scientific reports

Check for updates

OPEN

The emergence and spread of a novel sequence type of *Corynebacterium diphtheriae* in Vietnam

Le Huy Hoang[1,2,7], Luong Minh Hoa[1,7], Pham Thanh Hai[1], Do Bich Ngoc[1], Nguyen Tat Thang[3], Phan Thao Ly[1,4], Phuong Thi Lan Nguyen[1], Phan Trong Lan[1], Tran Nhu Duong[1], Nguyen Thi Thanh Huong[3], Masatomo Morita[5], Hoang Thi Thu Ha[1]✉ & Mitsutoshi Senoh[6]✉

Diphtheria remains a public health concern in Vietnam. This study used whole-genome sequencing to characterize the genetic diversity, virulence, and antimicrobial resistance profiles of 21 *Corynebacterium diphtheriae* isolates collected from 2013 to 2024. Phylogenetic relationships were determined using core-genome MLST (cgMLST) and SNP analysis. AMR was assessed via phenotypic testing (CLSI/EUCAST standards) and genotypic screening (staramr/ARIBA/ResFinder/diphtOscan). Vaccination histories were categorized based on the Vietnam National Expanded Program on Immunization schedule. Analysis revealed a novel sequence type (ST1040) in 19 isolates and ST244 in the remaining two isolates. The ST1040 lineage showed that the nearest neighbors were contemporary isolates from China and India (863–1030 allele differences), suggesting a regional East/Southeast Asian evolutionary context. In contrast, the ST244 lineage, isolated in Vietnam in 2014 and 2015, exhibited strong genetic relatedness to an Austrian ST244 strain (isolated in 2018), suggesting a potential earlier circulation or origin of this lineage in Vietnam, or a shared global reservoir. All 21 isolates were Gravis biovar, carried the *tox* gene, and were toxigenic. Sulfonamide (*sul1*) and tetracycline (*tet(33)* resistance genes were prevalent in the ST1040 lineage, with consistent genotype-phenotype correlation. Among the 15 cases with interpretable vaccination data, none had completed the four-dose routine schedule prior to exposure. In 12 cases with date-verified records, vaccination was confirmed as reactive (administered post-infection) rather than pre-exposure. To effectively control diphtheria, this study emphasises the emergence of novel and drug-resistant strains of *C. diphtheriae* in Vietnam, which calls for improved genomic surveillance, proactive vaccination programs, and ongoing AMR monitoring.

Keywords Diphtheria, *Corynebacterium diphtheriae*, Phylogenomics, Antibiotic resistance, Vaccine history

Diphtheria, a preventable yet potentially deadly disease, continues to threaten populations worldwide, particularly in regions with limited access to healthcare. The enduring nature of this infectious disease is highlighted by recent outbreaks in South America, Africa, and Southeast Asia[1]. Vietnam, a country with a history of diphtheria cases, faces ongoing challenges in controlling the spread of this pathogen.

Toxigenic *Corynebacterium diphtheriae* (*C. diphtheriae*), the causative agent of diphtheria, produces diphtheria toxin (DT), a potent virulence factor encoded by the *tox* gene carried by a lysogenic bacteriophage integrated into the bacterial chromosome[2–4]. DT causes the formation of characteristic pseudomembranes in the throat, as well as systemic complications like myocarditis, neuropathy, and even death[5]. In addition, *C.*

[1]Bacteriology Department, National Institute of Hygiene and Epidemiology, 1 Yersin Street, Hanoi 10000, Vietnam. [2]Faculty of Medical Technology, Hanoi Medical University, Hanoi, Vietnam. [3]Ha Giang Center for Disease Control and Prevention, Ha Giang, Vietnam. [4]School of Medicine and Pharmacy, The University of Danang, Danang, Vietnam. [5]Department of Bacteriology I, National Institute of Infectious Diseases, Japan Institute for Health Security, Tokyo, Japan. [6]Department of Bacteriology II, National Institute of Infectious Diseases, Japan Institute for Health Security, 4-7-1, Gakuen, Musashimurayama, Tokyo 208-0011, Japan. [7]Le Huy Hoang and Luong Minh Hoa contributed equally to this work. ✉email: htth@nihe.org.vn; senoh.m@jihs.go.jp

diphtheriae possesses several other virulence factors that aid in its pathogenesis, such as adhesins, pili, and iron-uptake systems[6]. Additionally, treatment and control efforts are made more difficult by the development of antimicrobial resistance (AMR) genes in *C. diphtheriae* strains[7].

Multilocus sequence typing (MLST) has been documented as a helpful tool for tracking the genetic diversity and geographic distribution of *C. diphtheriae.*[8]. This method involves sequencing specific housekeeping genes to determine a strain's sequence type (ST). Identifying novel STs and characterizing their relationship to known lineages can provide valuable insights into the epidemiology of diphtheria and guide public health interventions.

While studies from the 1990s established a baseline for *C. diphtheriae* lineages in Vietnam, genomic data from the last two decades remain sparse[9]. It is currently unclear whether the recent resurgence of diphtheria in the northern mountainous provinces stems from the persistent evolution of endemic lineages, such as those identified during sporadic outbreaks in the mid-2010s and 2020s, or the introduction of novel, more virulent or drug-resistant clones. Furthermore, the genomic characteristics of these contemporary isolates have not been systematically compared to historical Vietnamese strains, leaving a significant gap in our understanding of how the pathogen is evolving in response to local selective pressures and public health interventions.

To address these questions, this study utilized whole-genome sequencing (WGS) and antimicrobial susceptibility testing to: (i) Delineate the lineage structure and phylogenetic relationships between contemporary outbreak strains, historical Vietnamese isolates, and global reference clades; (ii) Characterize the genotypic and phenotypic AMR profiles of these isolates to evaluate the continued efficacy of first-line treatments; and (iii) Identify genomic markers associated with virulence and biovar-specific traits to assess the potential for clonal expansion within these lineages. By integrating high-resolution genomic findings with clinical metadata, we aim to provide a clearer understanding of the evolution and transmission of *C. diphtheriae* in Vietnam over the last two decades.

Materials and methods

Ethical considerations and approval

This study was conducted in accordance with the principles outlined in the Declaration of Helsinki. Clinical specimens were collected by Provincial Centers for Disease Control (CDCs) as part of routine diagnostic testing and mandatory national diphtheria surveillance and outbreak-response activities conducted under the authority of the Vietnamese Ministry of Health (MOH) (Law No. 03/2007/QH12). Specimens were subsequently transferred to the National Institute of Hygiene and Epidemiology (NIHE) in Vietnam for confirmatory testing and molecular characterization. Written informed consent was obtained from patients (or from parents or legal guardians for individuals under 18 years of age) at the time of specimen collection, in accordance with standard procedures for public health surveillance activities. No additional informed consent was required for genomic sequencing and analyses reported here, as sequencing was performed exclusively on bacterial isolates collected during routine surveillance, and no identifiable patient-level clinical or demographic data were used. All analyses were conducted on fully de-identified bacterial isolates in compliance with MOH regulations on infectious disease surveillance, data protection, and laboratory biosafety. As this work constituted molecular characterization of surveillance isolates as part of an official public health response, no external institutional review board approval was required. The molecular characterization protocol followed the WHO guidelines for "Laboratory testing for diphtheria in outbreak settings" and was reviewed and implemented by NIHE as the designated national reference laboratory.

Isolate collection and selection

Corynebacterium diphtheriae isolates were obtained from confirmed diphtheria cases as part of national surveillance in Vietnam. Local health authorities initially collected pseudomembrane samples at the site of reported diphtheria cases/outbreaks. These clinical samples were transported to the National Institute of Hygiene and Epidemiology (NIHE), Vietnam, using appropriate transport media (Amies or Cary-Blair media) for subsequent culturing and species identification by standard microbiological methods[10]. For this study, a total of 21 *C. diphtheriae* isolates representing 21 individual patients were selected for whole-genome sequencing to characterize circulating strains. Nineteen of these isolates (BH01-BH19) were collected from cases during outbreaks in Dien Bien and Ha Giang provinces between 2023 and 2024, following the described local collection and NIHE processing workflow. The remaining two isolates (BH20 and BH21), from previous outbreaks in Nghe An province (North Central Coast, 2014) and the Central Highlands (2015), respectively, were retrieved from the existing *C. diphtheriae* strain collection stored at NIHE. These two historical outbreak isolates were intentionally included to provide immediate genomic context for the contemporary isolates, allowing for a preliminary assessment of whether the strains from the current major outbreaks (BH01-BH19) represented a continuous local lineage or a distinct, recent emergence compared to previous confirmed outbreak events in Vietnam. Notably, this study focused exclusively on clinical isolates from confirmed diphtheria cases and did not include any carriage isolates from contact screening or samples from cutaneous diphtheria.

Antimicrobial susceptibility testing and toxigenic confirmation

All isolates were re-cultured on Blood Agar (BA, Oxoid, UK) and Cystine Tellurite Blood Agar medium (CTBA, Thermo Fisher Scientific Inc., USA). The isolates were reidentified using MALDI-TOF mass spectrometry (Bruker, Germany) and were confirmed to be toxigenic by the Elek test (UKHSA, Colindale, London, UK). Antimicrobial susceptibility testing was determined using Mueller-Hinton agar supplemented with 5% sheep blood and an inoculum to a 0.5 McFarland standard. Minimum Inhibitory Concentrations (MICs) were determined using the E-test for penicillin G, trimethoprim/sulfamethoxazole (Biodisk, Solna, Sweden), and tetracycline (bioMérieux, Marcy-l'Étoile, France). Due to resource limitations, susceptibility to erythromycin was determined by disk diffusion using a 15 µg disk. Results were interpreted according to the EUCAST 2023

recommendations (Version 13.1), clinical breakpoints specific to *C. diphtheriae*. Specifically, zone diameters and MICs for erythromycin, tetracycline, and trimethoprim/sulfamethoxazole were interpreted according to EUCAST v13.1 (2023) clinical breakpoints for *C. diphtheriae* (erythromycin: S ≥ 24 mm; tetracycline: S ≤ 1 mg/L; trimethoprim/sulfamethoxazole: S ≤ 0.5), respectively. For penicillin G, results were interpreted using CLSI M45 (3rd edition) breakpoints for *Corynebacterium* spp. (susceptible ≤ 0.12 µg/mL, intermediate 0.25–2 µg/mL, resistant ≥ 4 µg/mL), as this standard corresponds to the parenteral penicillin G treatment guidelines used locally. Quality control was maintained using *Streptococcus pneumoniae* ATCC 49,619. Phenotypic testing for rifampicin was not conducted; resistance to this agent was predicted genotypically based on mutations in the *rpoB* gene. While EUCAST provides MIC breakpoints for rifampicin, standardized phenotypic disk diffusion criteria for this species were not available for our laboratory's workflow during the study period.

DNA extraction and genome sequencing

Genomic DNA was extracted using the QIAamp® DNA miniKit (Qiagen, Hilden, Germany), and concentrations were determined using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). The QIAseq FX DNA library kit (Qiagen, Hilden, Germany) was used to prepare a genomic library from 100 ng of each genomic DNA. The prepared libraries were sequenced using the MiSeq™ Reagent Kit v3 (600 cycles) on a MiSeq™ sequencing platform (Illumina, San Diego, CA, USA), generating a 2 × 300 bp paired-end output.

Quality control and assembly

Trim Galore version 0.6.10 (Babraham Institute, Cambridge, CB22 3AT) command-line tools were used for quality control. This involved removing adapter sequences and then performing quality trimming by deleting bases with a Phred quality score of less than 30 (Q30) from the ends of the reads. Additionally, the first 20 bp were deleted from each read. SPAdes version v.4.2.0 with "-careful" and k-mer set to "-auto" was used for the de novo assembly[11]. This approach was chosen over reference-based mapping to enable unbiased detection of novel sequence types and potential mobile genetic elements that may be absent from the reference genome. Contigs with lengths under 500 bp were removed by the seqtk tool version 1.4 (https://github.com/lh3/seqtk). To ensure taxonomic accuracy, raw sequencing reads generated in this study and Nguyen Thi Nguyen et al. (2022)[9] were classified using Kraken2 v2.15 with the standard database[12]. The taxonomic identity of the downloaded assembled reference genomes was independently confirmed using KmerFinder v3.2 (command-line version) with the "bacteria" database[13–15]. This step confirmed that all newly sequenced isolates, downloaded raw reads, and genomes were correctly assigned to *C. diphtheriae* prior to downstream analyses.

Phylogenetic analysis based on core-genome MLST (cgMLST)

For a comprehensive phylogenetic analysis based on core-genome multilocus sequence typing (cgMLST), allele calling was performed using the official, publicly available cgMLST scheme for *C. diphtheriae*, curated and maintained by the Institut Pasteur and comprising 1305 core loci (accessible via BIGSdb-Pasteur). This standardized scheme replaces the initial de novo scheme to ensure comparability with published literature. The schema allele sequences were downloaded from BIGSdb-Pasteur (accessed October 4th, 2025) and processed with the PrepExternalSchema module of chewBBACA v3.3.4[16]. Allele calling was conducted using allelecall with default parameters, except for --cpu 28 and --bsr 0.6 (Blast Score Ratio threshold), to balance sensitivity and specificity in fragmented assemblies. Allele calling was conducted on a highly curated collection of 90 *C. diphtheriae* isolates to maximize phylogenetic resolution and regional relevance. This refined dataset includes all 21 contemporary Vietnamese isolates, three historical Vietnamese isolates (CIP107532, CIP107558, CIP107578), one vaccine strain (PW8, accession no. CP003216.1), and 65 strategically selected global reference strains (as detailed in Supplementary Table 3). These reference strains were chosen to maximize the representation of Asian lineages (e.g., Thailand, Malaysia, India) and core global lineages (e.g., ST3, ST8, ST9) while rigorously excluding phylogenetically redundant strains from distant territories. The core genome was defined as loci in ≥ 95% of the analyzed isolates. The resulting phylogenetic tree was visualized and annotated using iTOLv7 online (https://itol.embl.de/).

Comparative phylogenetic analysis with historical Vietnamese isolates

To further contextualize the evolutionary relationships of the contemporary Vietnamese *C. diphtheriae* isolates, a focused comparative phylogenetic analysis was performed against 54 historical *C. diphtheriae* genome sequences from Vietnam. These isolates were described in a recent study by Nguyen Thi Nguyen et al. (2022)[9], focusing on erythromycin-resistant strains from the 1990s, and are publicly available through the European Nucleotide Archive (ENA). The 54 additional historical Vietnamese genomes from the Nguyen Thi Nguyen et al. (2022) study were combined with the 21 *C. diphtheriae* isolates from the current research for this high-resolution comparative analysis. Crucially, the core-genome alignment for this specific analysis was performed independently of the preceding cgMLST analysis to maximize the number of core genomic positions for comparison. Core-genome alignment was carried out using Snippy v4.3.6 (https://github.com/tseemann/snippy), using *C. diphtheriae* strain PW8 as the reference genome. This genome was selected due to the standardization of locus definition and consistency in comparative literature. PW8 is the globally accepted, fully sequenced reference genome for *C. diphtheriae*, ensuring the consistent and reproducible calling of conserved genomic regions and facilitating direct comparison of SNP positions with existing studies in the field. A minimum read coverage of 10x (--mincov) and a minimum variant fraction of 90% (--minfrac) were required to call a site. Furthermore, a minimum VCF variant quality score of 100 (--minqual) was applied to filter out low-confidence calls and potential sequencing artifacts. These stringent parameters ensured that the resulting core-genome alignment consisted only of high-quality, reproducible SNP positions for subsequent phylogenetic reconstruction. A dedicated phylogenetic tree was generated by FastTree v2.1.11 (https://morgannprice.github.io/fasttree/) based on the identified SNPs from

the core-genome alignment of this focused dataset. The tree was constructed using the GTR + CAT model, applying the -gamma flag for Gamma20 likelihood rescaling to account for rate variation among sites. This approach ensured that the genetic distances among these closely related Vietnamese isolates were maximally resolved, preventing the underestimation of divergence that might occur if the analysis were restricted to the less variable cgMLST loci defined by the larger global dataset. Pairwise SNP distances between all isolates in this dataset were calculated using Snippy-core under the Snippy package, followed by snp-dists v0.8.2 (https://github.com/tseemann/snp-dists).

Antimicrobial resistance, virulence, biovar-associated genes, and sequence type identification

STs, AMR genes, and plasmid replicons were comprehensively detected using a multi-tool approach. Initial detection of sequence type, AMR genes, and plasmid replicons was performed with staramr v0.5.1 against the ResFinder, PointFinder, and PlasmidFinder databases with "-diphtheria_3" options[17]. To ensure accuracy and complement assembly-based detection, all raw reads were also screened against the ResFinder database using both ARIBA v2.14.7[18] and ResFinder v4.7.2[19]. All downstream genomic analyses, including sequence typing, biovar-associated genes, and identification of antimicrobial resistance and virulence genes, were performed using the diphtOscan tool[20]. To ensure a comprehensive characterization of the isolates, we utilized the -st (MLST), -t (*tox* allele), -res_vir (resistance/virulence), and -plus (extended genotyping) flags. The -plus option was specifically employed to expand the screening beyond the core *tox* gene to include all available virulence markers in the database. For all genomic screenings (staramr, ARIBA, ResFinder, and diphtOscan), we applied a stringent minimum identity threshold of 90%, a minimum length coverage of 60% and a minimum 80% overlap for point mutations to avoid false-positive calls from low-homology hits. Sequence types (STs) were determined by comparing the seven housekeeping gene alleles (*atpA*, *dnaE*, *dnaK*, *fusA*, *leuA*, *odhA*, and *rpoB*) against the *C. diphtheriae* MLST database hosted at BIGSdb-Pasteur (https://bigsdb.pasteur.fr/diphtheria/). The novel sequence type identified in this study was submitted to the database curator at the Institut Pasteur and was formally assigned the designation ST1040.

Vaccine history collection

Medical records and patient interviews were used to gather information about each participant's vaccination history, including the type of vaccine received, the number of doses, and the administration dates. Vaccination status was classified according to the National Expanded Program on Immunization (EPI) schedule, which includes a three-dose primary series in infancy followed by a booster dose at 18–24 months. Individuals were categorized as completely vaccinated (≥4 documented doses received prior to the outbreak year), partially vaccinated (1–3 documented doses), unvaccinated (0 documented doses), or of unknown vaccination status when no records were available and recall was unreliable; the latter group was analyzed separately to avoid misclassification. Vaccination events were further distinguished as routine immunization (scheduled childhood doses) or reactive vaccination (emergency doses administered as part of the outbreak response following the identification of the index case), allowing for the separation of pre-exposure immunity gaps from post-exposure public health interventions.

Results

To offer a thorough understanding of the phenotypic and genomic characteristics of the 21 *C. diphtheriae* isolates obtained from diphtheria cases in Vietnam between 2013 and 2024, this study compared them to historical and international reference strains.

Clinical context

All patients presented with typical respiratory diphtheria, including pseudomembranes, sore throat, fever, and cough. No cutaneous diphtheria cases or carriage isolates from contact screening were included in this study. The most frequently reported complication was acute myocarditis. No cases of muscle paralysis, renal failure, or neurological complications were reported. The majority were treated with antibiotics upon admission. According to the Ministry of Health guidelines for diphtheria treatment in Vietnam, these typically include standardized penicillin G, erythromycin, or azithromycin, generally administered for 14 days. Nineteen cases of diphtheria (90.4%) recovered and were discharged from the hospital after treatment. There were two deaths, one from Dien Bien and one from Ha Giang. Both were hospitalized in severe conditions, with high fever, difficulty breathing, and no prior treatment.

Regarding their epidemiological origin, 19 of the study isolates (BH01-BH19) were collected from confirmed diphtheria cases during two distinct outbreaks that occurred in Dien Bien (one strain, BH01, 2023) and Ha Giang (18 strains, BH02-BH19, 2023–2024). The remaining two isolates, BH20 and BH21, were from earlier outbreaks in Nghe An province and the Central Highlands (2015), respectively (Fig. 1).

Antimicrobial susceptibility testing

Five *C. diphtheriae isolates* (23.8%) with no resistance are susceptible to all antibiotics tested, and none of the isolates were found to be resistant to penicillin G and erythromycin. However, 76.2% of isolates resisted tetracycline and trimethoprim/sulfamethoxazole (Fig. 2).

cgMLST-based phylogenetic tree construction

Phylogenetic reconstruction using the Neighbor-Joining (NJ) method offers high-resolution visualization of strain relationships, illustrating the distinct clustering of contemporary Vietnamese isolates within the broader global context (Fig. 3). All 90 isolates belong to *C. diphtheriae* and formed a single, well-supported clade. The

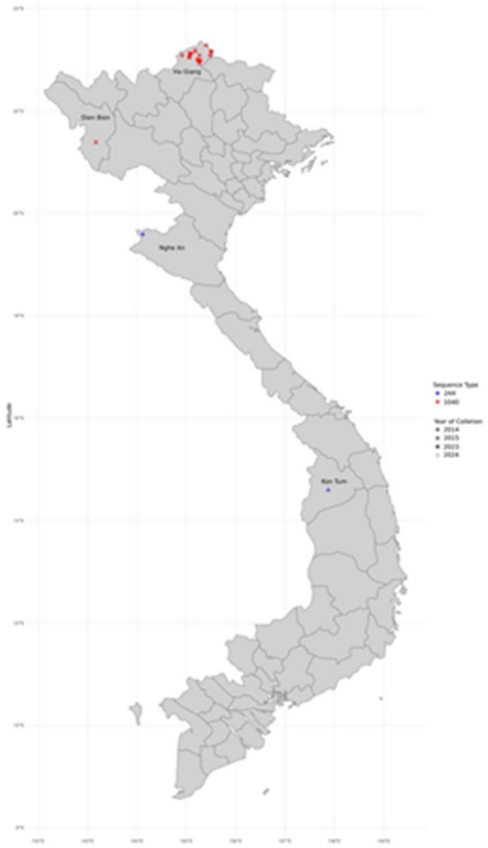

Fig. 1. Location of sample collection in Vietnam.

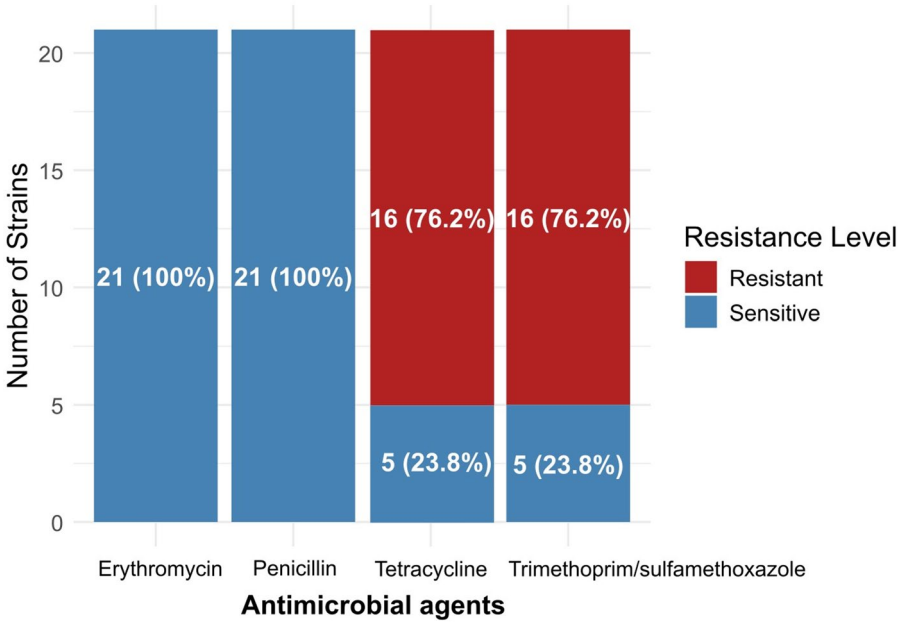

Fig. 2. Result of Antimicrobial susceptibility testing against 21 Vietnamese *C. diphtheriae* isolates.

tree was computed based on the multiple sequence alignment (MSA) of 1,030 core loci in ≥ 95% of the isolates. Pairwise allelic distance analysis (Supplementary Table 3) revealed that the dominant ST1040 clade (BH01–BH19), which includes 19 of the contemporary Vietnamese isolates, exhibited exceptional genetic homogeneity, with 0–4 allele differences among its members, confirming a recent, highly clonal expansion underlying the

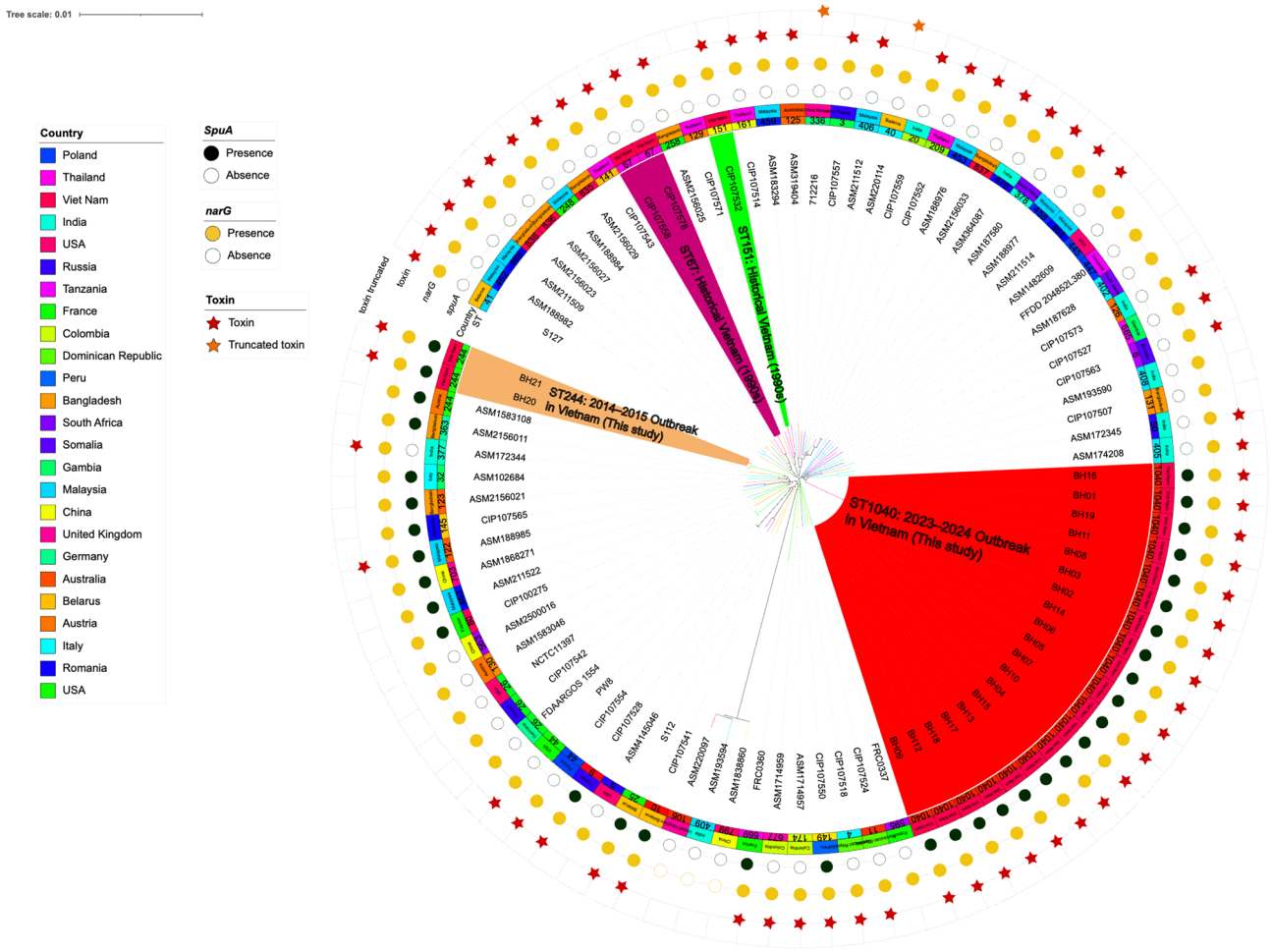

Fig. 3. Phylogenetic tree based on cgMLST. Filled circles indicate the presence of the gene, while unfilled circles indicate its absence.

current outbreak. ST1040 showed its closest genetic affinity to modern Asian strains, with 863–1030 allele differences from Chinese isolates (ST799, ST703) and Indian isolates (ST377, ST405), underscoring a strong regional phylogenetic link. In contrast, isolates BH20 and BH21 (ST244) differed by only one allele from each other and by 168–169 alleles from the Austrian reference strain ASM1583108 (also ST244), placing them firmly within a widespread Eurasian lineage. The three historical Vietnamese isolates (CIP107532/ST151; CIP107558 & CIP107578/ST67) were genetically distant from the current outbreak clone, showing 930–946 allele differences relative to ST1040, which reflects substantial genomic divergence over time. This clear phylogenetic and quantitative delineation supports the conclusion that the recent diphtheria outbreak in Vietnam is driven by a unique, recently emerged ST1040 clone embedded within the broader East Asian genotypic landscape.

Comparative phylogenetic analysis with historical Vietnamese isolates

To further investigate the genetic landscape of *C. diphtheriae* in Vietnam and contextualize the contemporary isolates (BH01-BH21), a focused whole-genome SNP-based phylogenetic analysis was performed by combining the 21 current isolates with 54 historical Vietnamese *C. diphtheriae* genomes (total $n = 75$). The resulting phylogenetic tree is presented in Fig. 4, and the corresponding pairwise SNP distance matrix is detailed in Supplementary Table 4. The phylogenetic structure in Fig. 4 clearly shows distinct clustering patterns within this combined Vietnamese dataset. The 21 contemporary isolates (BH01-BH21) maintained their resolution into two evolutionarily distinct clades, corresponding to the larger ST1040 clade (BH01-BH19) and the minor ST244 clade (BH20-BH21). Within the BH01-BH19 clade, the maximum pairwise SNP difference was observed to be 12 SNPs, with many isolates being identical or differing by only a few SNPs, consistent with a very recent, highly clonal emergence. Only seven SNPs separated BH20 and BH21, indicating a close relationship within the ST244 clade (Supplementary Table 4).

Crucially, the two main contemporary clades (ST1040 and ST244) are highly divergent, evidenced by pairwise SNP differences ranging from 23,187 to 23,194 SNPs (Supplementary Table 4). This extensive divergence confirms that these two groups reflect evolutionarily distinct lineages circulating in the Vietnamese context. The historical Vietnamese isolates from Nguyen et al. (2022) were primarily distantly related to the contemporary BH isolates, showing overall pairwise SNP differences ranging broadly from approximately 22,158 to 26,173 SNPs relative to

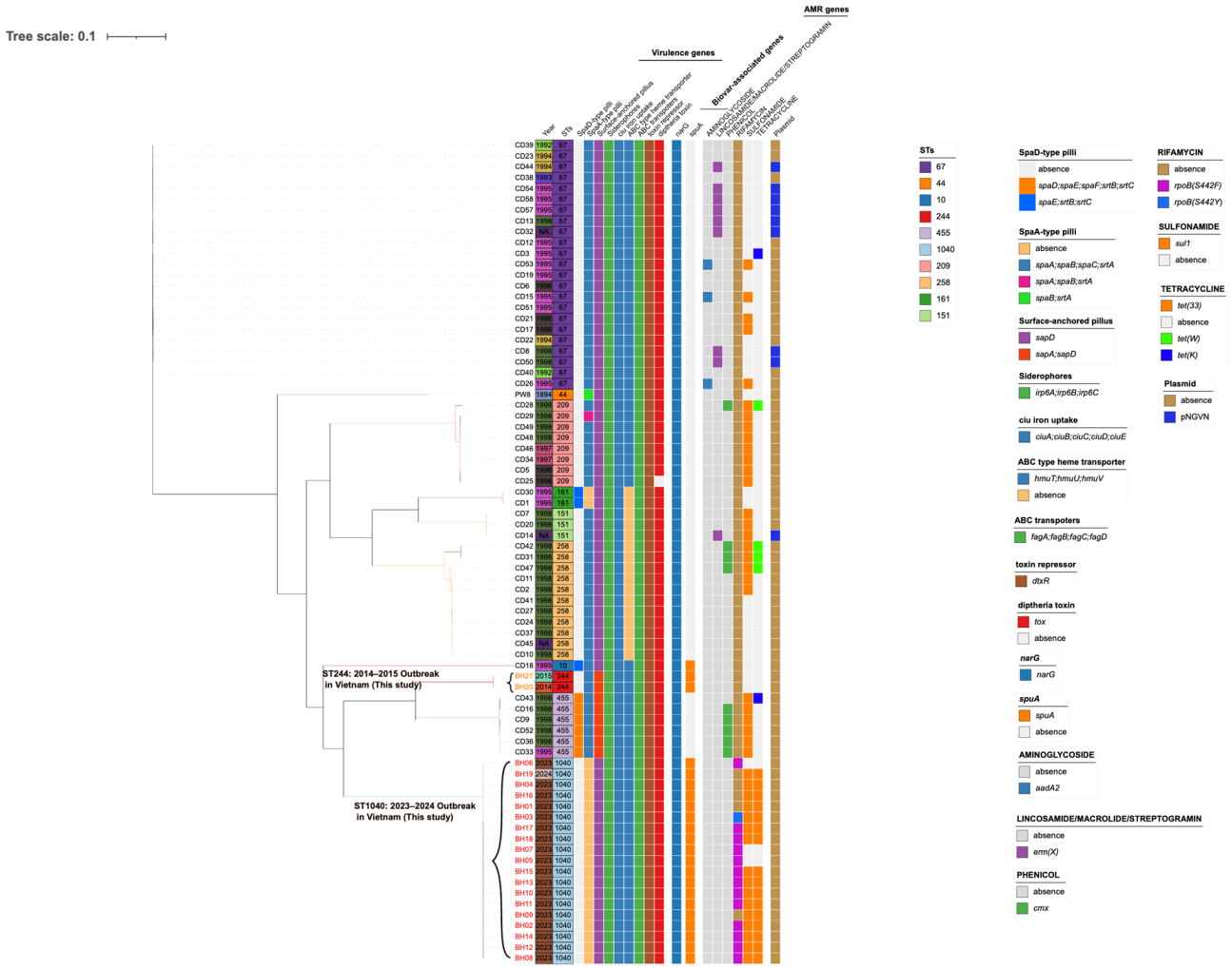

Fig. 4. Phylogenetic Tree of *Corynebacterium diphtheriae* Isolates with Associated Virulence and Antimicrobial Resistance Gene Profiles.

the BH01-BH21 isolates (Supplementary Table 4), indicating distinct long-term evolutionary trajectories. The historical isolates were observed to belong to various other distinct clades distributed across the tree (Fig. 4), further highlighting the genetic diversity of *C. diphtheriae* lineages prevalent in Vietnam during the 1990s.

Sequence types, biovar-associated genes, virulence factors, and antimicrobial resistance genes

Whole-genome analysis of the 21 contemporary *C. diphtheriae* clinical isolates (BH01-BH21), alongside 54 historical Vietnamese isolates and the reference strain PW8, revealed distinct genetic characteristics related to sequence types (STs), biovar associations, virulence potential, and antimicrobial resistance profiles (Fig. 4). Based on their phylogenetic grouping and the high degree of SNP difference observed in the focused analysis, the modern BH isolates were allocated to two distinct phylogenetic groups and corresponding STs: ST244 for the BH20 and BH21 clade and ST1040, a novel sequence type, for the BH01-BH19 clade. Nguyen et al. (2022) reported that the historical Vietnamese isolates (CD strains) showed a wide range of STs, such as ST258, ST67, ST151, ST455, ST209, ST161, and ST10, which reflected a more diverse genetic background prevalent in the 1990s. The ST44 strain PW8 serves as the reference strain. All contemporary clinical isolates (BH01-BH21) consistently harbored the *narG* gene, which is associated with nitrate reductase activity and characteristic of *C. diphtheriae* biovars. Crucially, they also possessed the *spuA* gene, a marker strongly linked to the Gravis biovar. Furthermore, all BH isolates carried the *tox* gene, indicating their toxigenic potential. Similar to the contemporary isolates, most historical CD isolates also possessed both *narG* and *tox* genes. However, among the historical CD isolates, the *spuA* gene was detected only in CD18, suggesting that the Gravis biovar association, as indicated by *spuA*, was not widespread among these historical strains. The reference strain PW8 carried *narG* and *tox*, but lacked *spuA*. In terms of other virulence-associated factors, genes involved in iron uptake, including the heme transporter (*hmuT*, *hmuU*, *hmuV*), ABC transporters (*fagA*, *fagB*, *fagC*, *fagD*), *ciu* iron uptake (*ciuA*, *ciuB*, *ciuC*, *ciuD*, *ciuE*), and siderophores (*irp6A*, *irp6B*, *irp6C*), were consistently detected across nearly all analyzed isolates (both contemporary BH and historical CD strains), underscoring their conserved roles in *C.*

diphtheriae pathogenesis. Variability was observed in pilus-encoding genes, which are essential for adherence. The BH01-BH19 clade predominantly possessed only the *sapD* gene (a surface-anchored pilus). In contrast, the BH20-BH21 clade harbored a complete SpaA-type pilus operon (*spa*, *spaB*, *spaC*, *srtA*) and *sapA* and *sapD*. The historical CD isolates displayed diverse pilus profiles in many, including the six phylogenetically intermediate strains, which carried the SpaA-type pilus (*spa*, *spaB*, *spaC*, *srtA*). Notably, the six intermediate CD strains (CD9, CD16, CD33, CD36, CD43, CD52) also contained the complete SpaD-type pilus operon (*spaD*, *spaE*, *spaF*, *srtB*, *srtC*). By contrast, CD1, CD18, and CD30 only possessed a partial set of SpaD-type pilus genes (*spaE*, *srtB*, *srtC*), indicating a more complex and varied pilus repertoire in some historical lineages. The reference strain PW8 contained *spaB* and *srtA*.

AMR genes and plasmids

Genomic analysis for AMR genes revealed distinct resistance profiles between the contemporary and historical isolates. Most BH01-BH19 isolates consistently carried the *sul1* gene, conferring resistance to sulfonamides, and the *tet(33)* gene, conferring resistance to tetracyclines. Variability in rifampicin resistance was observed within this clade, with *rpoB* mutations (specifically S442F and a unique S442Y mutation in BH03) identified in 14 isolates. Conversely, the BH20-BH21 clade lacked *sul1*, *tet(33)*, and *rpoB* mutations. All contemporary Vietnamese isolates (BH01-BH21) were notably free of detectable plasmids.

Among the historical Vietnamese isolates, *sul1* was also highly prevalent. A significant difference was the common detection of the *erm(X)* gene (conferring resistance to macrolides/lincosamides/streptogramins) in many CD strains, a gene absent in the contemporary BH isolates. Other tetracycline resistance genes, such as *tet(W)* and *tet(K)*, were also found in some historical CD isolates. Furthermore, genes conferring resistance to phenicols (*cmx*) and aminoglycosides (*aadA2*) were detected in a few CD isolates. The presence of pNGVN plasmids in several CD strains often correlated with the detection of *erm(X)*, suggesting plasmid-mediated macrolide resistance was a feature of these older strains.

Vaccination status and infection timing

Among the 21 patients, vaccination history was available for a subset of cases and was frequently incomplete (Supplementary Table 5). Fourteen individuals were documented to have received at least one dose of a diphtheria-toxoid–containing vaccine (DPT for children < 7 years, Td for individuals ≥ 7 years), with five receiving a single dose and nine receiving two doses. In all 12 cases with documented dose dates, vaccination occurred after specimen collection, confirming these as reactive post-exposure interventions rather than pre-existing immunity. One additional partially vaccinated case had received a single dose, but the timing was uncertain. No individual had completed the four-dose EPI schedule (3 primary + booster) prior to collection. Several infections occurred within days to weeks of reactive vaccination (e.g., BH03, BH12), consistent with exposure during the ~ 2-week window before seroconversion rather than primary vaccine failure. Due to incomplete vaccination records for some individuals, conclusions regarding baseline population immunity should be interpreted with caution.

Discussion

Our cgMLST analysis of 90 *Corynebacterium diphtheriae* isolates revealed two distinct co-circulating lineages in Vietnam. The 21 contemporary Vietnamese isolates resolved into two phylogenetically separate clades, including a dominant cluster of 19 isolates (BH01-BH19) assigned to a novel Sequence Type ST1040 and a smaller cluster of two isolates (BH20-BH21) belonging to ST244. Pairwise allelic distance analysis confirmed that these two groups are highly divergent, with 900–901 allele differences between ST1040 and ST244 isolates, supporting their classification as independent evolutionary lineages. The ST1040 clade exhibited exceptional genetic homogeneity with pairwise distances of only 0–4 alleles among its members, consistent with a recent highly clonal expansion underlying the current outbreak. Although ST1040 did not cluster closely with any international strain, all pairwise distances were ≥ 828 alleles; its nearest neighbors were contemporary Asian isolates. These included Chinese strains ST799 and ST703 (1030 and 943 alleles apart, respectively) and Indian strains ST377 and ST405 (863–918 alleles apart). This regional proximity suggests ST1040 likely emerged within the broader East and South Asian *C. diphtheriae* genotypic landscape rather than being imported from distant continents. In contrast, the Vietnamese ST244 isolates (BH20-BH21) showed a strong genetic link to an international reference strain. They differed by only 168–169 alleles from the Austrian ST244 isolate ASM1583108 (collected in 2018). Given that the Vietnamese ST244 isolates were collected earlier (2014 and 2015), this close relationship, coupled with identical ST assignment, suggests either an earlier circulation of this lineage in Vietnam followed by international spread or a shared yet uncharacterized global reservoir from which both regions acquired the strain. The contemporary outbreak strains showed substantial genomic divergence compared to historical Vietnamese isolates. The ST1040 isolates differed from the three historical Vietnamese strains (ST151 and ST67) by 930–946 alleles, reflecting a significant genetic shift over time and indicating that the current outbreak is driven by a newly emerged clone rather than the re-emergence of previously circulating lineages.

Together, these findings highlight the dynamic nature of *C. diphtheriae* population structure in Vietnam, with the recent emergence of a novel, highly clonal ST1040 lineage embedded within a regional Asian phylogenetic context, coexisting alongside a globally disseminated ST244 lineage. This underscores the importance of sustained genomic surveillance to track the origin, evolution, and spread of toxigenic *C. diphtheriae* in Southeast Asia.

Further contextualizing the contemporary isolates, our comparative phylogenetic analysis with 54 historical Vietnamese *C. diphtheriae* genomes from Nguyen Thi Nguyen et al. (2022) revealed their largely distinct long-term evolutionary trajectories (Supplementary Table 4). Despite some visual clustering in Fig. 4, the extensive pairwise SNP differences (ranging from approximately 22158 to 26173 SNPs) confirm that the contemporary

BH isolates represent genetically distinct lineages from these historical strains, indicating separate evolutionary paths over time. While a unique cluster of six historical strains (CD16, CD33, CD36, CD43, CD52, and CD9) was phylogenetically positioned between the BH01-BH19 and BH20-BH21 clades in Fig. 4, the substantial genetic distance observed underscores that these historical strains are not closely related to either contemporary BH clade. This positioning highlights a distinct historical cluster within the *C. diphtheriae* population of Vietnam during that period, rather than representing an intermediate evolutionary lineage directly linking the two highly divergent current BH clades. The pronounced genetic separation of the BH01-BH19 and BH20-BH21 clades themselves, evidenced by 23,187 to 23,194 SNPs, further reinforces that these two contemporary groups represent independently circulating lineages.

All contemporary clinical isolates (BH01-BH21) were confirmed to be toxigenic, consistently harboring the *tox* gene, and were associated with the Gravis biovar through the consistent presence of the *spuA* gene. This is consistent with worldwide patterns that show how common toxic strains are in this biovar[20]. Curiously, only CD18 of the historical CD isolates had the *spuA* gene, a marker closely associated with the Gravis biovar. However, the *narG* and *tox* genes were common among the historical Vietnamese isolates. This suggests a potential shift towards a higher prevalence of the Gravis biovar in more recent circulating strains. Our comprehensive analysis of pilus-encoding genes showed that the BH01-BH19 clade mostly carried the *sapD* gene. At the same time, the BH20-BH21 clade carried not only *sapA* and *sapD* but also a complete SpaA-type pilus operon. The pilus profiles of the historical CD isolates were much more varied. Some of them, such as the six intermediate phylogenetic strains, carried both SpaA-type and SpaD-type pilus operons. This observed diversity in pilus types across different lineages may reflect adaptations to host colonization or environmental factors.

Antimicrobial susceptibility testing revealed that while none of the contemporary isolates were resistant to penicillin G and erythromycin, a significant majority (76.2%) showed resistance to tetracycline and trimethoprim/sulfamethoxazole (Fig. 2). Genomic analysis further identified the consistent co-occurrence of the *sul1* gene (sulfonamide resistance) and the *tet(33)* gene (tetracycline resistance) in most BH01-BH19 isolates, which were notably absent in the BH20-BH21 clade. Importantly, for all contemporary Vietnamese isolates (BH01-BH21), the detected AMR genes (including *sul1* and *tet(33)*[33]*tet(33)* were consistently correlated with their corresponding phenotypic resistance profiles, indicating no genotype-phenotype mismatches in this study group. The presence of *rpoB* mutations (specifically S442F and a unique S442Y mutation in BH03) was associated with rifampicin resistance in 14 isolates. While such *rpoB* mutations are well-documented for conferring rifampicin resistance in other bacterial species like *Mycobacterium tuberculosis*[21], direct evidence in *C. diphtheriae* is more limited[22–24], underscoring the need for further functional validation. However, a limitation of this study is the inability to perform phenotypic MIC testing for rifampicin due to resource limitations. This makes it impossible to confirm directly how the identified *rpoB* mutations contribute to phenotypic resistance to rifampicin. All the current Vietnamese isolates (BH01–BH21) lack detectable plasmid replicons, indicating that the observed antimicrobial resistance is most likely chromosomally encoded. This highlights the necessity of continuous surveillance to track the establishment and dissemination of drug-resistant bacteria. Additionally, in the context of clinical treatment, the existence of antibiotic resistance genes in specific isolates merits careful attention. Penicillin G, erythromycin, or azithromycin given for 14 days are the mainstays of the Ministry of Health's standardized procedures for treating diphtheria in Vietnam. Our findings, consistent with phenotypic susceptibility to penicillin G and erythromycin in most isolates, support the continued efficacy of these first-line agents. However, the detection of genes conferring resistance to other antibiotics, specifically *tet(33)*[33]*tet(33)* (tetracycline resistance) and *sul1* (trimethoprim/sulfamethoxazole resistance), in some isolates is notable. While these drugs may not be the primary empirical choices for diphtheria, their presence highlights potential challenges for alternative treatment strategies or in cases of co-infections. Additionally, the identification of *rpoB* mutations (S442F and S442Y), although not phenotypically confirmed for rifampicin resistance in this study, emphasizes the importance of ongoing genomic surveillance for broader AMR patterns, particularly given the limited number of readily available first-line antimicrobial agents for *C. diphtheriae*. It is important to note that while the ST1040 lineage (BH01-BH19) exhibits clear evidence of clonal expansion based on core genome similarity (0–1 to 12 SNPs), variations in antimicrobial resistance genes (*sul1* and *tet(33)*[33]*tet(33)* and mutations (*rpoB*) were observed among isolates within this ST. This genetic heterogeneity does not contradict clonal expansion but rather reflects the dynamic nature of bacterial genomes. Even clonally related populations can acquire or lose mobile genetic elements carrying AMR genes through horizontal gene transfer, or develop de novo point mutations under selective pressures (e.g., antibiotic exposure) during their dissemination[25–28]. This highlights the ongoing adaptability and evolution of *C. diphtheriae*, even within recently expanded lineages.

Compared to historical CD isolates, a significant difference was the common detection of the *erm(X)* gene (macrolide/lincosamide/streptogramin resistance) in many CD strains, a gene absent in the contemporary BH isolates. Other tetracycline resistance genes like *tet(W)* and *tet(K)*, as well as *cmx* (phenicol resistance) and *aadA2* (aminoglycoside resistance), were also detected in some older strains. The correlation between *erm(X)* and pNGVN plasmids in several historical CD strains indicates that plasmid-mediated macrolide resistance was a feature of these older lineages, highlighting a shift in resistance mechanisms over time.

In contrast, none of the recent Vietnamese isolates harbored macrolide resistance genes or showed evidence of macrolide resistance. This finding contrasts with reports from other regions, including Indonesia, Brazil, and Canada, where phenotypic resistance or reduced susceptibility to erythromycin has been documented in *C. diphtheriae* isolates based on antimicrobial susceptibility testing, even when underlying genetic mechanisms were not always characterized[29–31]. The predominance of a fully susceptible population in this outbreak likely reflects the rapid clonal expansion of the novel ST1040 lineage, which appears to have emerged recently from a reservoir with minimal macrolide selective pressure. Although national treatment guidelines in Vietnam continue to recommend penicillin G or erythromycin as first-line therapy for diphtheria, the absence of macrolide resistance in ST1040 is clinically advantageous, supporting the continued effectiveness of erythromycin for outbreak

management. Nevertheless, the lack of resistance markers also suggests that this lineage has not yet acquired mobile genetic elements commonly associated with erm genes, underscoring the importance of ongoing genomic surveillance to detect any future emergence of antimicrobial resistance.

Our findings clarify the temporal relationship between infection and vaccination, distinguishing reactive public health responses from gaps in pre-exposure immunity. In all 12 cases with verifiable vaccination dates, diphtheria-toxoid-containing vaccines (DPT/Td) were administered after specimen collection, confirming that vaccination was a consequence, not a cause or means of preventing disease. Critically, no individual had completed the full four-dose EPI schedule (3 primary doses + booster) prior to infection, and several developed disease within days to weeks of receiving reactive doses, consistent with exposure during the window before seroconversion rather than vaccine failure. While post-exposure vaccination is recommended to reduce secondary transmission and protect contacts, the absence or incompleteness of primary vaccination before infection highlights persistent vulnerabilities in routine childhood immunization coverage, particularly among adolescents and young adults, for whom booster uptake is often suboptimal. This pattern is consistent with prior studies from sub-Saharan Africa, which have shown that delayed or incomplete primary vaccination leaves populations susceptible, despite the implementation of post-exposure prophylaxis[32,33]. Genomic analysis revealed that the 21 isolates segregated into two distinct sequence types: ST1040 (n = 19), comprising all cases from 2023 to 2024, and ST244 (n = 2), representing older, sporadic cases from 2014 to 2015 (BH20–21). All isolates carried the *tox* gene, confirming continued circulation of toxigenic *C. diphtheriae*. Notably, the entire ST1040 outbreak cluster consisted of individuals who were unvaccinated, partially vaccinated, or received reactive doses post-collection, with no evidence of breakthrough infection in fully vaccinated individuals. This concentration of under-immunized cases within a single clonal lineage strongly suggests that insufficient pre-exposure immunity, rather than strain virulence, is the primary driver of the 2023–2024 outbreak.

While our data robustly support the clonal expansion of a toxigenic ST1040 lineage and the predominance of under-immunized individuals among cases with documented histories, several limitations warrant cautious interpretation. First, phenotypic MIC testing for rifampicin was not performed. Resistance was instead inferred from *rpoB* mutations (S442F/S442Y). Although these mutations are mechanistically plausible based on established orthologs, they require functional validation specifically in *C. diphtheriae*. Second, the modest sample size of 21 isolates limits statistical power for subgroup analyses such as age-stratified AMR or vaccination effects. Third, isolates were derived exclusively from clinical, outbreak-associated respiratory cases, which introduces ascertainment bias. Consequently, our findings may not reflect the genomic or phenotypic diversity of *C. diphtheriae* in asymptomatic carriers, cutaneous infections, or non-outbreak settings. These populations are known to harbor distinct lineages and resistance profiles that were not captured in this study. Fourth, vaccination histories were incomplete for six individuals. Of the remaining 15 participants, only 12 had date-verified records in the National EPI system that enabled temporal classification into reactive versus pre-exposure doses. Thus, while no completely vaccinated individual was observed among the 15 with interpretable status, the true proportion of unvaccinated cases in the total outbreak remains uncertain. Claims regarding population-level immunity gaps are therefore descriptive and hypothesis-generating rather than definitive.

Importantly, our data support three primary conclusions: (i) the 2023–2024 outbreak was caused by a single clonal ST1040 lineage; (ii) all isolates carried tox and shared consistent genotype-phenotype correlations for tested antimicrobials, with the exception of rifampicin; and (iii) reactive vaccination dominated among cases with verifiable histories. However, these data do not support inferences about transmission directionality such as the identification of an index case, the national prevalence of ST1040, or comparative vaccine effectiveness. These analyses would require extensive contact tracing, population-based sampling, or serological data.

Together, these findings underscore that reactive vaccination, while essential for outbreak containment, cannot substitute for high-coverage routine immunization. To prevent future outbreaks in Vietnam, we recommend: (i) strengthening adolescent Td booster programs, particularly for ages 10–25, where coverage is often lowest; (ii) improving vaccination record-keeping and recall validation in outbreak settings; and (iii) integrating real-time genomic surveillance with immunization monitoring to rapidly identify transmission clusters and target interventions where immunity gaps exist.

Conclusion

This whole-genome sequencing study reveals that the 2023–2024 diphtheria resurgence in Vietnam was driven by the emergence of a novel, highly clonal lineage, ST1040, alongside the continued presence of the Eurasian-linked ST244. All isolates were toxigenic (biovar Gravis), confirming the circulation of virulent strains. While resistance to trimethoprim/sulfamethoxazole and tetracycline was prevalent and chromosomally mediated, the absence of macrolide resistance genes supports the continued use of erythromycin as a primary treatment.

The integration of genomic data with epidemiological records clarifies that most vaccinations were reactive post-exposure interventions rather than pre-existing immunity. Notably, no patient with available records had completed the four-dose National EPI schedule prior to infection. While these findings highlight significant immunity gaps, the incomplete records for a subset of cases necessitate a descriptive interpretation of these population-level vulnerabilities.

Ultimately, our results underscore the need to transition from reactive outbreak responses to proactive adolescent booster programs. By combining enhanced genomic surveillance with high-coverage routine immunization, public health authorities can better identify transmission clusters and mitigate future diphtheria outbreaks in Vietnam and the broader Southeast Asian region.

Data availability

The raw read sequences obtained in this study were deposited in the NCBI Sequence Read Archive under the BioProject PRJNA1134379 (accession numbers: SRR29773471 to SRR29773491).

Received: 26 November 2025; Accepted: 30 January 2026
Published online: 06 February 2026

References

1. Bennett, A. A. S. & Diphtheria, C. D. C. Yellow Book 2024 2024 [Available from: https://wwwnc.cdc.gov/travel/yellowbook/2024/infections-diseases/diphtheria
2. Murphy, J. R. Corynebacterium diphtheriae. Medical Microbiology 4th edition. (1996).
3. Holmes, R. K. Biology and molecular epidemiology of diphtheria Toxin and the Tox gene. *J. Infect. Dis.* **181** (Supplement_1), S156–S67 (2000).
4. Parveen, S., Bishai, W. R. & Murphy, J. R. Corynebacterium diphtheriae: diphtheria Toxin, the Tox Operon, and its regulation by Fe2(+) activation of apo-DtxR. *Microbiol. Spectr.* **7**(4), 10.1128/microbiolspec.gpp3-0063-2019 (2019).
5. Lamichhane, A. & Radhakrishnan, S. *Diphtheria. StatPearls. Treasure Island (FL): StatPearls Publishing Copyright © 2024* (StatPearls Publishing LLC., 2024).
6. Araújo, M. R. B. et al. Phenotypic and molecular characterization and complete genome sequence of a Corynebacterium diphtheriae strain isolated from cutaneous infection in an immunized individual. *Braz J. Microbiol.* **54** (3), 1325–1334 (2023).
7. Hennart, M. et al. Population genomics and antimicrobial resistance in Corynebacterium diphtheriae. *Genome Med.* **12** (1), 107 (2020).
8. Bolt, F. et al. Multilocus sequence typing identifies evidence for recombination and two distinct lineages of Corynebacterium diphtheriae. *J. Clin. Microbiol.* **48** (11), 4177–4185 (2010).
9. Nguyen Thi Nguyen, T. et al. Endemic erythromycin resistant Corynebacterium diphtheriae in Vietnam in the 1990s. *Microb. Genom.* **8**(10), mgen000861 (2022).
10. Leber, A. L. *Clinical Microbiology Procedures Handbook* (Wiley, 2020).
11. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using spades de Novo assembler. *Curr. Protocols Bioinf.* **70** (1), e102 (2020).
12. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with kraken 2. *Genome Biol.* **20** (1), 257 (2019).
13. Larsen, M. V. et al. Benchmarking of methods for genomic taxonomy. *J. Clin. Microbiol.* **52** (5), 1529–1539 (2014).
14. Hasman, H. et al. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J. Clin. Microbiol.* **52** (1), 139–146 (2014).
15. Clausen, P., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of Raw reads against redundant databases with KMA. *BMC Bioinform.* **19** (1), 307 (2018).
16. Silva, M. et al. ChewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microb. Genom.* **4**(3), e000166 (2018).
17. Bharat, A. et al. Correlation between phenotypic and in Silico detection of antimicrobial resistance in Salmonella enterica in Canada using Staramr. *Microorganisms* **10**, 292 (2022).
18. Hunt, M. et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genom.* **3**(10), e000131 (2017).
19. Bortolaia, V. et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75** (12), 3491–3500 (2020).
20. Hennart, M. et al. A global Corynebacterium diphtheriae genomic framework sheds light on current diphtheria reemergence. *Peer Community J.* **3**, e76 (2023).
21. Li, M. et al. RpoB mutations and effects on Rifampin resistance in Mycobacterium tuberculosis. *Infect. Drug Resist.* **14**, 4119–4128 (2021).
22. Crestani, C. et al. Corynebacterium Ramonii sp. nov., a novel toxigenic member of the Corynebacterium diphtheriae species complex. *Res. Microbiol.* **174** (7), 104113 (2023).
23. Ramdhan, N. D. et al. Genomic analysis of a novel nontoxigenic Corynebacterium diphtheriae strain isolated from a cancer patient. *New Microbes and New Infections*. **30**, 100544 (2019).
24. Fu, P., Jamison, M., La, S. & MacMillan, J. B. Inducamides A–C, chlorinated alkaloids from an RNA polymerase mutant strain of streptomyces Sp. *Org. Lett.* **16** (21), 5656–5659 (2014).
25. Trost, E. et al. Pangenomic study of Corynebacterium diphtheriae that provides insights into the genomic diversity of pathogenic isolates from cases of classical diphtheria, endocarditis, and pneumonia. *J. Bacteriol.* **194** (12), 3199–3215 (2012).
26. Jian, Z. et al. Antibiotic resistance genes in bacteria: Occurrence, spread, and control. *J. Basic Microbiol.* **61** (12), 1049–1070 (2021).
27. Wołkowicz, T., Zacharczuk, K. & Zasada, A. A. Genomic analysis of Corynebacterium diphtheriae strains isolated in the years 2007–2022 with a report on the identification of the first Non-Toxigenic Tox Gene-Bearing strain in Poland. *Int. J. Mol. Sci.* **24** (5), 4612 (2023).
28. Köser, C. U., Ellington, M. J. & Peacock, S. J. Whole-genome sequencing to control antimicrobial resistance. *Trends Genet.* **30** (9), 401–407 (2014).
29. Pereira, G. A. et al. Antimicrobial resistance among Brazilian Corynebacterium diphtheriae strains. *Mem. Inst. Oswaldo Cruz.* **103** (5), 507–510 (2008).
30. Husada, D. et al. First-line antibiotic susceptibility pattern of toxigenic Corynebacterium diphtheriae in Indonesia. *BMC Infect. Dis.* **19** (1), 1049 (2019).
31. Bernard, K. & Pacheco, A. L. In vitro activity of 22 antimicrobial agents against Corynebacterium and Microbacterium species referred to the Canadian National microbiology laboratory. *Clin. Microbiol. Newsl.* **37** (23), 187–198 (2015).
32. Nadella, P. et al. Determinants of delayed or incomplete diphtheria-tetanus-pertussis vaccination in parallel urban and rural birth cohorts of 30,956 infants in Tanzania. *BMC Infect. Dis.* **19** (1), 188 (2019).
33. Olulaja, O. N., Anjorin, E. T., Ekerin, O., Afolabi, O. T. & Inuojo, J. M. A looming epidemic: combating the recurrent outbreaks of diphtheria in Nigeria. *Pan Afr. Med. J.* **45**, 186 (2023).

Acknowledgements

We thank the Institut Pasteur teams for the curation and maintenance of BIGSdb-Pasteur databases at http://bigsdb.pasteur.fr/.

Author contributions

Conceptualization: HTTH, MS. Data curation: HTTH, LMH, PTH, DBN, NTT. Formal analysis: LHH, MM, MS. Investigation: HTTH, LMH, PTH, DBN, NTT, LHH, MM, MS. Methodology: HTTH, MS, LHH, MM.

Project administration: HTTH, MS. Resources: HTTH. Software: LHH, MM. Supervision: HTTH, MS. Visualization: LHH. Writing – original draft: LHH, HTTH, LMH, MS. Writing – review & editing: HTTH, MS, LHH, LMH, PTH, DBN, NTT, PTL, PTLN, TND, NTTH, MM.

Funding
This study was supported by the Research Program on Emerging and Re-emerging Infectious Diseases (JP-24fk0108683 and JP25fk0108683) from the Japan Agency for Medical Research and Development (AMED).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-38608-2.

Correspondence and requests for materials should be addressed to H.T.T.H. or M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.