Annals of Clinical Microbiology and Antimicrobials

RESEARCH

Open Access

Genomic characteristics, virulence potential, antimicrobial resistance profiles, and phylogenetic insights into *Nocardia cyriacigeorgica*

Chen Yang[1†], Yue-xin Zheng[2†], Hong-yi Gu[3†], Hong Chen[4†], Wei Li[5], Fang Li[5], Yu-wang Bi[6], Jing Chen[5], Fu-kun Wang[5], Qing-qing Sun[7], Han-bing Meng[7], Zuo-hao Wu[5], Shu Yu[8*], Jiang Gu[1*] and Yan Cheng[5*]

Abstract

Background *Nocardia cyriacigeorgica*, an opportunistic pathogen, is increasingly implicated in human infections. This pathogen predominantly causes pulmonary infections, leading to acute, subacute, or chronic necrotizing suppurative lesions, in severe cases, may progress to disseminated infections. Effective clinical diagnosis, prevention, and treatment strategies require a thorough understanding of its biological characteristics and pathogenic mechanisms. However, despite the rising incidence of nocardial diseases, research on the pathogenicity of *N. cyriacigeorgica* remains limited, primarily focusing on case reports and epidemiological studies. This study aimed to provide a comprehensive analysis of the genomic features, phylogenetic relationships, antimicrobial resistance profiles, and candidate virulence factors of *N. cyriacigeorgica* strains to inform future investigations into its pathogenesis.

Methods Whole-genome sequencing was conducted on five *N. cyriacigeorgica* strains isolated from patients with pulmonary infection at our hospital. This analysis utilized a combination of second-generation Illumina HiSeq and third-generation PacBio sequencing technologies. Additionally, publicly available genomic data from 58 strains in the National Center Biotechnology Information database were integrated, resulting in a dataset of 63 genomes. These genomes were subjected to comparative genomic analyses, including phylogenetic reconstruction, pan-genome evaluation, and gene distribution assessments.

Results Phylogenetic analysis identified five major clades within *N. cyriacigeorgica*. ANI analysis further subdivided clade B into five distinct subgroups. Pan-genome analysis revealed clade-specific orthogroups in the distribution of genes assigned to Clusters of Orthologous Groups, with clade A containing the highest number of clade-specific gene families. Comparative genomic analysis uncovered several potential pathogenic genes implicated in host

[†]Chen Yang, Yue-xin Zheng, Hong-yi Gu and Hong Chen have contributed equally to this work.

*Correspondence:
Shu Yu
ys13368146418@126.com
Jiang Gu
jianggu2012@163.com
Yan Cheng
chengyan092006free@aliyun.com
Full list of author information is available at the end of the article

cell invasion, phagosomal maturation arrest, and intracellular survival within macrophages, which were conserved across all analyzed strains. Notable differences in the distribution of enterobactin-encoding genes were observed among the clades. The *mce3C* gene also displayed variable distributions across clades; however, no correlation was established between its presence and strain source. Among the 63 strains, 27 were found to harbor both *mce3C* and *mce4F* genes, which were categorized into five distinct patterns. Furthermore, antibiotic resistance genes, including *VanSO*, *VanRO*, *erm(O)-lrm*, *srmB*, *ermH*, *bcl*, *bla1*, and *cmlR*, demonstrated clade-specific distribution patterns. Notably, the genes *erm(O)-lrm*, *srmB*, and *ermH* were associated with the isolation origin of the strains.

Conclusions This study provides a comprehensive evaluation of the genomic characteristics, potential virulence factors, antimicrobial resistance genes, and phylogenetic relationships of *N. cyriacigeorgica*. The findings offer valuable insights into the mechanisms underlying intracellular survival, replication within macrophages, and pathogen-host interactions in *N. cyriacigeorgica* infections. These results establish a foundation for future research into the pathogenesis and clinical management of *N. cyriacigeorgica*.

Keywords *Nocardia cyriacigeorgica*, Whole genome sequencing, Comparative genomics, Virulence factors, Pathogenic, Antibiotic resistance

Introduction

Nocardia, a genus of Gram-positive, weakly acid-fast branching bacilli, is widely distributed in various environmental niches globally, including soil, water, and aquatic ecosystems. It is classified as part of the aerobic actinomycetes group [1]. It is recognized as an opportunistic pathogen responsible for nocardiosis in both humans and animals. *Nocardia* spp. infect humans primarily through respiratory inhalation and injured skin [2]. To date, more than 100 *Nocardia* species have been identified (https://www.bacterio.net), with over 50 species implicated in human infections affecting both immunocompetent and immunosuppressed individuals [3]. Data from the Centers for Disease Control and Prevention (CDC) indicate that an estimated 500 to 1000 new cases of nocardiosis are reported annually in the United States (https://www.cdc.gov/nocardiosis/hcp/clinical-overview/). However, on a global scale, the precise incidence of nocardiosis remains uncertain due to the absence of standardized national reporting systems.

In contrast to the well-known and extensively studied *ESKAPE* pathogens—*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa*, and *Enterobacter* species [4, 5], *Nocardia* species are clinically uncommon bacteria. Their diagnosis and treatment remain challenging in clinical practice, highlighting a significant gap in both recognition and management. *Nocardia* spp. are associated with a wide spectrum of human diseases, including pulmonary infections, superficial cutaneous and subcutaneous infections, and systemic infections resulting from hematogenous dissemination [6, 7]. Hematogenous dissemination is a critical factor that markedly elevates mortality rates, with figures reaching as high as 85% among immunocompromised individuals [8–10]. Additionally, *Nocardia* species exhibit intracellular pathogenicity, enabling them to invade and survive within host cells, thereby causing refractory infections that are often challenging to treat [11, 12].

Nocardia cyriacigeorgica was first definitively identified in 2001 through the isolation and 16S rRNA gene sequencing of strain DSM 44484 T from the bronchial secretions of a patient with chronic bronchitis [13]. This intracellular pathogen has been implicated in a variety of infections, including pneumonia, brain abscesses, and infections of the kidney, heart, and eye [14–17]. Among the various *Nocardia* species, *N. cyriacigeorgica* is recognized as the most prevalent cause of human nocardiosis in regions such as North America [18, 19], Spain [20, 21], and Iran [22]. In China, its prevalence ranges from18.3 to 40.2% among nocardial infections [23–26]. Likewise, *N. cyriacigeorgica* has been frequently reported as a commonly isolated *Nocardia* species in other countries, including Thailand, Japan, Belgium, Australia, and South Africa [27–32]. In recent years, the incidence of nocardiosis has notably increased, posing a significant threat to public health. Addressing this growing challenge necessitates the development of effective clinical diagnostic, therapeutic, and preventive strategies, which, in turn, require a deeper understanding of the virulence mechanisms and biological properties of *N. cyriacigeorgica*. Current research on *N. cyriacigeorgica* has predominantly focused on case reports and epidemiological studies [14–17, 23]. Notably, previous studies have identified distinct virulence of *N. cyriacigeorgica* strains [33] and highlighted the role of heparin-binding hemagglutinin (HBHA) in adhesive and immunoregulatory functions [34]. Despite these findings, the biological characteristics, virulence factors,

antimicrobial resistance genotypes, and pathogenic mechanisms of *N. cyriacigeorgica* remains insufficiently understood.

Numerous molecular methods have been utilized to study *Nocardia* species, including 16S rRNA gene sequence analysis [35], sequencing of single copy gene such as *secA1*, *gyrB*, *hsp65*, and *ropB* [36–39], multilocus sequence analysis (MLSA) [40], matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF/MS) analysis [41]. However, each of these techniques is associated with certain limitations. Whole genome sequencing (WGS), a high-throughput sequencing technology, provide a powerful platform for investigating bacterial pathogenicity, drug resistance mechanisms, evolutionary trends, and phylogenetic relationships. WGS serves as a valuable tool for investigating genetic factors associated with pathogenicity. Additionally, it supports the development of innovative genetic and molecular approaches, which can improve clinical diagnostics and aid in the implementation of effective control strategies.

In this study, we determined and annotated the complete genome sequences of five clinical *N. cyriacigeorgica* strains isolated from pulmonary infection patients. To gain deeper insights, we performed a comparative genomic analysis of 63 strains, which included 58 publicly available strains retrieved from the National Center for Biotechnology Information (NCBI) database and the five clinical strains sequenced in this study. The analysis focused on key aspects such as genomic features, phylogenetic relationships, mobile genetic elements (MGEs), potential virulence factors, and antibiotic resistance genes (ARGs). The overarching goal was to investigate the phylogenetic relationships, biological properties, antimicrobial resistance genotypes, and potential virulence mechanisms of *N. cyriacigeorgica*.

Materials and methods

This study employed a combination of microbiological, molecular, and bioinformatics approaches to investigate the genomic and functional characteristics of *N. cyriacigeorgica*. The methodology included bacterial isolation and identification, Antimicrobial susceptibility testing (AST), WGS, genome assembly and annotation, and comparative genomic analyses. The detailed steps for each phase are described below.

Bacterial isolation, species identification, and antimicrobial susceptibility testing

Clinical samples were inoculated on the blood-containing medium at 35 °C for 3–7 days under aerobic conditions. Suspected *Nocardia* species were identified through Gram staining, modified acid-fast staining, and acid-fast staining, followed by confirmation via 16S rRNA gene sequencing. Isolates of *N. cyriacigeorgica* were preserved in brain heart infusion (BHI) broth supplemented with 25% glycerol at – 80 ℃ for subsequent studies.

The AST was performed in accordance with the guidelines outlined by the Clinical and Laboratory Standards Institute (CLSI, M24S-Ed2) for *Nocardia* spp. [42]. Briefly, the *Nocardia* strains were grinded in sterile 0.9% sodium chloride water and subjected to repeated vortexing until no visible particles or deposits remained. The resulting uniform bacterial suspension was then adjusted to match the turbidity of the 0.5 McFarland standard. Antimicrobial susceptibility testing was conducted using first-line recommended drugs, including amoxicillin-clavulanate (AMC), imipenem (IPM), minocycline (MIN), tobramycin (TOB), amikacin (AMK), ciprofloxacin (CIP), trimethoprim/sulfamethoxazole (STX), linezolid (LZD), ceftriaxone (CRO), and clarithromycin (CCR). The broth microdilution (BMD) method was used to determine minimum inhibitory concentrations (MICs), which were interpreted based on CLSI susceptibility breakpoints. Quality control for AST was ensured by employing *Escherichia coli* ATCC 35218, *Pseudomonas aeruginosa* ATCC 27853, and *Staphylococcus aureus* ATCC 29213 as reference strains.

Library construction and whole-genome sequencing

Following the identification and preservation of clinical isolates, genomic DNA was extracted for WGS to enable detailed genomic analyses.

The clinical isolates were cultured in BHI broth (Hopebio Technology, Qingdao, China) at 37 °C for 48–72 h with constant shaking at 180 rpm. Genomic DNA was extracted using the TIANamp Bacteria DNA Kit (TIANGEN, China) following the manufacturer's instructions. The extracted DNA was subjected to WGS using both the second-generation Illumina Novaseq 6000 platform (Illumina, San Francisco, CA, USA) and the third-generation PacBio Sequel IIe platform (Pacific Biosciences, San Francisco, CA, USA).

For Illumina sequencing, genomic DNA was fragmented into 400–500 bp segments using a Covaris M220 Focused Acoustic Shearer, following the manufacturer's protocol. The fragmented DNA was subsequently prepared into Illumina sequencing libraries, which were used for paired-end sequencing (2 × 150 bp) on the Illumina Novaseq 6000 platform. For PacBio sequencing, DNA fragments were purified, end-repaired, and ligated with SMRTbell sequencing adapters to generate libraries with an average size of approximately 10 kb, in accordance with the manufacturer's guidelines (Pacific Biosciences, CA).

Raw sequence data were filtered to exclude reads containing more than 40 bp of low-quality bases, more than 10 bp of ambiguous bases (N), or adapter sequences overlapping by more than 15 bp. High-quality reads were subsequently de novo assembled into contigs using SPAdes v3.8.0 [43]. Raw reads from both sequencing platforms were subjected to quality assessment and trimming to ensure the reliability and accuracy of downstream analyses.

Genome assembly and annotation

Sequence data generated from the PacBio Sequel IIe and Illumina Novaseq 6000 platforms were utilized for bioinformatics analysis. Raw Illumina sequencing reads from the paired-end library were quality-filtered using fastp v0.23.0. Clean short reads and HiFi reads from the PacBio platform were assembled into complete genomes using Unicycler v0.4.8 [44]. The assembly was subsequently polished using Pilon v1.22, which used short-read alignments to reduce the occurrence of small errors [45]. Coding sequences (CDs) on chromosomes and plasmids were predicted using Glimmer 3.02 and GeneMarkS 4.3, respectively [46, 47]. Gene and open reading frame (ORF) annotations were performed using Prodigal and Prokka with default parameters.

Comparative genomics analyses

A dataset of 63 genomes, comprising 58 publicly available complete genome sequences of *N. cyriacigeorgica* from the NCBI database and five clinical isolates sequenced in this study, was analyzed. Pangenome analysis was conducted using Roary and FastTree to identify core and accessory genomic elements. Average Nucleotide Identity (ANI) values across all genomes were calculated using the ANI calculator, as previously described [48], to assess overall genome similarity.

To evaluate genetic relatedness, a high-quality maximum likelihood (ML) phylogenetic tree was constructed based on the concatenation of 1164 conserved single-copy genes present in 99–100% of the strains (bootstraps, 1000), following the methodology outlined by Li et al. [48]. Orthogroups among all tested genomes were identified using OrthoFinder v2.4.0 with default parameters [49]. Functional annotation of orthogroups was performed using emapper v2.0.1 against the eggNOG v5.0 database [50].

Insertion sequence (IS) elements were predicted using ISEScan v1.7.2.2 [51]. Antimicrobial resistance genes (ARGs) were identified using Diamond v0.9.14 with a cutoff E-value of 1×10^{-6}, a minimum identity of 60%, and alignment against the Comprehensive Antibiotic Resistance Database (CARD) [52]. Virulence factors were identified using Diamond v0.9.14 with the same cutoff E-value, minimum identity, and coverage thresholds, utilizing the VFDB database (http://www.mgc.ac.cn/VFs/). Plasmid identification was performed using PlasFlow.

Results

Antimicrobial susceptibility profiles of clinical isolates

Five clinical isolates, identified as *N. cyriacigeorgica* through 16S rRNA gene sequencing, were obtained from patients diagnosed with pulmonary nocardiosis at Bethune International Peace Hospital between January 1, 2019, and December 31, 2020. The AST results for these isolates are presented in Table S1. All strains were susceptible to trimethoprim/sulfamethoxazole, amikacin, imipenem, linezolid, clarithromycin, tobramycin, and minocycline. However, resistance was observed against ciprofloxacin and amoxicillin-clavulanate. Additionally, 80% (4/5) of the isolates were susceptible to ceftriaxone.

Genomic characteristics of *N. cyriacigeorgica* strains

This comparative genomic study analyzed 63 *N. cyriacigeorgica* strains, including five clinical isolates sequenced in this study and 58 publicly available strains retrieved from the NCBI database. Among these, 50.8% (32/63) were identified as human pathogens, 12.7% (8/63) originated from environmental sources, and 36.5% (23/63) had no source information. The genomic features, including genome size, G+C content, number of CDSs, contig N50, rRNA, tRNA, and isolation sources, are summarized in Table 1.

The genome sizes of the analyzed strains ranged from 5.92 to 6.83 Mb, with an average size of approximately 6.42 Mb. The number of CDSs varied between 5,360 and 6,589, while the G+C content ranged from 66.92% to 68.47%, with an average of 68.2%. Notably, four strains had G+C contents below 68%. Members of clade B had more CDSs compared to other clades. Additionally, strains in clades A and B harbored more rRNA genes than those in other clades (Table 1 and Fig. 1).

Phylogenetic relationships and average nucleotide identity analysis

To evaluate the genetic relatedness of *N. cyriacigeorgica* strains, a maximum likelihood (ML) phylogenetic tree was constructed using the concatenation of 1,164 conserved single-copy core genes (Figure 2). The phylogenetic analysis revealed five major clades, with clade B further subdivided into five subgroups. Notably, 84.4% (27/32) of strains isolated from human sources

Table 1 The genomic features of 63 *Nocardia cyriacigeorgica* strains

Strain number	Strain	Clade	Isolated source (or host)	Years	Genome Size (Mb)	GC Content (%)	Contig N50	Gene Count	rRNA	tRNA	Submitted GenBank assembly	Assembly level	References
1	NKC	B	sputum	This study	6,182,120	68.5	6116347	5682	9	63	SAMN46383565	Complete Genome	This study
2	FMUBM48	B	sputum	2018	6,134,265	68.5	6134265	5622	9	61	GCA_027924465.1	Complete Genome	
3	CNM20110626	B	soil	2020	6,748,895	68.5	101705	6589	6	59	GCA_010858005.1	Scaffold	1、2
4	NKJ1	B	sputum	This study	6,323,420	68.32	6294401	5821	9	63	SAMN46383563	Complete Genome	This study
5	N-26	B	sputum	2020	6,204,092	68.5	1572682	5738	5	74	GCA_015477225.1	Contig	1、3
6	BJ06-0132	B	sputum	2020	6,385,560	68.5	732995	5861	7	61	GCA_015477975.1	Contig	1、3
7	GUH-2	B	kidney	2012	6,194,645	68.5	6194645	5633	9	61	GCA_000284035.1	Complete Genome	4、5
8	BJ06-0154	B	balf	2020	6,269,615	68.5	91619	5922	5	62	GCA_015477645.1	Contig	1、3
9	760185020	B	NA	2023	6,288,063	68.5	298041	5778	5	60	GCA_029868645.1	Contig	
10	FMUAM8	B	sputum	2022	6,272,366	68.5	6272366	5707	9	60	GCA_027924275.1	Complete Genome	
11	CNM20110649	B	soil	2020	6,294,681	67	66569	5875	6	56	GCA_010868115.1	Scaffold	1、2
12	CNM20110648	B	soil	2020	6,289,931	67	150755	5823	3	57	GCA_010868145.1	Scaffold	1、2
13	CNM20110639	B	soil	2020	6,250,323	67	59864	5881	6	57	GCA_010868155.1	Scaffold	1、2
14	915492502	B	NA	2023	6,178,923	67.5	181536	5514	5	56	GCA_029868385.1	Contig	
15	EML-446	B	environmental	2019	6,520,205	68	605338	6007	3	58	GCA_005863245.1	Contig	1、7
16	EML-1456	B	environmental	2019	6,830,276	68	154491	6288	2	68	GCA_005863295.1	Contig	7
17	BJ06-0121	B	sputum	2020	6,349,165	68.5	1402089	5769	6	63	GCA_015478165.1	Contig	1、3
18	BJ06-0062	B	sputum	2020	6,326,665	68.5	1741771	5736	5	63	GCA_015478435.1	Contig	1、3
19	BJ06-0118	B	sputum	2020	6,326,450	68.5	13592	6251	6	63	GCA_015478155.1	Contig	3
20	NKJ3	B	sputum	This study	6,497,056	68.37	6422549	5835	9	63	SAMN46383564	Complete Genome	This study
21	BJ06-0127	B	sputum	2020	6,370,672	68.5	4694306	5777	4	60	GCA_015478075.1	Contig	1、3
22	CNM20110624	B	soil	2020	6,445,564	68.5	106941	6109	5	60	GCA_010858045.1	Scaffold	1、2
23	BJ06-0133	B	sputum	2020	6,152,653	68.5	637687	5632	5	58	GCA_015477995.1	Contig	1、3
24	NKZ	A	sputum	This study	6,578,448	68.26	6549283	5913	9	61	SAMN46383566	Complete Genome	This study
25	BJ06-0142	A	sputum	2020	6,324,861	68.5	494265	5689	7	59	GCA_015477865.1	Contig	1、3
26	760185029	A	NA	2023	6,463,733	68.5	329,009	5904	3	59	GCA_029868465.1	Contig	
27	3012STDY6756504	A	NA	2019	6,476,621	68	6476621	5819	9	58	GCA_900683635.1	Complete Genome	3

Table 1 (continued)

Strain number	Strain	Clade	Isolated source (or host)	Years	Genome Size (Mb)	GC Content (%)	Contig N50	Gene Count	rRNA	tRNA	Submitted GenBank assembly	Assembly level	References
28	BJ06-0097	A	balf	2020	6,731,365	68	519092	6233	6	59	GCA_015478375.1	Contig	1、3
29	N-32	A	sputum	2020	6,630,200	68	569669	6063	7	61	GCA_015477145.1	Contig	1、3
30	N-18	A	sputum	2020	6,781,869	68	569662	6433	7	61	GCA_015477345.1	Contig	1、3
31	BJ06-0149	A	sputum	2020	6,577,733	68	1114157	5971	7	59	GCA_015477725.1	Contig	1、3
32	BJ06-0071	A	pus	2020	6,449,480	68	459070	5851	3	59	GCA_015478465.1	Contig	3
33	NKD	A	sputum	This study	6,640,201	68.17	6609818	5966	9	59	SAMN46383567	Complete Genome	This study
34	N-51	A	balf	2020	6,795,996	68	347971	6192	6	60	GCA_015476895.1	Contig	1、3
35	N-48	E	sputum	2020	6,075,410	68.5	833897	5515	6	58	GCA_015476905.1	Contig	1、3
36	N-43	E	sputum	2020	6,460,813	68	995316	5893	6	59	GCA_015476925.1	Contig	1、3
37	BJ06-0134	E	sputum	2020	6,428,271	68	1313346	5829	5	62	GCA_015477945.1	Contig	1、3
38	BJ06-0130	E	sputum	2020	6,427,423	68	1313346	5828	5	62	GCA_015478055.1	Contig	1、3
39	760185010	NA*	NA	2023	6,447,607	68.5	318386	5835	5	61	GCA_029868825.1	Contig	
40	MDA3732	D	Homo sapiens	2018	6,592,249	68	191022	6006	5	62	GCA_002933455.1	Contig	3、6
41	741007935	D	NA	2023	6,423,685	68.5	25792	6008	3	60	GCA_029868925.1	Contig	
42	760185019	D	NA	2023	6,692,998	68.5	370570	6118	5	62	GCA_029868665.1	Contig	
43	760185013	D	NA	2023	6,315,214	68.5	430292	5769	5	61	GCA_029868765.1	Contig	
44	760185011	D	NA	2023	6,487,247	68.5	502860	5894	5	62	GCA_029868805.1	Contig	
45	740327409	D	NA	2023	6,308,963	68.5	442312	5784	6	59	GCA_029869825.1	Contig	
46	760185018	D	NA	2023	6,532,988	68.5	315207	5955	6	60	GCA_029868685.1	Contig	
47	760185017	D	NA	2023	6,460,684	68.5	406670	5918	5	58	GCA_029868705.1	Contig	
48	MDA3349	D	Homo sapiens	2018	6,462,637	68.5	6320090	5886	9	63	GCA_002949635.1	Complete Genome	3、6
49	991511872	D	NA	2023	6,3793,13	68.5	396434	5790	5	61	GCA_029868265.1	Contig	
50	740413358	D	NA	2023	6,390,242	68.5	388901	5800	6	61	GCA_029867605.1	Contig	
51	740801120	C	NA	2023	6,413,898	68	312160	5820	3	58	GCA_029869675.1	Contig	
52	740801091	C	NA	2023	6,415,855	68	230439	5830	3	58	GCA_029869725.1	Contig	
53	740192944	C	NA	2023	6,415,331	68	290456	5823	4	58	GCA_029867685.1	Contig	
54	740413266	C	NA	2023	6,653,599	68	275099	6056	4	56	GCA_029867665.1	Contig	
55	740416236	C	NA	2023	6,403,878	68	281749	5813	4	58	GCA_029867565.1	Contig	
56	NBRC-100375	C	NA	2012	6,311,306	68	34812	5912	5	59	GCA_000308555.1	Contig	
57	DSM-44484	C	Homo sapiens	2019	6,262,819	68	185276	5750	3	58	GCA_005863225.1	Contig	5、7

Table 1 (continued)

Strain number	Strain	Clade	Isolated source (or host)	Years	Genome Size (Mb)	GC Content (%)	Contig N50	Gene Count	rRNA	tRNA	Submitted GenBank assembly	Assembly level	References
58	740800516	C	NA	2023	6,351,183	68	218747	5804	5	59	GCA_029867425.1	Contig	
59	760185012	C	NA	2023	6,269,970	68.5	341207	5735	5	59	GCA_029868785.1	Contig	
60	BJ06-0109	NA*	sputum	2020	5,915,323	68.5	315496	5360	3	56	GCA_015478255.1	Contig	1、3
61	NBC-00369	NA*	soil	2024	6,506,653	68.5	6506653	5882	9	61	GCA_036086475.1	Complete Genome	8
62	BJ06-0147	NA*	sputum	2020	6,695,080	68	579911	6208	6	64	GCA_015477705.1	Contig	1、3
63	112071522	NA*	NA	2023	6,416,351	68.5	372374	5882	5	66	GCA_029868005.1	Contig	

NA: strains had no source information

NA*: 760185010, BJ06-0109, NBC-00369, BJ06-0147, and 112071522 did not cluster with any of the five major clades [53, 58, 71, 72, 82–85]

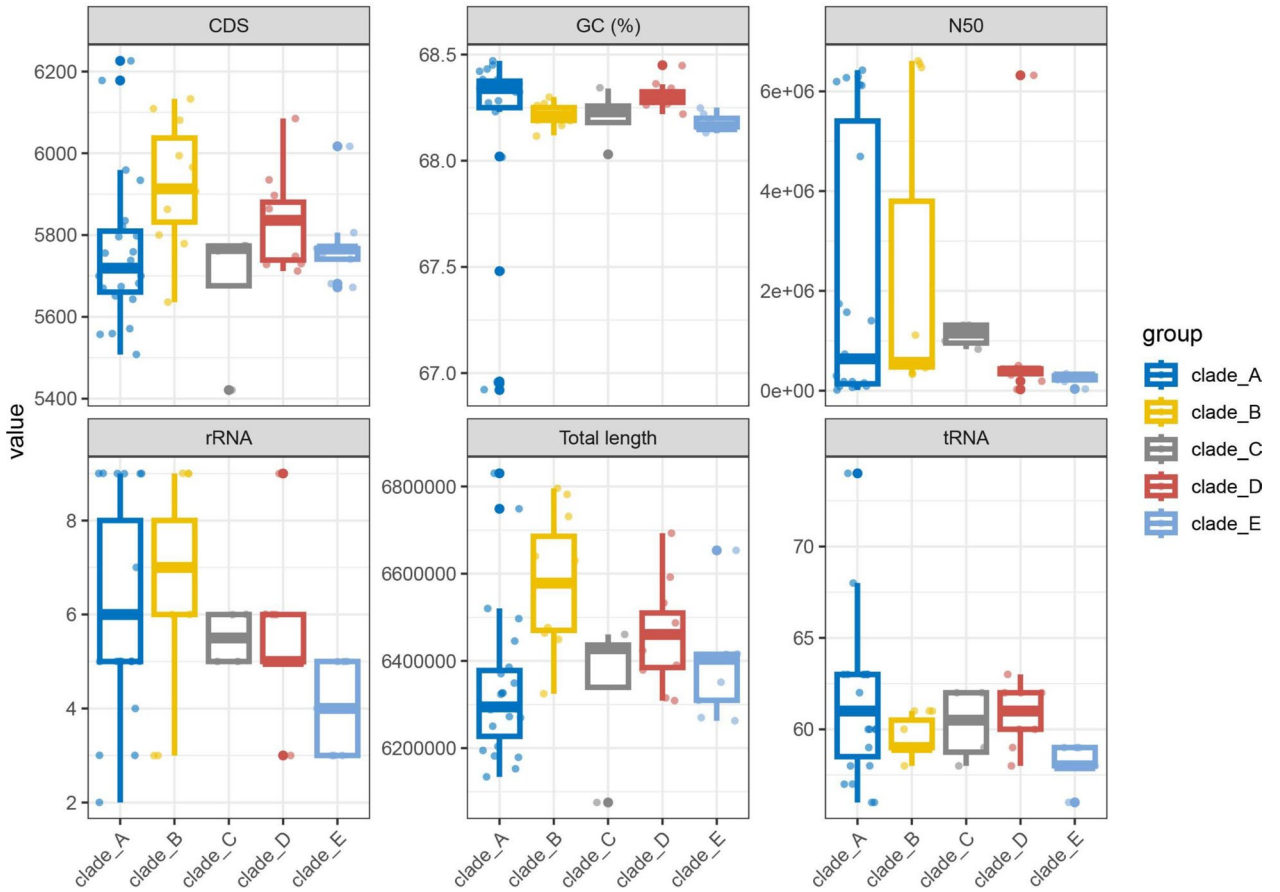

Fig. 1 Genomic features of the investigated strains in this study, including total length, CDS, GC content, N50, rRNA, and tRNA

were distributed among clades A, B, and E, whereas clades C and D predominantly comprised strains with no detailed source information. Five strains (BJ06-0109, BJ06-0147, NBC-00369, 112071522, 760185010) did not cluster with any of the five major clades. Genes with a presence higher than 30% from the accessory genomes were further analysed by DPAC (Dynamic profile analysis for clusters) and hclust (Hierarchical clustering), and the clustering results were consistent with those based on core genes (Figure S1 and Figure S2).

Genomes are classified as belonging to the same species if their pairwise ANI value is ≥95%. The pairwise ANI comparisons for each genome are presented in Table S2. Clade B strains were further divided into five subgroups, consistent with the phylogenetic tree findings (Figure 3). ANI values between genomes within the five subgroups of clade B were below 95%, and clade B strains exhibited ANI values of < 95% when compared to strains in other clades. In contrast, ANI values among strains in clades A, C, D, and E exceeded 95%, indicating high genome sequence identity within these clades.

Pan-genome analysis across five clades

The pan-genome analysis of 63 *N. cyriacigeorgica* strains identified a total of 33,900 genes, with only 1,164 (3.4%) classified as core genes shared by 99% of the strains. The remaining 32,736 genes (96.6%) were categorized as accessory genes, including 305 soft-core genes (present in 95% to < 99% of strains), 7253 shell genes (present in 15% to < 95% of strains), and 25,178 cloud genes (present in < 15% of strains) (Fig. 4).

The analysis further identified 2507 orthogroups distributed across all five clades. Among these, clade A contained the highest number of unique orthogroups (292), followed by clade B (34), clade D (13), clade E (13), and clade C (10). Clades A and B shared 48 orthogroups, suggesting some functional overlap (Fig. 5A). Functional annotation of clade-specific orthogroups, based on the eggNOG database, revealed that a significant proportion of genes in all clades were annotated as "function unknown" (Fig. 5B), highlighting the need for further functional studies.

The indicspecies R package was performed to statistically analyze the associate COGs in each clade,

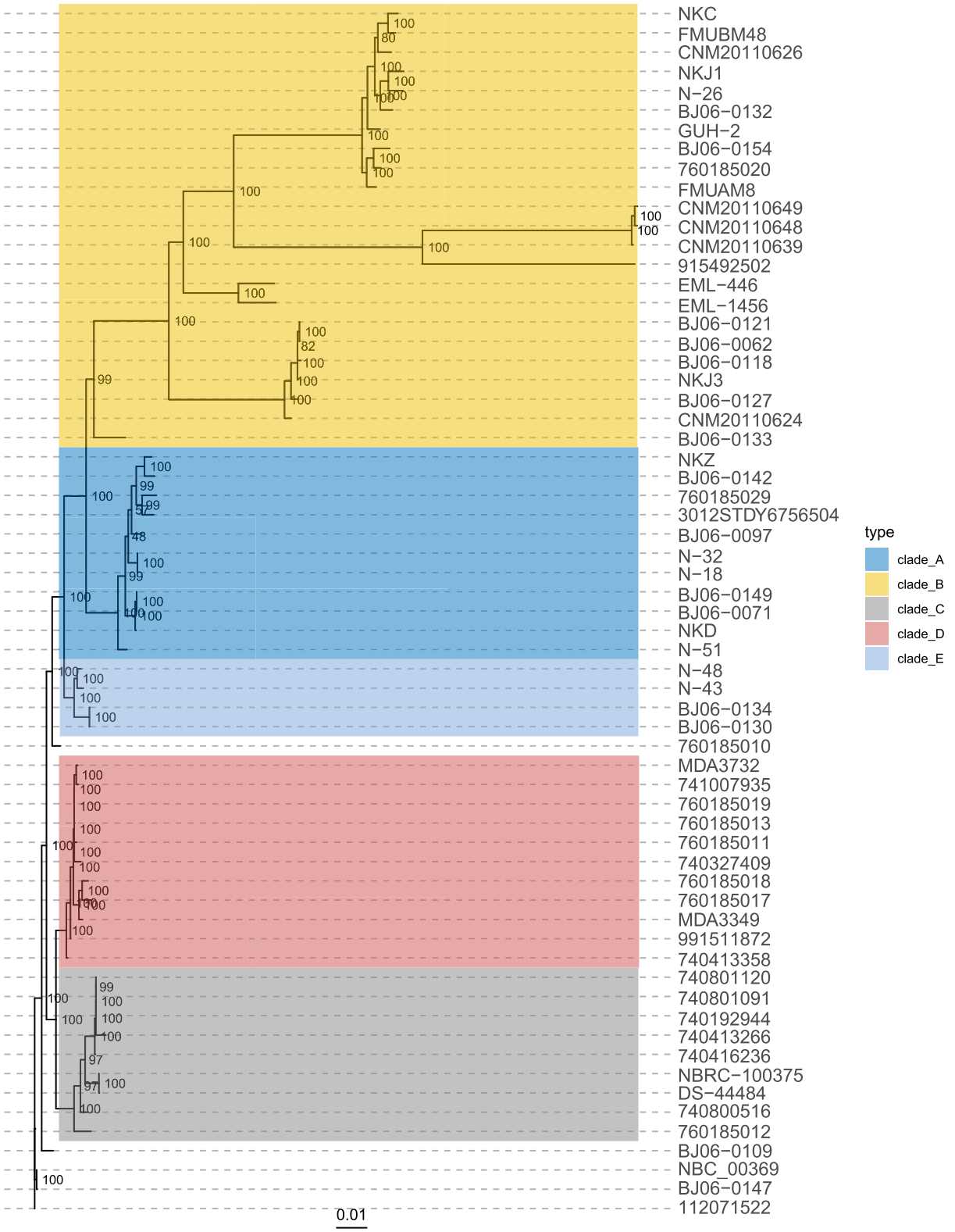

Fig. 2 The phylogenetic tree of *N.cyriacigeorgica* constructed based on the concatenation of the nucleotide sequence of 1164 conserved single-copy core genes

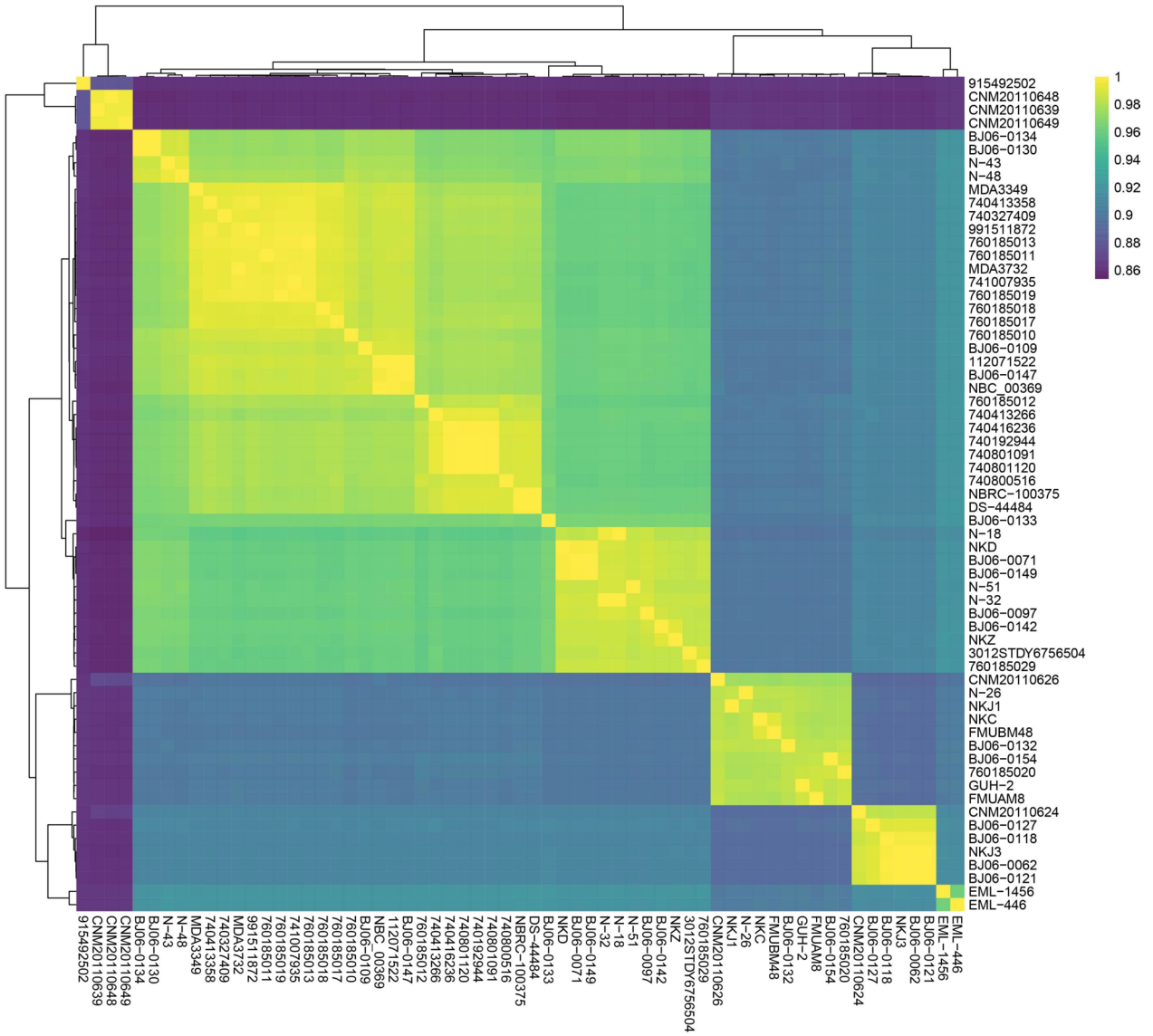

Fig. 3 Heat map of pairwise ANI analyses for 63 *N.cyriacigeorgica* genomes

considering a *P*-value < 0.01 as statistically significant. As shown in Fig. 5B and Table S6, the distribution of genes assigned to Clusters of Orthologous Groups (COG) categories varied among the clades. Clade A, B C, and D exhibited the most diverse set of functional COG categories, with a notable enrichment of genes involved in "replication, recombination, and repair", "transcription", and "amino acid transport and metabolism". However, compare to the clade A, B, C, and D, clade E contained fewer functional COGs associated with "amino acid transport and metabolism". Clade D and E were enriched with genes linked to "replication, recombination and repair" and "inorganic ion transport and metabolism". Notably, clade A harbored unique genes associated with "coenzyme transport and metabolism" which were absent in other clades, while lacking genes related to "signal transduction mechanisms" and "defense mechanisms". These findings suggest that the functional gene repertoire of each clade reflects its potential ecological and pathogenic adaptations.

Distribution of mobile genetic elements in the *N. cyriacigeorgica* genome

The mobile genetic elements (MGEs), such as ISs and plasmids, play a pivotal role in shaping genome structure, driving bacterial evolution, and conferring adaptive advantages. The presence of MGEs was

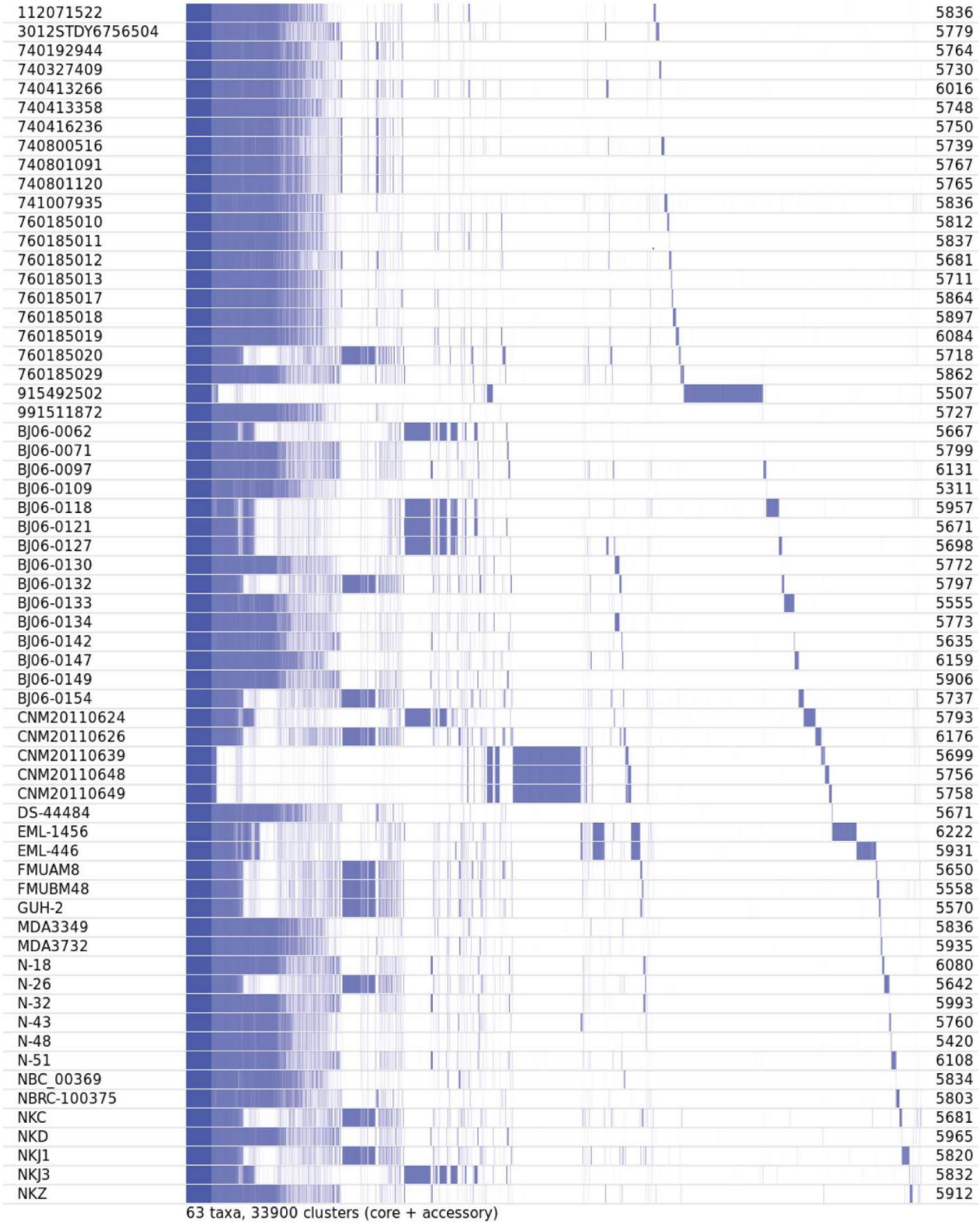

Fig. 4 Heat map of pan-genome analysis of 63 *N.cyriacigeorgica*

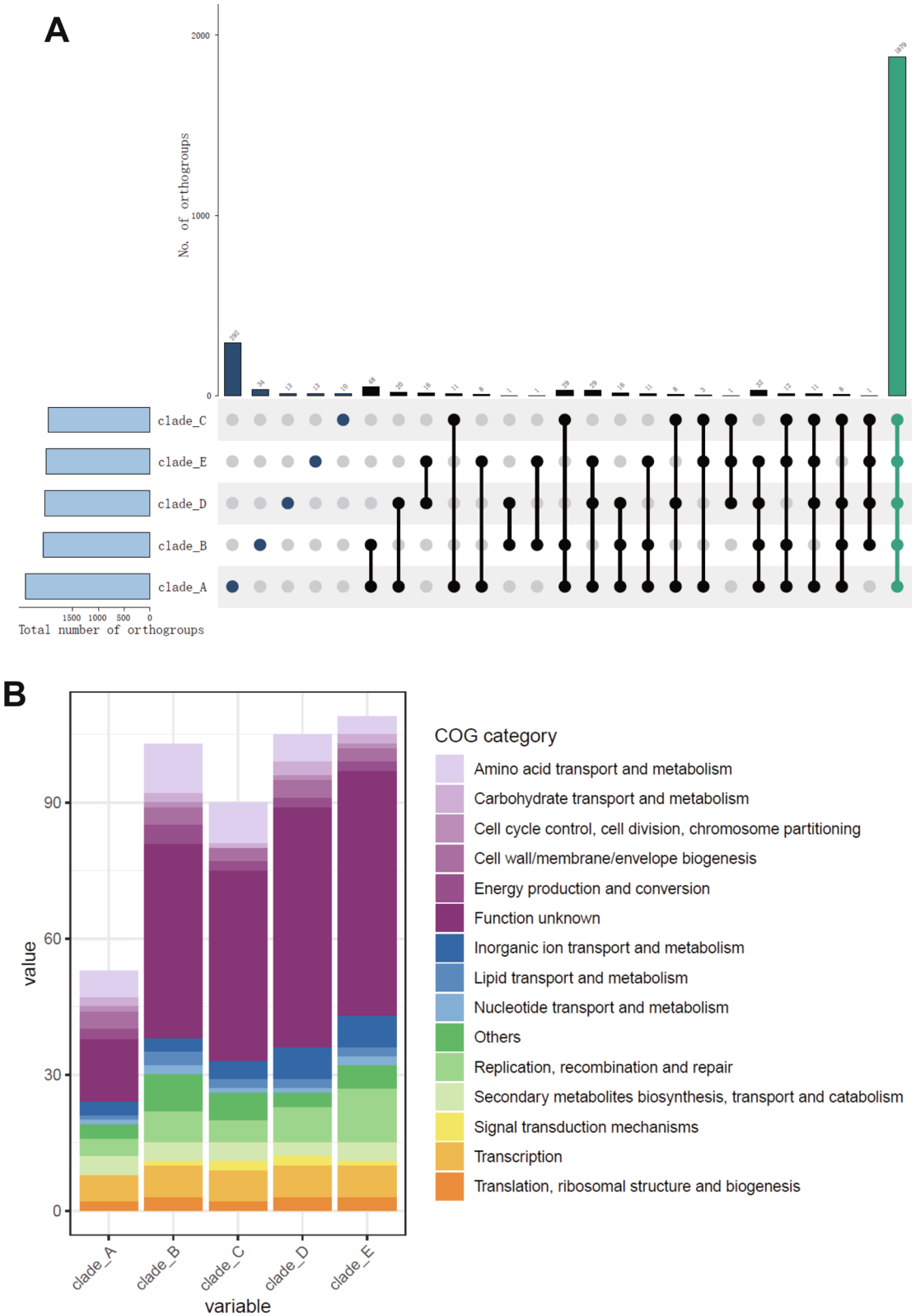

Fig. 5 **A** Distribution of the number of orthogroups in the *N.cyriacigeorgica* genomes. Clade-specific orthogroups are indicated in blue. **B** The associate functional COGs to specific clades

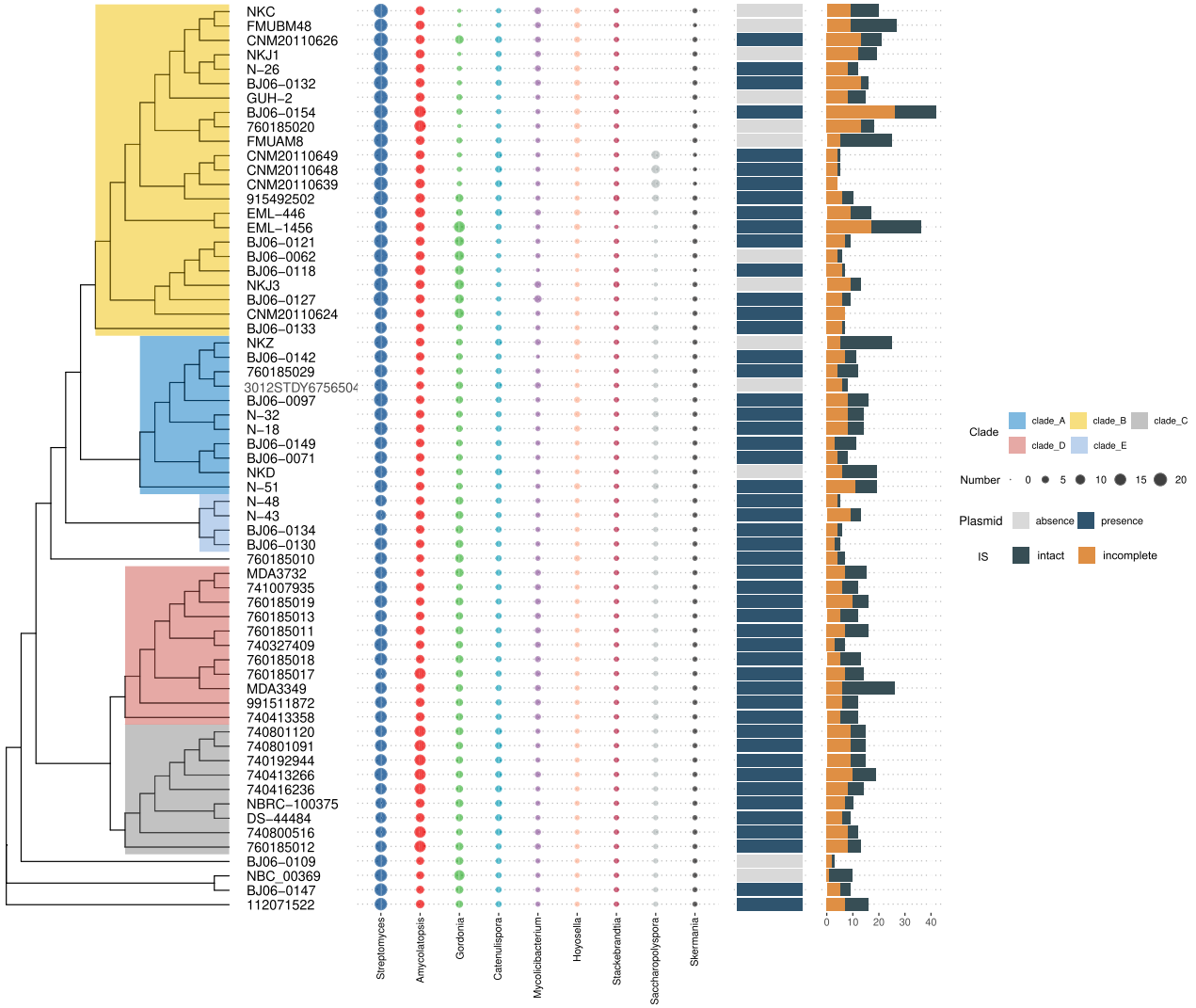

Fig. 6 Distribution of insertion sequences and plasmids in the 63 *N.cyriacigeorgica* genomes

analyzed across the 63 *N. cyriacigeorgica* strains, with their distributions shown in Figs. 6 and 7.

Plasmids were absent in 20.6% (13/63) of the strains, including 10 clinical isolates, 2 strains with no detailed source information, and 1 strain from soil (Fig. 6). Among the identified IS elements, 15 distinct IS families were detected, with the IS3 family being the most prevalent across the strains (Fig. 7). These findings highlight the variability in MGE content among *N. cyriacigeorgica* strains and highlight the potential role of IS elements in mediating genomic plasticity and adaptation.

Putative virulence-associated genes and their distribution in the *N. cyriacigeorgica* genome

To investigate potential virulence factors and assess the pathogenic characteristics of *N. cyriacigeorgica*, BLASTP searches of its CDSs were conducted against the Virulence Factor Database (VFDB) using stringent criteria (E-value < 1E-6, sequence length overlap > 60%, sequence similarity > 60%). A total of 208 candidate virulence-associated genes were identified, encompassing a wide range of functions critical for bacterial pathogenesis. These include roles in adherence and invasion, secretion systems, stress adaptation, nutrient acquisition, toxin production, modulation of phagocyte function, and intracellular survival within macrophages (Table 2).

The mammalian cell entry (*mce*) gene family, known for encoding proteins that facilitate *Mycobacterium tuberculosis* invasion and survival within macrophages, was represented by nine clusters in *N. cyriacigeorgica*. Notably, the *mce1* and *mce2* clusters were absent in all strains, while *mce5*, *mce6*, and *mce9* were universally present. The remaining *mce* genes (*mce3*, *mce4*, *mce7*,

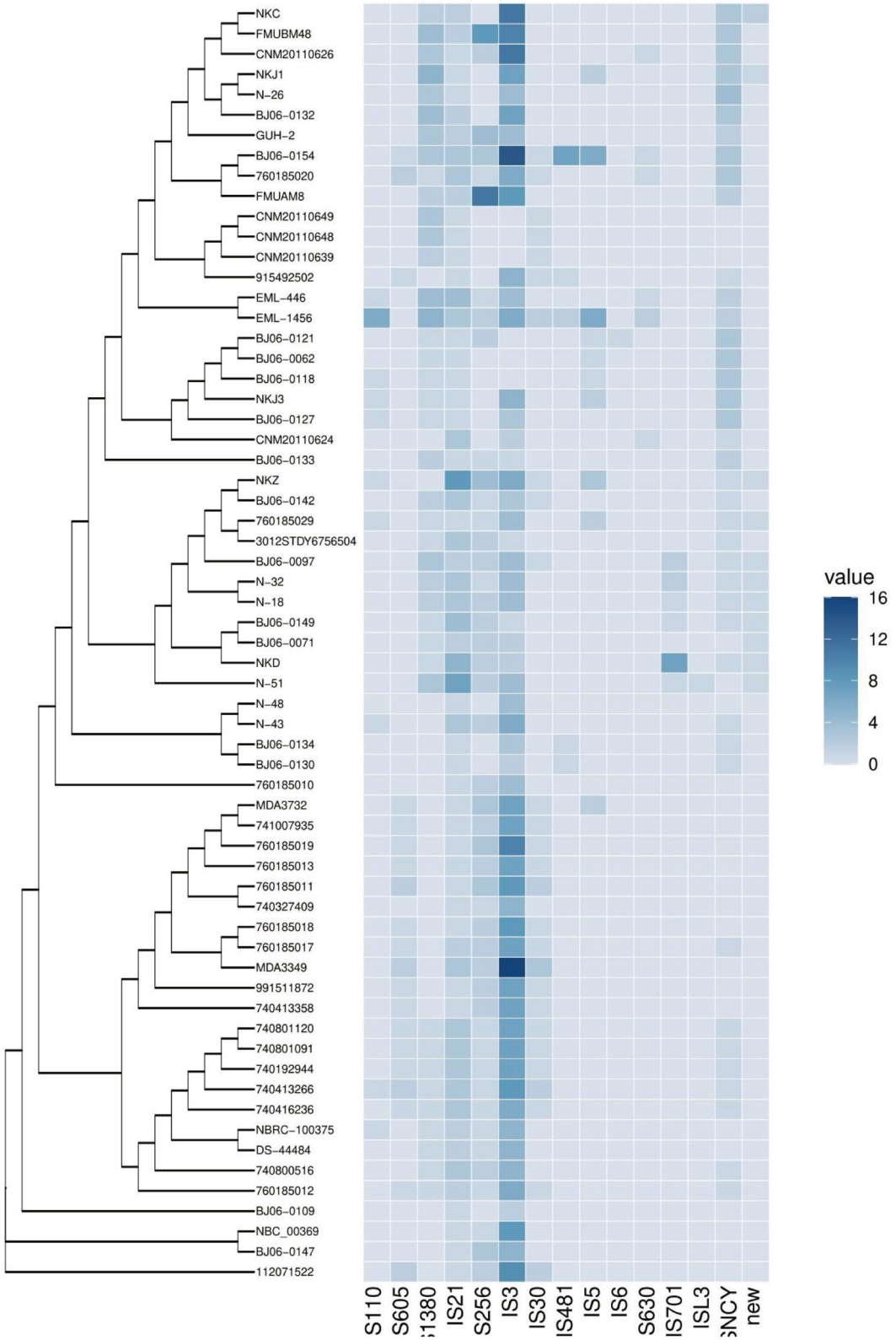

Fig. 7 Members of the IS families identified in the analysed strains

Table 2 Virulence factor candidates in the *Nocardia cyriacigeorgica* according to the Virulence Factor Database (VFDB)

Category	Gene name	Gene ID	Description
Adherence	clpC	VFG000079	Endopeptidase Clp ATP-binding chain C, is required for adhesion and promoting early escape form the phagosome of macrophages
	dnaK	VFG050290	Chaperone protein
	eno	VFG005576	Phosphopyruvate hydratase
	groEL	VFG012095	Chaperone protein
	htpB	VFG001855	60 K heat shock protein HtpB
	plr/gapA	VFG005358	Glyceraldehyde-3-phosphate dehydrogenase
	tufA	VFG046465	Elongation factor Tu
Invasion	Mce3	VFG010120	Mammalian cell entry family proteins
	Mce4	VFG010259	Mammalian cell entry family proteins
	Mce5	VFG010296	Mammalian cell entry family proteins
	Mce6	VFG010326	Mammalian cell entry family proteins
	Mce7	VFG010381	Mammalian cell entry family proteins
	Mce8	VFG010393	Mammalian cell entry family proteins
	Mce9	VFG010427	Mammalian cell entry family proteins
Stress survival	ahpC	VFG001404	Putative alkylhydroper oxide reductase
	clpP	VFG000077	Serine protease involved in proteolysis and is required for growth under stress conditions
	katA	VFG037028	Catalase, protects against reactive oxygen species
	katG	VFG001396	Catalase peroxidase
	sodA	VFG001421	Putative superoxide dismutase, important for intracellular survival and transmission
	sodC	VFG001380	Putative superoxide dismutase, important for intracellular survival and transmission
Nutritional/metabolic factor	irtB	VFG024069	ABC transporter ATP-binding protein
	carB	VFG047708	Carbamoyl phospate synthase large subunit
	ctpV	VFG031419	Copper exporter
	dahP	VFG044426	DAHP synthetase
	entB	VFG048419	Enterobactin, involved in Iron uptake
	fbpC	VFG001811	Secreted antigen 85-C, is an important immunomodulatory component of the cell wall
	hemL	VFG013203	Glutamate-1-semialdehyde-2,1-aminomutase
	irp5	VFG016174	Yersiniabactin biosynthetic protein irp5
	mbtA	VFG001817	Salicyl-AMP ligase + salicyl-S-ArCP synthetase
	pchH	VFG001267	Pyochelin
	mbtB	VFG001403	Phenyloxazoline synthase
	mbtC	VFG001823	Polyketide synthetase
	mbtE	VFG001821	Peptide synthetase
	mbtF	VFG001820	Peptide synthetase
	mbtG	VFG001819	Lysine-N-oxygenase
	mbtH	VFG001818	Putative protein
	mbtI	VFG009561	Isochorismate synthase
	narG	VFG001391	Nitrate reductase alpha subunit
	narH	VFG001814	Nitrate reductase beta subunit
	narI	VFG001816	Nitrate reductase gamma subunit
	narJ	VFG001815	Nitrate reductase delta subunit
Immune modulation	adhD	VFG030297	MymA operon, involved in mycolic acid synthesis
	gndA	VFG048830	6-phosphogluconate dehydrogenase
	mps1	VFG029717	GPL locus
	wcbT	VFG002546	8-amino-7-oxononanoate synthase, involved in antiphagocytosis
	ptpA	VFG031511	Protein-tyrosine phosphatase, involved in phagosome arresting
	ndk	VFG031473	Nucleoside diphosphate kinase, involved in phagosome arresting
Regulation	devR/dosR	VFG001824	Two component transcriptional regulator, LuxR family
	ideR	VFG001406	Positive regulator of oxidative stress responses

Table 2 (continued)

Category	Gene name	Gene ID	Description
	relA	VFG001826	Probable GTP pyrophosphokinase, associated with stationary phase adaptation and long-term survival
	phoP	VFG001386	Possible two component system response transcriptional positive regulator
Effector delivery system	eccB3	VFG022047	Type VII secretion system protein
	eccC4	VFG022232	Type VII secretion system protein
	espR	VFG021908	Type VII secretion system protein
	ricA	VFG045340	Rab2 interacting conserved protein A
Motility	flmH	VFG038840	Polar flagella
Exotoxin	pgi	VFG038132	Endotoxin, glucose-6-phosphate isomerase
	rtxB	VFG038918	RTX toxin transporter
	rtxE	VFG007023	RTX toxin transporter

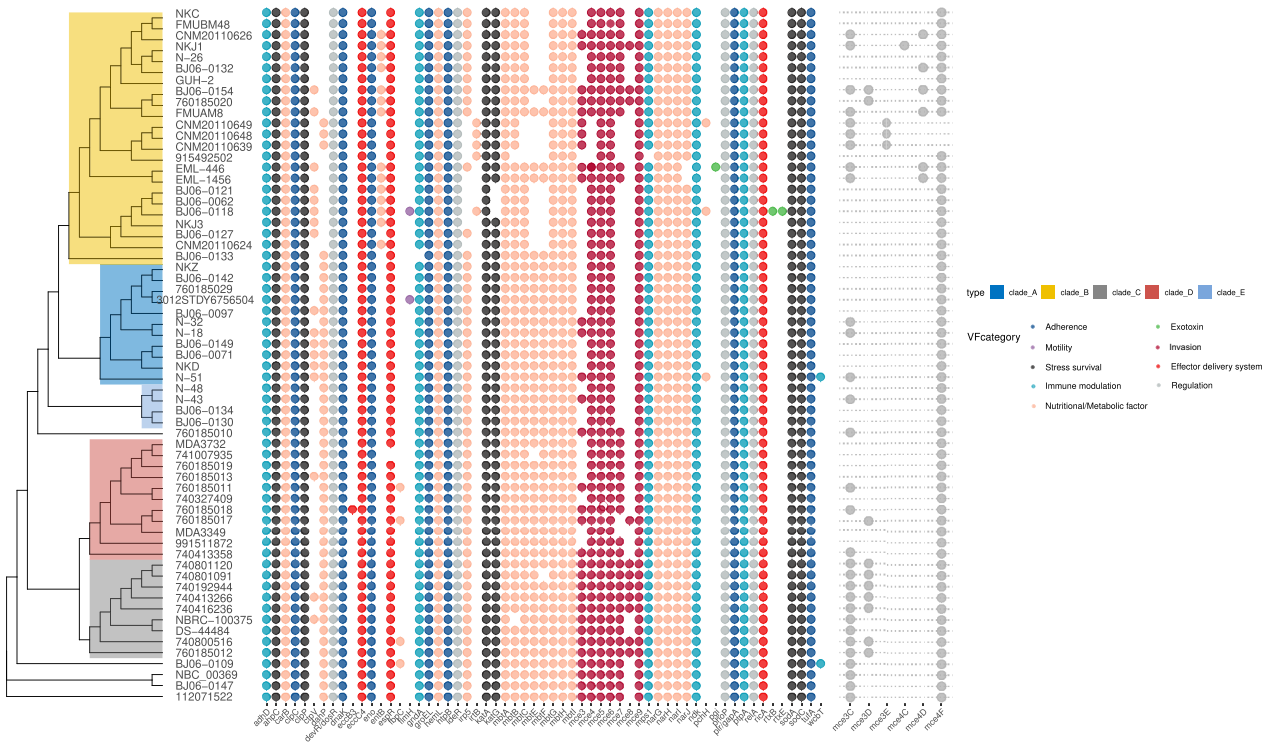

Fig. 8 Putative virulence-associated genes in the genomes of *N.cyriacigeorgica*. The distribution of the *mce3* and *mce4* gene family are shown on the right side. Each dot indicates the presence of the gene

and *mce8*) exhibited distinct distribution patterns across the 63 analyzed strains (Fig. 8 and Table S3).

The *mce4* gene was present in nearly all strains, except for three strains from oil sources in clade B. A specific subunit, *mce4F*, was detected in 95.2% (60/63) of strains. In contrast, *mce3* was identified in 32 strains, of which only 10 were recognized as human pathogens. All strains in clade C carried *mce3*, compared to 43.5% (10/23) in clade B, 27.3% (3/11) in clade A, 25% (1/4) in clade E, and 36.4% (4/11) in clade D. Subunits mce3C and *mce3D* were detected in 100% (9/9) and 77.8% (7/9) of clade C strains, respectively, while *mce3E* was identified exclusively in three oil-derived strains from clade B.

The *mce7* gene was predominantly detected in clades D and C, but it was absent in clades A and E. *mce8* was identified in 11 strains, including 3 strains from clade B, 1 from clade D, and 7 from clade C. Analysis of the *mce* gene family revealed that 27 strains carrying both *mce3C*

and *mce4F* were categorized into five distinct distribution patterns, as summarized in Table S4.

The *fbpC* gene, encoding antigen 85C (Ag85C), a protein involved in *Mycobacterium tuberculosis* cell wall assembly, was detected in only 6.3% (4/63) of strains. Genes associated with oxidative and nitrosative stress responses (*katA*, *sodC*, *sodA*, *narG*, *narH*, *narI*), iron import (*ideR*), and antioxidant defense (*ahpC*) were universally present across all strains. Similarly, genes encoding the Type VII secretion system (*eccC4*) and the effector delivery system (*ricA*) were also identified in all strains.

Genes critical for intracellular survival, such as *ndk* (nucleoside diphosphate kinase) and *ptpA* (protein tyrosine phosphatase A), which facilitate macrophage phagosomal arrest, were consistently detected in all strains. Interestingly, enterobactin-encoding genes (*entB*), responsible for siderophore-mediated iron acquisition, were found in 60.9% (14/23) of clade B strains but were absent in other clades. In contrast, yersiniabactin biosynthetic protein-related genes (*irp5*) were detected in all strains from clades A, C, D, and E, but were present in only 56.5% (13/23) of clade B strains. These findings highlight the conserved nature of key virulence-associated genes in *N. cyriacigeorgica*, while also revealing clade-specific variations that may influence pathogenicity and host interactions.

Comprehensive identification and distribution of antibiotic resistance genes in the *N. cyriacigeorgica* genome

A systematic screening for ARGs in the *N. cyriacigeorgica* pan-genome, conducted using BLASTP against the Comprehensive Antibiotic Resistance Database (CARD), identified 268 resistance genes spanning at least 12 distinct antibiotic classes. These genes encode a variety of resistance mechanisms, including efflux pumps, β-lactamases, and plasmid-mediated methyltransferases, which confer resistance through antibiotic inactivation, active efflux, and target modification (Table 3).

Resistance genes targeting multiple antibiotic classes were observed across all analyzed strains. These included cephalosporins (*AST-1*), aminoglycosides (*kdpE*), fluoroquinolones (*SoxR*, *gyrB*), glycopeptides (*rpoC*), macrolides (*oleC*, *oleB*, *carA*), tetracyclines (*tetA*(58), *tetB*(58)), sulfonamides (*folC*, *folP*), rifamycins (*rbpA*, *rpoB2*), and fosfomycin (*murA*, *AbaF*) (Fig. 9 and Table S5). Among these, β-lactamase-encoding genes, such as *CTX-M* and *KPC*, were highly prevalent, underscoring the widespread potential for β-lactam resistance in *N. cyriacigeorgica*.

While β-lactamase genes such as *CTX-M* and *KPC* were broadly distributed, certain β-lactamase genes displayed clade-specific patterns. For example, the β-lactamase genes *bcl* and *bla1* were restricted to clade B, being present in 43.5% (10/23) and 21.7% (5/23) of strains, respectively. These genes were absent in strains from other clades, suggesting localized selective pressures influencing β-lactam resistance within clade B.

Resistance genes for vancomycin (*VanSO* and *VanRO*) were widely distributed across clades A, C, D, and E, but were present in only 43.5% (10/23) of clade B strains. Conversely, the macrolide resistance gene *erm(O)-Irm* was detected in 43.5% (10/23) of clade B strains but was absent in other clades.

The macrolide efflux pump gene *srmB* exhibited a distinct distribution, being present in 78.9% (30/38) of strains from clades A, B, and E, but absent in clades C and D. The plasmid-mediated macrolide resistance gene *ermH* was identified in 69.6% (16/23) of clade B strains but was not detected in clades A, C, D, or E. These findings highlight the heterogeneous distribution of macrolide resistance mechanisms among clades.

Rifamycin resistance genes (*Rox-nf* and *Rox-sv*) and transposon-mediated streptogramin resistance genes (*VatF*) were exclusively identified in four strains from clade B. Additionally, chromosome-encoded chloramphenicol resistance genes (*cmlR*) were detected in 21.7% (5/23) of clade B strains but were present in all strains from other clades. These clade-specific distributions suggest that localized environmental or clinical pressures may have driven the evolution of these resistance mechanisms.

Discussion

Consistent with previous research [53], our analysis revealed that *N. cyriacigeorgica* consists of five distinct clades, as determined by the concatenation of 1164 conserved single-copy core genes. Within clade B, five subgroups were further delineated using pairwise ANI analysis. However, inconsistencies between the phylogenomic tree and ANI values were noted for specific strains. Specifically, three strains isolated from oil samples (CNM20110649, CNM20110648, CNM20110639) within clade B demonstrated high homology with each other but exhibited low genetic relationships between other stains. This observation indicates a notable evolutionary divergence between these three isolates and other clade B strains.

Low ANI values between clade B strains and those from other clades suggest high genomic diversity within clade B, potentially representing distinct species. Interestingly, five strains (BJ06-0109, BJ06-0147, NBC-00369, 112071522, 760185010) could not be assigned to any clades in the phylogenomic tree. Despite this, ANI values exceeding 95% between these strains and strains from clades A, C, D, and E suggest high genomic similarity.

Table 3 Antibiotic resistance genes in the *Nocardia cyriacigeorgica* genome according to the Comprehensive Antibiotic Resistance Database (CARD)

Resistance gene type	Gene name	Gene ID	Description
β-lactam antibiotics	CTX-M	ARO:3001961	Cephalosporin
	AST-1	ARO:3004740	Broad-spectrum beta-lactamase gene
	KPC	ARO:3002312	Carbapenem; cephalosporin
	Bcl	ARO:3002877	Cephalosporin
	Bla1	ARO:3000090	Beta-lactamase gene
Aminoglycoside antibiotics	kdpE	ARO:3003841	Transcriptional activator that is part of the two-component system
Fluoroquinolone antibiotics	SoxR	ARO:3003381	Antibiotic efflux; antibiotic target alteration
	gyrA	ARO:3007473	Point mutation of *Mycobacterium tuberculosis* gyrA resulted in the lowered affinity between fluoroquinolones and gyrA
	gyrB	ARO:3003304	Point mutation in *Mycobacterium tuberculosis* gyrB resulting in fluoroquinolone resistance, antibiotic target alteration
Glycopeptide antibiotics	VanSO	ARO:3002941	Vancomycin resistance operon genes, which can synthesize peptidoglycan with modified C-terminal D-Ala-D-Ala to D-alanine-D-lactate
	VanRO	ARO:3002930	Vancomycin resistance operon genes, which can synthesize peptidoglycan with modified C-terminal D-Ala-D-Ala to D-alanine-D-lactate
	rpoC	ARO:3004681	Point mutations in the rpoC region of *Clostridioides difficile* which confer resistance to vancomycin
Macrolide antibiotics	oleC	ARO:3003748	ABC transporter isolated from *Streptomyces antibioticus* and is involved in oleandomycin secretion
	oleB	ARO:3003036	ABC-F subfamily protein in *Streptomyces antibioticus* and is involved in oleandomycin secretion
	erm(O)-lrm	ARO:3004652	
	carA	ARO:3002817	ABC-F subfamily protein involved in macrolide resistance
	srmB	ARO:3002828	Macrolide efflux pump genes
	lmrC	ARO:3002881	ABC-F subfamily protein that confers resistance to lincosamides in *Streptomyces lincolnensis* and *Lactococcus lactis*
	ErmH	ARO:3002823	Plasmid-mediated methyltransferase found in *Streptomyces thermotolerans*
	ErmR	ARO:3000594	Plasmid-mediated methyltransferase found in *Streptomyces thermotolerans*
Tetracycline antibiotics	tetA(58)	ARO:3003980	Tetracycline efflux pump
	tetB(58)	ARO:3003981	Tetracycline efflux pump
Sulfonamide antibiotics	folC	ARO:3004157	Point mutations in the dihydrofolate synthetase folC gene shown clinically to confer resistance to p-aminosalicylic acid or other aminosalicylates
	folP	ARO:3003389	Point mutations in *Streptococcus pyogenes* dihydropteroate synthase folP prevent sulfonamide antibiotics from inhibiting its role in folate synthesis, thus conferring sulfonamide resistance
Rifamycin antibiotics	Rox-nf	ARO:3007210	Rifampin monooxygenase that inactivates rifampin
	Rox-sv	ARO:3007209	Class A flavoprotein monooxygenase that confers resistance to rifamycin antibiotics
	Iri	ARO:3002884	Monooxygenase that confers resistance to rifampin
	rbpA	ARO:3000245	The enzymatic inactivation of rifampin by phosphorylation at the 21-OH position
	rpoB2	ARO:3000501	Expression of the rpoB2 variant results in replacement of rifampin sensitivity with rifampin resistance
Streptogramin antibiotics	VatF	ARO:3003744	Transposon-mediated acetyltransferase
	VatE	ARO:3002844	Transposon-mediated acetyltransferase
Fosfomycin antibiotics	murA	ARO:3003784	Chlamydia murA confers intrinsic resistance to fosfomycin
	AbaF	ARO:3004573	Expression of abaF in *E. coli* resulted in increased resistance to fosfomycin
Chloramphenicol antibiotics	cmlR	ARO:3002690	Chromosome-encoded chloramphenicol phoshotransferase
Nitroimidazole antibiotic	msbA	ARO:3003950	Multidrug resistance transporter homolog

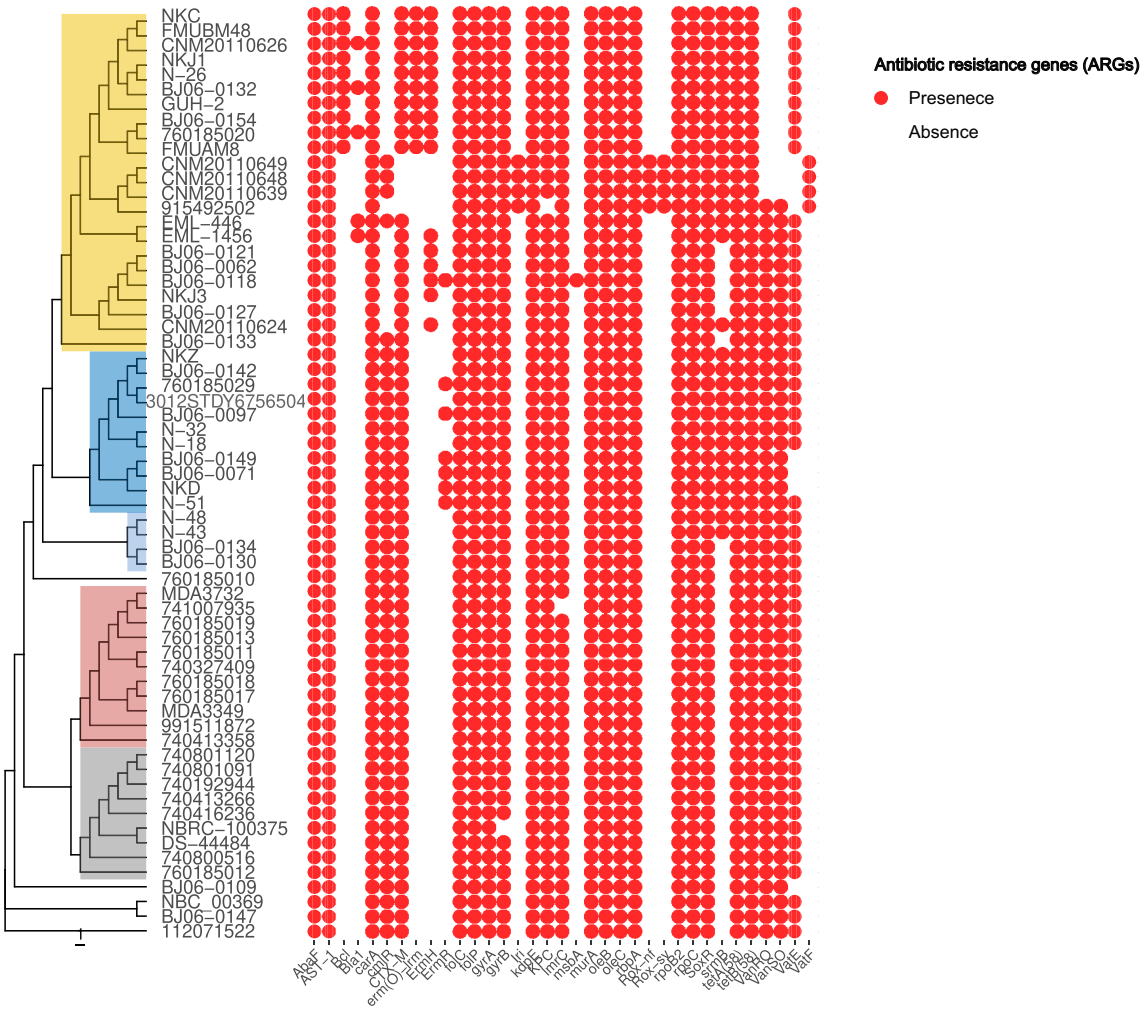

Fig. 9 Antibiotic resistance genes in the genomes of *N.cyriacigeorgica*. Each red dot indicates the presence of the gene

These inconsistencies likely arise from methodological differences: phylogenomic trees rely on conserved single-copy core genes, whereas ANI analysis incorporates whole-genome comparisons, including rapidly evolving genes. To further investigate evolutionary characteristics, clade-specific orthologous groups were analyzed using the eggNOG database. Functional differences among unique genes in the five clades suggest their potential roles in environmental adaptation. However, the presence of numerous orthologous groups with "unknown function" highlights the need for additional research to elucidate the evolutionary features of *N. cyriacigeorgica*.

Nocardia species exhibit diverse antimicrobial resistance profiles. Previous studies [54, 55] have classified *N. cyriacigeorgica* within the type VI drug resistance pattern, characterized by susceptibility to amikacin, cefamandole, cefotaxime, ceftriaxone, and imipenem, but resistance to amoxicillin-clavulanic acid, ampicillin, ciprofloxacin, clarithromycin, and erythromycin. In this study, the antimicrobial susceptibility profiles of five clinical strains largely aligned with the type VI pattern, except for clarithromycin. Imipenem resistance has been reported at high rates in *N. cyriacigeorgica*, with one study documenting an 86.7% resistance rate (52/60 strains) [53]. Similarly, McGuinness et al. [56] observed elevated resistance to imipenem in *N. cyriacigeorgica* strains. In contrast, all five clinical strains analyzed in our study exhibited complete sensitivity to imipenem (100%, 5/5). This discrepancy may result from geographical variation or the limited sample size in our analysis.

Antibiotic resistance represents a critical global public health challenge, with the proliferation of ARGs contributing to the rise of multidrug-resistant pathogens. In the genomes of the analyzed *N. cyriacigeorgica* strains, numerous ARGs were identified, corresponding to four primary resistance mechanisms: (1) mutations or regulatory changes in target enzymes, (2) efflux pump-encoding

genes, (3) inhibition of protein or cell wall synthesis, and (4) antibiotic inactivation.

Sulfonamides, particularly trimethoprim-sulfamethoxazole (TMP-SMX), remain the cornerstone of treatment for *Nocardia* infections [57]. However, prolonged or suboptimal antibiotic exposure increases the risk of resistance development. TMP-SMX resistance is often linked to mutations in the *folP* and *folA* genes, which encode dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR), respectively. A prior study demonstrated that adaptive mutations in *folP* and *folP2* reduced TMP-SMX sensitivity in eight strains of *Nocardia nova* and two strains of *N. cyriacigeorgica* [58]. Furthermore, resistant *folP* variants have been identified on MGEs or plasmids, facilitating interspecies transmission [59–61]. Although detailed resistance phenotypes for all tested strains in our study remain unresolved, the detection of the *folP* gene in all strains suggests a widespread prevalence of TMP-SMX resistance genes. This finding underscores the importance of monitoring resistance mechanisms in *N. cyriacigeorgica* to ensure effective clinical management and mitigate the spread of resistance.

Plasmid-encoded resistance genes, such as *ermH* and *ermR*, which confer macrolide resistance, were identified in the majority of strains within clade B (69.6%, 16/23) and clade A (54.5%, 6/11), respectively. This observation suggests the potential for the evolution and dissemination of resistance genes among strains from diverse sources via plasmids. Horizontal gene transfer, mediated by genetic elements such as plasmids and ISs, has been shown to significantly contribute to the dissemination of antibiotic resistance among bacteria [62, 63]. In our study, most strains carried plasmids, and a wide variety of IS elements were detected. These findings underscore the importance of preventing the horizontal transfer of ARGs among *Nocardia* strains to mitigate the worsening of antimicrobial resistance.

Despite the presence of various resistance genes across all strains, differences in the distribution of specific genes, including *VanSO*, *VanRO*, *erm(O)-Irm*, *srmB*, *ermH*, *bcl*, *bla1*, and *cmlR*, were observed. Particularly noteworthy, we found that 80% (8/10) of strains carrying the macrolide resistance gene *erm(O)-Irm* in clade B, over 60% (19/30) of strains carrying *srmB* in clades B, A, and E, and 75% (12/16) of strains carrying *ermH* in clade B were isolated from human specimens. These findings suggest that the antimicrobial resistance genotypes of *N. cyriacigeorgica* strains may be associated with their isolation origin. The observed variation in resistance genes among strains from different sources highlights the potential for distinct resistance mechanisms in clinical strains, warranting further investigation.

The virulence mechanisms of *N. cyriacigeorgica* remain poorly characterized. Known virulence factors in *Nocardia* species include catalase, superoxide dismutase (SOD), secreted toxins, and cell wall proteins [64, 65]. However, the virulence factors vary significantly among different pathogenic *Nocardia* species [9]. In this study, 208 potential virulence genes were identified, which may contribute to *N. cyriacigeorgica* pathogenesis and its interactions with host cells. Among these, the Mce family proteins, recognized as key virulence factors in *M. tuberculosis* and *Nocardia* spp., plays a critical role in host cell invasion and intracellular survival within macrophages [66–68]. The *mce* gene family comprises nine clusters (*mce1–9*), with mce4F confirmed as essential for bacterial pathogenesis in *M. tuberculosis* [69]. For instance, Mce3C as a surface protein has been shown to promote mycobacterial adhesion to and invasion of macrophages [70]. Phylogenetic and pan-genome analyses of 141 *Nocardia* genomes revealed the absence of *mce1* and *mce2* in 27 *N. cyriacigeorgica* strains, while highlighting the pathogenic roles of *mce3C* and *mce4F* in *N. keratitis* [71]. Our findings demonstrated that *mce3*, *mce4*, *mce5*, *mce6*, *mce7*, *mce8*, and *mce9* were present in most tested strains, whereas mce1 and *mce2* were absent. Further analysis revealed that all strains, except for three environmental strains from oil, harbored *mce4F*. Additionally, 27 of the 63 strains carried both *mce4F* and *mce3C*. The widespread presence of *mce4F* suggests its potential as a reliable detection marker. Although *mce3C* exhibited varied distributions across the five clades, no correlation was observed between the *mce3C* gene and strain origin, possibly due to incomplete source information for some strains. Further investigation is required to clarify the roles of *mce4F* and *mce3C* in the pathogenic mechanisms of *N. cyriacigeorgica*.

Antioxidant proteins such as catalase and SOD are believed to counteract the oxidative killing mechanisms of phagocytes [64]. Consistent with previous studies on *Nocardia* species, including *N. seriolae* UTF1, *N. brasiliensis* HUJEG-1, and *N. cyriacigeorgica* GUH-2 [72–74], our study identified two catalase genes (*katA* and *katG*) and two SOD genes (*sodA* and *sodC*) in the majority of *N. cyriacigeorgica* genomes. Intracellular pathogens must also adapt to low-oxygen conditions to survive within host cells. In our study, nitrosative stress-related genes (*narG*, *narH*, and *narI*) were present in all analyzed strains. However, *nirB* and *nirD* were not detected, contrasting with previous genomic research on *N. cyriacigeorgica* GUH-2 [72]. The importance of catalase, SOD, and nitrate reductase genes in the pathogenesis of *N. brasiliensis* has been demonstrated in studies of genomic changes associated with the loss of virulence after 200 continuous subcultures in mice [75]. Based on

our comparative genomic results, *katA*, *katG*, *sodA*, *sodC*, *narG*, *narH*, and *narI* may serve as potential targets for further investigation into the pathogenic mechanisms of *N. cyriacigeorgica*.

When engulfed by phagocytes, intracellular pathogens are enclosed within phagosomes, which subsequently fuse with vesicles to form phagolysosomes for the degradation of ingested particles [76]. Previous studies have shown that *Nocardia* can survive within host cells and evade immune responses by inhibiting phagosome-lysosome fusion and reducing intracellular acid phosphatase levels in macrophages [12]. PtpA and Ndk proteins have been implicated in arresting macrophage phagosomal maturation, enabling pathogens to evade immune clearance and establish persistent infections [77]. Our study identified the presence of the *ptpA* and *ndk* genes in all strains, which are associated with phagosome arrest. These findings underscore the necessity of further investigation into the roles of *ptpA* and *ndk* as potential candidate genes in phagosomal maturation arrest. Future studies could utilize gene knockout experiments, confocal microscopy analysis, and multi-omics approaches to investigate their functions.

The *entB* gene, responsible for enterobactin production, has been identified as a virulence factor in enterobacteria [78–80]. Notably, Han et al. demonstrated that the siderophore virulence gene *entB* plays a critical role in significantly enhancing the virulence of carbapenem-resistant *K. pneumoniae* (CRKP) strains [81]. In our study, *entB* was detected in over 60% (14/23) of strains within clade B, with the majority of these strains (64.3%, 9/14) originating from clinical sources. This observation indicates clade-specific variations in *entB* distribution. To date, limited research has been conducted on the role of *entB* as a siderophore virulence gene in intracellular pathogens. Further studies are warranted to elucidate its contribution to the virulence of *N. cyriacigeorgica*.

Conclusions

Phylogenetic analysis based on single-copy genes identified five major clades within *N. cyriacigeorgica*. This finding enhances our understanding of the species' genetic diversity and evolutionary relationships. Additionally, plasmid-mediated ARGs were identified within the *N. cyriacigeorgica* chromosome, underscoring the potential risk of horizontal transmission of drug resistance genes via MGEs. Comparative genomic analysis identified several pathogenic genes that may play critical roles in the virulence of *N. cyriacigeorgica*. However, the precise functions and mechanisms of these genes in pathogenicity remain unclear. These findings emphasize the need for further comprehensive studies to investigate the molecular basis of *N. cyriacigeorgica* pathogenesis and its implications for clinical treatment strategies.

Limitations

This comparative genomic study has several limitations that warrant acknowledgment. First, the lack of detailed source information for some strains from the NCBI database may have introduced bias, potentially affecting the accuracy of the phylogenetic and genomic findings. Second, although ISs and plasmids within the *N. cyriacigeorgica* chromosome were analyzed, potential virulence or ARGs on plasmids were not investigated. Such an investigation would offer a more comprehensive assessment of the risks linked to horizontal gene transfer of these elements.

Third, while the *mce3* and *mce4* gene families were evaluated, the pathogenicity of strains carrying both *mce3C* and *mce4F* genes was not explored. This represents a significant gap in understanding the roles of these genes in virulence and host–pathogen interactions. Future research should prioritize functional studies, including gene knockout experiments and in vivo models, to clarify the roles of *mce3C* and *mce4F* in the pathogenic mechanisms of *N. cyriacigeorgica*.

Addressing these limitations will enhance our understanding of the genomic features, resistance mechanisms, and pathogenic potential of *N. cyriacigeorgica*. Such insights are essential for developing effective strategies to combat infections caused by this opportunistic pathogen.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12941-025-00791-x.

Supplementary material 1: Figure S1. Genes with a frequency higher than 30% from the accessory genomes were further analysed by hclust (Hierarchical clustering).

Supplementary material 2: Figure S2. Genes with a frequency higher than 30% from the accessory genomes were further analysed by DPAC (Dynamic profile analysis for clusters).

Supplementary material 3.

Supplementary material 4.

Supplementary material 5.

Supplementary material 6.

Supplementary material 7.

Supplementary material 8.

Acknowledgements
Not applicable.

Author contributions
All authors contributed to this work. All authors have read and agreed to the published version of the manuscript. YC, JG, and SY designed the study. YC, JG, SY, CY, Y-x Z, H-y G, HC, WL, FL, Y-w B, JC, F-k W, Q-q S, H-b M, and Z-h W

performed the experiments and interpreted the data. YC, JG, SY, CY, Y-x Z, H-y G, and HC wrote the first draft of the paper. YC, JG, and SY reviewed and approved the final report.

Funding
This research was funded by the Hebei Natural Science Foundation (No.H2022505004), the National Natural Science Foundation of China (No.31200142), and the Hebei Medical Science Research Project (No.20211116).

Data availability
The sequencing data in the study are publicly available. These data can be found here: the National Library of Medicine (http://www.ncbi.nlm.nih.gov/bioproject/1214743) under the accession numbers PRJNA1214743.

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics Committee of Bethune International Peace Hospital (No.2024-KY-409).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]National Engineering Research Center of Immunological ProductsDepartment of Microbiology and Biochemical PharmacyCollege of Pharmacy, Army Medical University, Chongqing 400038, China. [2]Department of General Surgery, The Sixth Medical Center of PLA General Hospital, Beijing 100048, China. [3]Department of Public Affairs Management, Tianjin Medical University, Tianjin 300203, China. [4]Department of Clinical Pharmacy, Bethune International Peace Hospital, Shijiazhuang 050081, China. [5]Department of Clinical Laboratory, Bethune International Peace Hospital, Shijiazhuang 050081, China. [6]Department of Information, Bethune International Peace Hospital, Shijiazhuang 050081, China. [7]Department of Basic Medical Laboratory, Bethune International Peace Hospital, Shijiazhuang 050081, China. [8]Department of Laboratory Medicine, People's Hospital of Chongqing Hechuan, Chongqing 401520, China.

Received: 30 January 2025 Accepted: 27 March 2025
Published online: 05 April 2025

References

1. Brown-Elliott BA, Brown JM, Conville PS, Wallace RJ. Clinical and laboratory features of the Nocardia spp. based on current molecular taxonomy. Clin Microbiol Rev. 2006;19(2):259–82.
2. Traxler RM, Bell ME, Lasker B, Headd B, Shieh WJ, McQuiston JR. Updated review on Nocardia species: 2006–2021. Clin Microbiol Rev. 2022;35(4):e0002721.
3. Conville PS, Brown-Elliott BA, Smith T, Zelazny AM. The complexities of Nocardia taxonomy and identification. J Clin Microbiol. 2018. https://doi.org/10.1128/JCM.01419-17.
4. Roch M, Sierra R, Andrey DO. Antibiotic heteroresistance in ESKAPE pathogens, from bench to bedside. Clin Microbiol Infect. 2023;29(3):320–5.
5. Venkateswaran P, Vasudevan S, David H, Shaktivel A, Shanmugam K, Neelakantan P, Solomon AP. Revisiting ESKAPE Pathogens: virulence, resistance, and combating strategies focusing on quorum sensing. Front Cell Infect Microbiol. 2023;13:1159798.
6. Beaman BL, Beaman L. Nocardia species: host-parasite relationships. Clin Microbiol Rev. 1994;7(2):213–64.
7. Wang HK, Sheng WH, Hung CC, Chen YC, Lee MH, Lin WS, Hsueh PR, Chang SC. Clinical characteristics, microbiology, and outcomes for patients with lung and disseminated nocardiosis in a tertiary hospital. J Formos Med Assoc. 2015;114(8):742–9.
8. Martinez-Barricarte R. Isolated Nocardiosis, an unrecognized primary immunodeficiency? Front Immunol. 2020;11:590239.
9. Mehta HH, Shamoo Y. Pathogenic Nocardia: a diverse genus of emerging pathogens or just poorly recognized? PLoS Pathog. 2020;16(3): e1008280.
10. McNeil MM, Brown JM. The medically important aerobic actinomycetes: epidemiology and microbiology. Clin Microbiol Rev. 1994;7(3):357–417.
11. Lerner PI. Nocardiosis. Clin Infect Dis. 1996;22(6):891–903.
12. Davis-Scibienski C, Beaman BL. Interaction of Nocardia asteroides with rabbit alveolar macrophages: association of virulence, viability, ultrastructural damage, and phagosome-lysosome fusion. Infect Immun. 1980;28(2):610–9.
13. Yassin AF, Rainey FA, Steiner U. Nocardia cyriacigeorgici sp. nov. Int J Syst Evol Microbiol. 2001;51(4):1419–23.
14. Saunier F, Grange S, Rigaill J, Lutz MF, Gagneux-Brunon A, Botelho-Nevers E. Bacteremia and adrenal gland abscess due to Nocardia cyriacigeorgica: a case report and review. BMC Infect Dis. 2022;22(1):966.
15. Gabay S, Yakubovsky M, Ben-Ami R, Grossman R. Nocardia cyriacigeorgica brain abscess in a patient on low dose steroids: a case report and review of the literature. BMC Infect Dis. 2022;22(1):635.
16. Akcaglar S, Yilmaz E, Heper Y, Alver O, Akalin H, Ener B, Tore O, Ersoy C, Imamoglu S. Nocardia cyriacigeorgica: pulmonary infection in a patient with Basedow-Graves disease and a short review of reported cases. Int J Infect Dis. 2008;12(3):335–8.
17. Alshammari K, Al Hothaly B, Alrabiah F. Case of Nocardia cyriacigeorgica infection of the eye in a granulomatosis with polyangiitis patient. Cureus. 2020;12(10): e11178.
18. Schlaberg R, Huard RC, Della-Latta P. Nocardia cyriacigeorgica, an emerging pathogen in the United States. J Clin Microbiol. 2008;46(1):265–73.
19. Gupta S, Grant LM, Powers HR, Kimes KE, Hamdi A, Butterfield RJ, Gea-Banacloche J, Vijayvargiya P, Hata DJ, Meza Villegas DM, et al. Invasive Nocardia infections across distinct geographic regions United States. Emerg Infect Dis. 2023;29(12):2417–25.
20. Ercibengoa M, Camara J, Tubau F, Garcia-Somoza D, Galar A, Martin-Rabadan P, Marin M, Mateu L, Garcia-Olive I, Prat C, et al. A multicentre analysis of Nocardia pneumonia in Spain: 2010–2016. Int J Infect Dis. 2020;90:161–6.
21. Valdezate S, Garrido N, Carrasco G, Medina-Pascual MJ, Villalon P, Navarro AM, Saez-Nieto JA. Epidemiology and susceptibility to antimicrobial agents of the main Nocardia species in Spain. J Antimicrob Chemother. 2017;72(3):754–61.
22. Hashemi-Shahraki A, Heidarieh P, Bostanabad SZ, Hashemzadeh M, Feizabadi MM, Schraufnagel D, Mirsaeidi M. Genetic diversity and antimicrobial susceptibility of Nocardia species among patients with nocardiosis. Sci Rep. 2015;5:17862.
23. Wang H, Zhu Y, Cui Q, Wu W, Li G, Chen D, Xiang L, Qu J, Shi D, Lu B. Epidemiology and antimicrobial resistance profiles of the Nocardia species in China, 2009 to 2021. Microbiol Spectr. 2022;10(2): e0156021.
24. Yang J, Ren HT, Wang J, Dong AY, Chen YL, Hu DX, Wang H, Ji HM, Wen HN, Chen Y, et al. Clinical characteristics, susceptibility profiles, and treatment of nocardiosis: a multicenter retrospective study in 2015–2021. Int J Infect Dis. 2023;130:136–43.
25. Han Y, Cheng M, Li Z, Chen H, Xia S, Zhao Y, Wang Y, He W, Wang W. Clinical characteristics and drug resistance of Nocardia in Henan, China, 2017–2023. Ann Clin Microbiol Antimicrob. 2024;23(1):23.
26. Wei M, Xu X, Yang J, Wang P, Liu Y, Wang S, Yang C, Gu L. MLSA phylogeny and antimicrobial susceptibility of clinical Nocardia isolates: a multicenter retrospective study in China. BMC Microbiol. 2021;21(1):342.
27. Kageyama A, Hoshino Y, Yazawa K, Poonwan N, Takeshita N, Maki S, Mikami Y. Nocardia cyriacigeorgica is a significant pathogen responsible for nocardiosis in Japan and Thailand. Mycopathologia. 2005;160(1):15–9.
28. Kageyama A, Yazawa K, Ishikawa J, Hotta K, Nishimura K, Mikami Y. Nocardial infections in Japan from 1992 to 2001, including the first report of infection by Nocardia transvalensis. Eur J Epidemiol. 2004;19(4):383–9.
29. Poonwan N, Mekha N, Yazawa K, Thunyaharn S, Yamanaka A, Mikami Y. Characterization of clinical isolates of pathogenic Nocardia strains and related actinomycetes in Thailand from 1996 to 2003. Mycopathologia. 2005;159(3):361–8.
30. Wauters G, Avesani V, Charlier J, Janssens M, Vaneechoutte M, Delmee M. Distribution of nocardia species in clinical samples and their routine rapid identification in the laboratory. J Clin Microbiol. 2005;43(6):2624–8.

31. Tan YE, Chen SC, Halliday CL. Antimicrobial susceptibility profiles and species distribution of medically relevant Nocardia species: results from a large tertiary laboratory in Australia. J Glob Antimicrob Resist. 2020;20:110–7.
32. Lowman W, Aithma N. Antimicrobial susceptibility testing and profiling of Nocardia species and other aerobic actinomycetes from South Africa: comparative evaluation of broth microdilution versus the Etest. J Clin Microbiol. 2010;48(12):4534–40.
33. Han L, Ji X, Xu S, Fan S, Wang C, Wei K, Wang X, Song H, Zheng N, Sun L, et al. Microbiological profile of distinct virulence of Nocardia cyriacigeorgica strains in vivo and in vitro. Microb Pathog. 2020;142:104042.
34. Ji X, Zhang X, Sun L, Hou X, Song H, Han L, Xu S, Li H, Qiu X, Li M, et al. The heparin-binding hemagglutinin of Nocardia cyriacigeorgica GUH-2 stimulates inflammatory cytokine secretion through activation of nuclear factor kappaB and mitogen-activated protein kinase pathways via TLR4. Front Cell Infect Microbiol. 2020;10:3.
35. Kong F, Chen SC, Chen X, Sintchenko V, Halliday C, Cai L, Tong Z, Lee OC, Sorrell TC. Assignment of reference 5'-end 16S rDNA sequences and species-specific sequence polymorphisms improves species identification of Nocardia. Open Microbiol J. 2009;3:97–105.
36. Kong F, Wang H, Zhang E, Sintchenko V, Xiao M, Sorrell TC, Chen X, Chen SC. secA1 gene sequence polymorphisms for species identification of Nocardia species and recognition of intraspecies genetic diversity. J Clin Microbiol. 2010;48(11):3928–34.
37. Carrasco G, Valdezate S, Garrido N, Medina-Pascual MJ, Villalon P, Saez-Nieto JA. gyrB analysis as a tool for identifying Nocardia species and exploring their phylogeny. J Clin Microbiol. 2015;53(3):997–1001.
38. Carrasco G, Valdezate S, Garrido N, Villalon P, Medina-Pascual MJ, Saez-Nieto JA. Identification, typing, and phylogenetic relationships of the main clinical Nocardia species in spain according to their gyrB and rpoB genes. J Clin Microbiol. 2013;51(11):3602–8.
39. Yin X, Liang S, Sun X, Luo S, Wang Z, Li R. Ocular nocardiosis: HSP65 gene sequencing for species identification of Nocardia spp. Am J Ophthalmol. 2007;144(4):570–3.
40. McTaggart LR, Richardson SE, Witkowska M, Zhang SX. Phylogeny and identification of Nocardia species on the basis of multilocus sequence analysis. J Clin Microbiol. 2010;48(12):4525–33.
41. Toyokawa M, Ohana N, Ueda A, Imai M, Tanno D, Honda M, Takano Y, Ohashi K, Saito K, Shimura H. Identification and antimicrobial susceptibility profiles of Nocardia species clinically isolated in Japan. Sci Rep. 2021;11(1):16742.
42. Woods GL, Brown-Elliott BA, Conville PS, Desmond EP, Hall GS, Lin G, Pfyffer GE, Ridderhof JC, Siddiqi SH, Wallace RJ, et al. In Susceptibility testing of mycobacteria, Nocardiae, and other aerobic actinomycetes. 2nd ed. Wayne: CLSI Standards: Guidelines for Health Care Excellence; 2011.
43. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
44. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6): e1005595.
45. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11): e112963.
46. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010;11:119.
47. Besemer J, Borodovsky M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 2005;33:W451-454.
48. Li Z, Lu X, Wang D, Liang WL, Zhang J, Li J, Xu J, Pang B, Kan B. Genomic comparison of serogroups O159 and O170 with other Vibrio cholerae serogroups. BMC Genom. 2019;20(1):241.
49. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):157.
50. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–14.
51. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics. 2017;33(21):3340–7.
52. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AV, Cheng AA, Liu S, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48(D1):D517–25.
53. Xu S, Wei M, Li G, Li Z, Che Y, Han L, Jia W, Li F, Li D, Li Z. Comprehensive analysis of the Nocardia cyriacigeorgica complex reveals five species-level clades with different evolutionary and pathogenicity characteristics. mSystems. 2022;7(3): e0140621.
54. Conville PS, Witebsky FG. Organisms designated as Nocardia asteroides drug pattern type VI are members of the species Nocardia cyriacigeorgica. J Clin Microbiol. 2007;45(7):2257–9.
55. Wallace RJ Jr, Steele LC, Sumter G, Smith JM. Antimicrobial susceptibility patterns of Nocardia asteroides. Antimicrob Agents Chemother. 1988;32(12):1776–9.
56. McGuinness SL, Whiting SE, Baird R, Currie BJ, Ralph AP, Anstey NM, Price RN, Davis JS, Tong SY. Nocardiosis in the tropical northern territory of Australia 1997–2014. Open Forum Infect Dis. 2016;3(4):ofw208.
57. Wilson JW. Nocardiosis: updates and clinical overview. Mayo Clin Proc. 2012;87(4):403–7.
58. Mehta H, Weng J, Prater A, Elworth RAL, Han X, Shamoo Y. Pathogenic Nocardia cyriacigeorgica and Nocardia nova evolve to resist trimethoprim-sulfamethoxazole by both expected and unexpected pathways. Antimicrob Agents Chemother. 2018;62(7):e00364-e418. https://doi.org/10.1128/AAC.00364-18.
59. Metcalf BJ, Chochua S, Gertz RE Jr, Li Z, Walker H, Tran T, Hawkins PA, Glennen A, Lynfield R, Li Y, et al. Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. Clin Microbiol Infect. 2016;22(12):1002 e1001-1002 e1008.
60. Huovinen P. Resistance to trimethoprim-sulfamethoxazole. Clin Infect Dis. 2001;32(11):1608–14.
61. Cornick JE, Harris SR, Parry CM, Moore MJ, Jassi C, Kamng'ona A, Kulohoma B, Heyderman RS, Bentley SD, Everett DB. Genomic identification of a novel co-trimoxazole resistance genotype and its prevalence amongst Streptococcus pneumoniae in Malawi. J Antimicrob Chemother. 2014;69(2):368–74.
62. Yao S, Yu J, Zhang T, Xie J, Yan C, Ni X, Guo B, Cui C. Comprehensive analysis of distribution characteristics and horizontal gene transfer elements of bla(NDM-1)-carrying bacteria. Sci Total Environ. 2024;946:173907.
63. Yin Y, Chen H, Li S, Gao H, Sun S, Li H, Wang R, Jin L, Liu Y, Wang H. Daptomycin resistance in methicillin-resistant *Staphylococcus aureus* is conferred by IS256 insertion in the promoter of mprF along with mutations in mprF and walK. Int J Antimicrob Agents. 2019;54(6):673–80.
64. Beaman BL, Black CM, Doughty F, Beaman L. Role of superoxide dismutase and catalase as determinants of pathogenicity of Nocardia asteroides: importance in resistance to microbicidal activities of human polymorphonuclear neutrophils. Infect Immun. 1985;47(1):135–41.
65. Black CM, Beaman BL, Donovan RM, Goldstein E. Intracellular acid phosphatase content and ability of different macrophage populations to kill Nocardia asteroides. Infect Immun. 1985;47(2):375–83.
66. Arruda S, Bomfim G, Knights R, Huima-Byron T, Riley LW. Cloning of an M. tuberculosis DNA fragment associated with entry and survival inside cells. Science. 1993;261(5127):1454–7.
67. Ji X, Tan X, Hou X, Si C, Xu S, Tang L, Yuan X, Li Z. Cloning, expression, invasion, and immunological reactivity of a mammalian cell entry protein encoded by the mce1 operon of Nocardia farcinica. Front Microbiol. 2017;8:281.
68. Ji X, Zhang X, Sun L, Hou X, Song J, Tan X, Song H, Qiu X, Li M, Tang L, et al. Mce1C and Mce1D facilitate N. farcinica invasion of host cells and suppress immune responses by inhibiting innate signaling pathways. Sci Rep. 2020;10(1):14908.

69. Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, Sassetti CM. Characterization of mycobacterial virulence genes through genetic interaction mapping. Proc Natl Acad Sci U S A. 2006;103(31):11760–5.
70. Zhang Y, Li J, Li B, Wang J, Liu CH. Mycobacterium tuberculosis Mce3C promotes mycobacteria entry into macrophages through activation of beta2 integrin-mediated signalling pathway. Cell Microbiol. 2018;20(2): e12800.
71. Guo X, Zhang Z, Chen Q, Wang L, Xu X, Wei Z, Zhang Y, Chen K, Wang Z, Lu X, et al. Whole genome sequencing highlights the pathogenic profile in Nocardia keratitis. Invest Ophthalmol Vis Sci. 2024;65(3):26.
72. Zoropogui A, Pujic P, Normand P, Barbe V, Belli P, Graindorge A, Roche D, Vallenet D, Mangenot S, Boiron P, et al. The Nocardia cyriacigeorgica GUH-2 genome shows ongoing adaptation of an environmental Actinobacteria to a pathogen's lifestyle. BMC Genom. 2013;14:286.
73. Yasuike M, Nishiki I, Iwasaki Y, Nakamura Y, Fujiwara A, Shimahara Y, Kamaishi T, Yoshida T, Nagai S, Kobayashi T, et al. Analysis of the complete genome sequence of Nocardia seriolae UTF1, the causative agent of fish nocardiosis: The first reference genome sequence of the fish pathogenic Nocardia species. PLoS ONE. 2017;12(3): e0173198.
74. Vera-Cabrera L, Ortiz-Lopez R, Elizondo-Gonzalez R, Ocampo-Candiani J. Complete genome sequence analysis of Nocardia brasiliensis HUJEG-1 reveals a saprobic lifestyle and the genes needed for human pathogenesis. PLoS ONE. 2013;8(6): e65425.
75. Gonzalez-Carrillo C, Millan-Sauceda C, Lozano-Garza HG, Ortiz-Lopez R, Elizondo-Gonzalez R, Welsh O, Ocampo-Candiani J, Vera-Cabrera L. Genomic changes associated with the loss of Nocardia brasiliensis virulence in mice after 200 in vitro passages. Infect Immun. 2016;84(9):2595–606.
76. Vieira OV, Botelho RJ, Grinstein S. Phagosome maturation: aging gracefully. Biochem J. 2002;366(Pt 3):689–704.
77. Forrellad MA, Klepp LI, Gioffre A, Sabio y Garcia J, Morbidoni HR, de la Paz Santangelo M, Cataldi AA, Bigi F. Virulence factors of the Mycobacterium tuberculosis complex. Virulence. 2013;4(1):3–66.
78. Hu D, Li Y, Ren P, Tian D, Chen W, Fu P, Wang W, Li X, Jiang X. Molecular epidemiology of hypervirulent carbapenemase-producing Klebsiella pneumoniae. Front Cell Infect Microbiol. 2021;11:661218.
79. Hwang JH, Park JS, Bae TW, Hwang JH, Lee J. Community-acquired solitary brain abscesses caused by Hypervirulent Klebsiella pneumoniae in a healthy adult. Microorganisms. 2024;12(5):894.
80. May T, Okabe S. Enterobactin is required for biofilm development in reduced-genome *Escherichia coli*. Environ Microbiol. 2011;13(12):3149–62.
81. Han R, Niu M, Liu S, Mao J, Yu Y, Du Y. The effect of siderophore virulence genes entB and ybtS on the virulence of Carbapenem-resistant Klebsiella pneumoniae. Microb Pathog. 2022;171:105746.
82. Carrasco G, Monzón S, San Segundo M, et al. Molecular characterization and antimicrobial susceptibilities of *Nocardia* species isolated from the soil; a comparison with species isolated from humans. Microorganisms. 2020;8(6):900. https://doi.org/10.3390/microorganisms8060900.
83. Xu S, Li Z, Huang Y, et al. Whole genome sequencing reveals the genomic diversity, taxonomic classification, and evolutionary relationships of the genus *Nocardia*. PLoS Negl Trop Dis. 2021;15(8): e0009665. https://doi.org/10.1371/journal.pntd.0009665.
84. Vautrin F, Bergeron E, Dubost A, et al. Genome sequences of three *Nocardia cyriacigeorgica* strains and one *Nocardia asteroides* strain. Microbiol Resour Announc. 2019;8(33):e00600-e619. https://doi.org/10.1128/MRA.00600-19.
85. Jorgensen TS, Mohite OS, Sterndorff EB, et al. A treasure trove of 1034 actinomycete genomes. Nucleic Acids Res. 2024;52(13):7487–503. https://doi.org/10.1093/nar/gkae523.

Publisher's Note