

# NGHIÊN CỨU ĐẶC ĐIỂM HỆ GENE CỦA 243 MẪU VI KHUẨN *Neisseria gonorrhoeae* PHÂN LẬP TẠI VIỆT NAM

Nguyễn Tiến Đạt<sup>1</sup>, Vương Thị Hương<sup>1</sup>, Trịnh Thị Xuân<sup>2</sup>, Nguyễn Cường<sup>3,\*</sup>

<sup>1</sup>Công ty TNHH LOBI Việt Nam

<sup>2</sup>Khoa Công nghệ thông tin, Đại học Mở Hà Nội

<sup>3</sup>Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

## TÓM TẮT

Tình trạng kháng kháng sinh ở loài vi khuẩn *Neisseria gonorrhoeae* gây bệnh lậu ở người, đang là một mối nguy hại sức khỏe toàn cầu theo WHO. Khu vực Đông Á và Đông Nam Á là nguồn chính xuất hiện và lây lan các chủng vi khuẩn này trong nhiều thập kỷ gần đây, trong đó có Việt Nam. Mặc dù vậy, các nghiên cứu về đặc điểm hệ gene của loài vi khuẩn này ở Việt Nam còn rất hạn chế. Đa số các nghiên cứu ở Việt Nam về vi khuẩn này trong 5 năm trở lại đây tập trung vào nghiên cứu xu hướng kháng kháng sinh của chúng dựa vào các kĩ thuật vi sinh và có rất ít nghiên cứu phân tích về đặc điểm hệ gene của *N. gonorrhoeae*. Do đó, chúng tôi xây dựng một quy trình bao gồm 4 nội dung cần thiết để phân tích các đặc điểm di truyền của 243 trình tự thô của vi khuẩn *N. gonorrhoeae* ở Việt Nam bao gồm: Lắp ráp, dự đoán, chú giải hệ gene; dự đoán các gene độc lực và gene kháng kháng sinh; xác định ST và vẽ cây phân loài dựa vào các house-keeping gene; phân tích Pangenome. Kết quả thu được 241/243 hệ gene lắp ráp và chú giải phù hợp (kích thước hệ gene 2,074-2,307 Mb; số lượng gene 2.024-3.112; tỷ lệ GC 52,27%-52,70%) có độ tương đồng cao với hệ gene tham chiếu ASM1303007v1 (kích thước hệ gene 2,2Mb; số lượng gene 2.283; tỷ lệ GC 52,5%). Đồng thời, phát hiện được 43 gene độc lực và 11 gene kháng kháng sinh. Trong đó, có 39/43 gene độc lực và 7/11 gene kháng kháng sinh xuất hiện ở tất cả các hệ gene lắp ráp. Kết quả phân tích pangenome bao gồm: 1600/4270 (37,47%) core gene, 1925/4270 (45,08%) accessory gene và 745/4270 (17,45%) unique gene. Kết quả phân nhóm ST cho biết 240/241 mẫu này thuộc 41 nhóm ST khác nhau với mẫu ST7371 chiếm tỷ lệ cao nhất (11,11%). Kết quả xây dựng cây phân loài cho thấy có 2 nhánh được phân chia từ gốc, nhánh 1 với 117 hệ gene và nhánh 2 với 123 hệ gene.

**Từ khóa:** Cây phân loại, đặc điểm hệ gene, *Neisseria gonorrhoeae*, Tin sinh học, tình trạng kháng kháng sinh.

## MỞ ĐẦU

Vào năm 2019, Tổ chức WHO đã đưa ra một cảnh báo về 10 nguy cơ ảnh hưởng lớn đến sức khỏe mà nhân loại đang phải đối mặt (WHO, 2019). Trong đó phải kể đến sự xuất hiện ngày càng nhiều của các chủng vi khuẩn kháng kháng sinh do sự lạm dụng nhiều loại thuốc trong ngăn ngừa và điều trị các bệnh nhiễm trùng, nhiễm khuẩn. *N. gonorrhoeae* là một loài vi khuẩn gây bệnh lậu ở người, lây qua đường tình dục (STIs). Theo WHO, mỗi năm có khoảng 82,4 triệu ca nhiễm trùng do *N. gonorrhoeae* trên toàn thế giới, xếp thứ 3 trong nhóm STIs trên phạm vi toàn cầu (WHO, 2024). Tuy nhiên, việc quản lý và điều trị nhiễm trùng *N. gonorrhoeae* ngày càng trở nên phức tạp do sự xuất hiện tình trạng kháng kháng sinh (AMR). Hiện nay, *N. gonorrhoeae* đã phát triển để kháng tất cả các nhóm kháng sinh truyền thống được sử dụng để điều trị nhiễm trùng bệnh lậu (Adamson *et al.*, 2020).

Adamson và các cộng sự Việt Nam đã thực hiện một nghiên cứu về 409 chủng *N. gonorrhoeae* phân lập từ 2017- 2019 và phát hiện chúng đã kháng với penicillin, tetracycline, ciprofloxacin và đang có xu hướng giảm dần độ nhạy cảm với ceftriaxone, azithromycin. Đây là hai loại kháng sinh được khuyến cáo sử dụng khi bị nhiễm khuẩn *N. gonorrhoeae* mang lại hiệu quả khi mà các kháng sinh khác đã không thể tiêu diệt được loài vi khuẩn này (Adamson *et al.*, 2020). Điều này chứng tỏ mức độ nguy hiểm của loài vi khuẩn này ngày càng tăng ở Việt Nam. Thêm vào đó, số lượng các nghiên cứu về *N. gonorrhoeae* trong những năm gần đây rất ít và thường chỉ tập trung vào kiểm tra độ nhạy cảm với kháng sinh bằng các kĩ thuật vi sinh chứ chưa đi sâu vào nghiên cứu đặc điểm hệ gene.

Vào năm 2023, Adamson và các cộng sự Việt Nam đã tiến hành một nghiên cứu khác về phát hiện các chủng *N. gonorrhoeae* kháng ceftriaxon. Kết quả nhấn mạnh rằng: Ceftriaxon, liệu pháp duy nhất còn lại trong điều trị lậu cầu đang bị đe dọa. Và việc phát hiện các chủng này là có thể sự xuất hiện của gene *penA-60.001* trong hệ gene của chúng (Adamson *et al.*, 2024). Vì vậy, các biện pháp can thiệp dựa vào đặc điểm hệ gene rất quan trọng nhằm phát hiện trình trạng kháng thuốc và giảm thiểu sự lây truyền thêm của *N. gonorrhoeae* một cách chính xác và nhanh chóng. Do đó, việc tìm hiểu về các đặc điểm di truyền để dự đoán và giám sát tình trạng kháng kháng sinh của *N. gonorrhoeae* là cần thiết.

Nhằm thực hiện mục tiêu này, chúng tôi lấy dữ liệu giải trình tự của 243 mẫu *N. gonorrhoeae* phân lập ở Việt Nam, được lưu trữ trên SRA và phân tích các đặc điểm di truyền của chúng bằng các quy trình Tin sinh học. Nội dung nghiên cứu quy trình bao gồm: Lắp ráp, dự đoán, chú giải hệ gene; dự đoán các gene độc lực và gene kháng kháng sinh; xác định ST và vẽ cây phân loài dựa vào các house-keeping gene và phân tích Pangenome. Các kết quả này có thể ứng dụng trong dự đoán, giám sát và các nghiên cứu liên quan về *N. gonorrhoeae* nói riêng cũng như nhiều loài vi khuẩn khác nói chung trong tương lai.

## DỮ LIỆU VÀ PHƯƠNG PHÁP

### Dữ liệu

Trong nghiên cứu này, chúng tôi sử dụng 243 bộ dữ liệu WGS của vi khuẩn *N. gonorrhoeae* được phân lập ở Việt Nam. Trong đó, 229 mẫu thu thập vào năm 2011 và 2015-2016 ở Bệnh viện Da liễu, Hà Nội (Lan *et al.*, 2020). Còn lại 14 mẫu phân lập trong năm 2019-2020 ở Hồ Chí Minh và Đà Nẵng (Trinh *et al.*, 2022). Toàn bộ dữ liệu WGS này được tải về từ cơ sở dữ liệu SRA với mã truy cập là PRJEB34425 và PRJEB45627.

### Phương pháp

#### Lắp ráp, dự đoán và chú giải hệ gen

Dữ liệu WGS với định dạng FastQ được đánh giá bằng FastQC v0.12.1 và tiền xử lý bằng Trimmomatic v0.39 với các tham số như sau: SLIDINGWINDOW:4:20 (Loại bỏ các vị trí bases có chất lượng trung bình thấp hơn 20), MINLEN:100 (Loại bỏ các đoạn trình tự ngắn hơn 100bp) và loại bỏ các adapters. Dữ liệu trình tự sau tinh sạch được lắp ráp bằng phần mềm Shovill v1.1.0 với tham số mặc định. Kết quả lắp ráp được đánh giá chất lượng bằng phần mềm QUAST v5.0.2 (Gurevich *et al.*, 2013) với tham số mặc định và đánh giá độ hoàn thiện bằng BUSCO v5.6.1 (Manni *et al.*, 2021) với cơ sở dữ liệu bacteria\_odb10. Các hệ gene có độ hoàn thiện cao (>90% Complete) được giữ lại và thực hiện dự đoán, chú giải gene bằng công cụ Prokka v1.14.6 với các tham số mặc định.

#### Dự đoán gene độc lực và gene kháng kháng sinh

Dựa vào các thông tin về gene độc lực và gene kháng kháng sinh trên cơ sở dữ liệu lần lượt là VFDB (<http://www.mgc.ac.cn/VFs/main.htm>) và CARD (<https://card.mcmaster.ca/>), công cụ Abricate v1.0.1 (Seemann, 2023) được sử dụng để phát hiện các đoạn trình tự có mức độ tương đồng cao (>80%) so với các gene trong cơ sở dữ liệu. Các gene này sau đó được thống kê và phân nhóm theo các nhóm độc lực trên VFDB và nhóm kháng sinh trên CARD.

#### Phân tích đại hệ gene Pan-Genome

Các kết quả chú giải (định dạng gff) từ Prokka v1.14.6 (Seemann, 2014) được tổng hợp lại và phân tích pangenome bằng công cụ Roary (Page *et al.*, 2015). Sau đó, số lượng các nhóm genes như: Core gene, Unique gene và new gene được thống kê theo số lượng hệ gene phân tích và trực quan hóa bằng python.

#### Xây dựng cây phân loài dựa trên ST

Các contigs thu được từ kết quả lắp ráp được sử dụng làm đầu vào cho công cụ mlst v2.23 dựa vào thông tin phân loại ST trên cơ sở dữ liệu pubMLST (<https://pubmlst.org/neisseria>). Phương pháp này dựa vào việc xác định sequence type của 7 house-keeping gene của *Neisseria* (*adk*, *abcZ*, *aroZ*, *fumC*, *pgm*, *gdh*, *pdhC*) để phân nhóm ST. Sau đó, công cụ MAFFT v7.526 (Kato *et al.*, 2002) được dùng cho việc đóng hàng đa trình tự. Kết quả đóng hàng được đưa vào FastTree v2.1.11 để vẽ cây phân loài và iTOL v6 được sử dụng để trực quan hóa hình dạng cây.

## KẾT QUẢ VÀ THẢO LUẬN

### Kết quả lắp ráp, dự đoán và chú giải hệ gene

Hai trăm bốn mươi ba bộ dữ liệu hệ gene của *N. gonorrhoeae* được tinh sạch bằng Trimmomatic và lắp ráp bằng Shovill, thu được 243 hệ gen với kích thước từ 2,074-2,307Mb (trung bình 2,153 ± 0,027) với số lượng contig từ 68-1.375 (trung bình 103 ± 98) với chỉ số N50 từ 2.178-80.821 (trung bình 55.604 ± 10.133).

Kết quả dự đoán gene xác định số lượng gene của 243 hệ gene vi khuẩn *N. gonorrhoeae* dao động từ 2.024-3.112 (trung bình 2.125 ± 79) gene. Thêm vào đó, các thống kê về số lượng RNA bao gồm: 3-4 rRNA, 48-61 tRNA và 1 tmRNA. Tỷ lệ GC khoảng 52,27%-52,70%. Kết quả lắp ráp, dự đoán và chú giải hệ gene này phù hợp với hệ gen tham chiếu (ASM1303007v1) có kích thước hệ gene là 2,2Mb, số lượng gene 2.283, tỷ lệ GC 52,5%.

Độ toàn vẹn hệ gene được đánh giá bằng phần mềm BUSCO cho thấy tỷ lệ hoàn thiện của 241/243 mẫu *N. gonorrhoeae* đều trên 90% (Complete). Hai mẫu còn lại là ERR3623126 và ERR3623149 có độ hoàn thiện thấp hơn 90%, lần lượt là 59.7% và 75% bị loại bỏ khỏi các phân tích tiếp theo.

**Kết quả dự đoán gene độc lực và gene kháng kháng sinh**

Các gene độc lực là các gene có khả năng mã hóa thành các protein giúp vi khuẩn xâm nhập, bám dính, sinh sản và gây hại cho tế bào vật chủ, dẫn đến nhiễm khuẩn. Kết quả xác định gene độc lực trong cơ sở dữ liệu VFDB cho thấy có 43 gene độc lực tồn tại trong 241 hệ gene *N. gonorrhoeae*. Trong đó, 39/43 gene được phát hiện ở hầu như tất cả các mẫu, chiếm 88%. Trong 39 gene chung này, có 18 gene thuộc nhóm Adherence (Bám dính), chiếm số lượng nhiều nhất. Ngoài ra, số lượng các gene thuộc nhóm khác bao gồm: 6 gene thuộc Stress survival, 9 gene thuộc nhóm Nutritional/metabolic factor, 5 gene thuộc Antimicrobial activity/Competitive advantage, 1 gene thuộc nhóm Immune modulation. 4/43 gene còn lại chỉ tồn tại ở một số mẫu (Tỷ lệ chưa tới 10%). Đặc biệt có xuất hiện gene *porB* là một gene được phân vào nhóm Invasion với tỷ lệ 6.6%, một nhóm mới so với các nhóm ở trên. Các kết quả thống kê chi tiết được trình bày ở Bảng 1.

**Bảng 1. Thống kê số lượng các gene độc lực trong 241 mẫu *N. gonorrhoeae***

Nhóm	Tên gene	Số lượng	Tỷ lệ
Adherence	<i>pilX, pilK, pilJ, pill, pilH, pilZ, pilT2, pilP, pilO, pilN, pilM, pilW, pilF, pilD, pilG, pilV, pilT</i>	241	100%
	<i>pilU</i>	240	99.6%
	<i>pilQ</i>	2	0.8%
	<i>pilE</i>	9	3.7%
Stress Survival	<i>recN, mntA, mntB, mntC, katA, msrA/B (pilB)</i>	241	100%
Nutritional/metabolic factor	<i>fbpA, fbpB, fbpC, lbpA, lbpB, hpuA, hpuB, hmbR, tbpA</i>	241	100%
	<i>tbpB</i>	2	0.8%
Antimicrobial activity	<i>mtrC, mtrD, mtrE, farA, farB</i>	241	100%
Immune modulation	<i>nspA</i>	241	100%
Invasion	<i>porB</i>	16	6.6%

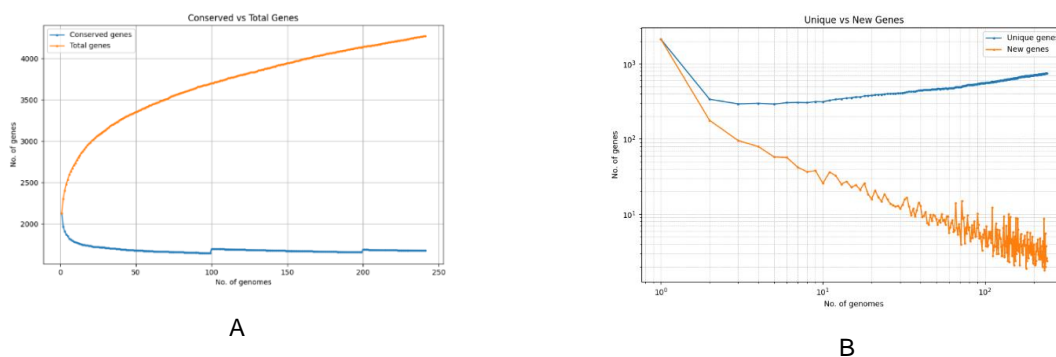
Trong 241 mẫu *N. gonorrhoeae*, chúng tôi phát hiện được 11 gene kháng kháng sinh theo cơ sở dữ liệu CARD. Đồng thời, có 7/11 (63.64%) gene có ở tất cả các mẫu với tỷ lệ xuất hiện là 100%. Các gene này kháng với macrolide, penam và free fatty acids. Các gene còn lại chia làm 2 nhóm, bao gồm: Nhóm kháng tetracycline có 97 mẫu; nhóm kháng cephalosporin, monobactam, penam, penem với 80 mẫu. Thông tin về tên gene và số lượng mẫu có ở Bảng 2.

**Bảng 2. Thống kê số lượng các gene kháng kháng sinh trong 241 mẫu *N. gonorrhoeae***

Tên gene	Số lượng mẫu	Tỷ lệ (%)	Loại thuốc kháng
<i>tetM</i>	97	40,25%	Tetracycline
<i>mtrC</i>	241	100%	Macrolide, penam
<i>mtrD</i>	241	100%	Macrolide, penam
<i>mtrE</i>	241	100%	Macrolide, penam
<i>farA</i>	241	100%	Free fatty acids
<i>farB</i>	241	100%	Free fatty acids
<i>macA</i>	241	100%	Macrolide
<i>macB</i>	241	100%	Macrolide
TEM-1	43	17,84%	Cephalosporin, monobactam, penam, penem
TEM-135	36	14,94%	Cephalosporin, monobactam, penam, penem
TEM-206	1	0,41%	Cephalosporin, monobactam, penam, penem

**Kết quả phân tích Pangenome**

Pangenome là tập hợp toàn bộ gene của tất cả các cá thể trong một loài. Cấu trúc của pangenome bao gồm: Core genome (Các gene chung của tất cả các cá thể và có liên quan đến các chức năng cơ bản và thiết yếu cho sự sống), Accessory genome (Các gene hiện diện ở một số cá thể và có liên quan đến đặc điểm thích nghi cụ thể của từng nhóm nhỏ), Unique genome (Các gene tồn tại ở 1 chủng duy nhất, mang tính đặc trưng cho từng cá thể). Trong 4270 gene được phát hiện ở 241 mẫu *N. gonorrhoeae*, có 1675 (37,47%) core gene, 1925 (45,08%) accessory gene và 745 (17,45%) unique gene.



**Hình 1. A-Mối quan hệ giữa core genes và số lượng mẫu; B-Mối quan hệ giữa unique genes và new genes với số lượng mẫu**

Các mối quan hệ giữa các nhóm genes với số lượng hệ genes được thể hiện ở Hình 1. Khi số lượng hệ genes tăng thì số lượng các gene bảo tồn (Core genome) có xu hướng giảm dần trong khoảng 25 hệ gene đầu và hầu như không thay đổi khi phân tích thêm vào các hệ gene tiếp theo. Tương tự, số lượng unique gene (Tổng hợp các gene duy nhất trong các hệ gene phân tích) có xu hướng giảm ở các hệ gene đầu và tăng dần khi phân tích thêm càng nhiều hệ gene. Ngược lại, số lượng new genes (Số lượng gene mới so với các hệ gene phân tích trước đó) giảm dần theo cấp số nhân. Điều này chứng tỏ rằng các gene đặc trưng riêng cho từng cá thể *N. gonorrhoeae* xuất hiện ngày càng nhiều, giúp vi khuẩn thích nghi với nhiều loại kháng sinh khác nhau.

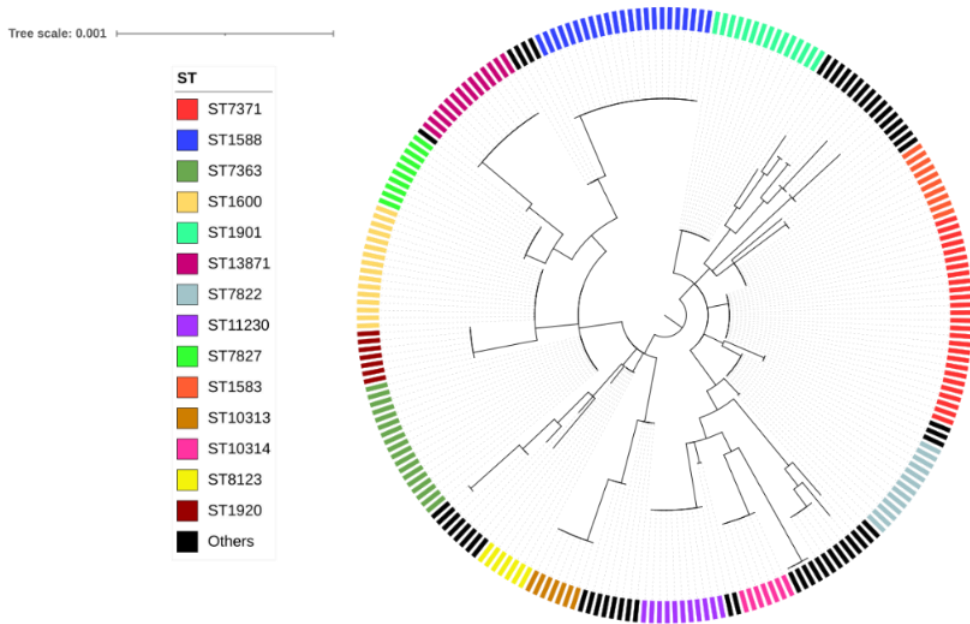
**Kết quả xây dựng cây phân loài dựa trên ST**

Trong 241 mẫu *N. gonorrhoeae* thực hiện phân tích, có 1 mẫu là ERR3623233 không xác định được nhóm ST. Kết quả chi tiết về các nhóm ST và số lượng mẫu thuộc từng nhóm có ở Bảng 3. Đồng thời, các kết quả phân tích tỷ lệ của từng nhóm cũng được thể hiện ở Hình 2.

**Bảng 3. Thống kê số lượng ST trong 243 mẫu *N. gonorrhoeae***

STT	Số lượng	Tỷ lệ	ST	Số lượng	Tỷ lệ	ST	Số lượng	Tỷ lệ
ST7371	27	11,11%	ST7360	5	2,06%	ST7367	2	0,82%
ST1588	23	9,47%	ST6714	5	2,06%	-	1	0,82%
ST7363	18	7,41%	ST1596	4	1,65%	ST11710	1	0,41%
ST1600	16	6,58%	ST8153	4	1,65%	ST12516	1	0,41%
ST1901	14	5,76%	ST7823	3	1,23%	ST12520	1	0,41%
ST13871	14	5,76%	ST8109	3	1,23%	ST1904	1	0,41%
ST7822	13	5,35%	ST1902	3	1,23%	ST16471	1	0,41%
ST11230	11	4,53%	ST8143	2	0,82%	ST1921	1	0,41%
ST7827	10	4,12%	ST10241	2	0,82%	ST1925	1	0,41%
ST1583	10	4,12%	ST9900	2	0,82%	ST1927	1	0,41%
ST10313	7	2,88%	ST6960	2	0,82%	ST11231	1	0,41%
ST10314	7	2,88%	ST11173	2	0,82%	ST1963	1	0,41%
ST8123	7	2,88%	ST12518	2	0,82%	ST1580	1	0,41%
ST1920	7	2,88%	ST12514	2	0,82%	ST1582	1	0,41%

Kết quả cây phân loài theo ST cho thấy các nhóm ST được phân theo từng cụm với đơn vị độ dài của cây là 0.001. Trong đó, Ngoài ra, có thể thấy được nhóm ST có khoảng cách xa nhất so với gốc là ST8153 (thuộc nhóm others) và nhóm ST8123 (màu vàng) có khoảng cách gần nhất so với gốc. Từ gốc, cây phân loài chia làm 2 nhánh: Nhánh 1 chứa 117 hệ gene với đa số các hệ gene thuộc 6 nhóm ST: ST1901, ST7371, ST10314, ST1583, ST7822 và ST11230; nhánh 2 chứa 123 hệ gene. Thêm vào đó, nhánh 1 cũng là nhánh có sự đa dạng về ST hơn so với nhánh 2 vì số lượng các hệ gene thuộc nhóm others (màu đen) chiếm số lượng nhiều hơn hẳn. Trong đó, nhóm ST7371 chiếm tỷ lệ nhiều nhất ở nhánh 1 có mối quan hệ gần nhất với nhóm ST7822 (màu xám) và nhóm ST1588 chiếm tỷ lệ nhiều nhất ở nhánh 2 có khoảng cách gần nhất với nhóm ST1596 (nhóm others).



Hình 2. Cây phân loại theo MLST

## KẾT LUẬN

243 bộ dữ liệu giải trình tự Illumina từ SRA của loài vi khuẩn *N. gonorrhoeae* được sử dụng trong nghiên cứu đặc điểm hệ gene thu được 241 hệ gene phù hợp sau khi lắp ráp. Các hệ gene này có kích thước từ 2,074-2,307 Mb, tỷ lệ GC khoảng 52,27%-52,70% và số lượng gene 2.024-3.112. Các số liệu này phù hợp với hệ gene tham chiếu của loài vi khuẩn này với mã ID là ASM1303007v1 trên NCBI. Số lượng gene độc lực và kháng kháng sinh phát hiện lần lượt là 43 và 11. Trong đó, có 39/43 và 7/11 gene xuất hiện ở tất cả các mẫu. Do đó, 241 hệ gene này đều có khả năng kháng các loại kháng sinh bao gồm: macrolide, penam, free fatty acids và có thể gây độc theo nhiều cơ chế khác nhau như: Adherence, stress survival, nutritional/metabolic factor, antimicrobial activity, immune modulation. Đồng thời, phát hiện được 1675/4270 (37,47%) core gene, 1925/4270 (45,08%) accessory gene và 745/4270 (17,45%) unique gene. Xu hướng của core gene và unique gene giảm dần ở các mẫu ban đầu và thay đổi rất ít khi tăng số lượng mẫu lên. Mặt khác, số lượng new gene giảm theo cấp số nhân khi tăng lượng mẫu. Kết quả phân tích ST cho thấy có 240/241 hệ gene xác định được ST. Trong đó, tỷ lệ ST7371 là nhiều nhất với 11,11%. Cây phân loài xây dựng bằng sự khác biệt giữa 7 house-keeping gene chia các mẫu thành 2 nhánh: nhánh 1 (117 hệ gene) với độ đa dạng về ST hơn so với nhánh 2 (123 hệ gene).

**Lời cảm ơn:** Nghiên cứu này được tài trợ bởi Công ty TNHH LOBI Việt Nam.

## TÀI LIỆU THAM KHẢO

- Adamson PC, Le HV, Le HHL, Le GM, Nguyen TV, Klausner JD (2020) 'Trends in antimicrobial resistance in *Neisseria gonorrhoeae* in Hanoi, Vietnam, 2017–2019', *BMC Infect Dis*, 20(1), p. 809. Available at: <https://doi.org/10.1186/s12879-020-05532-3>.
- Adamson PC, Hieu VN, Nhung PH, Whiley DM, Chau TM (2024) 'Ceftriaxone resistance in *Neisseria gonorrhoeae* associated with the penA-60.001 allele in Hanoi, Viet Nam', *Lancet Infect Dis*, 24(6), pp. e351–e352. Available at: [https://doi.org/10.1016/S1473-3099\(24\)00230-5](https://doi.org/10.1016/S1473-3099(24)00230-5).
- Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) 'QUAST: quality assessment tool for genome assemblies', *Bioinformatics*, 29(8), pp. 1072–1075. Available at: <https://doi.org/10.1093/bioinformatics/btt086>.
- Katoh K, Misawa K, Kuma K, Miyata T (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', *Nucleic Acids Res*, 30(14), pp. 3059–3066. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC135756/> (Accessed: 2 July 2024).
- Lan PT, Golparian D, Ringlander J, Van HL, Van TN, Unemo M (2020) 'Genomic analysis and antimicrobial resistance of *Neisseria gonorrhoeae* isolates from Vietnam in 2011 and 2015–16', *J Antimicrob Chemoth*, 75(6), pp. 1432–1438. Available at: <https://doi.org/10.1093/jac/dkaa040>.
- Manni M, Berkeley MR, Seppely M, Zdobnov EM (2021) 'BUSCO: Assessing Genomic Data Quality and Beyond', *Current Protocols*, 1(12), p. e323. Available at: <https://doi.org/10.1002/cpz1.323>.
- Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S (2015) 'Roary: rapid large-scale prokaryote pan genome analysis', *Bioinformatics*, 31(22), pp. 3691–3693. Available at: <https://doi.org/10.1093/bioinformatics/btv421>.

Seemann T (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics*, 30(14), pp. 2068–2069. Available at: <https://doi.org/10.1093/bioinformatics/btu153>.

Seemann T (2023) 'ABRicate'. Available at: <https://github.com/tseemann/abricate> (Accessed: 14 September 2023).

Trinh TM, Nguyen TT, Le TV, Nguyen TT, Ninh DT, Duong BH, Nguyen MV, Kesteman T, Pham LT, Doorn HR (2022) 'Neisseria gonorrhoeae FC428 Subclone, Vietnam, 2019–2020 - Volume 28, Number 2—February 2022 - Emerging Infectious Diseases journal - CDC'. Available at: <https://doi.org/10.3201/eid2802.211788>.

WHO (2019) *Ten health issues WHO will tackle this year*. Available at: <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> (Accessed: 10 July 2024).

WHO (2024) *Gonorrhoea (Neisseria gonorrhoeae infection)*. Available at: [https://www.who.int/news-room/fact-sheets/detail/gonorrhoea-\(neisseria-gonorrhoeae-infection\)](https://www.who.int/news-room/fact-sheets/detail/gonorrhoea-(neisseria-gonorrhoeae-infection)) (Accessed: 10 July 2024).

## STUDY ON THE GENOMIC CHARACTERISTICS OF 243 *NEISSERIA GONORRHOEAE* ISOLATES IN VIETNAM

Nguyễn Tiến Đạt<sup>1</sup>, Vương Thị Hương<sup>1</sup>, Trinh Thi Xuan<sup>2</sup>, Nguyễn Cường<sup>3,\*</sup>

<sup>1</sup>LOBI Vietnam Ltd.

<sup>2</sup>Faculty of Information Technology, Hanoi Open University

<sup>3</sup>Institute of Information Technology, Vietnam Academy of Science and Technology

### SUMMARY

Currently, antibiotic resistance *Neisseria gonorrhoeae*, which causes gonorrhea in humans, is a global health threat according to the WHO. The East Asia and Southeast Asia regions have been the main sources of emergence and spread of these bacterial strains in recent decades, including Vietnam. However, studies on the genetic characteristics of this bacteria in Vietnam are still very limited. Most studies in Vietnam on this bacteria in the past five years have focused on researching its antibiotic resistance trends using microbiological techniques, with very few studies analyzing the genetic characteristics of *N. gonorrhoeae*. Therefore, we have developed a protocol comprising four essential components to analyze the genetic characteristics of 243 *N.gonorrhoeae* raw sequences in Vietnam, including: genome assembly, gene prediction and gene annotation; prediction of virulence genes and antibiotic resistance genes; determination of ST and phylogenetic tree construction based on housekeeping genes; pangenome analysis. The results obtained 241/243 assembled and annotated *N. gonorrhoeae* genomes (genome size 2.074-2.307Mb; number of genes 2,024-3,112; GC content 52.27%-52.7%) with high similarity to reference genome ASM1303007v1 (genome size 2.2Mb; number of genes 2,283; GC content 52.5%). Additionally, 43 virulence genes and 11 antibiotic resistance genes were identified. Among them, 39/43 virulence genes and 7/11 antibiotic resistance genes were present in all assembled genomes. Pangenome analysis results include: 1600/4270 (37.47%) core genes, 1925/4270 (45.08%) accessory genes, and 745/4270 (17.45%) unique genes. The ST grouping results showed that 240/241 samples belonged to 41 different ST groups, with ST7371 being the most prevalent (11.11%). The phylogenetic tree construction results revealed two distinct clades, with clade 1 comprising 117 genomes and clade 2 comprising 123 genomes.

**Keywords:** Phylogenetic tree, genomic characteristics, *Neisseria gonorrhoeae*, bioinformatics, antibiotic resistance.

\* Author for correspondence: Tel: +84-916110333; Email: [ncuong@ioit.ac.vn](mailto:ncuong@ioit.ac.vn)