyping. In contrast, the Nextflow implementation enables reproducible, scalable workflows for multi-sample batch processing in high-performance computing (HPC) and containerized environments. It leverages modular pipeline design with support for automated database setup, quality control, conversion, BLAST alignment, and results parsing.

AMRScan helps bridge the gap between lightweight exploratory analysis and production-ready surveillance pipelines, making it suitable for both research and public health genomics applications.

2 Statement of Need

While several large-scale AMR detection platforms such as ResFinder (Zankari et al. 2012) exist, many are resource-intensive or require complex installations. AMRScan addresses the need for a minimal, transparent, and reproducible toolkit that can be used flexibly in small labs, clinical settings, or large-scale surveillance workflows.

The inclusion of a pure Nextflow implementation enables high-throughput, multi-sample analyses in cloud and HPC environments, while the standalone R script remains accessible to users in data science, microbiology, and epidemiology. Both versions use shared components (e.g., a BLAST parsing script) to ensure consistency and reproducibility of results.

3 Usage Guidance

AMRScan provides two usage modes, tailored to user needs:

- **R script (`AMRScan_standalone.R`)**: Best suited for small datasets, single-sample analysis, quick local tests, educational use, and lightweight environments without workflow managers.
- **Nextflow workflow (`main.nf`)**: Designed for large-scale, automated analyses, this version excels in multi-sample settings, HPC/cloud infrastructure, and environments where reproducibility, parallelism, and containerization are priorities.

This flexible dual-mode implementation ensures that AMRScan can serve both teaching/demo scenarios and production-grade bioinformatics pipelines.

4 Implementation

- The R script `scripts/AMRScan_standalone.R` encapsulates the entire pipeline in a linear script-based format.
- The Nextflow workflow `workflow/main.nf` organizes the same logic into modular processes:
  - `DownloadCARD`, `MakeBLASTdb`, `ConvertFASTQ`, `RunBLAST`, and `ParseResults`
- The shared R script `scripts/parse_blast.R` is used for post-BLAST result summarization.
- Both implementations are documented, testable, and validated using mock NGS input.

5 Example Dataset and Demonstration

The example data used to validate AMRScan was obtained from a study by Munim et al. (2024) on multidrug-resistant *Klebsiella pneumoniae* isolated from poultry in Noakhali, Bangladesh (Munim et al. 2024). The assembled genome was downloaded from GenBank (GCA_037966445.1).

For antimicrobial resistance gene detection, we used the protein homolog model from the Comprehensive Antibiotic Resistance Database (CARD) (Jia et al. 2017), version Broadstreet v4.0.1, available at: https://card.mcmaster.ca/download/0/broadstreet-v4.0.1.tar.bz2

A sample output summary is shown below:

5.1 Top AMR Hits Summary

Query	Subject	Identity	Length	Evalue	Bitscore	Annotation
JBBPBW010000028.1	OprA	40.839	453	0	252	OprA *Pseudomonasaeruginosa*
JBBPBW010000035.1	LAP-2	100.000	285	0	587	LAP-2 *Enterobactercloacae*
JBBPBW010000001.1	SHV-11	100.000	286	0	581	SHV-11 *Klebsiellapneumoniae*
JBBPBW010000010.1	eptB	99.303	574	0	1109	eptB *Klebsiellapneumoniaesubsp.rhinoscleromatis*
JBBPBW010000104.1	dfrA14	98.726	157	0	327	dfrA14 *Escherichiacoli*
JBBPBW010000011.1	YojI	83.912	547	0	885	YojI *Escherichiacolistr.K* − 12*substr.MG*1655

6 Software Repository

The source code for AMRScan is freely available on GitHub at:
https://github.com/biosciences/AMRScan

7 Acknowledgements

The author thanks collaborators at University of Sydney for feedback on early concepts.

References

Altschul, Stephen F, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Jia, Baofeng, Aliyu R Raphenya, Brian Alcock, Nicholas Waglechner, Peng Guo, Katelyn K Tsang, Brendan A Lago, et al. 2017. "CARD 2017: Expansion and Model-Centric Curation of the Comprehensive Antibiotic Resistance Database." *Nucleic Acids Research* 45 (D1): D566–73. https://doi.org/10.1093/nar/gkw1004.

Munim, Md Adnan, Afroza Akter Tanni, Md Mobarok Hossain, Kallyan Chakma, Adnan Mannan, SM Rafiqul Islam, Jully Gogoi Tiwari, and Shipan Das Gupta. 2024. "Whole Genome Sequencing of Multidrug-Resistant Klebsiella Pneumoniae from Poultry in Noakhali, Bangladesh: Assessing Risk of Transmission to Humans in a Pilot Study." *Comparative Immunology, Microbiology and Infectious Diseases* 114: 102246. https://doi.org/10.1016/j.cimid.2024.102246.

Zankari, Ea, Henrik Hasman, Salvatore Cosentino, Martin Vestergaard, Simon Rasmussen, Ole Lund, Frank M Aarestrup, and Mette V Larsen. 2012. "Identification of Acquired Antimicrobial Resistance Genes." *Journal of Antimicrobial Chemotherapy* 67 (11): 2640–44. https://doi.org/10.1093/jac/dks261.