

Ứng dụng công cụ tin sinh AMROMICS vào phân tích tự động dữ liệu giải trình tự toàn bộ hệ gen vi khuẩn

Application of AMROMICS bioinformatics tool to automatic analysis of bacterial whole genome sequencing data

Nguyễn Thị Ngọc Anh*, Hoàng Thu Trang**,
Lê Phương Thảo Quyên***, Trần Tiến Mạnh***,
Trần Trung Kiên***, Bùi Đức Nam***,
Hoàng Nhật Đức***, Hồ Hữu Thọ*,****

*Viện NC Y Dược học QS, Học viện Quân y,
**Đại học KHTN, Đại học Quốc gia Hà Nội,
***Học viện Quân y,
****Bệnh viện Quân y 103

Tóm tắt

Mục tiêu: Nghiên cứu kết quả bước đầu ứng dụng công cụ tin sinh AMROMICS trong phân tích tự động hệ gen của vi khuẩn kháng kháng sinh. *Đối tượng và phương pháp:* Phân tích toàn bộ hệ gen của 14 chủng vi khuẩn *E. coli* và chủng *E. coli* K-12 MG1655 được công bố trên cơ sở dữ liệu NCBI bằng công cụ tin sinh AMROMICS. *Kết quả:* Công cụ tự động phân tích toàn bộ hệ gen của 15 mẫu vi khuẩn trong thời gian ngắn, cùng lúc đưa ra kết quả về cấu trúc hệ gen, kiểu trình tự, gen độc lực, gen kháng với giao diện đơn giản và dễ sử dụng. Đa số các chủng nghiên cứu có thể là vi khuẩn đa kháng và mang gen kháng với các kháng sinh thông dụng như quinolone và aminoglycoside. *Kết luận:* Công cụ AMROMICS có tiềm năng ứng dụng trong lâm sàng để nhanh chóng xác định gen kháng kháng sinh và gen độc lực của vi khuẩn, trên cơ sở đó có thể ứng dụng trong điều trị và kiểm soát nhiễm khuẩn hiệu quả.

Từ khóa: AMROMICS, phân tích giải trình tự, toàn bộ hệ gen, vi khuẩn, kháng kháng sinh.

Summary

Objective: To investigate initially the application of the AMROMICS bioinformatics tool in the automatic analysis of bacterial genomes. *Subject and method:* Whole-genome analysis of 14 *E. coli* isolates and the standard *E. coli* K-12 MG1655 strain published on NCBI database using the AMROMICS bioinformatics tool. *Result:* AMROMICS automatically analyzed the entire genome of 15 bacterial samples in a short time, and simultaneously showed genome structure, sequence type, virulence gene, and resistance gene with a simple and easy-to-use interface. Most strains could be multi-drug resistant bacteria and harbor genes associated with resistance to commonly used antibiotics such as cephalosporin, quinolone and aminoglycoside. *Conclusion:* The AMROMICS tool has the potential for clinical application, with the ability to quickly determine drug-resistant genes and virulence genes, AMROMICS can be applied in developing effective treatments regimens and infection control measures.

Keywords: AMROMICS, sequencing analysis, whole genome, bacteria, drug resistance.

Ngày nhận bài: 17/6/2022, ngày chấp nhận đăng: 7/7/2022

Người phản hồi: Hồ Hữu Thọ, Email: hohuutho@vmmu.edu.vn - Bệnh viện Quân y 103, Học viện Quân y

1. Đặt vấn đề

Hiện nay, vi khuẩn kháng kháng sinh đang trở thành một vấn đề toàn cầu, đặc biệt là ở các nước đang phát triển. Tình trạng kháng kháng sinh đang ngày càng phức tạp hơn, đặc biệt là khi xuất hiện các chủng đa kháng, đa kháng diện rộng, và kháng toàn bộ với các kháng sinh đang sử dụng hiện nay.

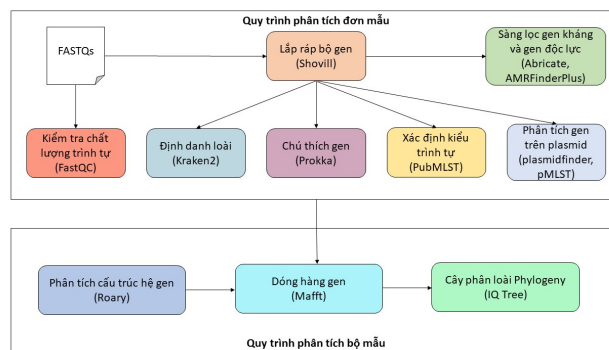
Cơ chế kháng kháng sinh của vi khuẩn rất phức tạp, là tổng hợp của nhiều cơ chế khác nhau, thông thường là do sự xuất hiện của các gen và các đột biến liên quan đến kháng thuốc ở vi khuẩn. Phương pháp giải trình tự gen thế hệ mới và phân tích toàn bộ hệ gen vi khuẩn đang có xu hướng được nghiên cứu và sử dụng rộng rãi để hiểu rõ tình trạng và cơ chế kháng kháng sinh ở vi khuẩn, dự đoán xu thế kháng thuốc ở vi khuẩn, cũng như xây dựng phác đồ điều trị chính xác và hiệu quả. Tuy nhiên phân tích giải trình tự thế hệ mới còn nhiều khó khăn, yêu cầu nhân lực có trình độ cao và máy móc hiện đại. Từ dữ liệu giải trình tự thô, cần nhiều bước xử lý bằng các công cụ tin sinh khác nhau như kiểm tra chất lượng trình tự, lắp ráp các đoạn gen, chú thích gen, đối chiếu nguồn dữ liệu gen độc lực, gen kháng kháng sinh và xây dựng cây phân loài. Quy trình này cần nhiều thời gian và có thể có độ chính xác không cao nếu lựa chọn công cụ phân tích không phù hợp.

Để giải quyết các vấn đề trên, công cụ tin sinh AMROMICS đã được phát triển trong khuôn khổ đề tài "Giải quyết vấn đề vi khuẩn kháng kháng sinh bằng công nghệ giải trình tự DNA và phân tích dữ liệu lớn" được tài trợ bởi quỹ VINIF với mã số đề tài VINIF.2019.DA11. Công cụ này cho phép tự động phân tích dữ liệu giải trình tự toàn bộ hệ gen vi khuẩn, đưa ra tất cả các thông tin cơ bản về bộ gen của vi khuẩn như định danh loài, gen độc lực và các thông tin liên quan đến kháng thuốc.

AMROMICS hoạt động dựa trên nguyên lý kết hợp các công cụ phân tích tin sinh hiện đại đang được sử dụng phổ biến trên thế giới (Hình 1). Sau khi tiếp nhận dữ liệu Assembly hoặc dữ liệu giải trình tự thô của từng mẫu, công cụ tiến hành kiểm tra chất lượng trình tự mẫu trực tiếp bằng công cụ FastQC (<https://github.com/s-andrews/FastQC>). Sau đó, công cụ sử dụng Shovill (<https://github.com/tseemann/>

shovill) để lắp ráp trình tự bộ gen. Việc lắp ráp mẫu sau đó được chú thích bằng Prokka (<https://github.com/tseemann/prokka>) để xác định các gen mã hóa protein và các trình tự RNA vận chuyển. Tiếp theo, các trình tự đã được lắp ráp được sàng lọc gen kháng kháng sinh và gen độc lực bằng công cụ Abricate (<https://github.com/tseemann/abricate>) và cơ sở dữ liệu AMRFinderPlus (<https://github.com/ncbi/amr>). Song song, công cụ cũng sử dụng PubMLST (<https://github.com/tseemann/mlst>) để xác định kiểu trình tự (Sequence Type- ST) của loài và Kraken2 để định danh loài (<https://github.com/DerrickWood/kraken2>).

Sau khi tất cả các mẫu đơn trong bộ mẫu đã được phân tích xong, AMROMICS tiếp tục phân tích bộ mẫu bằng Roary (<https://github.com/sanger-pathogens/Roary>) để phân tích cấu trúc hệ gen của bộ mẫu. Đối với mỗi cụm gen mã hóa, các trình tự gen được thực hiện đóng hàng gen bằng MAFFT (<https://github.com/GSLBiotech/mafft>) để xác định sự thay đổi protein và các đột biến nucleotide. Sau cùng, công cụ sử dụng Iqtree (<https://github.com/iqtree/iqtree2>) để tạo cây phân loài.



Hình 1. Quy trình phân tích tự động của công cụ tin sinh AMROMICS

Trong nghiên cứu này, chúng tôi tiến hành bước đầu nghiên cứu ứng dụng công cụ tin sinh AMROMICS trong phân tích dữ liệu giải trình tự toàn bộ hệ gen của vi khuẩn *E. coli*. Dựa trên dữ liệu giải trình tự toàn bộ hệ gen, công cụ tự động xác định được cấu trúc hệ gen, kiểu gen và các gen độc lực của nhóm vi khuẩn nghiên cứu. Ngoài ra, công cụ dự đoán được tính kháng kháng sinh của các vi khuẩn nghiên cứu dựa trên sự có mặt của các gen và các đột biến liên quan đến tính kháng kháng sinh.

2. Vật liệu và phương pháp

2.1. Vật liệu

Trình tự toàn bộ hệ gen (WGS) của 14 chủng vi khuẩn *E. coli* được phân lập từ mẫu máu của bệnh nhân tại Hà Nội và 01 chủng vi khuẩn *E. coli* K-12 MG1655 (Mã truy cập: U00096) làm mẫu đối chứng. 14 chủng này được giải trình tự toàn bộ hệ gen bằng công nghệ Illumina và kết quả giải trình tự đã được công bố trên cơ sở dữ liệu NCBI với các mã truy cập (tên mẫu) như sau:

BGPF01000000 (MH14-093M);
 BHWJ01000000 (MH15-214M);
 BHWL01000000 (MH15-262M);
 BFCO01000000 (MH13-051M);
 BGNL01000000 (MH13-017M);
 BLSL01000000 (MH15-170M);
 BGPA01000000 (MH14-070D);
 BGPM01000000 (MH14-236M);
 BGPO01000000 (MH16-327M);
 BLSM01000000 (MH15-171M);
 BHWK01000000 (MH15-257M);
 BGPE01000000 (MH14-089M);
 BHC01000000 (MH15-166M);
 BGNC01000000 (MH16-401M).

2.2. Phương pháp

Dữ liệu giải trình tự hệ gen được phân tích tự động bằng công cụ tin sinh AMROMICS tích hợp trên nền tảng website <https://platform.dev.amromics.org/>. Dữ liệu giải trình tự được tải về máy tính, sau đó được đưa lên nền tảng website của công cụ. Phân tích được tự động thực hiện sau khi ấn nút "Run analysis" trên website. Kết quả phân tích cấu trúc hệ gen (pangenome), kiểu trình tự (MLST), các gen kháng và gen độc lực được tự động trích xuất dưới dạng bảng và đồ họa trực quan.

3. Kết quả

3.1. Kết quả phân tích từ AMROMICS

Chỉ sau một bước nhấp chuột (Run analysis), công cụ hoàn thành phân tích tự động toàn bộ hệ

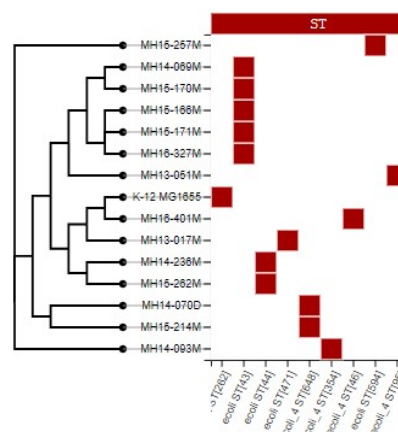
gen của nhóm vi khuẩn nghiên cứu. Kết quả phân tích từng mẫu thu được sau khoảng 30 phút; còn đối với phân tích bộ mẫu 15 chủng *E. coli*, công cụ chỉ mất khoảng 3-4 tiếng để có thể đưa ra được kết quả. Đối với từng mẫu, công cụ đưa ra các thông số thống kê lắp ráp gen (Assembly stat), kiểu trình tự (Sequence type), gen kháng kháng sinh, gen độc lực và biểu đồ gen (Genome browser). Kết quả phân tích bộ mẫu bao gồm kiểu cấu trúc gen của bộ mẫu (pangenome), đóng hàng gen (Genes alignment) (Hình 3), các thông tin về gen kháng, gen độc lực và kiểu trình tự dưới dạng biểu đồ (Heatmap) (Hình 2).

3.2. Kết quả phân tích cấu trúc hệ gen

Kết quả phân tích cho thấy, có 28,31% các gen chính (core genes) có mặt trong toàn bộ 15 chủng, 31,22% số gen có mặt trong 2 hoặc hơn 2 chủng nhưng không có mặt ở toàn bộ các chủng (shell genes) và 40,47% số gen chỉ có mặt trong 1 chủng (cloud genes).

Phân tích MLST cho thấy các chủng nghiên cứu thuộc 9 phân loại ST, trong đó, ST43 xuất hiện nhiều nhất ở 5/15 chủng phân tích (Hình 2).

Tổng số gen độc lực phát hiện được lên tới 152 gen. Trong đó, các gen độc lực *entB*, *entC*, *entD*, *entE*, *entF*, *entS*, *fepA*, *fepB*, *fepC*, *fepD*, *fepG*, *fimF*, *fimH*, *fimG*, *fes*, *ompA* có mặt ở tất cả các chủng được phân tích. Các chủng vi khuẩn *E. coli* MH15-171, MH16-327, MH13-051 có số lượng gen độc lực cao nhất, lên tới 75 gen.



Hình 2. Kết quả kiểu trình tự của 14 chủng nghiên cứu và chủng chuẩn

3.3. Gen kháng thuốc

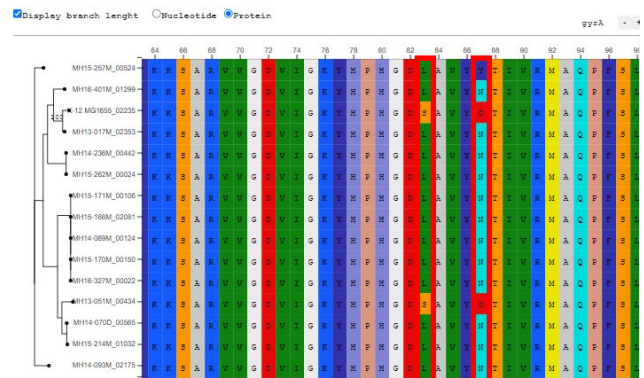
Có tổng cộng 47 gen AMR được xác định liên quan đến tính kháng với 18 loại kháng sinh (Hình 4). Hầu hết các chủng đều mang gen kháng với các kháng sinh phổ biến như β -lactam (100%), cephalosporin (13/14), nhóm aminoglycoside (12/14), và quinolone (7/14). 13/14 (92,9%) chủng mang nhiều gen kháng thuốc và có thể là vi khuẩn đa kháng, chỉ có duy nhất 1 chủng mang gen kháng với một kháng sinh (Hình 3). Đặc biệt, chủng MH16-401M mang gen kháng với colistin (*mcr-1*)- một kháng sinh có độc lực cao ít được sử dụng; chủng MH15-262M mang gen kháng với 15 kháng sinh khác nhau và chủng này còn mang gen *ble-Sh* liên quan đến tính kháng với bleomycin- một dòng kháng sinh được dùng trong điều trị ung thư.

Đối với nhóm quinolone, ngoài các gen kháng thuốc, AMROMICS còn tìm được các đột biến liên quan kháng thuốc trên kết quả đóng hàng gen (Hình 4). 12/14 chủng (92,86%) có sự thay thế amino acid S83L và D87N trên gen *gyrA*, 2 đột biến này đã được chứng minh liên quan đến tính kháng quinolone với nồng độ ức chế tối thiểu tăng lên 10 lần so với chủng nhạy cảm [4].

	β -lactam	Cephalosporin	SSS	TET	TMP	GM	STR	AN	KAN	TM	CHL	FF	Quinolone	Macrolide	RIM	FOS	CL	Bleomycin
K12-MG1655																		
MH13-051M																		
MH15-170M																		
MH16-327M																		
MH15-171M																		
MH14-069M																		
MH15-166M																		
MH14-070D																		
MH14-093M																		
MH15-257M																		
MH16-401M																		
MH14-236M																		
MH15-214M																		
MH13-017M																		
MH15-262M																		

Hình 3. Kiểu kháng thuốc của 15 chủng vi khuẩn nghiên cứu

SSS: Sulfonamide; TET: Tetracycline; TMP: Trimethoprim; GM: Gentamicin; STR: Streptomycin; AN: Amikacin; KAN: Kanamycin; TM: Tobramycin; CHL: Chloramphenicol; FF: Flofenicol; RIM: Rifamycin; FOS: Fosfomycin; CL: Colistin.



Hình 4. Hình ảnh kết quả đóng hàng gen *gyrA*

Thay thế amino acid S83L và D87N liên quan đến tính kháng quinolone.

4. Bàn luận

Từ kết quả giải trình tự, công cụ đưa ra toàn bộ kết quả phân tích hệ gen chỉ trong vài giờ đồng hồ với giao diện đơn giản, dễ sử dụng. Việc phân tích được thực hiện trên website, không sử dụng đến các câu lệnh phức tạp, không cần thiết bị có bộ nhớ và dung lượng lưu trữ lớn. Ngoài cấu trúc hệ gen và các thông tin liên quan như kiểu trình tự, gen độc lực, công cụ còn đưa ra thông tin các gen liên quan đến tính kháng của vi khuẩn đối với tất cả các kháng sinh hiện hành, kể cả những kháng sinh ít sử dụng như colistin.

Sự xuất hiện các gen ESBL trên hầu hết chủng nghiên cứu cho thấy mức độ kháng kháng sinh của *E. coli* đang ở mức báo động, đặc biệt ở nhóm cephalosporin thế hệ 3, 4 - một nhóm kháng sinh mạnh và được sử dụng phổ biến trên lâm sàng cho vi khuẩn gram âm như *E. coli*. Theo nghiên cứu của Hoàng Quỳnh Hương và cộng sự, *E. coli* chiếm tỷ lệ cao nhất trong các vi khuẩn gây nhiễm khuẩn huyết (66%) với tỷ lệ kháng thuốc cao, 73,1% với cefazolin và 51,6% với ceftriaxone, và 53,7% với cefotaxim [1].

Hầu hết các chủng vi khuẩn nghiên cứu có thể là vi khuẩn đa kháng, kết quả này cũng tương tự như một số nghiên cứu trước với tỷ lệ đa kháng ở vi khuẩn *E. coli* lên tới 91,46% [2]. Ngoài các chủng đa kháng, hiện nay các chủng vi khuẩn đa kháng diện rộng và toàn kháng cũng đã xuất hiện ở Việt Nam [3]. Sự có mặt của gen *ble-Sh* liên quan đến tính kháng bleomycin đã được chứng minh có thể tăng khả năng sống sót và sinh trưởng của vi khuẩn *E. coli*

[5, 6]. Như vậy, nguy cơ lây lan trên diện rộng của chủng vi khuẩn mang gen này là rất lớn. Thực trạng trên là một thách thức rất lớn cho quy trình quản lý nhiễm khuẩn tại bệnh viện và trong cộng đồng. Việc phát hiện và sàng lọc sớm những chủng vi khuẩn kháng thuốc là vô cùng quan trọng, nên được phát triển và ứng dụng một cách phổ biến trong nghiên cứu và điều trị lâm sàng.

5. Kết luận

Nghiên cứu này đã ứng dụng thành công công cụ tin sinh AMROMICS trong phân tích tự động dữ liệu giải trình tự toàn bộ hệ gen vi khuẩn *E. coli*. Với khả năng tự động, đơn giản và nhanh chóng xác định được cấu trúc hệ gen và dự đoán được tính kháng kháng sinh của vi khuẩn, công cụ có tiềm năng ứng dụng trong lâm sàng để điều trị và kiểm soát nhiễm khuẩn hiệu quả. Kết quả nghiên cứu cũng cho thấy tình trạng kháng kháng sinh ngày càng phức tạp với sự xuất hiện của các chủng vi khuẩn mang nhiều gen kháng với các kháng sinh khác nhau và gen kháng với kháng sinh có độc lực cao. Việc xác định nhanh chóng và chính xác tính kháng của vi khuẩn gây bệnh là vô cùng cấp thiết trong lâm sàng để điều trị bệnh hiệu quả và kiểm soát sự lây lan của vi khuẩn kháng thuốc trong bệnh viện cũng như trong cộng đồng.

Tài liệu tham khảo

1. Hoàng Quỳnh Hương và Nguyễn Thanh Hằng (2021) *Nghiên cứu tình trạng kháng kháng sinh của một số chủng vi khuẩn Enterobacteriaceae gây nhiễm khuẩn huyết phân lập được tại bệnh viện đa khoa tỉnh Thái Bình năm 2018-2019*. Tạp chí Y học Việt Nam, tr. 498.
2. Nguyễn Thị Tuyền, Lê Nữ Xuân Thanh & Ngô Viết Quỳnh Trâm (2021) *Tình hình kháng kháng sinh của các chủng E. Coli và đặc điểm gen mã hóa carbapenemase của các chủng E. coli kháng carbapenem phân lập được tại Bệnh viện Trường Đại học Y - Dược Huế*. Tạp chí Y dược học 11.
3. Bộ Y tế, Việt Nam xuất hiện siêu vi khuẩn kháng tất cả kháng sinh, bác sĩ bắt lực. Available at: https://moh.gov.vn/chuong-trinh-muc-tieu-quoc-gia/-/asset_publisher/7ng11fEWgASC/content/viet-nam-xuat-hien-sieu-vi-khuan-khang-tat-ca-khang-sinh-bac-si-bat-luc. (Truy cập: 3rd June 2022)
4. Varughese LR et al (2018) *Analytical profiling of mutations in quinolone resistance determining region of gyrA gene among UPEC*. PLoS One 13.
5. Blot M, Meyer J & Arber W (1991) *Bleomycin-resistance gene derived from the transposon Tn5 confers selective advantage to Escherichia coli K-12*. Proc. Natl. Acad. Sci. U. S. A. 88: 9112.
6. Blot M, Hauer B & Monnet G (1994) *The Tn5 bleomycin resistance gene confers improved survival and growth advantage on Escherichia coli*. Mol. Gen. Genet. 242: 595-601.